RĪGAS TEHNISKĀ UNIVERSITĀTE

Emīls Bolmanis

CEĻĀ UZ INDUSTRIJU 4.0: REĀLLAIKA MONITORINGS, HIBRĪDĀ MODELĒŠANA UN PROGNOZĒJOŠĀ VADĪBA *PICHIA PASTORIS* FERMENTĀCIJĀS

Promocijas darbs

TOWARDS INDUSTRY 4.0: REAL-TIME MONITORING, HYBRID MODELING AND PREDICTIVE CONTROL IN *PICHIA PASTORIS* FERMENTATIONS

Doctoral Thesis

RTU Izdevniecība
RTU Press
Rīga 2026

RIGA TECHNICAL UNIVERSITY
Faculty of Natural Sciences and Technology
Institute of Biomaterials and Bioengineering

RĪGAS TEHNISKĀ UNIVERSITĀTE
Dabaszinātņu un tehnoloģiju fakultāte
Biomateriālu un bioinženierijas institūts

Emīls Bolmanis
Doctoral Student of the Study Programme "Chemistry, Materials Science and Engineering"
Doktora studiju programmas "Ķīmija, materiālzinātne un tehnoloģijas" doktorants

TOWARDS INDUSTRY 4.0: REAL-TIME MONITORING, HYBRID MODELING AND PREDICTIVE CONTROL IN *PICHIA PASTORIS* FERMENTATIONS

Doctoral Thesis

CEĻĀ UZ INDUSTRIJU 4.0: REĀLLAIKA MONITORINGS, HIBRĪDĀ MODELĒŠANA UN PREDIKTĪVĀ VADĪBA *PICHIA PASTORIS* FERMENTĀCIJĀS

Promocijas darbs

Scientific supervisors / Zinātniskie vadītāji
Senior Researcher *Dr. biol.* ANDRIS KAZĀKS
Professor *Dr. sc. ing*. JURIS VANAGS

Consultant / Konsultants
Professor *Dr. sc. ing.*
VYTAUTAS GALVANAUSKAS

RTU Press / RTU izdevniecība
Riga 2026 / Rīga 2026

Bolmanis, E. Towards Industry 4.0: Real-Time Monitoring, Hybrid Modeling and Predictive Control in *Pichia Pastoris* Fermentations. Doctoral Thesis. Riga: RTU Press, 2026. 205 p.

Published in accordance with the decision of the Promotion Council "RTU P-01" of 9 December 2026, Minutes No. 04030-9.1/77.

The Doctoral Thesis was developed at the Latvian Biomedical Research and Study Centre, Riga Technical University, and the Latvian State Institute of Wood Chemistry.

This work has been supported by the European Social Fund within the Project No. 8.2.2.0/20/I/008 "Strengthening of PhD students and academic personnel of Riga Technical University and BA School of Business and Finance in the strategic fields of specialization" and by the EU Recovery and Resilience Facility within Project No. 5.2.1.1.i.0/2/24/I/CFLA/003 "Implementation of consolidation and management changes at Riga Technical University, Liepaja University, Rezekne Academy of Technology, Latvian Maritime Academy and Liepaja Maritime College for the progress towards excellence in higher education, science and innovation" academic career doctoral grant. I acknowledge Riga Technical University's HPC Center for providing access to their computing infrastructure.

Cover image from www.shutterstock.com

DOCTORAL THESIS PROPOSED TO RIGA TECHNICAL UNIVERSITY FOR PROMOTION TO THE SCIENTIFIC DEGREE OF DOCTOR OF SCIENCE

To be granted the scientific degree of Doctor of Science (PhD), the present Doctoral Thesis has been submitted for defence at the open meeting of RTU Promotion Council on 29 April 2026, at 14:00 at the Faculty of Natural Sciences and Technology of Riga Technical University, Paula Valdena street 3, Room 272.

OFFICIAL REVIEWERS

Associate Professor Dr. sc. ing. Linda Mežule
Riga Technical University

Professor Dr. biol. Uldis Kalnenieks
University of Latvia, Latvia

Professor PhD Rui Manuel Freitas Oliveira
NOVA University Lisbon, Portugal

DECLARATION OF ACADEMIC INTEGRITY

I hereby declare that the Doctoral Thesis submitted for review to Riga Technical University for promotion to the scientific degree of Doctor of Science (PhD) is my own. I confirm that this Doctoral Thesis has not been submitted to any other university for promotion to a scientific degree.

Emīls Bolmanis ……. (signature)
Date: ……

The Doctoral Thesis has been prepared as a collection of thematically related scientific publications with summaries in Latvian and English. The Doctoral Thesis compiles five scientific publications written in English. The total volume of the Doctoral Thesis is 205 pages, including appendices.

ANNOTATION

This Doctoral Thesis advances the field of recombinant *Pichia pastoris* (reclassified as *Komagataella phaffii*) fermentation engineering through the development and integration of enhanced strategies for real-time monitoring, predictive modeling, and process control, with a strong emphasis on data-driven bioprocessing aligned with Industry 4.0 principles. Using recombinant *P. pastoris* strains producing human hepatitis B core antigen (HBcAg), leghemoglobin (LegH), and bacteriophage Qβ coat protein particles as case studies, the work addresses key technological challenges in fed-batch bioprocesses.

The first part of the thesis focuses on the validation and real-time enhancement of sensor systems for biomass, methanol, and exhaust gas monitoring. Signal processing algorithms were developed to improve data quality and sensor reliability during fermentation.

The second part investigates mechanistic, data-driven, and hybrid modeling approaches, evaluating their predictive accuracy and robustness. Transfer learning was employed to accelerate hybrid model development by leveraging historical fermentation data, thereby reducing experimental time and effort.

The third part presents the implementation of advanced control strategies. A classical proportional-integral (PI) controller was used for online residual methanol regulation, while a novel hybrid model predictive control (MPC) framework was developed for real-time biomass growth trajectory tracking. The MPC system demonstrated robustness under process variability, confirming its suitability for intelligent control in biomanufacturing.

The Thesis is presented as a collection of thematically linked four original research articles and one review, collectively contributing to the modernization of fermentation engineering practices. It includes summaries in English and Latvian, 15 figures, one scheme, three tables, and five appendices, comprising a total of 205 pages.

CONTENTS

ABBREVIATIONS 6
GENERAL OVERVIEW OF THE THESIS 7
Introduction 7
Aims and Objectives 9
Theses to Defend 9
Scientific Novelty 10
Practical Significance 10
Structure and Volume of the Thesis 10
Publications and Approbation of the Thesis 11
MAIN RESULTS OF THE THESIS 14
1. Literature Survey 14
2. Real-Time Fermentation Monitoring 17
2.1. Cell biomass measurement 17
2.2. Methanol concentration measurement 19
2.3. Reactor exhaust gas composition analysis 20
2.4. Biomass sensor signal anomaly detection and removal 21
3. *P. pastoris* Fermentation Modeling 25
3.1. Mechanistic modeling 25
3.2. Data-driven modeling 27
3.3. Hybrid modeling 29
3.4. Model performance comparison 31
3.5. Transfer learning 35
4. Fermentation Control 38
4.1. Methanol set-point control 38
4.2. Model predictive control 40
CONCLUSIONS 44
REFERENCES 45

Appendix I: Bolmanis, E.; Dubencovs, K.; Suleiko, A.; Vanags, J. Model Predictive Control – A Stand Out among Competitors for Fed-Batch Fermentation Improvement. *Fermentation* **2023**, *9*, 3.

Appendix II: Bolmanis, E.; Uhlendorff, S.; Pein-Hackelbusch, M.; Galvanauskas, V.; Grigs, O. Anomaly Detection and Removal Strategies for *In-Line* Permittivity Sensor Signal Used in Bioprocesses. *Front. Bioeng. Biotechnol.* **2025***, 13*.

Appendix III: Bolmanis, E.; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-level production of recombinant HBcAg virus-like particles in a mathematically modelled *P. pastoris* GS115 Mut$^+$ bioreactor process under controlled residual methanol concentration. *Bioprocess Biosyst. Eng.* **2022**, *45*, 9.

Appendix IV: Bolmanis, E.; Bogans, J.; Akopjana, I.; Suleiko, A.; Kazaka, T.; Kazaks, A. Production and Purification of Soy Leghemoglobin from *Pichia pastoris* Cultivated in Different Expression Media. *Processes* **2023**, *11*, 11.

Appendix V: Bolmanis, E.; Galvanauskas, V.; Grigs, O.; Vanags, J.; Kazaks, A. Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development. *Fermentation* **2025***, 11*, 7.

ABBREVIATIONS

AICc	corrected Akaike information criterion
ADAM	adaptive moment estimation optimizer
AI	artificial intelligence
ANN	artificial neural network
AOX1	alcohol oxidase 1 promoter
CER	carbon dioxide evolution rate
DCW	dry cell weight
DNN	deep neural network
DO	dissolved oxygen
DRA	double rolling aggregate
FC	fully connected (layer)
GRAS	generally recognized as safe
HBcAg	human hepatitis B core antigen
Industry 4.0	the fourth industrial revolution
IQR	inter-quartile range
LegH	soy leghemoglobin
LeakyReLU	leaky rectified linear unit
LSTM	long short-term memory
MAD	median absolute deviation
MeOH	methanol
MPC	model predictive control
N_C	control horizon
N_P	prediction horizon
NRMSE	normalized root mean square error [%]
ODE	ordinary differential equation
OUR	oxygen uptake rate
P	product concentration
Pareto front	optimal trade-off curve in multi-objective optimization
PID	proportional-integral-derivative (controller)
PI	proportional-integral (controller)
Qβ	bacteriophage Q-beta coat protein virus-like particle
ReLU	rectified linear unit
RNN	recurrent neural network
RQ	respiratory quotient
S	substrate concentration
SCADA	supervisory control and data acquisition
Tanh	hyperbolic tangent activation function
V	volume
VLP	virus-like particle
X	cell biomass concentration

GENERAL OVERVIEW OF THE THESIS

Introduction

Recombinant protein production underpins a wide array of biotechnological applications, including the development of biopharmaceuticals, diagnostics, industrial enzymes, and synthetic biology [1, 2]. As demand for biologics continues to rise – driven by aging populations and advances in precision medicine – the global market is projected to reach $740 billion by 2031 [3]. Meeting this demand requires scalable, cost-effective biomanufacturing platforms featuring robust host strains, optimized expression systems, and advanced process control.

To illustrate the diversity and application potential of recombinant proteins, recombinant human hepatitis B core antigen (HBcAg), bacteriophage Qβ capsid protein (Qβ), and soy leghemoglobin (LegH) are notable examples with significant medical and industrial relevance. HBcAg is a self-assembling viral capsid protein extensively studied for virus-like particle (VLP) platforms, which show great promise in vaccine development, drug delivery, and immunotherapy due to their strong immunogenicity and safety [4, 5]. Similarly, Qβ capsid protein, another VLP-forming protein derived from bacteriophages, serves as a versatile scaffold in nanotechnology, vaccines, and diagnostics, thanks to its structural uniformity and modifiable surface [6–8]. In contrast, LegH, a plant-derived heme protein responsible for imparting meat-like flavor and aroma, has emerged as a key component in the rapidly growing market for plant-based meat alternatives, notably in products such as the Impossible™ Burger [9, 10]. The microbial expression of these proteins, both at the laboratory and industrial scale, offers a scalable and cost-effective platform for sustainable production, aligning with growing healthcare demands and environmental sustainability.

A central contributor to this scalable production is the microbial host system, and among the available options, *Pichia pastoris* is a well-established and versatile choice. Although taxonomically reclassified as *Komagataella phaffii*, since the experimental work, strain designations, and the majority of cited literature refer to the organism as *Pichia pastoris*, the designation *P. pastoris* is used throughout this Thesis for consistency.

Widely regarded as a workhorse in industrial biotechnology, this methylotrophic yeast combines rapid growth to high cell densities in defined media with tightly regulated, methanol-inducible promoters – most notably AOX1 – for strong and controllable gene expression [11–13]. It also features an efficient secretory pathway and supports essential eukaryotic post-translational modifications, such as disulfide bond formation and glycosylation, crucial for the proper function of many therapeutic proteins [14, 15]. It is generally recognized as safe (GRAS) status further highlights its suitability for pharmaceutical and industrial applications [16, 17].

To unlock the full potential of such microbial platforms, especially at scale, robust and efficient bioprocessing strategies are required. Fed-batch cultivation, the industry standard for microbial processes, enables high product yields through controlled substrate feeding and is widely used to produce amino acids, antibiotics, enzymes, and other biochemicals [18–20]. Its ability to mitigate large-scale challenges like mass and heat transfer by adjusting feed rates supports optimal mixing, oxygenation, and temperature control [21]. However, the success of

fed-batch processes depends on precise feed control, as over- or underfeeding can lead to substrate inhibition, oxygen limitation, or metabolic overflow, compromising culture performance [18, 22]. To address this, advanced modeling and control strategies are increasingly adopted to enhance process efficiency and scalability.

At the heart of these advanced strategies lie the three core pillars of fermentation engineering: monitoring, modeling, and control [23, 24]. Continuous monitoring through physical sensors and/or sampling provides data on biomass, substrate, and product levels, which modeling translates into predictive insights using mechanistic, data-driven, or hybrid approaches. These models support dynamic control strategies that adjust key process variables to maintain optimal conditions and ensure consistent product quality. By integrating these three pillars, fermentation processes can be precisely managed, enhancing scalability, robustness, and efficiency in recombinant protein production.

Historically, however, control strategies in *P. pastoris* fermentations have relied on empirical heuristics or simplified mechanistic models, limiting adaptability and predictive power [25, 26]. However, the path towards the 4th industrial revolution (Industry 4.0) has transformed bioprocessing into a data-rich discipline, where historical and real-time data streams can be systematically leveraged to enhance understanding, prediction, and decision-making [27]. In this context, intelligent hybrid modeling offers a robust solution to the nonlinear, dynamic nature of microbial systems, enabling adaptive, scalable, and efficient control strategies for modern biomanufacturing.

Among emerging strategies, hybrid modeling approaches – particularly those that integrate first-principles knowledge with machine learning components – are gaining traction for their ability to preserve process interpretability while enhancing predictive performance [27–29]. Among data-driven techniques, deep neural networks (DNNs), including recurrent neural networks (RNNs) and long short-term memory (LSTM) architectures, are especially suited to bioprocess applications due to their capacity to learn from temporal patterns and capture delayed system responses. These models have been successfully employed for state estimation, fault detection, process optimization, and soft sensing in various biomanufacturing contexts [30–32]. However, their black-box nature can limit interpretability and regulatory acceptance when used in isolation.

To translate these advanced modeling capabilities into actionable control, model predictive control (MPC) has emerged as a particularly powerful strategy. MPC offers a structured framework for managing multivariable, constrained, and time-varying systems [33–35]. By leveraging predictive models – mechanistic, data-driven, or hybrid – MPC can forecast future process behavior and compute optimal control actions in a receding horizon fashion. Its inherent ability to handle constraints and anticipate process disturbances makes MPC well-suited for fed-batch fermentation, where maintaining optimal substrate concentrations or microbial growth rate, minimizing oscillations, and maximizing productivity are critical [36–38]. In the context of *P. pastoris* fermentations, the integration of MPC with hybrid models holds strong potential for real-time optimization of substrate feeding and environmental conditions – enhancing process robustness, scalability, and performance. However, such applications remain largely unaddressed in the current scientific literature.

To bridge these gaps, this Thesis addresses key challenges in recombinant *P. pastoris* fermentation engineering – monitoring, modeling, and control – using HBcAg-, LegH-, and Qβ-producing strains as case studies. Real-time monitoring using biomass, methanol, and exhaust gas analyzer sensors is investigated to generate extensive datasets, while sensor signal quality is addressed through data processing techniques to reduce signal noise and an algorithm that detects and removes anomalies in biomass probe signals. For process modeling, mechanistic, data-driven, and hybrid models were developed and comparatively evaluated. Transfer learning was applied to successfully adapt the hybrid model to new datasets using historical process data. Finally, process control was demonstrated using a simple PI controller to regulate residual methanol levels, and a hybrid model-based MPC framework was implemented to track predefined cell growth trajectories.

Aims and Objectives

This thesis aims to advance recombinant *P. pastoris* fermentation engineering by developing and integrating enhanced strategies for process monitoring, modeling, and control. By aligning with Industry 4.0 principles, the work contributes to the transition toward intelligent, data-driven bioprocessing. Using HBcAg, LegH, and Qβ-producing strains as case studies, the Thesis has three specific aims.

1. Validate biomass, methanol, and exhaust gas sensors in *P. pastoris* fermentations to ensure high-quality real-time data. Develop real-time signal processing algorithms to enhance sensor signal quality and reliability.
2. Develop, evaluate, and compare mechanistic, data-driven, and hybrid modeling approaches. Investigate the use of transfer learning to accelerate hybrid model development by leveraging historical process data.
3. Implement and experimentally validate substrate feed rate control strategies, including a conventional PI controller for residual methanol regulation using online sensor feedback, and an advanced MPC controller for biomass growth trajectory tracking based on the developed hybrid process model.

Theses to Defend

1. Sensor signal quality is critical in fermentations, and effective real-time pre-processing is essential to ensure accurate process monitoring and control.
2. Hybrid modeling approaches outperform purely mechanistic and data-driven models in both predictive accuracy and robustness.
3. Transfer learning is an effective strategy in bioprocess engineering for reducing model training time and experimental effort by leveraging historical data.
4. Hybrid model-based MPC framework allows precise substrate feed control, facilitating the tracking of predefined growth trajectories in *P. pastoris* fermentations.

Scientific Novelty

The scientific novelty of this Thesis is reflected in three key areas – process monitoring, modeling, and control – each contributing to the advancement of intelligent *P. pastoris* fermentation engineering.

1. Real-time signal processing solutions were developed and applied to biomass, methanol and exhaust gas sensors, improving signal quality and enabling more reliable online monitoring in *P. pastoris* fermentations.
2. Mechanistic, data-driven, and hybrid models were systematically compared, demonstrating that hybrid models integrating neural networks with first-principles knowledge achieved the best predictive performance across HBcAg, LegH, and Qβ strains. Transfer learning was successfully applied to leverage historical data, reducing training time and experimental effort.
3. A hybrid model-based MPC framework was implemented to control the specific growth rate in real time, achieving trajectory tracking with 10.6 % NRMSE in experimental fermentations. The system demonstrated robustness under process variability, confirming the suitability of hybrid MPC for intelligent control in biomanufacturing.

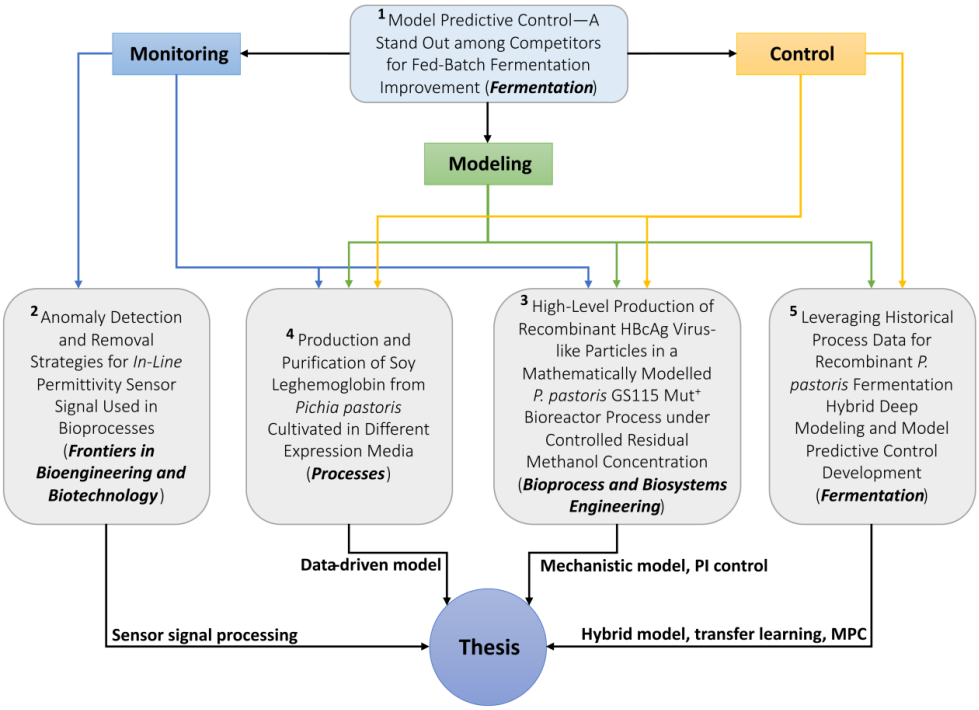
Practical Significance

The practical significance of this Thesis lies in its contributions to improving real-time bioprocess monitoring, predictive modeling, and closed-loop control in recombinant *P. pastoris* fermentations. Aligned with Industry 4.0 principles, the work contributes to the transition toward intelligent, data-driven bioprocessing.

1. Real-time signal processing methods improved the reliability of biomass, methanol, and gas sensor data, enabling more accurate online monitoring and decision-making during fermentation runs.
2. The hybrid modeling approach facilitated better process understanding and prediction, while transfer learning reduced experimental demand – offering practical tools for rapid model adaptation in industrial settings.
3. The hybrid MPC system enabled automated control of the specific growth rate, supporting consistent process performance and scalability for industrial protein production.

Structure and Volume of the Thesis

This Doctoral Thesis presents a collection of thematically linked publications that advance process monitoring, modeling, and control strategies for recombinant *P. pastoris* fed-batch fermentations. Emphasizing improved sensor signal quality, hybrid modeling with transfer learning, and intelligent MPC-based control, the Thesis includes four original research articles and one review, contributing to Industry 4.0-aligned data-driven bioprocessing (Scheme 1).



Scheme 1. Schematic representation of the Thesis structure.

Publications and Approbation of the Thesis

The Thesis results are reported in four original scientific publications. One review article has been published. The main results were presented at 3 conferences.

Scientific publications

1. **Bolmanis, E.**; Dubencovs, K.; Suleiko, A.; Vanags, J. Model Predictive Control – A Stand Out among Competitors for Fed-Batch Fermentation Improvement. *Fermentation* **2023**, *9*, 206, doi: 10.3390/fermentation9030206. [Scopus, WoS, Open Access, IF 5.123, Q1, CiteScore 5.3]
2. **Bolmanis, E.**; Uhlendorff, S.; Pein-Hackelbusch, M.; Galvanauskas, V.; Grigs, O. Anomaly Detection and Removal Strategies for In-Line Permittivity Sensor Signal Used in Bioprocesses. *Front. Bioeng. Biotechnol.* **2025**, *13*, doi: 10.3389/fbioe.2025.1609369. [Scopus, WoS, Open Access, IF 4.8, Q1, CiteScore 8.8]
3. **Bolmanis, E.**; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. pastoris* GS115 Mut+ Bioreactor Process under Controlled Residual Methanol Concentration.

Bioprocess Biosyst. Eng. **2022**, *45*, 1447–1463, doi: 10.1007/s00449-022-02754-4. [Scopus, WoS, IF 3.6, Q2, CiteScore 6.7]

4. **Bolmanis, E.**; Bogans, J.; Akopjana, I.; Suleiko, A.; Kazaka, T.; Kazaks, A. Production and Purification of Soy Leghemoglobin from *Pichia pastoris* Cultivated in Different Expression Media. *Processes* **2023**, *11*, 3215, doi: 10.3390/pr11113215 [Scopus, WoS, Open Access, IF 3.5, Q2, CiteScore 4.7]
5. **Bolmanis, E.**; Galvanauskas, V.; Grigs, O.; Vanags, J.; Kazaks, A. Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development. *Fermentation* **2025**, *11*, 411, doi: 10.3390/fermentation11070411 [Scopus, WoS, Open Access, IF 3.3, Q2, CiteScore 5.7]

Other scientific publications

1. Grigs, O.; **Bolmanis, E.**; Galvanauskas, V. Application of In-Situ and Soft-Sensors for Estimation of Recombinant *P. pastoris* GS115 Biomass Concentration: A Case Analysis of HBcAg (Mut^+) and HBsAg (Mut^S) Production Processes under Varying Conditions. *Sensors* **2021**, *21*, 1268, doi: 10.3390/s21041268.
2. Grigs, O.; Didrihsone, E.; **Bolmanis, E**. Investigation of a Broad-Bean-Based Low-Cost Medium Formulation for *Bacillus subtilis* MSCL 897 Spore Production. *Fermentation* **2023**, *9*, 4, doi: 10.3390/fermentation9040390.
3. Pentjuss, A.; **Bolmanis, E.**; Suleiko, A.; Didrihsone, E.; Suleiko, A.; Dubencovs, K.; Liepins, J.; Kazaks, A.; Vanags, J. *Pichia pastoris* Growth – Coupled Heme Biosynthesis Analysis Using Metabolic Modelling. *Sci. Rep.* **2023**, *13*, 15816, doi: 10.1038/s41598-023-42865-w.
4. Suleiko, A.; Dubencovs, K.; Kazaks, A.; Suleiko, A.; Daugavietis, J. E.; Didrihsone, E.; Liepins, J.; **Bolmanis, E.**; Grigs, O.; Vanags, J. Performance of Recombinant *Komagataella phaffii* in Plant-Based Meat Flavor Compound (Leghemoglobin) Production through Fed-Batch Fermentations. *Fermentation* **2024**, *10*, 1, doi: 10.3390/fermentation10010055.
5. **Bolmanis, E.**; Grigs, O.; Didrihsone, E.; Senkovs, M.; Nikolajeva, V. Pilot-Scale Production of *Bacillus subtilis* MSCL 897 Spore Biomass and Antifungal Secondary Metabolites in a Low-Cost Medium. *Biotechnol. Lett.* **2024**, *46*, 3, doi: 10.1007/s10529-024-03481-4.

Participation in scientific conferences

1. **Bolmanis, E.**; Ramm, S.; Pein-Hackelbusch, M.; Galvanauskas, V.; Grigs, O. Dielectric Permittivity Sensor Signal Anomaly Detection and Compensation Strategies in Yeast *P. pastoris* Fermentations. *83rd International Scientific Conference of the University of Latvia*. February 14, 2025, Riga, Latvia (*Oral presentation*).

2. Uhlendorff, S.; **Bolmanis, E.**; Pein-Hackelbusch, M.; Galvanauskas, V.; Grigs, O. Analysis of Anomaly Detection Techniques for *In-line* Permittivity Sensors in Bioprocesses. *8th European Congress of Applied Biotechnology (ECAB)*. September 8–10, 2025, Lisbon, Portugal (*Poster presentation*).
3. **Bolmanis, E.**; Galvanauskas, V.; Kazaks, A. Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling. *6th Congress of Baltic Microbiologists*. October 1–3, 2025, Riga, Latvia (*Oral presentation*).

Participation in other scientific events

1. **Bolmanis, E.;** Kazaks, A. Soy leghemoglobin (LegH) production in yeast *P. pastoris* in different cultivation media. *Informative seminar on the results of the Project "The development of an efficient pilot-scale leghemoglobin production technology, based on recombinant Pichia pastoris and Kluyveromyces lactis fed-batch fermentations (BioHeme)".* November 15, 2023, Riga, Latvia (*Oral presentation*).
2. **Bolmanis, E.;** Kazaks, A. Soy leghemoglobin (LegH) production in yeast *P. pastoris* in different cultivation media. *Informative seminar on the results of the Project "The development of an efficient pilot-scale leghemoglobin production technology, based on recombinant Pichia pastoris and Kluyveromyces lactis fed-batch fermentations (BioHeme)".* November 23, 2023, Riga, Latvia (*Oral presentation*).

MAIN RESULTS OF THE THESIS

1. Literature Survey

Publication:

- **Bolmanis, E.**; Dubencovs, K.; Suleiko, A.; Vanags, J. Model Predictive Control – A Stand Out among Competitors for Fed-Batch Fermentation Improvement. *Fermentation* **2023**, *9*, 206 [33].

Fed-batch cultivation has long been a cornerstone of industrial fermentation, widely adopted for producing a broad spectrum of high-value biotechnological products. It remains the dominant mode of operation in biopharmaceutical manufacturing, encompassing both marketed therapeutics and clinical-stage products [19, 20].

A central challenge in fed-batch operations is the control of substrate feed rate – a critical process variable that directly influences the specific growth rate, metabolic flux distribution, product titer, and batch-to-batch reproducibility [13, 33]. Optimizing the feed strategy is particularly complex due to the nonlinear and time-varying nature of microbial responses to dynamic environmental conditions.

Feed rate control strategies in fed-batch fermentations can be broadly classified into two categories: open-loop and closed-loop (feedback) control (Fig. 1.1), each with distinct advantages and limitations. Open-loop strategies are simple to implement but lack flexibility, as the feed profile is predefined and remains unchanged throughout the process. In contrast, closed-loop strategies incorporate a feedback element – such as real-time sensor data or model-based predictions – to continuously adjust the feed rate during cultivation. This dynamic adjustment enables greater robustness to disturbances and biological variability, improving process stability and reproducibility. In *P. pastoris* fermentations, an open-loop approach would involve calculating the substrate feed profile in advance, whereas a feedback-based strategy would allow the profile to evolve in response to real-time signals from the feedback element.

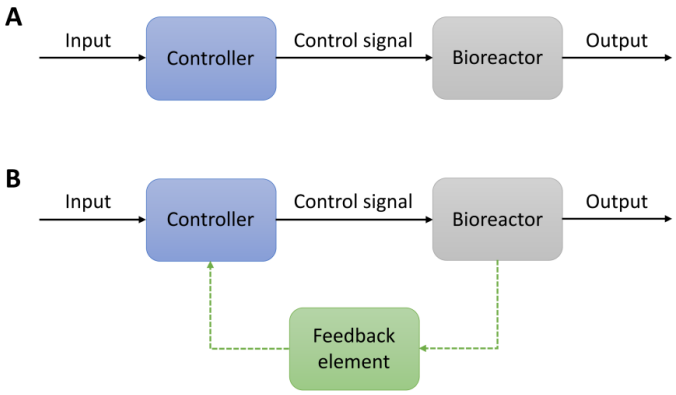


Fig. 1.1. Schematic of open-loop (A) and closed-loop (B) control architectures.

The selection of an appropriate control strategy must balance implementation complexity with expected performance, as this trade-off directly impacts the cost-effectiveness and consistency of the bioprocess [33]. Given the inherent variability of microbial systems – even under nominally constant operating conditions – feedback control is especially valuable. By continuously adapting substrate supply in response to real-time measurements, feedback-based strategies help maintain metabolic balance, enhance reproducibility, and ensure that nutrient availability aligns with cellular demands throughout the cultivation period [18].

Effective implementation of feedback control depends on the availability of a reliable feedback mechanism. Typically, such mechanisms combine real-time physical sensors (e.g., for dissolved oxygen (DO), substrate, or biomass), at-line or off-line analytical data (e.g., optical density or residual substrate concentration), and predictive process models. These components work in synergy to estimate internal states and guide control actions in real time [39].

The accuracy, responsiveness, and robustness of the feedback loop are critical determinants of controller performance. An inadequate or noisy feedback signal may lead to misguided control actions, resulting in overfeeding, reduced yields, or even process instability [40]. Therefore, integrating high-quality monitoring technologies and robust process models is essential to fully realize the benefits of feedback control in fed-batch fermentations.

At the heart of the feedback infrastructure are sensors and process models, which serve complementary roles in bioprocess monitoring and control. Physical sensors provide direct, real-time measurements of key process variables such as pH, temperature, DO, biomass (via dielectric spectroscopy or turbidity), and carbon sources like glucose or methanol [41, 42]. While these sensors are generally robust and easy to calibrate, they are limited in scope, may be costly, and are prone to fouling or drift – especially in large-scale applications [43, 44].

To overcome these limitations, process models have become increasingly important. These models – whether mechanistic, data-driven, or hybrid – can estimate unmeasured variables (e.g., specific growth or production rate) by integrating available measurements [45, 46]. Their key strengths lie in flexibility, cost-effectiveness, and the ability to infer otherwise unmeasurable process states. However, their reliability depends heavily on model structure and input data quality, necessitating regular recalibration to maintain long-term accuracy [47].

Bioprocess models span from mechanistic approaches – rooted in biochemical and physiological principles – to data-driven models such as statistical regressions or machine learning algorithms that capture empirical relationships within data. Mechanistic models offer interpretability and insight, but often require significant domain expertise and labor-intensive parameter estimation [23]. Data-driven models excel at modeling complex, nonlinear behaviors without detailed prior knowledge, but tend to lack transparency and depend on data quality [48].

To leverage the strengths of both paradigms, hybrid models are gaining traction in bioprocess engineering [29, 49, 50]. Hybrid modeling integrates mechanistic structure with data-driven flexibility to provide more accurate and generalizable representations of bioprocess dynamics, especially when complete mechanistic knowledge is lacking.

Among closed-loop control strategies, the proportional-integral-derivative (PID) controller remains the most widely implemented method in industrial fed-batch fermentation. This classical controller computes the error between the measured variable and the setpoint, then

adjusts the input based on the proportional (K_P), integral (K_I), and derivative (K_D) components [51]. PID controllers are typically deployed in indirect feedback configurations, adjusting feed rates based on secondary signals such as pH (pH-stat), DO (DO-stat), specific growth rate (μ-stat), or residual substrate. While relatively simple and robust, PID performance is often hampered by limited access to reliable real-time measurements for biological variables and by the nonlinear, time-varying behavior of microbial systems [33, 52].

To overcome these limitations, model predictive control (MPC) has emerged as a superior alternative. MPC utilizes a dynamic model to predict future system behavior and optimize control actions accordingly [34, 53]. In contrast to PID, which reacts to current deviations in a single variable, MPC can manage multiple variables concurrently, respect operational constraints, and better handle nonlinearity and process disturbances [33, 54]. Moreover, while PID tuning requires regular gain adjustment – often a laborious and sensitive process – MPC uses model-based parameters such as cost function weights and prediction horizons for tuning, enabling greater adaptability and reduced need for frequent adjustments. Recent studies further highlight that MPC can be deployed efficiently on standard industrial hardware, underscoring its practical applicability for real-time bioprocess control [33].

2. Real-Time Fermentation Monitoring

Publications:

- **Bolmanis, E**.; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. pastoris* GS115 Mut+ Bioreactor Process under Controlled Residual Methanol Concentration. *Bioprocess Biosyst. Eng.* **2022**, *45*, 1447–1463 [4].
- **Bolmanis, E.**; Bogans, J.; Akopjana, I.; Suleiko, A.; Kazaka, T.; Kazaks, A. Production and Purification of Soy Leghemoglobin from *Pichia pastoris* Cultivated in Different Expression Media. *Processes* **2023**, *11*, 3215 [56].
- **Bolmanis, E.**; Uhlendorff, S.; Pein-Hackelbusch, M.; Galvanauskas, V.; Grigs, O. Anomaly Detection and Removal Strategies for In-Line Permittivity Sensor Signal Used in Bioprocesses. *Front. Bioeng. Biotechnol.* **2025**, *13* [60].

Effective fermentation process monitoring is essential to ensure product quality, optimize yields, and maintain operational consistency by enabling the timely detection and control of biological and environmental variability. In *P. pastoris* cultivations, precise monitoring of key variables – such as biomass concentration, substrate availability, metabolic activity, and product formation – not only improves process understanding but also facilitates early fault detection and supports enhanced productivity. Moreover, high-resolution monitoring data are critical for the development and application of data-driven modeling approaches that can further refine process control and optimization.

This chapter explores the integration of physical sensors – such as biomass probes, methanol sensors, and exhaust gas analyzers – to enable continuous, non-invasive monitoring throughout the fermentation process. These sensors were employed in selected fermentation experiments to complement standard bioreactor measurements, including DO, pH, temperature, and stirrer speed. The resulting datasets provided a comprehensive view of process dynamics and cellular behavior, serving as a robust foundation for process analysis, real-time control strategies, and the development of hybrid and machine learning-based models.

Monitoring data collected during bioprocesses served as a critical foundation for understanding, optimizing, and modeling fermentation dynamics. Real-time measurements offered valuable insights into the physiological state of the culture throughout fermentations and provided the basis for the data-driven modeling (Chapter 3) and control strategies (Chapter 4) developed and evaluated in this thesis.

2.1. Cell biomass measurement

In situ biomass probes provide real-time, non-invasive measurement of cell density during fermentations, enabling continuous monitoring of microbial growth without the need for manual sampling. During the course of the Thesis research, two types of *in situ* biomass sensor probes were employed in select fermentation processes: an optical probe (*ASD19-EB-01, Optek-Danulat*) that measures culture turbidity, and a dielectric spectroscopy probe (*Incyte, Hamilton*).

These complementary technologies offered distinct advantages for monitoring biomass dynamics. The optical turbidity probe provided a fast and robust signal correlated with total biomass, though it included both viable and non-viable cells. In contrast, the dielectric spectroscopy probe selectively estimated viable biomass by measuring the electrical properties of intact cell membranes. Since key fermentation parameters – such as growth and production rates – are primarily influenced by viable cells, this measurement is more informative. However, it requires advanced reference data, such as viable cell counts, for accurate calibration, which were not available during these experiments. The signals from both biomass sensors, along with reference dry biomass measurements, are shown in Fig. 2.1.

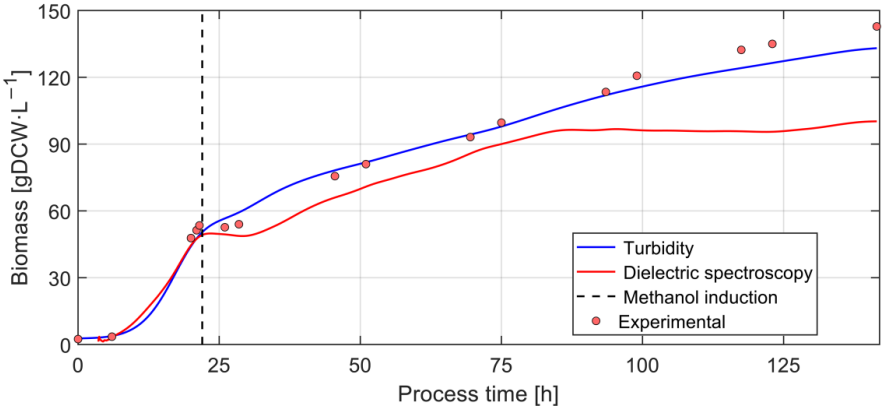


Fig. 2.1. Comparison of optical (turbidity) and dielectric spectroscopy sensor probe cell biomass measurements in a *P. pastoris* fermentation.

As shown in Fig. 2.1, the optical biomass sensor signal closely aligns with the experimentally measured cell biomass values, exhibiting a strong correlation ($R^2 = 0.99$). In contrast, the dielectric spectroscopy probe signal closely matches both the optical sensor and experimental biomass measurements up until methanol induction, after which it exhibits a sharp decline and eventually plateaus beyond approximately 85 hours of cultivation. This behavior reflects changes in cell viability. During the glycerol growth phase, cell viability remains high – close to 100 % – resulting in strong agreement among all three measurement methods [45, 55]. However, following methanol induction, the correlation weakens due to methanol's cytotoxic effects, which reduce the viable cell fraction during the adaptation phase. As adaptation progresses and growth resumes, the signals once again show similar trends, though the dielectric spectroscopy signal remains slightly lower, reflecting the presence of a non-viable cell population. Eventually, around 85 hours into fermentation, the dielectric signal plateaus, suggesting that cell growth and death rates have reached equilibrium. This plateau is not observed in the optical sensor or experimental measurements, as both continue to account for the total biomass, including the non-viable cell fraction.

Process monitoring using biomass probes not only supports cultivation control – by providing timely feedback for adjusting process parameters – but also serves as a rich source of data for developing and refining data-driven models [56, 57]. Notably, the combined use of

optical and dielectric spectroscopy probes offers a more comprehensive perspective on biomass composition by distinguishing between total and viable cell populations. This dual-sensor approach presents a novel opportunity for future research, where simultaneous integration of both signals could help account for the heterogeneity of the cell population [58]. Such an approach could enhance the predictive accuracy of hybrid models and enable more informed control strategies in recombinant *P. pastoris* fermentations.

2.2. Methanol concentration measurement

In addition to biomass monitoring, real-time methanol measurement was a critical component of the fermentation processes studied in this Thesis. For *P. pastoris* constructs driven by the AOX1 promoter, methanol serves a dual role as both a carbon source and an inducer of recombinant protein expression. Precise monitoring and control of its concentration in the culture medium are therefore essential, as excessive residual methanol can inhibit cell growth and negatively impact productivity. To track methanol dynamics, two different sensors were employed: a gas-phase sensor (*BCP-EtOH, BlueSens*) that measures methanol concentration in the reactor exhaust gas, and an *in situ* liquid-phase probe (*MeOH sensor, Raven Biotech*) that directly quantifies methanol levels within the culture broth. Figure 2.2 illustrates the performance of both sensors during fermentation.

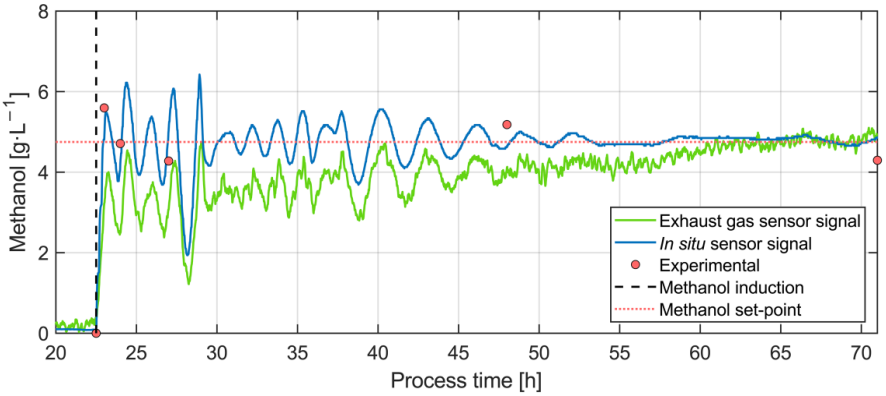


Fig. 2.2. Comparison of exhaust gas and *in situ* methanol sensor performance in a *P. pastoris* fermentation with residual methanol concentration control.

Figure 2.2 presents a detailed comparison of sensor performance throughout the fermentation process. The *in situ* methanol sensor clearly outperforms its exhaust gas-based counterpart in key aspects such as response time, accuracy, and signal quality. A key limitation of the exhaust gas sensor is its inherently noisy signal, which requires the application of filtering and smoothing techniques to extract meaningful trends – at the cost of introducing additional signal delay. In this study, a simple moving average filter with a window size of 10 was applied, significantly enhancing signal quality, reducing fluctuations by 63 % (from ± 0.27 to 0.10 [g·L^{-1}]), but further increasing the overall delay in the feedback signal [4, 59]. Moreover, this sensor does not directly measure conditions within the liquid culture medium; instead, it detects

methanol concentration in the gaseous phase above the liquid surface. Although its signal correlates with methanol levels in the broth, this correlation is subject to a noticeable time lag, which limits its utility for real-time control applications. Additionally, while the signal dynamics generally mirror those of the *in situ* sensor, the estimated concentrations are consistently lower – particularly during the early phase of methanol induction – and gradually stabilize toward the end of the fermentation. This discrepancy may indicate sensor signal drift or delayed equilibration between gas and liquid phases.

Accurate, real-time measurement of methanol concentration enables better control of feeding strategies, enhances process stability, and supports consistent recombinant protein production [12, 59]. Among the available technologies, *in situ* methanol sensors offer superior responsiveness and direct insight into the culture environment, making them especially valuable for process optimization and advanced control applications [59].

2.3. Reactor exhaust gas composition analysis

Reactor exhaust gas analysis is a key aspect of fermentation monitoring, offering real-time insights into microbial respiration and substrate utilization. By measuring critical gases such as oxygen (O_2) and carbon dioxide (CO_2), this approach enables the calculation of key metabolic rates, including the oxygen uptake rate (OUR), carbon dioxide evolution rate (CER), and respiratory quotient (RQ). These parameters are essential for evaluating cellular activity, identifying metabolic shifts, and informing process control strategies in both research and industrial bioprocesses. In this work, an exhaust gas analyzer (*BlueInOneFerm, BlueSens*) was used to continuously monitor O_2 and CO_2 concentrations in the bioreactor exhaust stream. These measurements provided valuable real-time data on respiratory dynamics throughout the fermentation. The typical exhaust gas profiles observed during a representative fermentation are shown in Fig. 2.3.

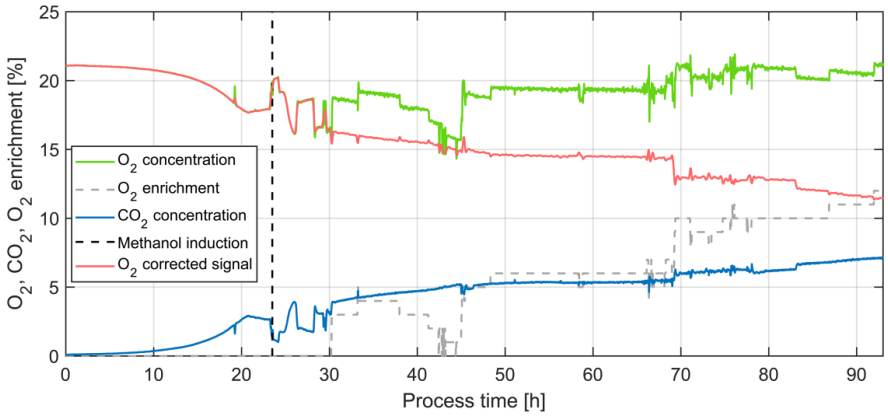


Fig. 2.3. O_2 and CO_2 concentrations in the bioreactor exhaust gas stream during fermentation and inlet air enrichment with pure O_2.

P. pastoris fermentations typically begin with a batch phase using glycerol as the carbon source, followed by a glycerol fed-batch phase to achieve sufficiently high biomass levels before methanol induction. Upon switching the feed to methanol, the cells require time to adjust their metabolism, a transition that is clearly reflected in the exhaust gas analyzer readings (Fig. 2.3). Immediately after initiating methanol feeding, a sharp drop in CO_2 concentration is observed, indicating reduced metabolic activity during adaptation. As CO_2 levels begin to rise again, this signals that the cells are adapting to methanol and resuming growth. These real-time exhaust gas dynamics can be leveraged to tailor the methanol feeding strategy for faster and more efficient adaptation, offering a more responsive alternative to the commonly used but often overly conservative three-step induction protocol.

When adequate oxygenation can no longer be maintained by increasing the agitation rate alone, the bioreactor system initiates enrichment of the inlet air with pure oxygen. While this is a standard procedure in high cell density fermentations, the resulting increase in inlet oxygen concentration also influences the readings of the exhaust gas analyzer. As shown in Fig. 2.3, each sudden enrichment step leads to a corresponding spike in the measured O_2 concentration in the exhaust gas.

To account for this effect and ensure accurate interpretation of respiratory activity, a correction factor should be applied to the O_2 readings based on the percentage of oxygen enrichment in the inlet air. By analyzing the corresponding increase in exhaust gas O_2 concentration with rising oxygen enrichment levels, a strong linear correlation was established ($R^2 = 0.99$). This relationship can be used to apply a correction term to the measured O_2 signal, based on the oxygen enrichment percentage, effectively compensating for the influence of inlet air enrichment (Fig. 2.3, O_2 corrected signal). During fermentations, CO_2 levels rise with increasing cell density, while O_2 levels decrease due to elevated cellular oxygen consumption and CO_2 production. This dynamic was not accurately reflected in the raw O_2 sensor signal, but was captured more reliably in the corrected signal. Notably, the skew introduced by O_2 enrichment significantly affects calculations of oxygen uptake rate (OUR) and respiratory quotient (RQ), potentially compromising batch performance.

2.4. Biomass sensor signal anomaly detection and removal

Due to the highly dynamic conditions in bioreactors during high cell density fermentations, *in situ* sensor signals can experience reduced quality or exhibit unexpected anomalies. As high-quality data is critical for effective data-driven modeling, these issues must be identified and addressed – ideally in real time – especially if the sensor signal is used in fermentation control.

Analysis of dielectric spectroscopy biomass sensor data revealed a concerning pattern: the permittivity signal displayed sudden, unexplained spikes and level shifts during the methanol induction phase (Fig. 2.4 A) [45, 60]. A similar, though less pronounced, trend was also observed in the turbidity sensor data. These anomalies – a largely unexplored challenge in bioprocessing – can significantly impact process performance, particularly when real-time sensor data is used for substrate feed control. To address this, a robust algorithm was developed to detect and correct signal anomalies in real-time, leveraging the previously collected experimental dataset [60].

Simple filtering methods, such as moving average smoothing, often fall short when addressing the complexity and variability inherent in real bioprocess data. To overcome these limitations, a structured three-step approach was developed: (1) signal preprocessing to reduce noise and remove contextual dependencies; (2) anomaly detection using threshold-based criteria; and (3) anomaly correction and validation.

1. **Signal preprocessing**
   To optimize real-time smoothing of the permittivity signal in *P. pastoris* fermentations, multiple filtering techniques were evaluated against a manually curated, noise-free reference signal. Performance was assessed using normalized root mean square error (NRMSE) and signal delay analysis, enabling the identification of filtering methods and parameters that achieved effective noise reduction without compromising signal fidelity or introducing excessive lag.
2. **Anomaly detection**
   A double rolling aggregate (DRA) transformer was applied to highlight signal deviations while simultaneously linearizing the signal and removing context-dependency. To determine suitable thresholds, both static and dynamic methods were evaluated, including manual threshold sweeps and statistical approaches such as the 3-sigma rule, median absolute deviation (MAD), and interquartile range (IQR), each tested across multiple window sizes. To identify the most robust detection method, each strategy was benchmarked against manually annotated signal anomalies by computing the F1-score.
3. **Signal correction and validation**
   Upon detection of an anomaly, the affected data point is corrected by replacing it with the mean of the 15 preceding values. A subsequent 15-minute validation window is applied to adjust the signal baseline, using the difference between pre- and post-anomaly levels as a dynamic correction term to ensure continuity and minimize signal drift.

The smoothing performance of various filtering methods was evaluated using NRMSE and signal delay, the latter calculated via cross-correlation between raw and filtered signals in real-time process simulations. Among the tested approaches, the Gaussian filter with a window size of 70 offered the best trade-off between noise reduction and responsiveness, achieving an average NRMSE of 4.56 ± 1.40 % (33 % reduction in signal noise) and an acceptable signal delay of 6.4 minutes (Fig. 2.4 A) [60]. Other methods either resulted in higher prediction errors, introduced longer delays, or – such as robust local regression – were computationally too demanding for real-time application.

The best anomaly detection performance was achieved using a static thresholding approach, which produced an F1-score of 0.79 (with window sizes $w_1 = 1$, $w_2 = 15$, and a threshold of 1.06 $pF \cdot cm^{-1}$). This method not only demonstrated strong detection capability but also required minimal computational resources, making it well-suited for real-time implementation. In contrast, dynamic thresholding methods underperformed due to their reliance on historical signal values, demonstrating F1-scores in the range of 0.31–0.47. As a result, sudden increases in signal volatility were not promptly reflected in the threshold, causing it to lag and remain too

low, leading to false positives during signal fluctuations (Fig. 2.4 B) [60]. It is likely that incorporating a predictive criterion is essential for these methods to perform at a comparable level.

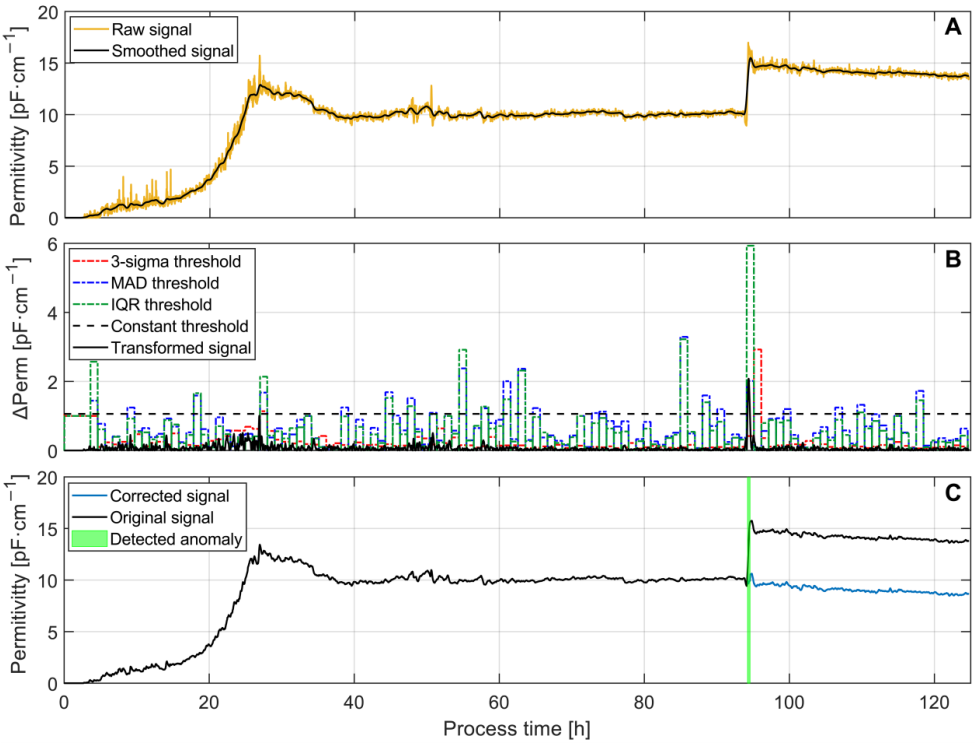


Fig. 2.4. Visual overview of the algorithm's performance: A – signal enhancement through preprocessing, B – comparison of anomaly detection method thresholds applied to the DRA-transformed permittivity signal, and C – performance in a real-time process simulation.

In the final correction step, identified anomalies in the permittivity signal are replaced with the mean of the 15 preceding values to prevent sharp spikes from distorting potential substrate feed rate calculations. Without correction, such artifacts could be misinterpreted as abrupt increases in viable biomass, prompting excessive feed rates that risk process instability or even batch failure. Each detected anomaly is followed by a 15-minute validation window during which corrections continue. If a typical signal spike is observed – marked by a rapid rise and a corresponding drop – both events are treated as a single anomaly to avoid redundant corrective actions, since the signal usually returns to baseline. The performance of the signal anomaly detection and removal algorithm is demonstrated in a fermentation simulation (Fig. 2.4 C).

The proposed three-step workflow was successfully applied to recombinant *P. pastoris* fermentation simulations, yielding accurate and stable sensor output despite disturbances. Using a static threshold of 1.06 pF·cm^{-1} and a DRA transformer (window sizes $w_1 = 1$, $w_2 = 15$), the approach achieved an F1-score of 0.79 (essentially a 79 % accuracy), demonstrating strong

anomaly detection performance. Its simplicity, low computational overhead, and adaptability make it well-suited for real-time monitoring and control across a wide range of bioprocesses and sensor signal types [60]. Nonetheless, while such signal processing significantly improves control reliability, it is equally important to identify and resolve the underlying causes of sensor anomalies to ensure long-term measurement integrity and process robustness.

3. *P. pastoris* Fermentation Modeling

Publications:

- **Bolmanis, E**.; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. pastoris* GS115 Mut+ Bioreactor Process under Controlled Residual Methanol Concentration. *Bioprocess Biosyst. Eng.* **2022**, *45*, 1447–1463 [4].
- **Bolmanis, E.**; Bogans, J.; Akopjana, I.; Suleiko, A.; Kazaka, T.; Kazaks, A. Production and Purification of Soy Leghemoglobin from *Pichia pastoris* Cultivated in Different Expression Media. *Processes* **2023**, *11*, 3215 [56].
- **Bolmanis, E.**; Galvanauskas, V.; Grigs, O.; Vanags, J.; Kazaks, A. Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development. *Fermentation* **2025**, *11*, 411 [57].

Process modeling is essential for understanding, optimizing, and controlling *P. pastoris* fermentations. Modeling approaches range from mechanistic models, which describe biological processes using biochemical and physiological principles, to data-driven models, such as statistical and machine learning techniques that infer empirical relationships from process data. Mechanistic models offer interpretability and insight into system behavior but require extensive domain knowledge and detailed parameterization [23]. Data-driven models, on the other hand, handle complex, nonlinear dynamics with minimal prior knowledge, though they depend heavily on data quality and often lack transparency [48]. To address these limitations, hybrid models that integrate mechanistic understanding with data-driven flexibility are increasingly adopted [29, 49, 50]. Effective modeling supports process development, scale-up, and real-time control, ultimately enhancing productivity, product quality, and reproducibility in industrial fermentations.

3.1. Mechanistic modeling

A mechanistic bioreactor model was developed using a dataset of *P. pastoris* fermentations producing HBcAg. Modeling results for the glycerol phase are not shown here, as the methanol induction phase is more critical for recombinant protein production; for a complete overview of the modeling results, including the glycerol phase, refer to the original article [4].

A macrokinetic process model was developed to capture the intracellular energy and metabolite balances during methanol metabolism, based on the formulation by Ren et al. [61]:

$$\begin{bmatrix} \frac{3}{1-\varphi} & \frac{3}{1-\varphi}-K_1 & 0 & 0 \\ \frac{5\varphi+1}{1-\varphi} & \frac{6\varphi}{1-\varphi}-K_1-4K_2 & 5 & -2 \\ -1 & -3K_1-K_2-\frac{1}{Y_{\mathrm{ATP}}} & 1 & 2P/O \\ 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} q_G \\ \mu \\ q_{Ac} \\ q_{O_2} \end{bmatrix} = \begin{bmatrix} q_{\mathrm{MeOH}} \\ 0 \\ mATP_{\mathrm{MeOH}} \\ 0 \end{bmatrix}, \tag{3.1}$$

where φ is the fraction of formaldehyde oxidized to formate, K_1 and K_2 are model parameters, Y_{ATP} is the ATP yield coefficient [g·mol^{-1}], *P*/*O* is the oxidative phosphorylation effectiveness coefficient, q_G is the specific glycolysis rate [mol·g^{-1}·h^{-1}], μ is the specific growth rate [h^{-1}], q_{Ac} is the specific acetyl-CoA production rate [mol·g^{-1}·h^{-1}], and q_{O2} denotes the specific oxygen uptake rate [mol·g^{-1}·h^{-1}].

The biomass specific growth rate (μ) was obtained by solving the system of equations using linear algebra methods. The specific methanol uptake rate (q_{MeOH}) was computed using a non-monotonically increasing function originally proposed by Jackson & Edwards [62]:

$$q_{\mathrm{MeOH}} = \frac{q_{\max} \times S}{K_S + S + (S^2/K_i)} \times M, \tag{3.2}$$

where q_{max} is the maximum specific methanol uptake rate [g·g^{-1}·h^{-1}], *S* is the residual methanol concentration in culture media [g·L^{-1}], K_S is the methanol saturation constant [g·L^{-1}], K_i is the methanol inhibition constant [g·L^{-1}], and *M* is the molar mass of methanol [g·mol^{-1}].

Product accumulation (q_P) was described using the Luedeking-Piret model, which relates product formation to both growth-associated and non-growth-associated mechanisms:

$$q_P = \mu \times Y_{PX}, \tag{3.3}$$

where μ is the specific cell growth rate [h^{-1}], and Y_{PX} is the specific product yield coefficient [g·g^{-1}].

The calculated growth, substrate uptake, and product formation rates were incorporated into bioreactor mass balance differential equations to simulate the dynamic behavior of biomass (*X*), residual methanol (*S*), culture volume (*V*), and product accumulation (*P*) throughout the fermentation process. Optimal model parameters were determined based on values reported in the scientific literature and refined through a parameter tuning procedure using fermentation data. The developed model successfully reproduced the dynamic behavior of key process variables during the methanol induction phase of *P. pastoris* fermentation [4]. The corresponding simulation results are shown in Fig. 3.1.

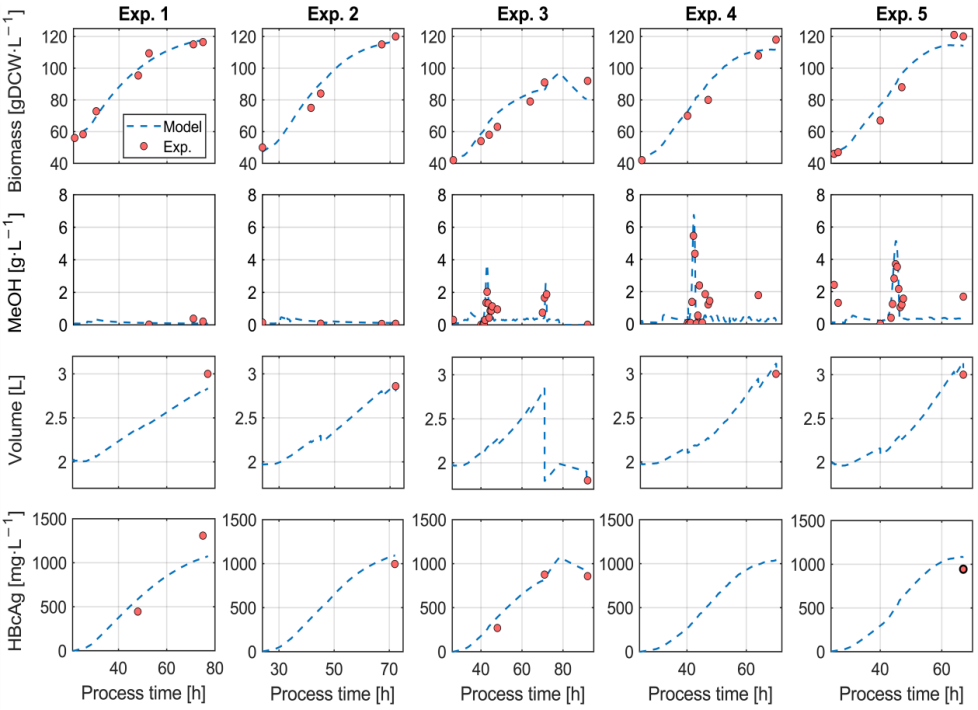


Fig. 3.1. Mechanistic modeling results, showing cell biomass, methanol concentration, culture volume and product concentration dynamics from five *P. pastoris* fermentations producing HBcAg.

The model achieved good accuracy for biomass (5.05 % NRMSE), reactor volume (5.65 %), and product concentration (8.57 %). However, it frequently underestimated residual methanol concentration after cellular adaptation, leading to reduced accuracy (20.83 %) [4]. Despite this, methanol accumulation during increased substrate feed rates was accurately captured. While further refinement is needed, this model represents a rare effort to simulate residual methanol dynamics in *P. pastoris* fermentations.

Finally, a sensitivity analysis was conducted to evaluate the model's robustness and to identify the most influential parameters. The results indicated that certain parameters exhibited high sensitivity, where even minor deviations significantly affected model accuracy. The greatest sensitivity was observed during the glycerol growth phase, where small errors tended to accumulate and amplify throughout the fermentation, leading to a compounding effect on model predictions [4]. These findings suggest that the model is best suited for application during the methanol induction phase, where it demonstrates greater stability and predictive reliability.

3.2. Data-driven modeling

Data-driven modeling leverages historical process data to uncover empirical relationships and predict system behavior without relying on detailed mechanistic knowledge [63]. This approach is particularly valuable in complex bioprocesses like *P. pastoris* fermentations, where

nonlinear dynamics and limited process understanding can hinder purely mechanistic modeling. By employing machine learning techniques, data-driven models can capture intricate input–output relationships, support real-time monitoring (soft sensors), and enhance predictive accuracy – provided that high-quality, representative datasets are available [56, 63].

An ANN-based soft sensor for cell biomass estimation was developed using only standard bioreactor measurements from two fermentation runs [56]. Predictor inputs included stirrer speed (RPM), DO (%), O_2 enrichment (%), pumped base, glycerol, methanol feeds (mL), and reactor volume (L), while real-time *in situ* turbidity biomass probe data served as the training target. The dataset was divided into 70 % for training, 15 % for testing, and 15 % for validation. To reduce abrupt fluctuations and noise in the developed biomass soft sensor, a Savitzky-Golay filter with a first-order polynomial and a frame length of 29 was applied. A two-layer feedforward neural network with 10 sigmoid-activated hidden neurons and a single linear output neuron was used for model training, as illustrated in Fig. 3.2.

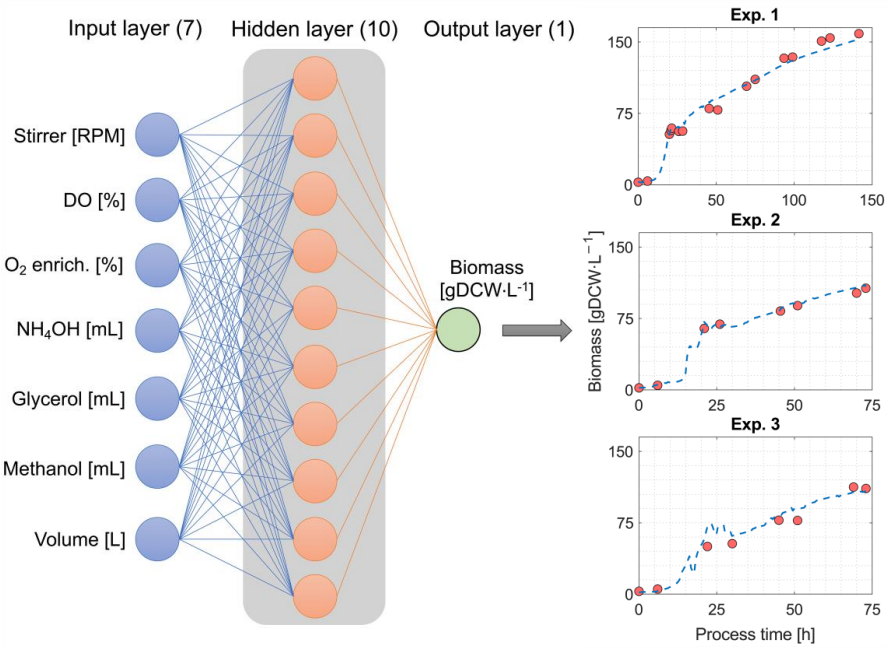


Fig. 3.2. Structure and the biomass modeling results of the developed ANN-based biomass soft sensor.

The developed data-driven model accurately captured the cell biomass dynamics in the evaluated cultivations, demonstrating a good fit across nearly all experiments (Fig. 3.2). The overall precision was estimated at 3.72 % NRMSE on the training dataset; however, due to a methodological oversight, its ability to generalize to unseen data was not evaluated. Notably, the soft sensor relies solely on standard bioreactor measurements and does not incorporate additional signals such as CO_2 concentrations [46, 64]. While this may slightly limit accuracy, it eliminates the need for extra instrumentation. Overall, the achieved performance can be

considered sufficient for application in recombinant *P. pastoris* cultivations as a complementary measurement to experimental sampling.

3.3. Hybrid modeling

Bioprocess monitoring and control continue to face significant challenges due to the complexity and nonlinearity of biological systems. In response, hybrid modeling approaches – combining the strengths of mechanistic and data-driven models – have emerged as powerful tools. These models play a key role in the digital transformation of biomanufacturing, particularly as machine learning grows in prominence for its ability to capture process dynamics without requiring complete system knowledge [29, 49]. Selecting an optimal neural network architecture is critical for achieving high model accuracy and generalization in deep learning applications.

A universal hybrid process model for recombinant *P. pastoris* fermentations was developed by leveraging a historical dataset comprising 17 fermentation runs conducted over the course of the Thesis research [57]. The general structure of the hybrid process model is illustrated in Fig. 3.3.

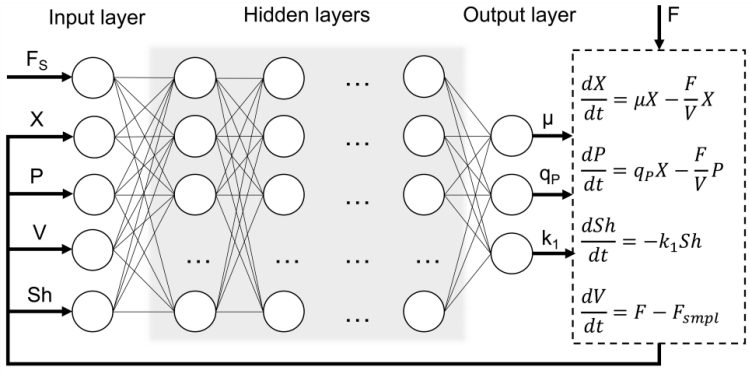


Fig. 3.3. General structure of the developed hybrid process model.

The model's input layer consists of five key variables: substrate (methanol) feed rate (F_S, mL·min^{-1}), dry cell biomass concentration (X, gDCW·L^{-1}), product concentration (P, mg·L^{-1}), culture volume (V, L), and an empirical shock factor (Sh). The shock factor, initialized as $Sh(0) = 1$, captures the cumulative toxic effect of methanol feeding on the cells and is treated as an unmeasured internal state. The model generates three outputs: specific growth rate (μ), specific production rate (qp), and the rate of change of the shock factor (k_1). These outputs are then passed to a parametric component formulated as a system of ordinary differential equations (ODEs), grounded in material balances and physicochemical assumptions. Within this structure, substrate feed rate (F_S), sampling rate (F_{smpl}) and volumetric flow rate (F) serve as the only external inputs.

A three-step strategy was employed to optimize the hybrid model's hidden layer architecture. First, Bayesian optimization rapidly explored key hyperparameters – such as layer

type (LSTM or fully connected (FC)), depth, activation functions, and node count – by training networks in parallel for 10 epochs with an elevated learning rate and selecting architectures based on validation loss and AICc to balance model fit and complexity. Next, a focused grid search refined these candidates by evaluating the remaining 200 combinations of activation functions (LeakyReLU, ReLU, Tanh, or none), LSTM units (1–5), and fully connected nodes (1–10), from which the top five models were selected. Finally, these top models were fully trained (20 000 iterations), and the best-performing model was selected; the use of a dropout layer was also assessed to enhance robustness and generalization.

Bayesian optimization efficiently identified promising hybrid model architectures by focusing the search on high-performing hyperparameter regions, significantly reducing the number of models evaluated. The top-ranked architectures consistently featured an LSTM layer followed by one FC layer, underscoring the value of sequential feature extraction and nonlinear output mapping. A subsequent exhaustive grid search refined the model selection by evaluating all feasible combinations of hidden units, node counts, and activation functions. Using both validation loss and AICc, five balanced models were selected from the Pareto front, achieving strong predictive accuracy with moderate complexity, thus ensuring generalizability and computational efficiency (Table 3.1).

Table 3.1

Summary of the best-performing network architectures

Hidden units	Nodes	Activation	Validation loss [%]	No. of parameters	AICc
3	5	LeakyReLU	7.28	146	1294
2	10	LeakyReLU	6.37	127	1155
2	9	Tanh	8.14	121	1236
2	**8**	**ReLU**	**4.93**	**115**	**998**
1	9	Tanh	8.27	76	1090

In the final optimization step, the five shortlisted architectures were each trained over 10 independent runs (20 000 iterations) to evaluate robustness and performance consistency under different random initializations (Table 3.1). The top-performing model featured two LSTM units, eight nodes in the FC layer, and a ReLU activation function, achieving the lowest validation loss (4.93 %) and AICc (998), indicating an optimal trade-off between predictive accuracy and simplicity (Fig. 3.4). Introducing dropout layers (0.1–0.5 probability) consistently degraded performance, suggesting that the model was already sufficiently regularized.

As shown in Fig. 3.4, the model effectively captures both biomass and product dynamics throughout the fermentation process. However, it does not fully account for the subtle decline in biomass concentration following methanol induction, which reflects the cellular adaptation to methanol metabolism. This results in a slight overestimation of biomass during the early methanol phase. Due to limited cell growth during the adaptation to methanol, biomass sampling was often omitted; however, such measurements are essential for the model to

accurately learn and reflect the growth stagnation characteristic of this phase. Despite the limited availability of experimental product measurements, the model also demonstrates strong performance in estimating product concentration.

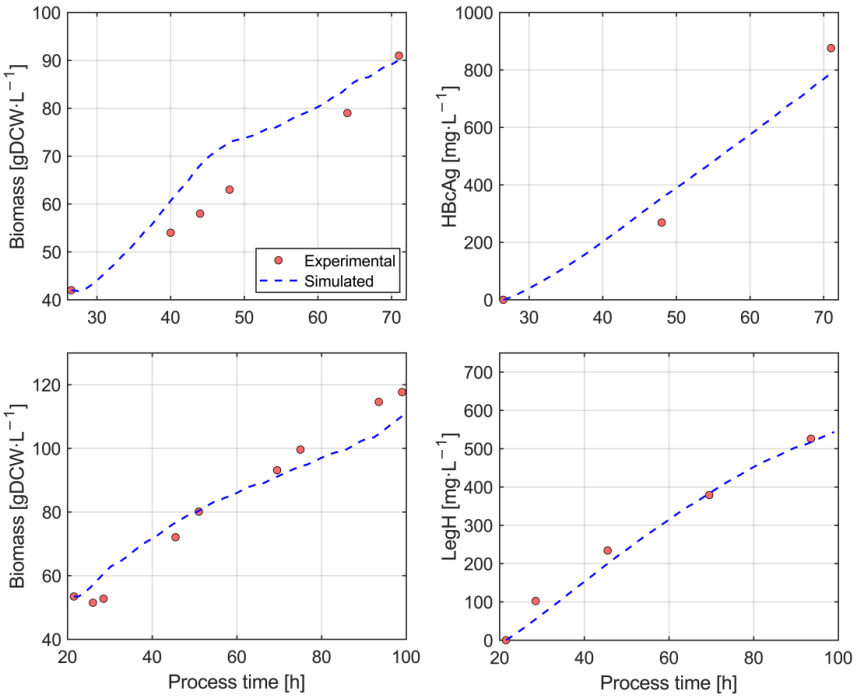


Fig. 3.4. Cell biomass and product concentrations estimated using the selected hybrid process model for HBcAg (upper row) and LegH (lower row).

3.4. Model performance comparison

To enable a fair and meaningful comparison between mechanistic, data-driven, and hybrid modeling approaches, all three model types were evaluated using the same dataset. While earlier sections demonstrated each model's capabilities individually – often using different subsets of data – direct comparisons require consistent training and testing conditions. Therefore, a unified dataset (20 experiments) was selected to ensure that all models operated on equivalent input-output information. This standardized benchmark enables an objective assessment of predictive accuracy, generalization performance, and model complexity across the different modeling paradigms.

Given that three distinct recombinant products were investigated during this Thesis research, product-specific parameters were optimized independently within each corresponding experimental subset to ensure accurate and unbiased product concentration estimation. Finally, to fully leverage the available data, 4-fold cross-validation was applied by randomly partitioning the dataset into four subsets, each with separate training and testing partitions. Each testing fold included two experiments producing HBcAg, two producing LegH, and one

producing Qβ, ensuring a balanced and representative evaluation across all product types. Model performance was reported as the mean ± standard deviation of the testing results across all cross-validation folds. The full experimental dataset is summarized in Table 3.2.

Table 3.2

Experimental dataset used for model comparison

Exp No.	Strain, product	Induction time [h]	Biomass [gDCW·L^{-1}]	Feed rate [mL·min^{-1}]	V_{end} [L]
1	GS115, HBcAg	65	37.5–101.6	0.12–0.78	2.85
2	GS115, HBcAg	45	40.6–113.5	0.12–1.00	3.09
3	GS115, HBcAg	43	41.2–120.1	0.12–0.98	3.13
4	GS115, HBcAg	50	59.2–120.1	0.12–0.36	2.54
5	GS115, HBcAg	51	41.4–96.6	0.12–0.36	2.87
6	GS115, HBcAg	48	49.1–120.0	0.12–0.50	2.88
7	GS115, HBcAg	43	53.7–101.5	0.12–0.36	2.74
8	GS115, HBcAg	54	44.1–84.0	0.12–0.56	2.75
9	X-33, LegH	65	55.4–123.2	0.12–0.36	2.57
10	X-33, LegH	46	49.5–95.4	0.12–0.60	2.98
11	X-33, LegH	65	48.9–111.2	0.12–0.36	2.85
12	X-33, LegH	50	45.3–101.3	0.12–0.36	2.61
13	X-33, LegH	45	52.9–103.1	0.12–0.36	2.55
14	X-33, LegH	46	45.1–101.3	0.12–0.36	2.52
15	X-33, LegH	65	51.0–101.7	0.12–0.36	2.66
16	X-33, LegH	46	50.6–92.4	0.12–0.60	3.00
17	X-33, Qβ	65	52.5–117.6	0.12–0.49	3.23
18	X-33, Qβ	48	49.3–117.2	0.12–1.00	3.40
19	X-33, Qβ	55	50.1–107.7	0.12–0.36	2.84
20	X-33, Qβ	52	52.9–112.6	0.12–0.87	3.45

Unsurprisingly, the modeling performance varied significantly across the different model types. The training and testing losses for each model type are summarized in Table 3.3.

Table 3.3

Model performance comparison, average precision (NRMSE)

Model	Metric	Biomass [%]	Product [%]	Average [%]
Mechanistic	Train	10.9 ± 2.0	19.6 ± 2.7	15.3 ± 2.2
	Test	13.1 ± 4.4	65.4 ± 33.8	**39.2 ± 18.9**
Data-driven	Train	7.9 ± 1.3	7.7 ± 2.2	7.8 ± 0.5
	Test	14.8 ± 2.3	41.7 ± 9.2	**28.2 ± 5.4**
Hybrid	Train	9.1 ± 1.9	11.7 ± 3.1	10.4 ± 2.5
	Test	9.1 ± 2.0	13.1 ± 5.5	**11.1 ± 2.6**

Table 3.3 illustrates several key differences between the three model types. All models demonstrated reasonable accuracy in predicting biomass concentration during training, with NRMSEs ranging from 7.9 % to 10.9 %. However, the hybrid model achieved the lowest test error for biomass prediction (9.1 %), indicating better generalization to unseen data compared to the mechanistic (13.1 %) and data-driven (14.8 %) models. This suggests that integrating mechanistic insights with data-driven learning offers improved robustness, especially in capturing biomass dynamics under varying process conditions.

In terms of product concentration estimation, the differences between model types became more pronounced. The hybrid model significantly outperformed the others with a test error of 13.1 %, closely aligned with its training performance (11.7 %), suggesting strong generalization and reliable learning of production dynamics. In contrast, the mechanistic model exhibited a substantial performance gap, with a relatively high training error (19.6 %) and a very large test error (65.4 %), reflecting poor adaptability to variability in experimental data and limited ability to capture recombinant protein expression patterns. The data-driven model performed better than the mechanistic one (41.7 %), but still lagged behind the hybrid approach, potentially due to its lack of process-specific prior knowledge.

The hybrid model demonstrated the best overall performance, with an average training error of 10.4 % and a test error of 11.1 %, indicating both accurate fitting and strong generalization. The mechanistic model had limited flexibility (39.2 % test error), while the data-driven model showed signs of overfitting, with low training error (7.8 %) but higher test error (28.2 %). These results highlight the hybrid model's ability to effectively combine mechanistic knowledge with data-driven learning, making it the most robust and reliable choice for modeling complex bioprocesses like recombinant *P. pastoris* fermentations.

Figure 3.5 illustrates the model performance in predicting biomass and product concentrations for a representative experiment from each recombinant protein producer. The hybrid model consistently delivers the most accurate results, effectively capturing both biomass growth and product accumulation trends without direct fitting to the data.

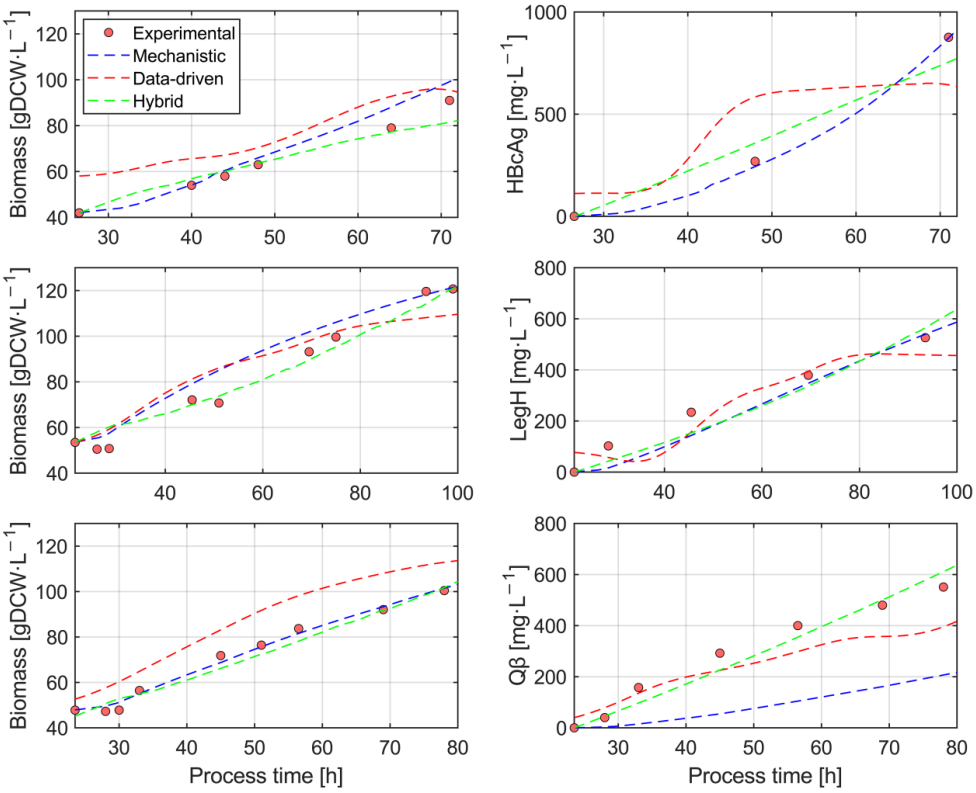


Fig. 3.5. Comparison of model performance in estimating biomass and product concentrations for one experiment from each recombinant protein producer.

To put these results into perspective, several factors must be considered – notably, the quality, diversity, and completeness of the dataset. The data encompassed three different *P. pastoris* constructs from two distinct strains, GS115 and wild-type X-33. While growth and production dynamics were generally similar, biological variability is expected. Furthermore, the experiments were conducted over several years, and although standardized protocols were followed, minor operational or technical inconsistencies may have occurred. Data scarcity – particularly for recombinant protein concentration – also played a significant role. Product quantification involved time- and labor-intensive purification procedures, often requiring multiple chromatography runs, which limited the frequency of measurements. As a result, the models sometimes struggled to accurately capture product dynamics. Collectively, these factors may contribute to increased modeling error.

Despite these challenges, the hybrid model demonstrated strong performance – accurately capturing both biomass and product dynamics across diverse conditions. Its ability to generalize was supported by the close alignment between training and testing losses, indicating that the model did not overfit and maintained predictive robustness on unseen data.

3.5. Transfer learning

Transfer learning is a machine learning technique that leverages knowledge gained from training a model on one task or dataset to improve performance on a related but distinct task or dataset [65]. In the context of hybrid bioprocess modeling, it enables the adaptation of a pretrained model to new products, strains, or process variants – accelerating training, reducing the need for extensive new data, and enhancing predictive accuracy. This is particularly valuable in bioprocess development, where generating high-quality experimental data is both time-consuming and labor-intensive, making transfer learning a practical strategy for improving model scalability, efficiency, and applicability across diverse bioproduction scenarios [57, 65].

In this study, transfer learning was applied to adapt a historical hybrid model – trained on 17 HBcAg and LegH fermentation experiments – to the Qβ production process using only two experiments (Exps. 17 and 18 from Table 3.2). The key idea was to use the historical process model as the initialization for the Qβ model training. By updating the LSTM layer weights at a reduced learning rate (0–1.0 relative to the rest of the network), the model retained previously learned temporal dynamics while still adapting to the new Qβ dataset. This strategy enabled the model to leverage general *P. pastoris* fermentation patterns captured from the historical data, thereby improving generalization and mitigating the limitations of training on such a small dataset.

First, the effect of adapting the historical hybrid model to the Qβ dataset was evaluated by comparing its test loss (Exp. 19 in Table 3.2) and variability against a baseline model trained from scratch with randomly initialized weights. The results show that initializing Qβ model training with the pre-trained historical process model, rather than training from scratch, led to markedly lower test loss values (mean 5.31 % vs. 9.90 %) with substantially reduced variability (std ± 0.34 vs. ± 7.38) (Fig. 3.6 A). In addition, training converged more quickly, requiring fewer iterations on average (8 820 vs. 13 070) (Fig. 3.6 B). These findings suggest that general process models, trained on related datasets, can serve as effective starting points for new hybrid model development in *P. pastoris* fermentations [57, 64]. By retaining prior process knowledge, such models not only reduce test loss but also shorten training time, which is particularly valuable when working with small datasets.

To achieve optimal transfer learning, the LSTM layer's relative training rate was determined through systematic screening. Because the LSTM layer encodes process temporal dynamics, adjusting its training rate allows the model to retain prior knowledge while adapting to a new dataset. The results indicate that training the LSTM layer at 0.6–0.8 of the learning rate of the remaining network yields the best performance, with an average test loss of 4.53 ± 0.20 % across ten repetitions (Fig. 3.6 C). These findings suggest that partially retraining memory layers, rather than fully freezing or fully retraining them, provides an effective balance between preserving prior temporal knowledge and capturing new process-specific features [57]. In this particular case, setting the LSTM learning rate in this range enables efficient transfer learning with minimal variability in test loss.

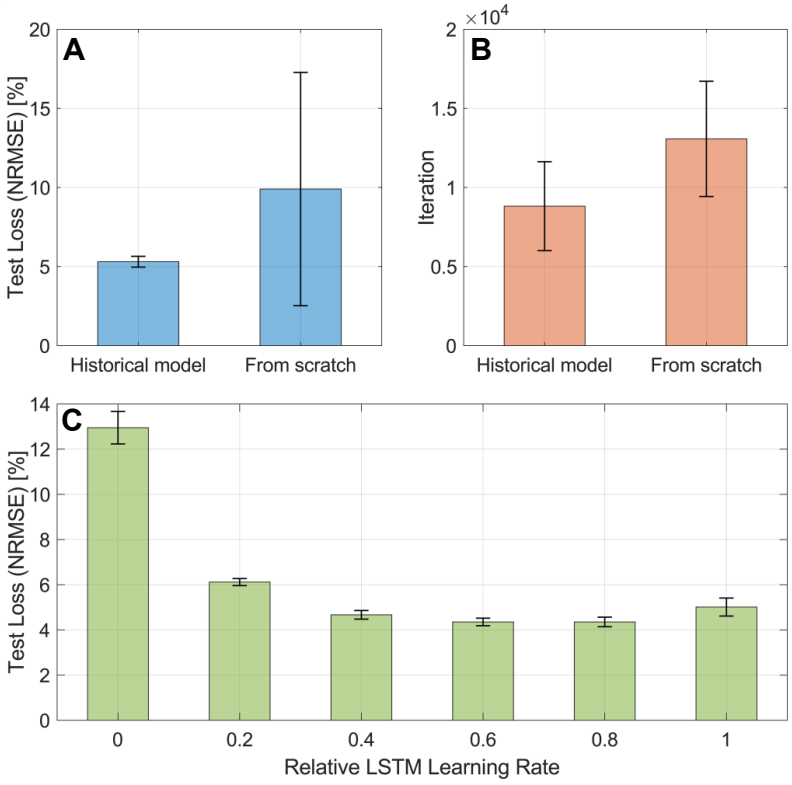


Fig. 3.6. Mean test losses (A) and number of training iterations (B) for the hybrid Qβ process model when adapted from a historical model or trained from scratch, and mean test loss as a function of the LSTM layer relative learning rate (C).

The final hybrid process model, adapted from the historical model with an LSTM relative learning rate of 0.6, exhibited strong predictive performance, accurately capturing the dynamics of key variables throughout the Qβ fermentation process. Overall, it achieved a test NRMSE of 4.35 %, with 3.16 % for biomass concentration and 5.64 % for product concentration, demonstrating reliable estimation of both process parameters (Fig. 3.7).

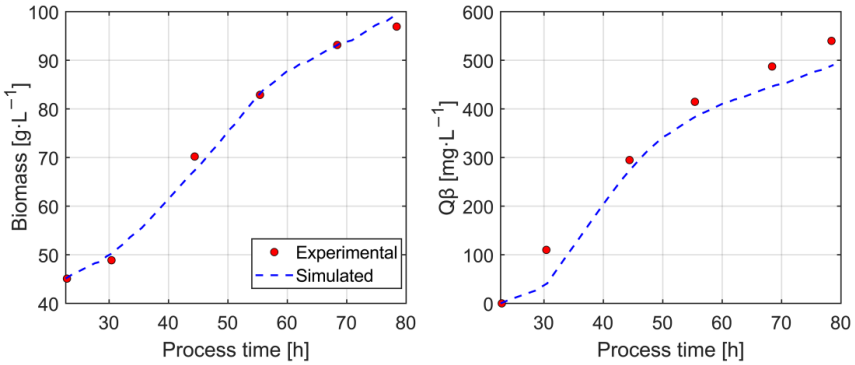


Fig. 3.7. Prediction of the dynamic profiles of cell biomass and product modeled with the trained hybrid process model.

As the biomass growth is often similar in most *P. pastoris* producers, the main source of variation when modeling fermentations across different producers arises from differences in product accumulation dynamics. Consequently, directly transferring a model trained on one producer to another without adaptation can lead to suboptimal predictions, as the learned representation of product kinetics may not generalize. To address this, additional strategies for transfer learning with partial layer freezing can be explored, such as including additional adapter modules, segmenting the hybrid model into separate subnetworks for each output (e.g., one module for biomass, another for product), or fine-tuning product-specific parameters. This approach helps retain generalizable process knowledge, such as conserved biomass growth dynamics, while allowing the model to capture producer-specific features, like strain-dependent product accumulation. This approach enables efficient adaptation to new producers even with limited experimental data, reducing the need to train entirely new models from scratch.

With only two experiments available, the training data are insufficient to fully explore the process parameter space. As a result, the model can perform well only under conditions similar to those encountered during training, while predictions for unobserved regions are likely to be unreliable due to limited extrapolation capability. Training on a new dataset while leveraging the historical process model allows the model to retain valuable knowledge from the more diverse historical dataset, thereby improving performance under process conditions not covered by the Qβ dataset, as demonstrated in this Thesis.

In summary, this study demonstrates that transfer learning using a historical hybrid model is an effective strategy for developing predictive models for new *P. pastoris* fermentation processes with limited experimental data. By initializing the Qβ model with the pre-trained historical model and partially retraining the LSTM layer at an optimized relative learning rate, the approach retained general temporal dynamics while adapting to strain-specific product accumulation. This method not only reduced test loss and variability compared to training from scratch but also accelerated convergence, yielding a final model capable of accurately predicting both biomass and product concentrations. These results highlight the value of leveraging prior knowledge through hybrid transfer learning to improve model generalization and efficiency in small-data bioprocess applications.

4. Fermentation Control

Publications:

- **Bolmanis, E**.; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. pastoris* GS115 Mut+ Bioreactor Process under Controlled Residual Methanol Concentration. *Bioprocess Biosyst. Eng.* **2022**, *45*, 1447–1463 [4].
- **Bolmanis, E.**; Galvanauskas, V.; Grigs, O.; Vanags, J.; Kazaks, A. Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development. *Fermentation* **2025**, *11*, 411 [57].

Substrate feed rate control is a critical aspect of fermentation process management, directly influencing cell growth, product formation, and overall process performance. Precise control of the substrate feed – such as glucose, glycerol, or methanol – ensures that microorganisms receive the optimal nutrient supply to maintain metabolic activity while avoiding substrate inhibition or nutrient limitation. Effective feed rate strategies help maintain desired growth rates, prevent accumulation of toxic by-products, and improve yield and productivity. Various control approaches, ranging from simple feed-forward schemes to advanced model-based and real-time feedback control systems, have been developed to optimize substrate delivery and stabilize fermentation dynamics [18, 33]. Robust substrate feed control is therefore essential for achieving reproducible and scalable bioprocesses.

4.1. Methanol set-point control

Residual methanol concentration control in *P. pastoris* fermentations is crucial for optimizing recombinant protein production while avoiding substrate inhibition or toxicity [12, 66]. Precise control ensures methanol is maintained at levels that support cell metabolism and protein expression without causing stress or excessive accumulation. Common strategies involve using online methanol sensors combined with control algorithms, such as simple PID loops, to dynamically adjust methanol feed rates and maintain optimal residual concentrations throughout the induction phase [33].

To investigate the effect of residual methanol concentration on recombinant protein biosynthesis, a series of experiments was conducted screening HBcAg production at residual methanol levels of 0.01 $g \cdot L^{-1}$, 1.0 $g \cdot L^{-1}$, and 2.0 $g \cdot L^{-1}$ [4]. A PI (proportional integral) controller was developed, using the signal from an off-gas methanol sensor (*BCP-EtOH, BlueSens*) to maintain constant methanol levels. The real-time methanol sensor signal was processed using the moving average filter described in Section 2.2. The control algorithm regulated methanol addition by stabilizing the sensor signal around the desired set-point through a feedback control equation incorporating the PI term:

$$F_{t+\Delta t} = F_t - \frac{V}{(S_0 - S)} \times \frac{dS}{dt} + K_p \left[(\varepsilon_t - \varepsilon_{t-1}) + \frac{\Delta t}{\tau_I} \varepsilon_t \right], \quad (4.1)$$

where F_t is the substrate feed rate at the current time t [mL·min^{-1}], V is the reactor volume [L], S_0 is the substrate concentration in the feed [g·L^{-1}], S is the substrate concentration in the reactor [g·L^{-1}], K_p is the proportional gain parameter [L^2·g^{-1}·h^{-1}], ε is the control error [g·L^{-1}], Δt is the time interval between steps [min], and τ_I is the integral time constant [min].

The integration of the model with the PI control algorithm allowed the parameters K_p and τ_I to remain constant throughout the cultivation. However, control performance was highly sensitive to the selected K_p value. This gain parameter, ranging between 0.02 and 0.05 [L^2·g^{-1}·h^{-1}], was adjusted based on the desired residual methanol concentration. In contrast, the integral time constant τ_I was fixed at 10 minutes across all experiments.

In Experiment 2, no residual methanol control was applied. The methanol feed rate was slightly increased to assess its impact on HBcAg productivity – providing baseline data for comparison. In Experiment 3, residual methanol was controlled at 1.0 [g·L^{-1}] starting at 40 hours using a PI controller, which maintained stability with an average deviation of ± 0.28 [g·L^{-1}] (28 % NRMSE) until 72 hours, when a 1 [L] cell harvest triggered a transient methanol spike. Although the controller initially adapted, methanol levels soon spiked again, indicating possible culture overfeeding.

In Experiment 4, the set-point was increased to 2.0 [g·L^{-1}], but the same control gain (Kp = 0.05 [L^2·g^{-1}·h^{-1}]) was retained, resulting in significant instability with methanol fluctuating between 1.0 and 3.0 [g·L^{-1}], with an average deviation of ± 1.26 [g·L^{-1}] (63 % NRMSE). Reducing the gain to Kp = 0.002 [L^2·g^{-1}·h^{-1}] in Experiment 5 significantly improved control performance. These results, illustrated in Fig. 4.1, demonstrate that the PI controller can maintain residual methanol levels reliably when properly tuned with an average deviation of ± 0.67 [g·L^{-1}]. However, its effectiveness was highly sensitive to the choice of control parameters – particularly the proportional gain. While the controller showed some adaptability to process disturbances, stability was compromised under poorly matched parameters or abrupt culture changes. These findings underscore the need for careful – and potentially adaptive – tuning of control settings to ensure robust methanol regulation across dynamic fermentation phases.

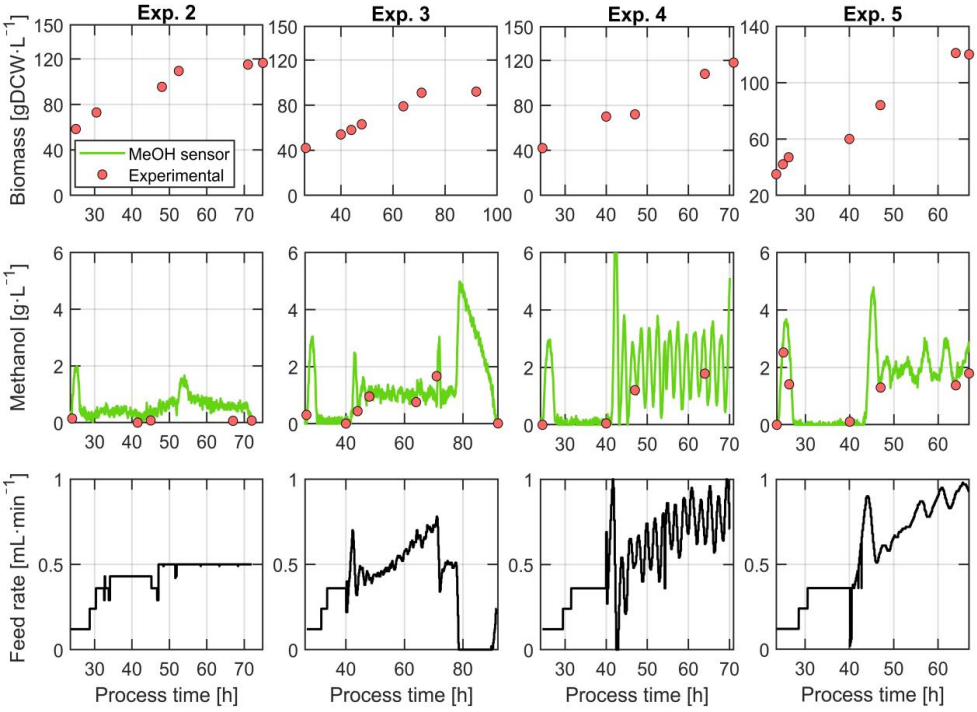


Fig. 4.1. Cell biomass, residual methanol concentration and methanol feed rate dynamics during *P. pastoris* fermentation experiments with residual methanol PI-control.

The PI-based methanol feed control algorithm proved effective for recombinant *P. pastoris* fermentations. When properly tuned, it enabled accurate control of residual methanol concentrations using feedback from the exhaust gas methanol sensor with an average set-point deviation of ± 0.28 – 0.67 [g·L^{-1}], corresponding to an NRMSE of 28–63 %. However, due to its high sensitivity to the tuning of control parameters, implementing an automated tuning procedure is recommended to ensure robust performance across different methanol set-points. Control accuracy could be further enhanced by incorporating an *in situ* methanol sensor probe, providing more direct and responsive measurement.

4.2. Model predictive control

Hybrid MPC systems integrate data-driven models, such as neural networks, with first-principles process knowledge to enable accurate prediction and real-time optimization in complex bioprocesses. These controllers leverage the strengths of both mechanistic understanding and machine learning to handle nonlinear dynamics, unmodeled disturbances, and measurement noise. In biomanufacturing, hybrid MPC is particularly well-suited for controlling fed-batch fermentations, where physiological variability and substrate-product interactions are difficult to capture with purely mechanistic models alone [67, 68].

An MPC framework was developed based on the hybrid process model to regulate cell growth near the maximum specific growth rate [57]. The MPC estimated the optimal substrate

feed rate, $F_S(t)$, required to track a predefined growth trajectory, $\mu_{set}(t)$. As the hybrid model is non-invertible – with F_S as input and μ as output – a numerical optimization using MATLAB's *fminbnd* function was applied at each control step within the bounds $F_S \in [0.36, 1.00]$ mL·min^{-1}:

$$\min_{Fs\in[0.36,1.00]\, mL\times \mathrm{min}^{-1}} \sum_{k=1}^{Np} [\mu(k) - \mu_{\mathrm{set}}(k)]^2, \tag{4.2}$$

subject to the hybrid model dynamics:

$$x(k+1) = f_{\mathrm{hybrid}}(x(k), F_S(k)), \tag{4.3}$$

where $x(k)$ denotes the state vector and $\mu(k)$ is the predicted growth rate at time step k. The control and prediction horizons were set to N_C = 1 hour and N_P = 12 hours, respectively. The hybrid model was simulated with a 1-minute sampling interval to ensure accurate predictions.

To maintain adaptability, the model was retrained ~3 times daily after each sampling using updated biomass measurements, $X_{meas}(t)$. Real-time process data – including substrate feed rate, base, and antifoam addition – were integrated into MATLAB, linked to the bioreactor SCADA system via an OPC server, enabling real-time closed-loop control.

MPC was initiated after methanol adaptation (8–10 h post-induction). The growth rate setpoint $\mu_{set}(t)$ was applied in a step-wise fashion: 0.04 h^{-1} (0–12 h), 0.02 h^{-1} (12–24 h), and 0.01 h^{-1} (24–36 h), balancing productivity and cellular stress.

To evaluate the practical applicability of the hybrid MPC framework, experimental validation was performed by controlling the feed rate in a real fermentation run. This enabled assessment of the system's ability to predict and regulate key bioprocess variables in real time (Fig. 4.2).

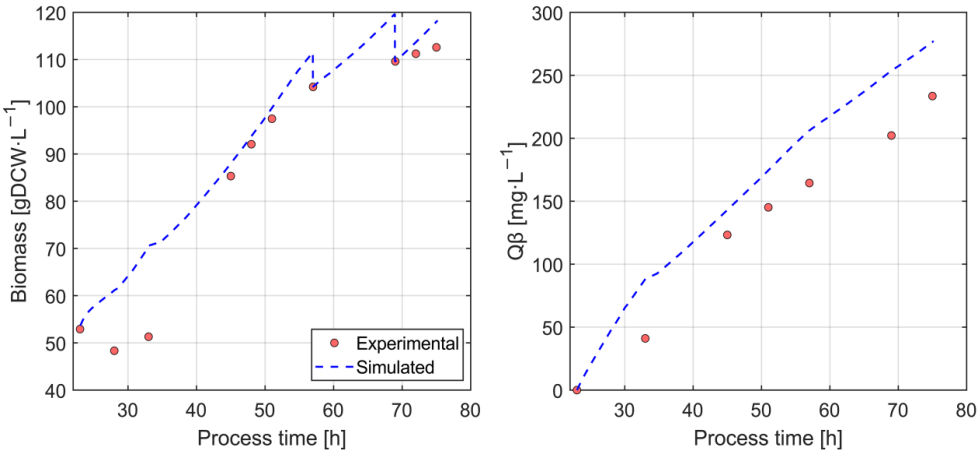


Fig. 4.2. Hybrid MPC predicted vs. experimentally determined biomass and product concentrations during experimental validation.

Biomass prediction accuracy was good; however, the model overestimated growth during the initial 8–12 hours post-induction – when cells adapt to methanol – and again toward the end of fermentation. Manual adjustments were applied during re-training based on offline sampling data, resulting in an overall biomass NRMSE of 6.5 %. Similarly, product concentration was consistently overestimated, yielding a moderate error of 14.6 % [57].

Despite these predictive limitations, the control performance of the hybrid MPC was robust. The system successfully generated feed profiles that maintained the desired specific growth rate, demonstrating effective regulation even in the presence of modeling inaccuracies. The controller tracking performance is illustrated in Fig. 4.3.

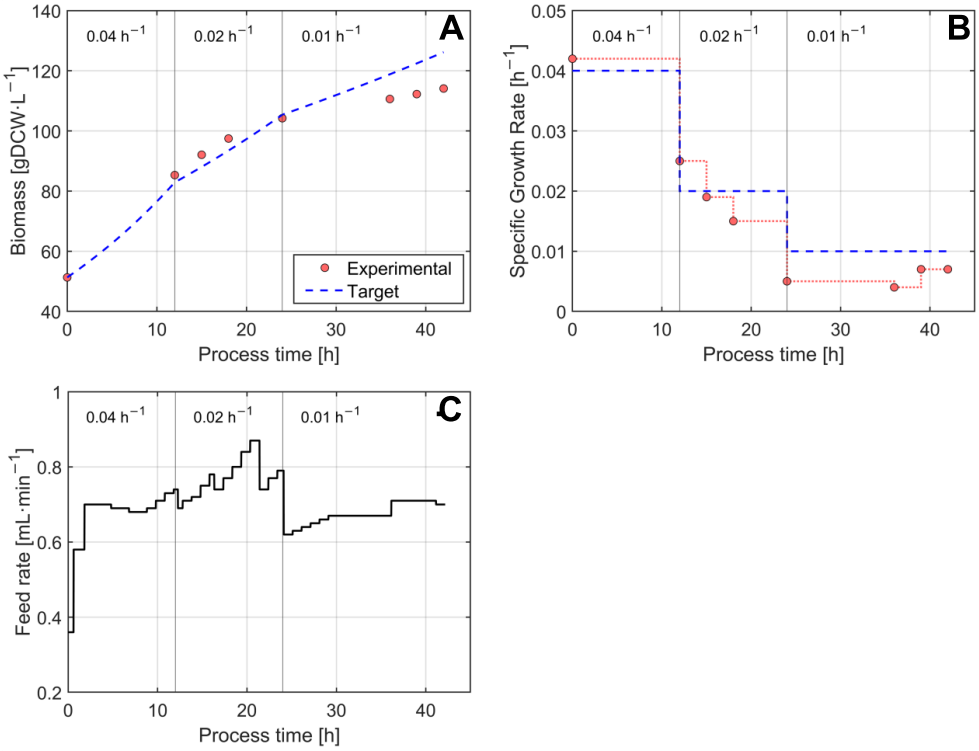


Fig. 4.3. Target vs experimental biomass growth (A), specific growth rate (B) and feed rate plots (C). Vertical lines indicate μ shifts.

The hybrid MPC system demonstrated strong performance in tracking the specific growth rate set-point throughout the fermentation with a tracking error of 10.6 % NRMSE. However, a slight deviation from the target biomass trajectory was observed during the final 12 hours, particularly after the growth rate was reduced to 0.01 h^{-1}. This deviation is likely attributable to the cytotoxic effects of methanol accumulation, which can inhibit cellular metabolism and biomass formation during the late fermentation phase. This trend is also reflected in the growth rate tracking plot, where the measured specific growth rate fell below the target in this period.

Performance of the MPC system could be further improved by conducting a more comprehensive exploration of the process design space. Systematic screening of a broader range of operating conditions – including feed rate trajectories, induction timings, and specific growth rate set-points – would help identify optimal control strategies. Such efforts would enhance the controller's ability to manage process variability and disturbances, thereby improving its robustness and adaptability across different fermentation scenarios. Additionally, a more comprehensive process model would support extended MPC applications, such as estimating optimal feed rates to maximize product yield.

Overall, despite minor discrepancies toward the end, the control system maintained accurate growth regulation for the majority of the fermentation, highlighting its robustness, reliability and potential application in controlling recombinant *P. pastoris* fermentations.

CONCLUSIONS

1. Real-time bioprocess sensor signal pre-processing significantly improves signal quality: methanol sensor deviation was reduced by 63 % and dielectric spectroscopy noise by 33 %. Additionally, the permittivity anomaly detection and removal algorithm achieved 79 % accuracy, supporting reliable sensor use in monitoring and control applications.
2. In modeling *P. pastoris* fermentation processes, hybrid models consistently outperformed mechanistic and data-driven approaches, achieving the lowest test NRMSE (11.1 ± 2.6 %) versus 39.2 ± 18.9 % and 28.2 ± 5.4 %, respectively.
3. Updating the historical hybrid process model with the Qβ dataset at a 0.6 relative learning rate outperformed training from scratch, lowering mean test loss and deviation (from 9.90 ± 7.38 % to 5.31 ± 0.34 %) in fewer iterations (8820 vs. 13 070) and achieving a test NRMSE of 4.35 %, enabling accurate predictions through transfer learning with limited data from just three experimental runs.
4. The PI-based feed rate controller achieved moderate methanol set-point control, with estimated NRMSE values of 28 % and 63 % at 1 [$g \cdot L^{-1}$] and 2 [$g \cdot L^{-1}$], respectively. Its tuning sensitivity emphasizes the need for automated parameter adjustment and direct *in situ* sensing to improve performance by enhancing sensor response and signal quality under dynamic fermentation conditions.
5. The hybrid MPC demonstrated robust control in *P. pastoris* fermentation process, maintaining the target specific growth rate with a 10.6 % NRMSE tracking error despite modeling inaccuracies. As MPC performance strongly depends on the underlying model, these findings highlight the need for high-quality models is crucial for successful MPC implementation in fermentation control and support further investigation of hybrid MPC for *P. pastoris* bioprocesses.

REFERENCES

1. Jayakrishnan, A.; Wan Rosli, W. R.; Tahir, A. R. M.; Razak, F. S. A.; Kee, P. E.; Ng, H. S.; Chew, Y. L.; Lee, S. K.; Ramasamy, M.; Tan, C. S.; et al. Evolving Paradigms of Recombinant Protein Production in Pharmaceutical Industry: A Rigorous Review. *Sci* **2024**, *6*.
2. Tripathi, N. K.; Shrivastava, A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Front. Bioeng. Biotechnol.* **2019**, *7*, doi:10.3389/fbioe.2019.00420.
3. Mordor Intelligence. Biologics Market Size & Share Analysis – Growth Trends and Forecast (2026–2031). **2025**.
4. Bolmanis, E.; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. Pastoris* GS115 Mut+ Bioreactor Process under Controlled Residual Methanol Concentration. *Bioprocess Biosyst. Eng.* **2022**, 45, 1447–1463, doi:10.1007/s00449-022-02754-4.
5. Letourneur, O.; Watelet, B. HBcAg Expression and Diagnostic and Therapeutic Uses. EP Patent No. EP1256804A1, **2001**.
6. Freivalds, J.; Dislers, A.; Ose, V.; Skrastina, D.; Cielens, I.; Pumpens, P.; Sasnauskas, K.; Kazaks, A. Assembly of Bacteriophage Qβ Virus-like Particles in Yeast *Saccharomyces Cerevisiae* and *Pichia Pastoris*. *J. Biotechnol.* **2006**, *123*, 297–303, doi:10.1016/j.jbiotec.2005.11.013.
7. Spohn, G.; Keller, I.; Beck, M.; Grest, P.; Jennings, G. T.; Bachmann, M. F. Active Immunization with IL-1 Displayed on Virus-like Particles Protects from Autoimmune Arthritis. *Eur. J. Immunol.* **2008**, *38*, 877–887, doi:10.1002/eji.200737989.
8. Vuitika, L.; Côrtes, N.; Malaquias, V. B.; Silva, J. D. Q.; Lira, A.; Prates-Syed, W. A.; Schimke, L. F.; Luz, D.; Durães-Carvalho, R.; Balan, A.; et al. A Self-Adjuvanted VLPs-Based Covid-19 Vaccine Proven Versatile, Safe, and Highly Protective. *Sci. Rep.* **2024**, *14*, doi:10.1038/s41598-024-76163-w.
9. Fraser, R. Z.; Shitut, M.; Agrawal, P.; Mendes, O.; Klapholz, S. Safety Evaluation of Soy Leghemoglobin Protein Preparation Derived from *Pichia Pastoris*, Intended for Use as a Flavor Catalyst in Plant-Based Meat. *Int. J. Toxicol.* **2018**, *37*, 241–262, doi:10.1177/1091581818766318.
10. Reyes, T. F.; Chen, Y.; Fraser, R. Z.; Chan, T.; Li, X. Assessment of the Potential Allergenicity and Toxicity of *Pichia* Proteins in a Novel Leghemoglobin Preparation. *Regulatory Toxicology and Pharmacology* **2021**, *119*, 104817, doi:10.1016/j.yrtph.2020.104817.
11. Cregg, J. M.; Cereghino, J. L.; Shi, J.; Higgins, D. R. Recombinant Protein Expression in *Pichia Pastoris*. *Mol. Biotechnol.* **2000**, *16*, 23–52, doi:10.1385/MB:16:1:23.
12. Potvin, G.; Ahmad, A.; Zhang, Z. Bioprocess Engineering Aspects of Heterologous Protein Production in *Pichia Pastoris*: A Review. *Biochem. Eng. J.* **2012**, *64*, 91–105, doi:10.1016/j.bej.2010.07.017.

13. Yang, Z.; Zhang, Z. Engineering Strategies for Enhanced Production of Protein and Bio-Products in *Pichia Pastoris*: A Review. *Biotechnol. Adv.* **2018**, *36*, 182–195, doi:10.1016/j.biotechadv.2017.11.002.
14. Karbalaei, M.; Rezaee, S. A.; Farsiani, H. Pichia Pastoris: A Highly Successful Expression System for Optimal Synthesis of Heterologous Proteins. *J. Cell. Physiol.* **2020**, *235*, 5867–5881, doi:10.1002/jcp.29583.
15. Macauley-Patrick, S.; Fazenda, M. L.; McNeil, B.; Harvey, L. M. Heterologous Protein Production Using the *Pichia Pastoris* Expression System. *Yeast* **2005**, *22*, 249–270, doi:10.1002/yea.1208.
16. U.S. Food and Drug Administration *GRAS Notice No. 204. Phospholipase C Enzyme Preparation from Pichia Pastoris Expressing a Heterologous Phospholipase C Gene*; **2006**.
17. Moraes, L. M. P. de; Marques, H. F.; Reis, V. C. B.; Coelho, C. M.; Leitão, M. de C.; Galdino, A. S.; Porto de Souza, T. P.; Piva, L. C.; Perez, A. L. A.; Trichez, D.; et al. Applications of the Methylotrophic Yeast *Komagataella Phaffii* in the Context of Modern Biotechnology. *Journal of Fungi* **2024**, *10*.
18. Mears, L.; Stocks, S. M.; Sin, G.; Gernaey, K. V. A Review of Control Strategies for Manipulating the Feed Rate in Fed-Batch Fermentation Processes. *J. Biotechnol.* **2017**, *245*, 34–46, doi:10.1016/j.jbiotec.2017.01.008.
19. Lindskog, E. K. The Upstream Process: Principal Modes of Operation; Elsevier Ltd., **2018**.
20. Lim, H. C.; Shin, H. S. Introduction to Fed-Batch Cultures. In *Fed-Batch Cultures*; Cambridge University Press, **2018**.
21. Villadsen, J.; Patil, K. R. Optimal Fed-Batch Cultivation When Mass Transfer Becomes Limiting. *Biotechnol. Bioeng.* **2007**, *98*, 706–710, doi:10.1002/bit.21451.
22. Grigs, O.; Galvanauskas, V.; Dubencovs, K.; Vanags, J.; Suleiko, A.; Berzins, T.; Kunga, L. Model Predictive Feeding Rate Control in Conventional and Single-Use Lab-Scale Bioreactors: A Study on Practical Application. *Chem. Biochem. Engineer. Quart.* **2016**, *30*, 47–60, doi:10.15255/CABEQ.2015.2212.
23. Mears, L.; Stocks, S. M.; Albaek, M. O.; Sin, G.; Gernaey, K. V. Mechanistic Fermentation Models for Process Design, Monitoring, and Control. *Trends Biotechnol.* **2017**, *35*, 914–924.
24. Cinar, A.; Parulekar, S. J.; Undey, C.; Birol, G. *Batch Fermentation*; CRC Press, **2003**.
25. Bailey, J.E. Mathematical Modeling and Analysis in Biochemical Engineering: Past Accomplishments and Future Opportunities. *Biotechnol. Prog.* **1998**, *14*, 8–20, doi:10.1021/bp9701269.
26. Noll, P.; Henkel, M. History and Evolution of Modeling in Biotechnology: Modeling & Simulation, Application and Hardware Performance. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 3309–3323, doi:10.1016/j.csbj.2020.10.018.
27. Helleckes, L. M.; Hemmerich, J.; Wiechert, W.; von Lieres, E.; Grünberger, A. Machine Learning in Bioprocess Development: From Promise to Practice. *Trends Biotechnol.* **2023**, *41*, 817–835, doi:10.1016/j.tibtech.2022.10.010
28. Pinto, J.; Mestre, M.; Ramos, J.; Costa, R. S.; Striedner, G.; Oliveira, R. A General Deep Hybrid Model for Bioreactor Systems: Combining First Principles with Deep Neural

Networks. *Comput. Chem. Eng.* **2022**, *165*, 107952, doi:10.1016/j.compchemeng.2022.107952.
29. Agharafeie, R.; Ramos, J. R. C.; Mendes, J. M.; Oliveira, R. From Shallow to Deep Bioprocess Hybrid Modeling: Advances and Future Perspectives. *Fermentation* **2023**, *9*, 1–22, doi:10.3390/fermentation9100922.
30. Pinto, J.; Ramos, J. R. C.; Costa, R. S.; Oliveira, R. Hybrid Deep Modeling of a GS115 (Mut^+) *Pichia Pastoris* Culture with State–Space Reduction. *Fermentation* **2023**, *9*, doi:10.3390/fermentation9070643.
31. Pinto, J.; Ramos, J. R. C.; Costa, R. S.; Rossell, S.; Dumas, P.; Oliveira, R. Hybrid Deep Modeling of a CHO-K1 Fed-Batch Process: Combining First-Principles with Deep Neural Networks. *Front. Bioeng. Biotechnol.* **2023**, *11*, doi:10.3389/fbioe.2023.1237963.
32. Ramos, J. R. C.; Pinto, J.; Poiares-Oliveira, G.; Peeters, L.; Dumas, P.; Oliveira, R. Deep Hybrid Modeling of a HEK293 Process: Combining Long Short-Term Memory Networks with First Principles Equations. *Biotechnol. Bioeng.* **2024**, *121*, 1554–1568, doi:10.1002/bit.28668.
33. Bolmanis, E.; Dubencovs, K.; Suleiko, A.; Vanags, J. Model Predictive Control – A Stand Out among Competitors for Fed-Batch Fermentation Improvement. *Fermentation* **2023**, *9*, 206, doi:10.3390/fermentation9030206.
34. Schwenzer, M.; Ay, M.; Bergs, T.; Abel, D. Review on Model Predictive Control: An Engineering Perspective. *J. Adv. Manuf. Technol.* **2021**, *117*, 1327–1349.
35. Jelsch, M.; Roggo, Y.; Kleinebudde, P.; Krumme, M. Model Predictive Control in Pharmaceutical Continuous Manufacturing: A Review from a User's Perspective. *Eur. J. Pharm.* **2021**, *159*, 137–142, doi:10.1016/j.ejpb.2021.01.003.
36. Craven, S.; Whelan, J.; Glennon, B. Glucose Concentration Control of a Fed-Batch Mammalian Cell Bioprocess Using a Nonlinear Model Predictive Controller. *J. Process. Control.* **2014**, *24*, 344–357, doi:10.1016/j.jprocont.2014.02.007.
37. Dubencovs, K.; Suleiko, A.; Sile, E.; Petrovskis, I.; Akopjana, I.; Suleiko, A.; Galvanauskas, V.; Tars, K.; Vanags, J. The Application of Adaptive Model Predictive Control for Fed-Batch *Escherichia Coli* BL21 (DE3) Cultivation and Biosynthesis of Recombinant Proteins. *Fermentation* **2023**, *9*, doi:10.3390/fermentation9121015.
38. Ashoori, A.; Moshiri, B.; Khaki-Sedigh, A.; Bakhtiari, M. R. Optimal Control of a Nonlinear Fed-Batch Fermentation Process Using Model Predictive Approach. *J. Process. Control.* **2009**, *19*, 1162–1173, doi:10.1016/j.jprocont.2009.03.006.
39. Reyes, S. J.; Durocher, Y.; Pham, P. L.; Henry, O. Modern Sensor Tools and Techniques for Monitoring, Controlling, and Improving Cell Culture Processes. *Processes* **2022**, *10*, 189, doi:10.3390/pr10020189.
40. Niu, S. S.; Xiao, D. *Process Control*; Springer International Publishing, **2022**.
41. Vojinović, V.; Cabral, J. M. S.; Fonseca, L. P. Real-Time Bioprocess Monitoring: Part I: *In Situ* Sensors. *Sens. Actuators B. Chem.* **2006**, *114*, 1083–1091, doi:10.1016/j.snb.2005.07.059.
42. Sundström, H.; Enfors, S. O. Software Sensors for Fermentation Processes. *Bioprocess Biosyst. Eng.* **2008**, *31*, 145–152, doi:10.1007/s00449-007-0157-5.

43. Panda, S.; Mehlawat, S.; Dhariwal, N.; Sanger, A.; Kumar, A. Comprehensive Review on Gas Sensors: Unveiling Recent Developments and Addressing Challenges. *Materials Science and Engineering: B* **2024**, *308*.
44. Biechele, P.; Busse, C.; Solle, D.; Scheper, T.; Reardon, K. Sensor Systems for Bioprocess Monitoring. *Eng. Life Sci.* **2015**, *15*, 469–488, doi:10.1002/elsc.201500014.
45. Grigs, O.; Bolmanis, E.; Galvanauskas, V. Application of *In-Situ* and Soft-Sensors for Estimation of Recombinant *P. Pastoris* GS115 Biomass Concentration: A Case Analysis of HBcAg (Mut^{+}) and HBsAg (Mut^{S}) Production Processes under Varying Conditions. *Sensors* **2021**, *21*, 1268, doi:10.3390/s21041268.
46. Beiroti, A.; Hosseini, S. N.; Norouzian, D.; Aghasadeghi, M. R.; Jajan, L. H. G. Development of Soft Sensors for Online Biomass Prediction in Production of Hepatitis B Vaccine. *Biointerface Res. Appl. Chem.* **2023**, *13*, 1–13, doi:10.33263/BRIAC132.194.
47. Mandenius, C. F.; Gustavsson, R. Mini-Review: Soft Sensors as Means for PAT in the Manufacture of Bio-Therapeutics. *J. Chem. Technol. Biotechnol.* **2015**, *90*, 215–227, doi:10.1002/jctb.4477.
48. Tsopanoglou, A.; Jiménez del Val, I. Moving towards an Era of Hybrid Modelling: Advantages and Challenges of Coupling Mechanistic and Data-Driven Models for Upstream Pharmaceutical Bioprocesses. *Curr. Opin. Chem. Eng.* **2021**, *32*, doi:10.1016/j.coche.2021.100691.
49. von Stosch, M.; Glassey, J. *Hybrid Modeling in Process Industries*; CRC Press, **2018**.
50. Schweidtmann, A. M.; Zhang, D.; von Stosch, M. A Review and Perspective on Hybrid Modeling Methodologies. *Digit. Chem. Eng.* **2024**, *10*, 100136, doi:10.1016/j.dche.2023.100136.
51. Roeva, O.; Slavov, T. PID Controller Tuning Based on Metaheuristic Algorithms for Bioprocess Control. *Biotechnol. Biotechnol. Equip.* **2012**, *26*, 3267–3277, doi:10.5504/BBEQ.2012.0065.
52. Nagy, Z. K. Model-Based Control of a Yeast Fermentation Bioreactor Using Optimally Designed Artificial Neural Networks. *Chem. Eng. J.* **2007**, *127*, 95–109, doi:10.1016/j.cej.2006.10.015.
53. Eaton, J. W.; Rawlings, J. B. Model-Predictive Control of Chemical Processes. *Chem. Eng. Sci.* **1992**, *47*, 705–720, doi:10.1016/0009-2509(92)80263-C.
54. Findeisen, R.; Allgöwer, F. An Introduction to Nonlinear Model Predictive Control. In Proceedings of the 21st Benelux Meeting on Systems and Control; Veldhoven, **2002**.
55. Horta, A. C. L.; Da Silva, A. J.; Sargo, C. R.; Cavalcanti-Montaño, I. D.; Galeano-Suarez, C. A.; Velez, A. M.; Santos, M. P.; Gonçalves, V. M.; Giordano, R. C.; Zangirolami, T. C. 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. *Braz. J. Chem. Eng.* **2015**, *32*, 821–829, doi:10.1590/0104-6632.20150324s00003534.
56. Bolmanis, E.; Bogans, J.; Akopjana, I.; Suleiko, A.; Kazaka, T.; Kazaks, A. Production and Purification of Soy Leghemoglobin from *Pichia Pastoris* Cultivated in Different Expression Media. *Processes* **2023**, *11*, 3215, doi:10.3390/pr11113215.

57. Bolmanis, E.; Galvanauskas, V.; Grigs, O.; Vanags, J.; Kazaks, A. Leveraging Historical Process Data for Recombinant *P. Pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development. *Fermentation* **2025**, *11*, 411, doi:10.3390/fermentation11070411.
58. Pentjuss, A.; Bolmanis, E.; Suleiko, A.; Didrihsone, E.; Suleiko, A.; Dubencovs, K.; Liepins, J.; Kazaks, A.; Vanags, J. Pichia Pastoris Growth – Coupled Heme Biosynthesis Analysis Using Metabolic Modelling. *Sci. Rep.* **2023**, *13*, 15816, doi:10.1038/s41598-023-42865-w.
59. Grigs, O.; Bolmanis, E.; Kazaks, A. HBsAg Production in Methanol Controlled *P. Pastoris* GS115 Mut^{S} Bioreactor Process. *Key Eng. Mater.* **2021**, *903*, 40–45, doi:10.4028/www.scientific.net/KEM.903.40.
60. Bolmanis, E.; Uhlendorff, S.; Pein-Hackelbusch, M.; Galvanauskas, V.; Grigs, O. Anomaly Detection and Removal Strategies for In-Line Permittivity Sensor Signal Used in Bioprocesses. *Front. Bioeng. Biotechnol.* **2025**, *13*, doi: 10.3389/fbioe.2025.1609369.
61. Ren, H. T.; Yuan, J. Q.; Bellgardt, K. H. Macrokinetic Model for Methylotrophic *Pichia Pastoris* Based on Stoichiometric Balance. *J. Biotechnol.* **2003**, *106*, 53–68, doi:10.1016/j.jbiotec.2003.08.003.
62. Jackson, J. V.; Edwards, V. H. Kinetics of Substrate Inhibition of Exponential Yeast Growth. *Biotechnol. Bioeng.* **1975**, *17*, 943–964, doi:10.1002/bit.260170702.
63. Zhu, X.; Rehman, K. U.; Wang, B.; Shahzad, M. Modern Soft-Sensing Modeling Methods for Fermentation Processes. *Sensors (Switzerland)* **2020**, *20*.
64. Wang, B.; Nie, Y.; Zhang, L.; Song, Y.; Zhu, Q. A Soft-Sensor Method for the Biochemical Reaction Process Based on LSTM and Transfer Learning. *Alex. Eng. J.* **2023**, *81*, 170–177, doi:10.1016/j.aej.2023.09.007.
65. Rogers, A. W.; Vega-Ramon, F.; Yan, J.; del Río-Chanona, E. A.; Jing, K.; Zhang, D. A Transfer Learning Approach for Predictive Modeling of Bioprocesses Using Small Data. *Biotechnol. Bioeng.* **2022**, *119*, 411–422, doi:10.1002/bit.27980.
66. Gurramkonda, C.; Adnan, A.; Gäbel, T.; Lünsdorf, H.; Ross, A.; Nemani, S. K.; Swaminathan, S.; Khanna, N.; Rinas, U. Simple High-Cell Density Fed-Batch Technique for High-Level Recombinant Protein Production with *Pichia Pastoris*: Application to Intracellular Production of Hepatitis B Surface Antigen. *Microb. Cell. Fact.* **2009**, *8*, 1–8, doi:10.1186/1475-2859-8-13.
67. Wu, Z.; Christofides, P. D.; Wu, W.; Wang, Y.; Abdullah, F.; Alnajdi, A.; Kadakia, Y. A Tutorial Review of Machine Learning-Based Model Predictive Control Methods. *Rev. Chem. Eng.* **2024**.
68. Rathore, A. S.; Mishra, S.; Nikita, S.; Priyanka, P. Bioprocess Control: Current Progress and Future Perspectives. *Life* **2021**, *11*, 557, doi:10.3390/life11060557.

Bolmanis, E. Ceļā uz Industriju 4.0: reāllaika monitorings, hibrīdā modelēšana un prognozējošā vadība *Pichia Pastoris* fermentācijās. Promocijas darbs. Rīga: RTU Izdevniecība, 2026. 205 lpp.

Publicēts saskaņā ar promocijas padomes "RTU P-01" 2026. gada 9. decembra lēmumu, protokols Nr. 04030-9.1/77.

Promocijas darbs izstrādāts Latvijas Biomedicīnas pētījumu un studiju centrā, Rīgas Tehniskajā universitātē un Latvijas Valsts koksnes ķīmijas institūtā.

Darbs izstrādāts ar Eiropas Sociālā fonda atbalstu darbības programmas "Izaugsme un nodarbinātība" 8.2.2. specifiskā atbalsta mērķa "Stiprināt augstākās izglītības institūciju akadēmisko personālu stratēģiskās specializācijas jomās" projektā Nr. 8.2.2.0/20/I/008 un ar Eiropas Savienības Atveseļošanas un noturības mehānisma atbalstu projektā Nr. 5.2.1.1.i.0/2/24/I/CFLA/003 "Konsolidācijas un pārvaldības izmaiņu ieviešana Rīgas Tehniskajā universitātē, Liepājas Universitātē, Rēzeknes Tehnoloģiju akadēmijā un Latvijas Jūras akadēmijā un Liepājas Jūrniecības koledžā virzībai uz izcilību augstākajā izglītībā, zinātnē un inovācijās" akadēmiskās karjeras doktorantūras granta ietvaros. Autors izsaka pateicību RTU *HPC* centram par piekļuves nodrošināšanu augstas veiktspējas skaitļošanas infrastruktūrai.

Vāka attēls no *www.shutterstock.com*.

PROMOCIJAS DARBS IZVIRZĪTS ZINĀTNES DOKTORA GRĀDA IEGŪŠANAI RĪGAS TEHNISKAJĀ UNIVERSITĀTĒ

Promocijas darbs zinātnes doktora (*Ph. D.*) grāda iegūšanai tiek publiski aizstāvēts 2026. gada 29. aprīlī plkst. 14 Rīgas Tehniskās universitātes Dabaszinātņu un tehnoloģiju fakultātē, Paula Valdena ielā 3, 272. auditorijā.

OFICIĀLIE RECENZENTI

Asoc. prof. *Dr. sc. ing*. Linda Mežule,
Rīgas Tehniskā universitāte

Profesors *Dr. biol*. Uldis Kalnenieks,
Latvijas Universitāte, Latvija

Profesors, *Ph. D. Rui Manuel Freitas Oliveira*,
Lisabonas NOVA Universitāte, Portugāle

APSTIPRINĀJUMS

Apstiprinu, ka esmu izstrādājis šo promocijas darbu, kas iesniegts izskatīšanai Rīgas Tehniskajā universitātē zinātnes doktora (*Ph. D.*) grāda iegūšanai. Promocijas darbs zinātniskā grāda iegūšanai nav iesniegts nevienā citā universitātē.

Emīls Bolmanis ……. (paraksts)
Datums: ……

Promocijas darbs sagatavots kā tematiski vienotu zinātnisko publikāciju kopa ar kopsavilkumu latviešu un angļu valodā. Tajā ietverti pieci zinātniskie oriģinālraksti angļu valodā. Promocijas darba kopējais apjoms, ieskaitot pielikumus, ir 205 lpp.

ANOTĀCIJA

Promocijas darbs sniedz ieguldījumu rekombinanto *Pichia pastoris* (pārklasificēts kā *Komagataella phaffii*) fermentāciju inženierijas attīstībā, izstrādājot un integrējot uzlabotas stratēģijas reāllaika procesu monitoringam, prognozējošai modelēšanai un procesa vadībai, uzsverot datos balstītu pieeju, kas atbilst ceturtās industriālās revolūcijas principiem. Kā pētāmos objektus izmantojot rekombinantās *P. pastoris* šūnu līnijas, kas producē cilvēka B hepatīta vīrusa kodola antigēnu (HBcAg), sojas leghemoglobīnu (LegH) un bakteriofāga Qβ apvalka proteīnu, tiek risināti būtiski tehnoloģiskie izaicinājumi fermentācijās ar substrāta piebarošanu (*fed-batch*).

Darba pirmajā daļā uzmanība pievērsta sensoru sistēmu validācijai un signālu kvalitātes uzlabojumiem šūnu biomasas, metanola koncentrācijas un bioreaktora izplūdes gāzu sastāva monitoringam reāllaikā. Tika izstrādāti sensoru signālu apstrādes algoritmi, lai uzlabotu iegūto datu kvalitāti un sensoru rādījumu ticamību fermentācijas laikā.

Otrajā daļā pētītas mehānistiskās, datos balstītās un hibrīdās modelēšanas pieejas, izvērtējot to prognozēšanas precizitāti un robustumu. Hibrīdo modeļu izstrādē tika izmantota zināšanu pārneses mācīšanās (*transfer learning*) pieeja, kas ļāva izmantot vēsturisko fermentāciju datus jauna produkta fermentācijas modeļa izveidei, samazinot eksperimentiem nepieciešamo laiku un resursus.

Visbeidzot, promocijas darba trešajā daļā atspoguļota uzlabotu procesa vadības stratēģiju ieviešana. Klasisks proporcionāli integrālais (PI) kontrolieris tika izmantots metanola koncentrācijas regulēšanai reāllaikā, savukārt uz hibrīdā procesa modeļa balstītais prognozējošās vadības (*MPC*) ietvars kontrolēja piebarošanas ātrumu, nodrošinot šūnu biomasas pieaugumu, saskaņā ar izvēlētu trajektoriju. *MPC* sistēma demonstrēja labu noturību mainīgajos procesa apstākļos, apliecinot tās piemērotību procesa vadībai fermentācijās.

Darbs strukturēts kā tematiski vienota četru oriģinālu zinātnisko rakstu un viena apskatraksta kopa, kas sniedz ieguldījumu fermentāciju inženierijas prakses modernizācijā. Tajā iekļauti kopsavilkumi angļu un latviešu valodā, 14 attēli, viena shēma, trīs tabulas un pieci pielikumi, kopējais apjoms – 205 lappuses.

SATURS

SAĪSINĀJUMI 54
PROMOCIJAS DARBA VISPĀRĒJS RAKSTUROJUMS 55
Ievads 55
Pētījuma mērķis un uzdevumi 57
Aizstāvamās tēzes 58
Zinātniskā novitāte 58
Praktiskā nozīme 59
Promocijas darba struktūra un apjoms 59
Promocijas darba aprobācija un publikācijas 60
PROMOCIJAS DARBA GALVENIE REZULTĀTI 63
1. Literatūras apskats 63
2. Fermentāciju reāllaika monitorings 66
2.1. Šūnu biomasas mērījumi 66
2.2. Metanola koncentrācijas mērījumi 68
2.3. Reaktora izejas gāzu sastāva analīze 69
2.4. Biomasas sensoru signāla anomāliju noteikšana un korekcija 70
3. *P. pastoris* fermentāciju modelēšana 74
3.1. Mehānistiskā modelēšana 74
3.2. Datos balstīta modelēšana 77
3.3. Hibrīdā modelēšana 78
3.4. Modeļu veidu salīdzinājums 80
3.5. Pārneses mācīšanās 84
4. Fermentācijas procesu vadība 87
4.1. Metanola koncentrācijas kontrole 87
4.2. Uz modeli balstīta prognozējošā vadība 89
SECINĀJUMI 93
LITERATŪRAS SARAKSTS 94

1. pielikums. Bolmanis, E., Dubencovs, K., Suleiko, A., Vanags, J. (2023). Model Predictive Control—A Stand Out among Competitors for Fed-Batch Fermentation Improvement. *Fermentation*, *9* (3), 206.

2. pielikums. Bolmanis, E., Uhlendorff, S., Pein-Hackelbusch, M., Galvanauskas, V., Grigs, O. (2025). Anomaly Detection and Removal Strategies for *In-Line* Permittivity Sensor Signal Used in Bioprocesses. *Front. Bioeng. Biotechnol., 13*.

3. pielikums. Bolmanis, E., Grigs, O., Kazaks, A., Galvanauskas, V. (2022). High-level production of recombinant HBcAg virus-like particles in a mathematically modelled *P. pastoris* GS115 Mut^+ bioreactor process under controlled residual methanol concentration. *Bioprocess Biosyst. Eng.*, *45* (9), 1447–1463.

4. pielikums. Bolmanis, E., Bogans, J., Akopjana, I., Suleiko, A., Kazaka, T., Kazaks, A. (2023). Production and Purification of Soy Leghemoglobin from *Pichia pastoris* Cultivated in Different Expression Media. *Processes*, *11* (11).

5. pielikums. Bolmanis, E., Galvanauskas, V., Grigs, O., Vanags, J., Kazaks, A. (2025). Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development. *Fermentation, 11* (7).

SAĪSINĀJUMI

AICc	Akaike informācijas kritērijs
ADAM	adaptīvais momenta novērtēšanas algoritms
AI	mākslīgais intelekts
ANN	mākslīgais neironu tīkls
AOX1	alkohola oksidāzes I promoters
CER	oglekļa dioksīda izdalīšanas ātrums
DCW	sausu šūnu masas svars
DNN	dziļais neironu tīkls
DO	izšķīdušā skābekļa koncentrācija
DRA	dubultā slīdošā vidējais
FC	pilnībā savienotais (slānis)
GRAS	atzīts par drošu lietošanai (*Generally Recognized As Safe*)
HBcAg	B hepatīta vīrusa kodola antigēns
Industrija 4.0	ceturtā industriālā revolūcija
IQR	interkvartiļu diapazons
LegH	sojas leghemoglobīns
LeakyReLU	rektificētā lineārā vienības aktivācija ar negatīvo vērtību caurlaidību
LSTM	ilgtermiņa-īstermiņas atmiņas tīkls
MAD	mediānas absolūtā novirze
MeOH	metanols
MPC	uz modeli bāzēta prognozējošā vadība (*Model Predictive Control)*
NC	vadības horizonts
NP	prognozes horizonts
NRMSE	normalizētā kvadrātsaknes vidējās kļūdas vērtība [%]
ODE	parastais diferenciālvienādojums
OUR	skābekļa patēriņa ātrums
P	produkta koncentrācija
Pareto fronte	optimālās kompromisa līknes daudz-mērķu optimizācijā
PID	proporcionāli–integrāli–diferenciālais (kontrolieris)
PI	proporcionāli–integrālais (kontrolieris)
Qβ	bakteriofāga Q-beta apvalka proteīna vīrusveidīgā daļiņa
ReLU	rektificētā lineārā vienība (aktivācijas funkcija)
RNN	neironu tīkls ar atgriezenisko saiti
RQ	elpošanas koeficients
S	substrāta koncentrācija
SCADA	uzraudzības un datu ieguves sistēma
Tanh	hiperboliskā tangensa aktivācijas funkcija
V	tilpums
VLP	vīrusveidīgā daļiņa
X	šūnu biomasas koncentrācija

PROMOCIJAS DARBA VISPĀRĒJS RAKSTUROJUMS

Ievads

Rekombinantie proteīni ir pamats plašam biotehnoloģisko lietojumu klāstam, tostarp biofarmaceitisko preparātu, diagnostikas līdzekļu, industriālo enzīmu un sintētiskās bioloģijas attīstībai [1, 2]. Sabiedrības novecošanās un precīzijas medicīnas attīstības ietekmē pieprasījums pēc bioloģiskajiem preparātiem turpina pieaugt. Plānots, ka biofarmaceitiskās industrijas globālā tirgus vērtība līdz 2031. gadam sasniegs 740 miljardu ASV dolāru [3]. Šī pieprasījuma apmierināšanai nepieciešamas mērogojamas un izmaksu ziņā efektīvas biotehnoloģiskās ražošanas platformas ar piemērotiem saimniekorganismiem, optimizētām ekspresijas sistēmām un inovatīviem procesa vadības risinājumiem.

Lai ilustrētu rekombinanto proteīnu daudzveidību un to lietojuma potenciālu, kā piemērus ar ievērojamu medicīnisko un industriālo nozīmi var minēt rekombinanta cilvēka B hepatīta vīrusa kodola antigēna (HBcAg) un bakteriofāga Qβ kapsīdas proteīna (Qβ) veidotās vīrusveidīgās daļiņas (*VLP*), kā arī sojas leghemoglobīnu (LegH). HBcAg ir vīrusa kapsīdas proteīns, kas plaši pētīts tieši kā *VLP* platforma. Pateicoties to augstajai imunogenitātei un drošībai, šīs platformas ir daudzsološas vakcīnu izstrādē, zāļu piegādē un imūnterapijā [4, 5]. Līdzīgi arī Qβ kapsīdas proteīns, vēl viens *VLP* veidojošs proteīns, kas iegūts no bakteriofāgiem, kalpo kā daudzpusīgs karkass nanotehnoloģijā, vakcīnās un diagnostikā, pateicoties tā strukturālajai viendabībai un modificējamai virsmai [6–8]. Savukārt LegH, augu izcelsmes hēma proteīns, kas piešķir gaļai tai raksturīgo garšu un aromātu, ir kļuvis par galveno komponentu strauji augošajā augu izcelsmes gaļas alternatīvu tirgū, īpaši tādos produktos kā *Impossible™ Burger* [9, 10]. Šo proteīnu mikrobiālā ekspresija gan laboratorijas, gan rūpnieciskā mērogā, piedāvā mērogojamu un izmaksu ziņā efektīvu platformu ilgtspējīgai ražošanai, kas atbilst pieaugošajām veselības aprūpes vajadzībām un vides ilgtspējas prasībām.

Galvenais elements šīs mērogojamās ražošanas nodrošināšanā ir mikroorganisms (saimniekorganisms), un no pieejamām ekspresijas sistēmām *Pichia pastoris* ir stabila, pārbaudīta un daudzpusīga izvēle. Lai gan taksonomiski pārklasificēts kā *Komagataella phaffii*, ņemot vērā to, ka eksperimentālais darbs, celmu apzīmējumi un lielākā daļa citētās literatūras atsaucas uz šo organismu kā *Pichia pastoris*, šajā promocijas darbā konsekvences nolūkos viscaur tiek lietots apzīmējums *P. pastoris*.

Plaši atzīts par vienu no galvenajiem "darba zirgiem" industriālajā biotehnoloģijā, šis metilotrofais raugs apvieno spēju strauji augt un sasniegt augstu šūnu koncentrāciju arī definēta sastāva barotnēs ar stingri regulētām, metanola inducējamām promoteru sistēmām, sevišķi AOX1, kas nodrošina spēcīgu un kontrolējamu gēnu ekspresiju [11–13]. Turklāt šim organismam ir arī efektīva sekrēcijas sistēma, un tas spēj nodrošināt būtiskas eikariotu pēctranslācijas modifikācijas, piemēram, disulfīdu saišu veidošanu un glikozilāciju, kas ir kritiska daudzu terapeitisko proteīnu pareizai funkcionalitātei [14, 15]. Tā statuss "atzīts par drošu lietošanai" (*GRAS*) vēl vairāk izceļ šī organisma piemērotību farmaceitiskajām un industriālajām lietojuma jomām [16, 17].

Lai pilnībā utilizētu šādu mikrobiālo ekspresijas platformu potenciālu, īpaši lielā mērogā, ir nepieciešamas robustas un efektīvas kultivācijas stratēģijas. Kultivācija ar substrāta piebarošanu (*fed-batch*) kļuvusi par nozares standartu mikroorganismu fermentācijās. Šī pieeja spēj nodrošināt augstu produkta iznākumu, pateicoties kontrolētai substrāta padevei, un tiek plaši izmantota aminoskābju, antibiotiku, enzīmu un citu bioķīmisko savienojumu ražošanai [18–20]. Regulējot substrāta piebarošanas ātrumu, iespējams efektīvi pārvarēt liela mēroga ražošanas radītos izaicinājumus, piemēram, nodrošināt efektīvu masas apmaiņu un siltuma pārnesi, veicinot optimālu maisīšanu, skābekļa pieejamību un stabilu temperatūras kontroli [21]. Tomēr *fed-batch* fermentāciju panākumi lielā mērā ir atkarīgi no precīzas piebarošanas ātruma kontroles, jo pārmērīga vai nepietiekama substrāta pievade var izraisīt tā trūkumu vai uzkrāšanos augšanu inhibējošā līmenī, skābekļa deficītu vai pārmērīgu metabolītu akumulāciju, kas negatīvi ietekmē kultūru [18, 22]. Šo problēmu risināšanai arvien plašāk tiek ieviestas uzlabotas procesu modelēšanas un vadības stratēģijas, lai paaugstinātu procesu efektivitāti un mērogojamību.

Šo uzlaboto stratēģiju pamatā ir trīs galvenie fermentācijas inženierijas balsti: procesa monitorings; modelēšana; vadība [23, 24]. Nepārtraukts procesa monitorings, izmantojot fiziskos sensorus un/vai paraugu ņemšanu, nodrošina datus par šūnu biomasas, substrāta un produkta līmeni, ko modelēšana pārvērš prognozējošās atziņās, izmantojot mehānistiskās, datos balstītas vai hibrīdās pieejas. Šie modeļi balsta dinamiskās vadības stratēģijas, kas pielāgo galvenos procesa mainīgos parametrus, lai uzturētu optimālus apstākļus un nodrošinātu pastāvīgu produkta kvalitāti. Integrējot šos trīs balstus, fermentācijas procesus iespējams precīzi pārvaldīt, tādējādi uzlabojot mērogojamību, robustumu un efektivitāti rekombinanto proteīnu ražošanā.

Vēsturiski *P. pastoris* fermentāciju vadības stratēģijas galvenokārt balstījušās empīriskās sakarībās vai vienkāršotos mehānistiskos modeļos, kas ierobežoja to pielāgojamību un prognozēšanas spēju [25, 26]. Savukārt tiekšanās uz ceturto industriālo revolūciju (Industrija 4.0) ir pārvērtusi fermentācijas procesus par datiem bagātu disciplīnu, kur vēsturisko datu kopas un reāllaika datu plūsmas iespējams sistemātiski izmantot, lai padziļinātu izpratni, uzlabotu prognozēšanas spēju un atbalstītu lēmumu pieņemšanu [27]. Šajā kontekstā inteliģenta hibrīdā modelēšana piedāvā robustu risinājumu mikroorganismu kultivāciju nelineārajai un dinamiskajai dabai, ļaujot ieviest adaptīvas, mērogojamas un efektīvas vadības stratēģijas mūsdienu biotehnoloģiskajā ražošanā.

Starp jaunajām pieejām arvien lielāku nozīmi iegūst hibrīdā modelēšana – īpaši tās pieejas, kas apvieno biotehnoloģisko procesu pamatprincipus ar mašīnmācīšanās komponentēm, jo tās spēj saglabāt procesa interpretējamību, vienlaikus uzlabojot prognozēšanas precizitāti [27–29]. Starp datos balstītajām metodēm dziļie neironu tīkli (*DNN*), tostarp ar atgriezenisko saiti (*RNN*) un ilgtermiņa-īstermiņas atmiņas arhitektūras (*LSTM*), ir īpaši piemēroti biotehnoloģisko procesu lietojumiem, pateicoties spējai apgūt laika gaitā mainīgās sakarības un uztvert aizkavētās sistēmas reakcijas. Šie modeļi veiksmīgi izmantoti sistēmas stāvokļa noteikšanai, kļūdainu vērtību detektēšanai, procesu optimizācijai un kā matemātiskie (*software*) sensori dažādos biotehnoloģiskās ražošanas kontekstos [30–32]. Tomēr to "melnās kastes" daba var ierobežot interpretējamību un lietojumu, īpaši, ja tie tiek lietoti izolēti.

Lai šīs uzlabotās modelēšanas iespējas pārvērstu risinājumos, kas īstenojami procesu vadībā, īpaši efektīva stratēģija ir uz modeli bāzēta prognozējošā vadība (*MPC*). *MPC* piedāvā strukturētu ietvaru vairāku mainīgo, ierobežotu un laikā mainīgu sistēmu vadībai [33–35]. Izmantojot mehānistiskos, datos balstītos vai hibrīdos prognozējošos modeļus, *MPC* spēj paredzēt procesa turpmāko attīstību un aprēķināt optimālas vadības darbības, izmantojot attālinošā horizonta pieeju. Tās spēja ņemt vērā noteiktos parametru ierobežojumus un paredzēt procesa traucējumus, padara *MPC* īpaši piemērotu *fed-batch* fermentācijām, kur kritiski ir uzturēt optimālu substrāta koncentrāciju vai mikrobu augšanas ātrumu, minimizēt parametru svārstības un maksimizēt produktivitāti [36–38]. *P. pastoris* fermentāciju kontekstā *MPC* integrācija ar hibrīdajiem procesa modeļiem sniedz augstu potenciālu substrāta piebarošanas un vides apstākļu reāllaika optimizācijai, uzlabojot procesa robustumu, mērogojamību un efektivitāti. Tomēr šādas lietojuma jomas pašreizējā zinātniskajā literatūrā nav pietiekami dziļi pētītas.

Lai pārvarētu minētās nepilnības, promocijas darbā risināti galvenie izaicinājumi rekombinanto *P. pastoris* fermentāciju inženierijā, kas saistīti ar procesa monitoringu, modelēšanu un vadību, izmantojot HBcAg, LegH un Qβ producējošos celmus kā pētāmos objektus. Reāllaika procesa monitoringa ietvaros tiek izmantoti biomasas, metanola un izplūdes gāzu sensori, lai ģenerētu plašas datu kopas. Savukārt sensoru signālu kvalitāte tiek uzlabota, lietojot datu apstrādes tehnikas, lai mazinātu signāla troksni un algoritmu, kas atklāj un noņem anomālijas biomasas zondes signālos. Fermentācijas procesa modelēšanai tika izstrādāti mehānistiskie, datos balstītie un hibrīdie modeļi, un to prognozēšanas spēja tika savstarpēji salīdzināta. Lai izveidoto hibrīdo procesa modeli sekmīgi adaptētu jaunam producentam, tika lietota pārneses mācīšanās (*transfer learning*) metode, izmantojot vēsturisko procesu datu kopu. Visbeidzot, procesa vadības kontekstā, tika ieviests gan vienkāršs PI kontrolieris metanola koncentrācijas regulēšanai, gan demonstrēts uz hibrīdu procesa modeli balstīts *MPC* risinājums, lai efektīvi nodrošinātu šūnu biomasas koncentrācijas pieaugumu, sekojot iepriekš definētai šūnu augšanas trajektorijai.

Pētījuma mērķis un uzdevumi

Promocijas darba mērķis ir sniegt ieguldījumu rekombinantu *P. pastoris* fermentāciju jomas attīstībā, izstrādājot un integrējot uzlabotas procesa monitoringa, modelēšanas un vadības stratēģijas. Darbs saskaņā ar Industrijas 4.0 principiem veicina pāreju uz inteliģentu, datos balstītu biotehnoloģisko ražošanu. Kā izpētes objekti izmantoti HBcAg, LegH un Qβ producējošie celmi.

Promocijas darba mērķi

1. Validēt biomasas, metanola un reaktora izejas gāzu sensorus *P. pastoris* fermentācijās, lai nodrošinātu augstas kvalitātes tiešsaistes datu ieguvi. Izstrādāt reāllaika signāla apstrādes algoritmus sensoru signālu kvalitātes un uzticamības uzlabošanai.

2. Izstrādāt, novērtēt un salīdzināt mehānistiskas, datos balstītas un hibrīdas modelēšanas pieejas. Izpētīt pārneses mācīšanās izmantošanu, lai paātrinātu hibrīdo procesa modeļu izstrādi, balstoties vēsturisko procesu datos.
3. Ieviest un eksperimentāli validēt substrāta piebarošanas vadības stratēģijas, tostarp tradicionālu PI kontrolieri metanola koncentrācijas regulācijai, izmantojot tiešsaistes sensoru atgriezenisko saiti, kā arī uz hibrīda procesa modeļa balstītu *MPC* ietvaru izvēlētas šūnu biomasas pieauguma dinamikas trajektorijas nodrošināšanai.

Aizstāvamās tēzes

1. Sensoru signālu kvalitātei fermentācijās ir būtiska nozīme, un efektīva reāllaika priekšapstrāde ir nepieciešama, lai nodrošinātu precīzu procesa monitoringu un vadību.
2. Hibrīdās modelēšanas pieejas pārspēj mehānistiskos un datos balstītos modeļus gan prognozēšanas precizitātes, gan robustuma ziņā.
3. Pārneses mācīšanās ir efektīva stratēģija bioprocesu inženierijā, samazinot modeļa apmācības laiku un eksperimentālos resursus, balstoties līdzīgu procesu vēsturiskajos datos.
4. Uz hibrīda procesa modeļa bāzēts *MPC* ietvars ļauj precīzi regulēt substrāta padevi, nodrošinot iepriekš definētu šūnu augšanas trajektoriju izsekošanu *P. pastoris* fermentācijās.

Zinātniskā novitāte

Zinātniskā novitāte šajā promocijas darbā atspoguļota trīs galvenajās jomās – procesa monitoringā, modelēšanā un vadībā, katrai no tām sniedzot ieguldījumu inteliģentu *P. pastoris* fermentāciju inženierijas attīstībā.

1. Izstrādāti un demonstrēti reāllaika signālu apstrādes risinājumi biomasas, metanola un reaktora izejas gāzu sensoriem, uzlabojot signālu kvalitāti un nodrošinot uzticamāku tiešsaistes monitoringu *P. pastoris* fermentācijās.
2. Mehānistiskie, datos balstītie un hibrīdie procesa modeļi tika sistemātiski salīdzināti, parādot to, ka hibrīdie modeļi, kas apvieno neironu tīklus ar fizikāliem pamatprincipiem, sniedz visaugstāko prognozēšanas precizitāti HBcAg, LegH un Qβ celmu fermentācijās. Demonstrēts pārneses mācīšanās lietojums, izmantojot vēsturisko procesu datu kopu, samazinot jaunā modeļa apmācības laiku un eksperimentālo resursu patēriņu.
3. Ieviests uz hibrīda procesa modeļa bāzēts *MPC* ietvars, lai reāllaikā kontrolētu īpatnējo augšanas ātrumu, demonstrējot augšanas trajektorijas izsekošanu ar 10,6 % kļūdu (*NRMSE*). Sistēma demonstrēja noturību procesa mainīguma apstākļos, apliecinot hibrīdā *MPC* piemērotību inteliģentai vadībai biotehnoloģiskajā ražošanā.

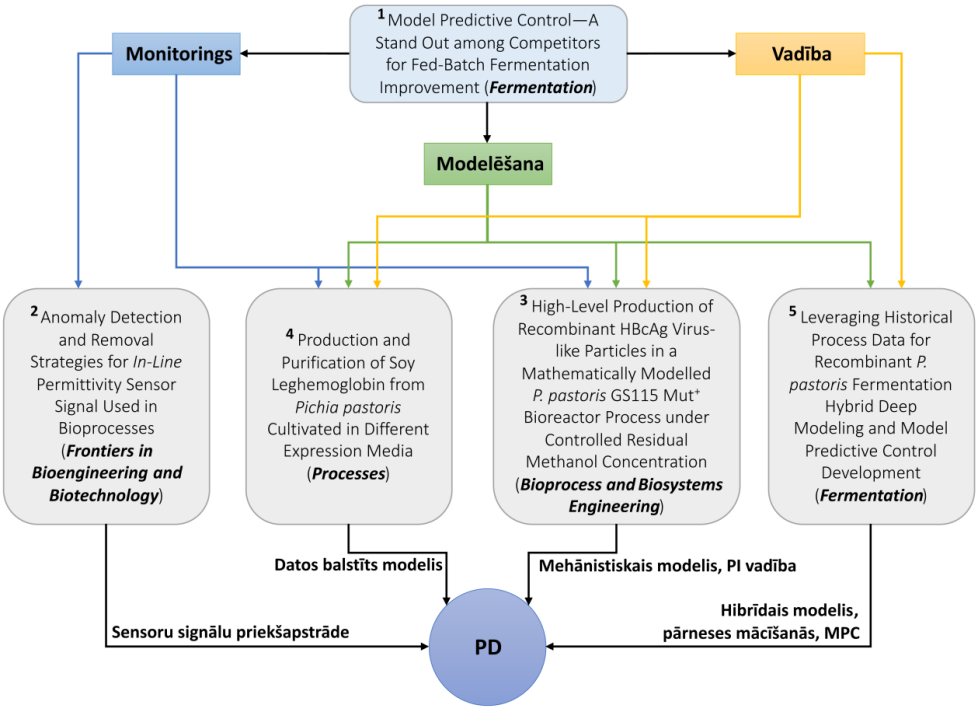
Praktiskā nozīme

Promocijas darba praktiskā nozīme ir tās ieguldījumā reāllaika bioprocesu monitoringa, prognozējošās modelēšanas un procesa vadības uzlabošanā rekombinantās *P. pastoris* fermentācijās. Pieskaņojoties Industrijas 4.0 principiem, darbs veicina pāreju uz inteliģentu un datos balstītu bioprocesu inženieriju.

1. Reāllaika signālu apstrādes metodes uzlaboja biomasas, metanola un izejas gāzu sensoru datu uzticamību, ļaujot precīzāk veikt procesu tiešsaistes monitoringu un lēmumu pieņemšanu fermentācijās.
2. Hibrīdā modelēšanas pieeja uzlaboja procesa izpratni un prognozēšanu, savukārt pārneses mācīšanās samazināja eksperimentālo slodzi, piedāvājot praktiskus rīkus ātrai modeļa adaptācijai industriālā vidē.
3. Hibrīdā *MPC* sistēma ļāva automatizēti kontrolēt šūnu īpatnējo augšanas ātrumu, sekmējot konsekventu procesa darbību un mērogojamību industriālai proteīnu ražošanai.

Promocijas darba struktūra un apjoms

Promocijas darbs veidots kā tematiski vienotu publikāciju kopa, kas sniedz ieguldījumu procesa monitoringa, modelēšanas un vadības stratēģijas rekombinantās *P. pastoris* fermentācijās ar piebarošanu. Darbs ietver četras zinātniskās oriģinālpublikācijas un vienu apskatrakstu, uzsverot uzlabotu tiešsaistes sensoru signālu kvalitāti, hibrīdo modelēšanu ar pārneses mācīšanos un inteliģentu hibrīdā *MPC* bāzētu procesa vadību, sniedzot ieguldījumu nozares attīstībā, saskaņā ar Industrijas 4.0 principiem datos balstītā biotehnoloģiskajā ražošanā (1. shēma).



1. shēma. Promocijas darba (PD) struktūras shematisks attēlojums.

Promocijas darba aprobācija un publikācijas

Promocijas darba rezultāti ir publicēti četros oriģinālos zinātniskajos rakstos un vienā apskatrakstā. Galvenie rezultāti tika prezentēti trīs konferencēs.

Zinātniskās publikācijas

1. **E. Bolmanis**, K. Dubencovs, A. Suleiko, and J. Vanags, "Model Predictive Control – A Stand Out among Competitors for Fed-Batch Fermentation Improvement", *Fermentation*, vol. 9, 206, 2023, doi: 10.3390/fermentation9030206 [Scopus, WoS, Open Access, IF 5,123, Q1, CiteScore 5,3].
2. **E. Bolmanis**, S. Uhlendorff, M. Pein-Hackelbusch, V. Galvanauskas, and O. Grigs, "Anomaly Detection and Removal Strategies for In-Line Permittivity Sensor Signal Used in Bioprocesses", *Front. Bioeng. Biotechnol.*, vol. 13, 2025, doi: 10.3389/fbioe.2025.1609369 [Scopus, WoS, Open Access, IF 4,8, Q1, CiteScore 8,8].
3. **E. Bolmanis**, O. Grigs, A. Kazaks, V. Galvanauskas, "High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. pastoris* GS115 Mut+ Bioreactor Process under Controlled Residual Methanol Concentration",

Bioprocess Biosyst. Eng., vol. 45, pp. 1447–1463, 2022, doi: 10.1007/s00449-022-02754-4 [Scopus, WoS, IF 3,6, Q2, CiteScore 6,7].

4. **E. Bolmanis**, J. Bogans, I. Akopjana, A. Suleiko, T. Kazaka, and A. Kazaks, "Production and Purification of Soy Leghemoglobin from *Pichia pastoris* Cultivated in Different Expression Media", *Processes*, vol. 11, 3215, 2023, doi: 10.3390/pr11113215 [Scopus, WoS, Open Access, IF 3,5, Q2, CiteScore 4,7].

5. **E. Bolmanis**, V. Galvanauskas, O. Grigs, J. Vanags, and A. Kazaks, "Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development", *Fermentation*, vol. 11, 411, 2025, doi: 10.3390/fermentation11070411 [Scopus, WoS, Open Access, IF 3,3, Q2, CiteScore 5,7].

Citas zinātniskās publikācijas

1. O. Grigs, **E. Bolmanis**, and V. Galvanauskas, "Application of *In-Situ* and Soft-Sensors for Estimation of Recombinant *P. pastoris* GS115 Biomass Concentration: A Case Analysis of HBcAg (Mut^+) and HBsAg (Mut^S) Production Processes under Varying Conditions", *Sensors*, vol. 21, 1268, 2021, doi: 10.3390/s21041268.

2. O. Grigs, E. Didrihsone, and **E. Bolmanis**, "Investigation of a Broad-Bean Based Low-Cost Medium Formulation for *Bacillus subtilis* MSCL 897 Spore Production", *Fermentation*, vol. 9, 4, 2023, doi: 10.3390/fermentation9040390.

3. A. Pentjuss, **E. Bolmanis**, A. Suleiko, E. Didrihsone, A. Suleiko, K. Dubencovs, J. Liepins, A. Kazaks, and J. Vanags, "*Pichia pastoris* Growth—Coupled Heme Biosynthesis Analysis Using Metabolic Modelling", *Sci. Rep.*, vol. 13, 15816, 2023, doi: 10.1038/s41598-023-42865-w.

4. A. Suleiko, K. Dubencovs, A. Kazaks, A. Suleiko, J.E. Daugavietis, E. Didrihsone, J. Liepins, **E. Bolmanis**, O. Grigs, and J. Vanags, "Performance of Recombinant *Komagataella phaffii* in Plant-Based Meat Flavor Compound (Leghemoglobin) Production through Fed-Batch Fermentations", *Fermentation*, vol. 10, 1, 2024, doi: 10.3390/fermentation10010055.

5. **E. Bolmanis**, O. Grigs, E. Didrihsone, M. Senkovs, and V. Nikolajeva, "Pilot-scale production of *Bacillus subtilis* MSCL 897 spore biomass and antifungal secondary metabolites in a low-cost medium", *Biotechnol. Lett.*, vol. 46, 3, 2024, doi: 10.1007/s10529-024-03481-4.

Dalība zinātniskajās konferencēs

1. **E. Bolmanis**, S. Ramm, M. Pein-Hackelbusch, V. Galvanauskas, and O. Grigs, "Dielectric permittivity sensor signal anomaly detection and compensation strategies in yeast *P. pastoris* fermentations", *Latvijas Universitātes 83. starptautiskā zinātniskā konference*. 14. Februāris, 2025, Rīga, Latvija (*Mutiskā prezentācija*).

2. S. Uhlendorff, **E. Bolmanis**, M. Pein-Hackelbusch, V. Galvanauskas, and O. Grigs, "Analysis of Anomaly Detection Techniques for *In-line* Permittivity Sensors in Bioprocesses", *8th European Congress of Applied Biotechnology (ECAB)*. 8.-10. Septembris, 2025, Lisabona, Portugāle (*Stenda referāta prezentācija*).
3. **E. Bolmanis**, V. Galvanauskas, and A. Kazaks, "Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling", *6. Baltijas mikrobiologu kongress*. 1.–3. oktobris, 2025, Rīga, Latvija (*Mutiskā prezentācija*).

Dalība citos zinātniskos pasākumos

1. **E. Bolmanis** un A. Kazaks, "Soy leghemoglobin (LegH) production in yeast *P. pastoris* in different cultivation media", *Informatīvs seminārs par projekta "The development of an efficient pilot-scale leghemoglobin production technology, based on recombinant Pichia pastoris and Kluyveromyces lactis fed-batch fermentations (BioHeme)" rezultātiem.* 15. novembris, 2023, Rīga, Latvija (*Mutiskā prezentācija*).
2. **E. Bolmanis** un A. Kazaks, "Soy leghemoglobin (LegH) production in yeast *P. pastoris* in different cultivation media", *Informatīvs seminārs par projekta "The development of an efficient pilot-scale leghemoglobin production technology, based on recombinant Pichia pastoris and Kluyveromyces lactis fed-batch fermentations (BioHeme)" rezultātiem.* 23. novembris, 2023, Rīga, Latvija (*Mutiskā prezentācija*).

PROMOCIJAS DARBA GALVENIE REZULTĀTI

1. Literatūras apskats

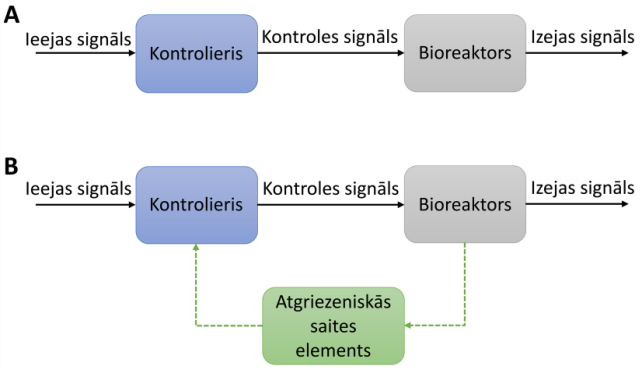
Publikācija

- **Bolmanis, E.**; Dubencovs, K.; Suleiko, A.; Vanags, J. Model Predictive Control – A Stand Out among Competitors for Fed-Batch Fermentation Improvement. *Fermentation* **2023**, *9*, 206 kā [33].

Fermentācija ar substrāta piebarošanu (*fed-batch*) jau ilgu laiku ir bijusi rūpniecisko fermentācijas procesu stūrakmens un tikusi plaši lietota augstas pievienotās vērtības biotehnoloģisko produktu ražošanā. Tā joprojām ir dominējošā pieeja biofarmaceitiskajā ražošanā, aptverot gan tirgū esošos terapeitiskos līdzekļus, gan klīnisko pētījumu stadijā esošos produktus [19, 20].

Viens no galvenajiem izaicinājumiem kultivācijas procesos ar piebarošanu ir substrāta piebarošanas ātruma kontrole – būtisks procesa parametrs, kas tieši ietekmē īpatnējo augšanas ātrumu, šūnu metabolismu, produkta titru un procesa atkārtojamību [13, 33]. Mikroorganismu augšanas nelineārais un laikā mainīgais raksturs apvienojumā ar dinamiskajiem vides apstākļiem bioreaktorā piebarošanas stratēģiju optimizēšanu padara īpaši sarežģītu.

Substrāta piebarošanas ātruma kontroles stratēģijas fermentācijās var iedalīt divās kategorijās – atvērtās un slēgtās cilpas (atgriezeniskās saites) vadībā (1.1. att.). Katrai no tām ir savas priekšrocības un ierobežojumi. Atvērtās cilpas stratēģijas ir vienkārši realizējamas, taču tām trūkst elastības, jo piebarošanas profils tiek noteikts iepriekš un procesa gaitā netiek mainīts. Savukārt slēgtās cilpas stratēģijas ietver atgriezeniskās saites elementu, piemēram, reāllaika sensoru datus vai modeļos balstītas prognozes, kas ļauj nepārtraukti pielāgot piebarošanas ātrumu kultivēšanas laikā. Šāda dinamiska pielāgošanās nodrošina lielāku noturību pret traucējumiem un bioloģisko mainīgumu, uzlabojot procesa stabilitāti un atkārtojamību. *P. pastoris* fermentācijās atvērtās cilpas pieeja nozīmētu substrāta piebarošanas profila aprēķināšanu iepriekš, savukārt atgriezeniskajā saitē balstīta stratēģija ļautu profilam attīstīties, reaģējot uz reāllaika signāliem no atgriezeniskās saites elementa.



1.1. att. Shematisks atvērtās (A) un aizvērtās (B) cilpas kontroles arhitektūras attēlojums.

Atbilstošas vadības stratēģijas izvēlei jānodrošina līdzsvars starp realizācijas sarežģītību un sagaidāmo veiktspēju, jo šis kompromiss tiešā veidā ietekmē izmaksu efektivitāti un atkārtojamību [33]. Ņemot vērā mikroorganismu sistēmu raksturīgo mainīgumu pat pie nomināli nemainīgiem darbības apstākļiem, atgriezeniskās saites vadība ir īpaši nozīmīga. Nepārtraukti pielāgojot substrāta padevi, balstoties reāllaika mērījumos, atgriezeniskajā saitē balstītas stratēģijas palīdz uzturēt metabolisko līdzsvaru, uzlabot procesa atkārtojamību un nodrošināt to, ka barības vielu pieejamība visā kultivēšanas procesā atbilst šūnu vajadzībām [18].

Efektīva atgriezeniskās saites vadības ieviešana ir atkarīga no uzticama atgriezeniskās saites mehānisma pieejamības. Parasti šādi mehānismi apvieno reāllaika fiziskos sensorus (piemēram, izšķīdušā skābekļa (*DO*), substrāta vai biomasas noteikšanai), *at-line* vai *off-line* analītiskos mērījumus (piemēram, kultūras optisko blīvumu vai substrāta koncentrāciju) un prognozējošos procesa modeļus. Šie komponenti darbojas sinerģijā, lai novērtētu iekšējos stāvokļus un reāllaikā vadītu regulācijas mehānismus [39].

Atgriezeniskās saites cilpas precizitāte, reaģētspēja un robustums ir kritiski faktori, kas nosaka kontroliera veiktspēju. Nepietiekams vai ar troksni piesārņots atgriezeniskās saites signāls var izraisīt kļūdainas vadības darbības, kas savukārt var radīt pārmērīgu substrāta padevi, samazinātus produkcijas ražīguma rādītājus vai pat rezultēties procesa utilizācijā [40]. Tāpēc augstas kvalitātes monitoringa tehnoloģiju un robustu procesa modeļu integrācija ir būtiska, lai pilnībā izmantotu atgriezeniskās saites vadības priekšrocības fermentācijās ar substrāta piebarošanu.

Atgriezeniskās saites infrastruktūras centrā atrodas sensori un procesa modeļi, kas pilda savstarpēji papildinošas funkcijas biotehnoloģisko procesu monitoringā un vadībā. Fiziskie sensori nodrošina galveno procesa parametru, piemēram, pH, temperatūras, *DO*, biomasas (mērot kultūras dielektrisko spektroskopiju vai duļķainību), un oglekļa avotu, piemēram, glikozes vai metanola, tiešus reāllaika mērījumus [41, 42]. Lai gan šie sensori parasti ir stabili un viegli kalibrējami, to lietojuma diapazons mēdz būt ierobežots, tie var būt dārgi, kā arī pakļauti aizsērēšanai vai signāla nobīdei – īpaši liela mēroga ražošanā [43, 44].

Lai pārvarētu šos ierobežojumus, arvien lielāku nozīmi iegūst procesa modeļi. Šie modeļi – mehānistiskie, datos balstītie vai hibrīdie – spēj novērtēt nemērītos lielumus (piemēram, šūnu īpatnējo augšanas vai produkcijas ātrumu), sekmīgi integrējot pieejamos mērījumus [45, 46]. To galvenās priekšrocības ir elastīgums, izmaksu efektivitāte un spēja prognozēt citādi nemērāmus procesa stāvokļus. Tomēr to uzticamība lielā mērā ir atkarīga no modeļa struktūras un ievades datu kvalitātes, tāpēc ilgtermiņa precizitātes uzturēšanai nepieciešama regulāra rekalibrācija [47].

Bioprocesu modeļi variē no mehānistiskām pieejām, balstītām bioķīmijas un fizioloģijas principos, līdz datos balstītiem modeļiem, piemēram, statistiskajām regresijām vai mašīnmācīšanās algoritmiem, kas spēj uztvert datu kopā esošās empīriskas sakarības. Mehānistiskie modeļi nodrošina augstāku interpretācijas pakāpi un dziļāku izpratni, taču bieži prasa ievērojamas nozares zināšanas un intensīvu darbu modeļa parametru noteikšanai [23]. Datos balstītie modeļi izceļas ar spēju modelēt sarežģītas un nelineāras dabas sakarības bez

detalizētām iepriekšējām zināšanām, taču tiem trūkst caurspīdīguma, un to precizitāte lielā mērā ir atkarīga no datu kvalitātes [48].

Efektīvi izmantojot abu iepriekšminēto modeļu veidu stiprās puses, hibrīdie modeļi kļūst arvien populārāki bioprocesu inženierijā [29, 49, 50]. Hibrīdā modelēšana apvieno mehānistisko modeļu struktūru ar datos balstīto modeļu elastīgumu, nodrošinot precīzākas un vispārināmākas bioprocesu dinamikas reprezentācijas, īpaši tad, kad trūkst pilnīgas mehānistiskās izpratnes.

Slēgtās cilpas vadības stratēģiju vidū proporcionāli-integrālā-atvasinātā (PID) kontroliera metode joprojām ir visplašāk lietotā rūpnieciskajās fermentācijās ar piebarošanu. Šis klasiskais kontrolieris aprēķina kļūdu starp izmērīto lielumu un uzdotās vērtības līmeni, pēc tam pielāgo ievadi (substrāta piebarošanas ātrumu), ņemot vērā proporcionālo (K_P), integrālo (K_I) un atvasināto (K_D) komponenti [51]. PID kontrolieri parasti tiek izmantoti netiešās atgriezeniskās saites konfigurācijās, pielāgojot substrāta piebarošanas ātrumu, balstoties sekundāros signālos, piemēram, pH (*pH-stat*), *DO* (*DO-stat*), īpatnējā augšanas ātrumā (μ-*stat*) vai substrāta koncentrācijā. Lai gan tie ir salīdzinoši vienkārši un robusti, PID kontrolieru veiktspēju bieži ierobežo pieejamo uzticamo reāllaika bioloģisko lielumu mērījumu trūkums un mikroorganismu sistēmu nelineārā un laikā mainīgā daba [33, 52].

Lai pārvarētu šos ierobežojumus, kā labāka alternatīva ir izvirzījusies uz modeli bāzēta prognozējošā vadība (*MPC*). *MPC* izmanto dinamisku procesa modeli, lai prognozētu sistēmas uzvedību nākotnē un attiecīgi optimizētu vadības darbības [34, 53]. Atšķirībā no PID, kas reaģē uz pašreizējām novirzēm vienā mainīgajā lielumā, *MPC* spēj vienlaikus pārvaldīt vairākus mainīgos, ievērot noteiktos darbības ierobežojumus un efektīvāk tikt galā ar mikroorganismiem raksturīgo nelinearitāti un procesa novirzēm [33, 54]. Turklāt PID kontrolierim nepieciešama regulāra parametru pielāgošana, kas bieži vien ir laikietilpīgs un pret neparedzamām novirzēm jutīgs process, savukārt *MPC* izmanto modeļos balstītus parametrus, piemēram, funkciju koeficientus un prognozes horizontus, kas nodrošina lielāku pielāgojamību un samazina nepieciešamību pēc biežām regulēšanas korekcijām. Turklāt jaunākajos pētījumos tiek uzsvērts, ka *MPC* var efektīvi darbināt uz standarta rūpnieciskās aparatūras, apliecinot tā praktisko piemērotību reāllaika bioprocesu vadībā [33].

2. Fermentāciju reāllaika monitorings

Publikācijas

- **Bolmanis, E**.; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. pastoris* GS115 Mut+ Bioreactor Process under Controlled Residual Methanol Concentration. *Bioprocess Biosyst. Eng.* **2022**, *45*, 1447–1463 kā [4].
- **Bolmanis, E.**; Bogans, J.; Akopjana, I.; Suleiko, A.; Kazaka, T.; Kazaks, A. Production and Purification of Soy Leghemoglobin from *Pichia pastoris* Cultivated in Different Expression Media. *Processes* **2023**, *11*, 3215 kā [56].
- **Bolmanis, E.**; Uhlendorff, S.; Pein-Hackelbusch, M.; Galvanauskas, V.; Grigs, O. Anomaly Detection and Removal Strategies for In-Line Permittivity Sensor Signal Used in Bioprocesses. *Front. Bioeng. Biotechnol.* **2025**, *13* kā [60].

Efektīvs fermentācijas procesa monitorings ir būtisks, lai optimizētu ražošanas iznākumu un nodrošinātu nemainīgu produkta kvalitāti, ļaujot savlaicīgi atklāt un kontrolēt bioloģiskās un procesa apstākļu svārstības. Kultivējot *P. pastoris*, precīza galveno mainīgo lielumu, piemēram, biomasas koncentrācijas, substrāta pieejamības, metabolisma aktivitātes un produktu veidošanās uzraudzība, ne tikai uzlabo procesa izpratni, bet arī veicina agrīnu kļūmju identificēšanu un atbalsta paaugstinātu produktivitāti. Turklāt augstas izšķirtspējas monitoringa dati ir būtiski datos balstītu modelēšanas pieeju izstrādē un lietošanā, kam ir potenciāls vēl vairāk uzlabot procesa vadību un optimizāciju.

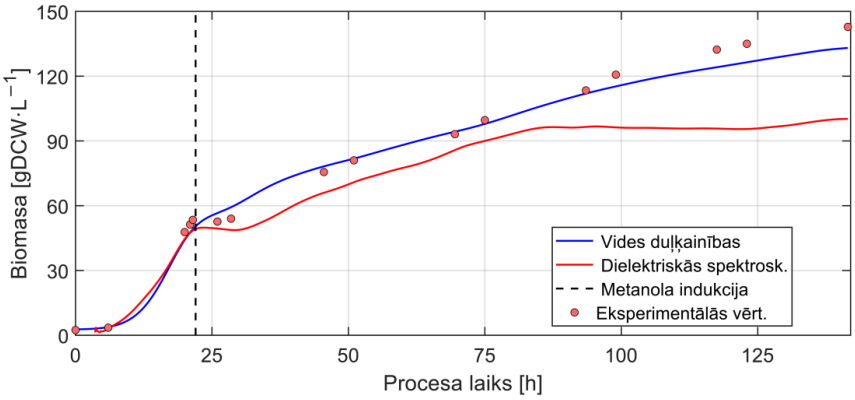
Šajā nodaļā aplūkota tādu fizisko sensoru kā biomasas un metanola sensoru un izplūdes gāzu analizatora integrācija bioreaktora sistēmā, lai nodrošinātu nepārtrauktu, neinvazīvu reāllaika uzraudzību visā fermentācijas procesā. Šie sensori tika izmantoti atsevišķos fermentācijas eksperimentos, lai papildinātu standarta bioreaktora mērījumus, tostarp *DO*, pH, temperatūru un maisītāja ātrumu. Iegūtās datu kopas nodrošināja visaptverošu ainu par procesu dinamiku un šūnu uzvedību fermentācijās, kas kalpoja par stabilu pamatu ne tikai procesu analīzei, bet arī reāllaika vadības stratēģiju ieviešanai, kā arī hibrīdo un mašīnmācīšanās modeļu izstrādei.

Fermentāciju laikā ievāktie monitoringa dati kalpoja par kritisku pamatu fermentācijas dinamikas izprašanai, optimizēšanai un modelēšanai. Reāllaika mērījumi sniedza vērtīgu ieskatu kultūras fizioloģiskajā stāvoklī fermentācijas laikā un nodrošināja pamatu datos balstītai modelēšanai (3. nodaļa) un vadības stratēģijām (4. nodaļa), kas izstrādātas un izvērtētas šajā promocijas darbā.

2.1. Šūnu biomasas mērījumi

In situ biomasas zondes sniedz reāllaika neinvazīvu šūnu koncentrācijas mērījumu fermentācijās, ļaujot nepārtraukti uzraudzīt mikroorganismu augšanu bez manuālas paraugu ņemšanas. Šī promocijas darba izstrādes gaitā izvēlētos fermentācijas procesos tika izmantotas divu veidu *in situ* biomasas sensoru zondes – optiskā zonde (*ASD19-EB-01, Optek-Danulat*), kas mēra kultūras duļķainību, un dielektriskās spektroskopijas zonde (*Incyte, Hamilton*).

Šīs savstarpēji papildinošās tehnoloģijas sniedza būtiskas priekšrocības biomasas dinamikas uzraudzībai. Optiskā vides duļķainības zonde nodrošināja ātru un stabilu signālu, kas korelēja ar kopējo biomasu, lai gan tā ietvēra gan dzīvotspējīgo, gan mirušo šūnu frakcijas. Turpretī dielektriskā spektroskopijas zonde selektīvi novērtēja dzīvotspējīgo biomasu, mērot tikai neskartu šūnu membrānu elektriskās īpašības. Ņemot vērā to, ka galvenos fermentācijas parametrus, piemēram, augšanu un ražošanas ātrumu, galvenokārt ietekmē dzīvotspējīgās šūnas, šis mērījums ir informatīvāks. Tomēr precīzai kalibrēšanai ir nepieciešami uzlaboti references mērījumi, piemēram, dzīvotspējīgu šūnu skaits, kas nebija pieejams šo eksperimentu laikā. Abu biomasas sensoru signāli kopā ar references sausās biomasas mērījumiem redzami 2.1. attēlā.



2.1. att. Optiskā (vides duļķainības) un dielektriskās spektroskopijas sensoru šūnu biomasas koncentrācijas mērījumu salīdzinājums *P. pastoris* fermentācijā.

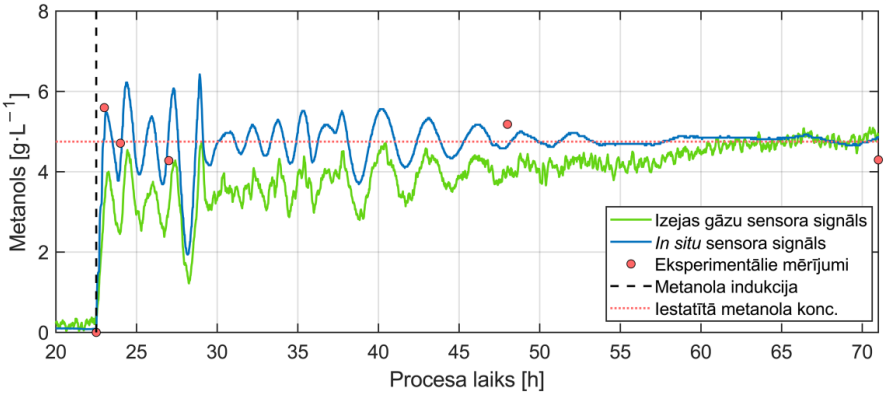
Kā redzams 2.1. attēlā, optiskā biomasas sensora signāls cieši sakrīt ar eksperimentāli noteiktajām šūnu biomasas vērtībām, demonstrējot spēcīgu korelāciju ($R^2 = 0,99$). Turpretī dielektriskās spektroskopijas zondes signāls līdz metanola indukcijai cieši seko gan optiskā sensora, gan eksperimentālajiem biomasas mērījumiem, pēc kuras tas uzrāda strauju kritumu un pēc aptuveni 85. kultivācijas stundas sasniedz plato, vairs būtiski nepieaugot. Šī uzvedība atspoguļo izmaiņas šūnu dzīvotspējā. Glicerīna augšanas fāzē šūnu dzīvotspēja joprojām ir augsta (tuvu 100 %), un tā rezultātā starp visām trim mērīšanas metodēm pastāv cieša korelācija [45, 55]. Tomēr pēc metanola indukcijas korelācija samazinās metanola citotoksiskās ietekmes dēļ, kas samazina dzīvotspējīgo šūnu frakciju adaptācijas fāzē (pirmajās stundās pēc indukcijas). Adaptācijai progresējot un atsākoties augšanai, signāli atkal uzrāda līdzīgas tendences, lai gan dielektriskās spektroskopijas signāls joprojām ir nedaudz zemāks, atspoguļojot mirušo šūnu populācijas klātbūtni. Aptuveni 85. procesa stundā dielektriskā sensora signāls sasniedz plato, kas liecina, ka šūnu augšanas un mirstības rādītāji ir sasnieguši līdzsvaru. Šis plato netiek novērots optiskajā sensorā vai eksperimentālajos mērījumos, jo abos gadījumos tiek ņemta vērā kopējā biomasa, tostarp mirušo šūnu frakcija.

Procesu monitorings, izmantojot biomasas zondes, ne tikai atbalsta fermentācijas procesa vadību, nodrošinot savlaicīgu atgriezenisko saiti procesa parametru pielāgošanai, bet arī kalpo

kā bagātīgs datu avots datos balstītu procesa modeļu izstrādei un pilnveidošanai [56, 57]. Proti, optisko un dielektrisko spektroskopijas zonžu kombinēta izmantošana sniedz visaptverošāku biomasu sastāva perspektīvu, nošķirot kopējās un dzīvotspējīgās šūnu populācijas. Šī divu sensoru pieeja sniedz jaunu iespēju turpmākiem pētījumiem, kur abu signālu vienlaikus integrācija varētu palīdzēt ņemt vērā šūnu populācijas neviendabīgumu [58]. Šāda pieeja varētu uzlabot hibrīdo modeļu prognozējamo precizitāti un nodrošināt informētākas procesu vadības stratēģijas rekombinantās *P. pastoris* fermentācijās.

2.2. Metanola koncentrācijas mērījumi

Papildus biomasas mērījumiem šajā promocijas darbā pētīto fermentācijas procesu būtiska sastāvdaļa bija arī metanola koncentrācijas mērīšana reāllaikā. *P. pastoris* konstrukcijām, kas utilizē AOX1 promoteru, metanols kalpo gan kā oglekļa avots, gan kā rekombinanto proteīnu ekspresijas inducētājs. Tāpēc ir būtiski precīzi uzraudzīt un kontrolēt tā koncentrāciju kultivācijas vidē, jo pārmērīga metanola koncentrācija var kavēt šūnu augšanu un negatīvi ietekmēt produktivitāti. Lai izsekotu metanola uzkrāšanās dinamiku, tika izmantoti divi dažādi sensori – gāzes fāzes sensors (*BCP-EtOH, BlueSens*), kas mēra metanola koncentrāciju reaktora izejas gāzēs, un *in situ* šķidrās fāzes zonde (*MeOH sensor, Raven Biotech*), kas tieši kvantificē metanola līmeni barotnē. 2.2. attēlā redzams abu sensoru veiktspējas salīdzinājums fermentācijas laikā.



2.2. att. Izejas gāzu un *in situ* metanola sensoru signālu salīdzinājums *P. pastoris* fermentācijā ar metanola koncentrācijas kontroli.

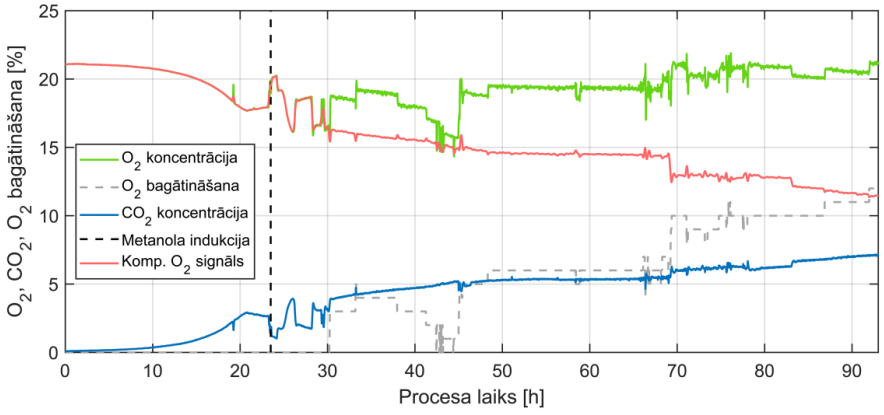
2.2. attēlā redzams detalizēts sensoru veiktspējas salīdzinājums fermentācijas procesa gaitā. *In situ* metanola sensors acīmredzami pārspēj tā izplūdes gāzēs balstīto analogu tādos galvenajos aspektos kā atbildes laiks, precizitāte un signāla kvalitāte. Galvenais izplūdes gāzu sensora ierobežojums ir tā trokšņainais signāls, kamdēļ nepieciešams lietot signāla trokšņa filtrēšanas metodes, lai gūtu skaidru priekšstatu par signāla tendencēm. Taču šīs apstrādes metodes negatīvi ietekmē jau tā salīdzinoši ilgo signāla atbildes laiku. Šajā pētījumā tika izmantots vienkāršs slīdošais vidējais filtrs (loga izmērs – 10), ievērojami uzlabojot signāla kvalitāti, samazinot tā svārstības par 63 % (no ± 0,27 līdz 0,10 [g·L^{-1}]), bet vēl vairāk paildzinot

kopējo sensora signāla atbildes laiku [4, 59]. Turklāt šis sensors tieši nemēra apstākļus šķidrajā barotnē; tā vietā tas nosaka metanola koncentrāciju gāzveida fāzē virs šķidruma virsmas. Lai gan tā signāls korelē ar metanola līmeni barotnē, šī korelācija ir pakļauta ievērojamai laika nobīdei, kas ierobežo tā lietderību reāllaika procesa vadības lietojumos. Turklāt, lai gan signāla dinamika būtībā atspoguļo *in situ* sensora dinamiku, aplēstās koncentrācijas pastāvīgi ir zemākas, jo īpaši metanola indukcijas sākumposmā, un pakāpeniski stabilizējas fermentācijas beigās. Šī neatbilstība var liecināt par sensora signāla novirzi vai aizkavētu mērījuma līdzsvarošanu starp gāzes un šķidruma fāzēm.

Precīzs reāllaika metanola koncentrācijas mērījums ļauj labāk kontrolēt piebarošanas stratēģijas, uzlabo procesa stabilitāti un atkārtojamību, sekmējot rekombinanto proteīnu ražošanu [12, 59]. Starp pieejamajām tehnoloģijām *in situ* metanola sensori nodrošina izcilu reaģētspēju un tiešu ieskatu kultūras vidē, padarot tos īpaši vērtīgus procesu optimizācijai un uzlabotos procesa vadības lietojumos [59].

2.3. Reaktora izejas gāzu sastāva analīze

Reaktora izejas gāzu sastāva analīze ir svarīgs fermentācijas uzraudzības aspekts, kas sniedz reāllaika ieskatu par mikrobu respirāciju un substrāta patēriņa ātrumu. Mērot svarīgo gāzu – skābekļa (O_2) un oglekļa dioksīda (CO_2) – koncentrācijas, šī pieeja ļauj aprēķināt šūnu metabolisma ātrumus, tostarp skābekļa uzņemšanas ātrumu (*OUR*), oglekļa dioksīda evolūcijas ātrumu (*CER*) un elpošanas koeficientu (*RQ*). Šie parametri ir būtiski šūnu aktivitātes novērtēšanai, vielmaiņas izmaiņu noteikšanai un procesu vadības stratēģiju realizācijai gan laboratorijas, gan rūpnieciskās ražošanas procesos. Šajā darbā tika izmantots izplūdes gāzu analizators (*BlueInOneFerm, BlueSens*), lai pastāvīgi uzraudzītu O_2 un CO_2 koncentrāciju bioreaktora izejas gaisa plūsmā. Šie mērījumi sniedza vērtīgus reāllaika datus par šūnu respiratoro dinamiku fermentācijas laikā. Tipiskie izplūdes gāzu profili, kas novēroti reprezentatīvas fermentācijas laikā, redzami 2.3. attēlā.



2.3. att. O_2 un CO_2 koncentrācijas bioreaktora izejas gāzu plūsmā un ieejas gaisa bagātināšanas ar O_2 fermentācijas procesā.

P. pastoris fermentācija ierasti sākas ar *batch* fāzi, kurā par oglekļa avotu kalpo glicerīns, kam seko fāze ar papildu glicerīna piebarošanu, lai sasniegtu pietiekami augstu biomasas līmeni pirms metanola indukcijas. Pārslēdzot substrāta padevi uz metanolu, šūnām nepieciešams laiks, lai pielāgotu metabolismu, kas skaidri atspoguļojas izplūdes gāzu analizatora rādījumos (2.3. att.). Uzreiz pēc metanola padeves sākšanas tiek novērots straujš CO_2 koncentrācijas kritums, kas liecina par samazinātu metabolisko aktivitāti adaptācijas laikā. Kad CO_2 līmenis atkal sāk pieaugt, tas norāda, ka šūnas pielāgojas metanolam un atsāk augšanu. Šo reāllaika izplūdes gāzu dinamiku var izmantot, lai pielāgotu metanola piebarošanas stratēģiju ātrākai un efektīvākai šūnu adaptēšanai, piedāvājot elastīgāku alternatīvu parasti izmantotajam, bet bieži vien pārāk konservatīvam trīspakāpju indukcijas protokolam.

Ja pietiekamu skābekļa piesātinājumu vairs nevar uzturēt, palielinot tikai maisīšanas ātrumu, bioreaktora sistēma sāk ieplūdes gaisa bagātināšanu ar tīru skābekli. Lai gan šī ir standarta procedūra augsta šūnu blīvuma fermentācijās, rezultātā pieaugošā skābekļa koncentrācija ieejā ietekmē arī izplūdes gāzu analizatora rādījumus. Kā redzams 2.3. attēlā, katrs pēkšņas bagātināšanas solis izraisa atbilstošu O_2 koncentrācijas lēcienu izejas gāzēs.

Lai kompensētu šo ietekmi un nodrošinātu precīzu elpošanas aktivitātes interpretāciju, O_2 rādījumiem jāpiemēro kompensācijas koeficients, pamatojoties uz skābekļa bagātināšanas procentuālo daudzumu bioreaktora ieplūdes gāzu maisījumā. Analizējot atbilstošo izplūdes gāzu O_2 koncentrācijas pieaugumu ar pieaugošu skābekļa bagātināšanas līmeni, tika noteikta spēcīga lineāra korelācija ($R^2 = 0,99$). Šo attiecību var izmantot, lai izmērītajam O_2 signālam piemērotu korekcijas terminu, pamatojoties uz skābekļa bagātināšanas procentuālo daudzumu, kas efektīvi kompensē ieplūdes gaisa bagātināšanas ietekmi (2.3. att., komp. O_2 signāls). Fermentācijas laikā CO_2 līmenis palielinās līdz ar šūnu blīvumu, bet O_2 līmenis samazinās paaugstināta šūnu skābekļa patēriņa un CO_2 ražošanas dēļ. Šī dinamika netika precīzi atspoguļota neapstrādātā O_2 sensora signālā, bet tika precīzi atainota kompensētajā signālā. Novirze, ko rada O_2 bagātināšana, būtiski ietekmē skābekļa uzņemšanas ātruma (*OUR*) un elpošanas koeficienta (*RQ*) aprēķinus, potenciāli apdraudot procesa stabilitāti un iznākumu.

2.4. Biomasas sensoru signāla anomāliju noteikšana un korekcija

Ņemot vērā ļoti dinamiskos apstākļus bioreaktorā augsta šūnu blīvuma fermentāciju laikā, *in situ* sensoru signāliem var būt pazemināta kvalitāte (troksnis) vai novērojamas neparedzētas signālu anomālijas. Ņemot vērā to, ka augstas kvalitātes datiem ir izšķiroša nozīme efektīvā datos balstītā modelēšanā, šīs sensoru signāla kvalitātes problēmas nepieciešams identificēt un risināt vēlams reāllaikā, jo īpaši, ja sensora signāls tiek izmantots fermentācijas vadībai.

Dielektriskās spektroskopijas biomasas sensora datu analīze atklāja interesantu sakarību – metanola indukcijas fāzes laikā permitivitātes signālā parādīja pēkšņi, neizskaidrojami signāla lēcieni un kritumi (pīķi) un signāla līmeņa nobīdes (2.4. A att.) [45, 60]. Līdzīga tendence, lai gan mazāk izteikta, bija vērojama arī optiskā vides duļķainības sensora datos. Šīs anomālijas – lielā mērā neizpētīta problēma fermentācijās – var būtiski ietekmēt procesa veiktspēju, jo īpaši, ja substrāta piebarošanas vadībai tiek izmantoti reāllaika sensoru dati. Lai to novērstu, izmantojot iepriekš ievākto eksperimentālo datu kopu, tika izstrādāts robusts algoritms signāla anomāliju noteikšanai un koriģēšanai reāllaikā [60].

Vienkāršas filtrēšanas metodes, piemēram, slīdošais vidējais, bieži vien nav pietiekamas, lai tiktu galā ar sarežģītību un mainīgumu, kas novērojams reālu bioprocesu datos. Lai pārvarētu šos ierobežojumus, tika izstrādāta strukturēta trīspakāpju pieeja: 1) signāla priekšapstrāde, lai samazinātu troksni un novērstu kontekstuālās atkarības; 2) anomāliju noteikšana, izmantojot sliekšņos balstītus kritērijus; 3) anomāliju korekcija un validācija.

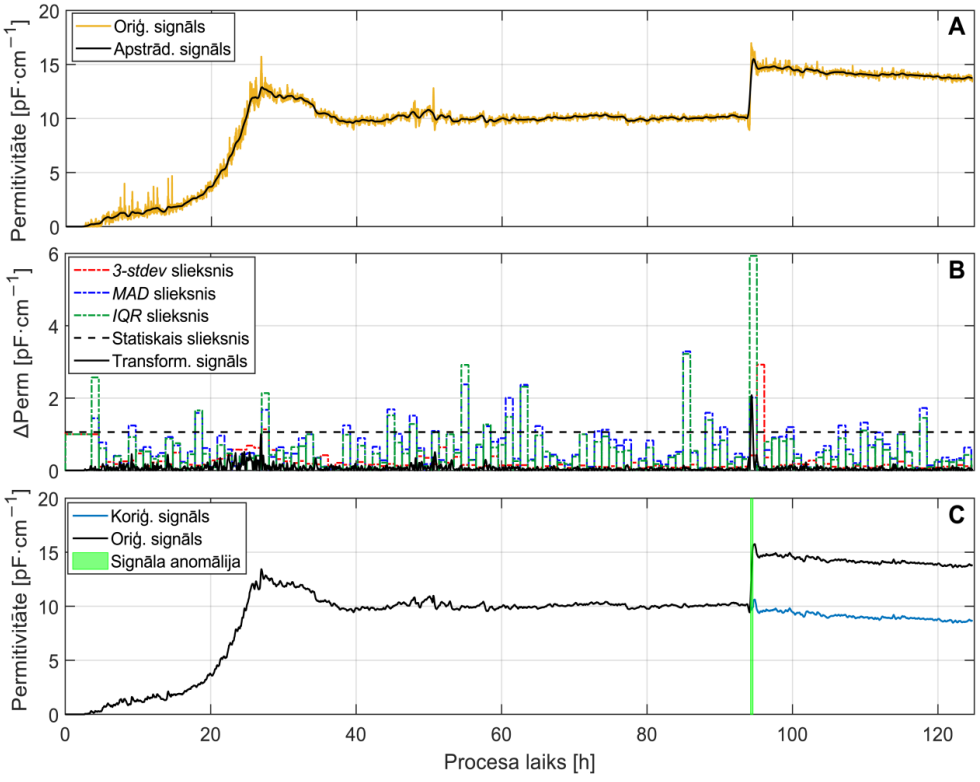
1. **Signāla priekšapstrāde**
   Lai optimizētu permittivitātes signāla trokšņu filtrēšanu *P. pastoris* fermentācijā reāllaikā, tika izvērtētas vairākas filtrēšanas metodes attiecībā pret manuāli atlasītu beztrokšņa references signālu. Veiktspēja tika novērtēta, izmantojot normalizētu vidējo kvadrātisko kļūdu (*NRMSE*) un signāla aizkaves analīzi, kas ļāva identificēt filtrēšanas metodes un parametrus, kas efektīvi samazināja troksni, neapdraudot signāla precizitāti un neradot pārmērīgu nobīdi atbildes laikā.
2. **Anomāliju noteikšana**
   Lai izceltu signāla novirzes, vienlaikus linearizējot signālu un noņemot kontekstuālo atkarību, tika izmantots dubultā slīdošā vidējā (*DRA*) metode. Lai noteiktu piemērotas sliekšņa robežvērtības, tika izvērtētas gan statiskās, gan dinamiskās metodes, tostarp manuāla sliekšņa vērtības izvēle un statistiskas pieejas, piemēram, trīs standartnoviržu likums, mediānas absolūtā novirze (*MAD*) un interkvartiļu diapazons (*IQR*), un katra no tām tika pārbaudīta vairākos logu izmēros. Lai noteiktu visstabilāko noteikšanas metodi, katra stratēģija tika salīdzināta ar manuāli anotētām signāla anomālijām, aprēķinot F1 testa rezultātu.
3. **Signāla korekcija un validācija**
   Konstatējot anomāliju, ietekmētais datu punkts tiek koriģēts, aizstājot to ar vidējo no 15 iepriekšējām vērtībām. Tam seko 15 minūšu validācijas logs, lai noteiktu signāla jauno bāzes līniju, izmantojot starpību starp līmeņiem pirms un pēc anomālijas kā dinamisku korekcijas terminu, lai nodrošinātu nepārtrauktību un samazinātu signāla novirzi.

Dažādu filtrēšanas metožu veiktspēja tika novērtēta, izmantojot *NRMSE* un signāla atbildes laika aizkaves laiku, kas tika aprēķināts, izmantojot krustenisko korelāciju starp neapstrādātiem un filtrētiem signāliem reāllaika procesa simulācijās. Starp pārbaudītajām pieejām Gausa filtrs ar loga izmēru 70 piedāvāja vislabāko kompromisu starp trokšņa samazināšanu un reaģētspēju, sasniedzot vidējo *NRMSE* 4,56 ± 1,40 % (33 % signāla trokšņa samazinājums) un pieņemamu signāla atbildes aizkavi par 6,4 minūtēm (2.4. A. att.) [60]. Citas metodes vai nu nespēja sasniegt tik augstu signāla trokšņa samazinājumu, radot ilgāku aizkaves laiku, vai arī skaitļošanas ziņā bija pārāk prasīgas, lai tās varētu izmantot reāllaikā.

Visaugstākā anomāliju noteikšanas efektivitāte tika sasniegta, izmantojot statiska sliekšņa pieeju, kas F1 testā sasniedza vērtējumu 0,79 (loga izmēri $w_1 = 1$, $w_2 = 15$, slieksnis – 1,06 $pF \cdot cm^{-1}$). Šī metode ne tikai demonstrēja spēcīgu signāla anomāliju noteikšanas spēju, bet arī prasīja minimālus skaitļošanas resursus, padarot to piemērotu ieviešanai reāllaikā. Turpretī dinamiskās sliekšņa aprēķināšanas metodes nebija pietiekami efektīvas, jo tās lielā mērā ir atkarīgas no vēsturiskajām signāla vērtībām, uzrādot F1 vērtējumus diapazonā no 0,31 līdz

0,47. Tā rezultātā pēkšņi signāla lēcieni vai kritumi netika nekavējoties atspoguļoti dinamiskajā slieksnī, izraisot tā nobīdi un palikšanu pārāk zemu, rezultējot kļūdaini pozitīvos signālos permitivitātes signāla svārstību laikā (2.4. B att.) [60]. Iespējams, ka nepieciešama kāda papildu prognozējoša kritērija iekļaušana, lai šīs metodes darbotos salīdzināmā līmenī.

Pēdējā korekcijas posmā konstatētās novirzes permitivitātes signālā tiek aizstātas ar vidējo no 15 iepriekšējām vērtībām, lai nepieļautu, ka straujas signāla fluktuācijas negatīvi ietekmē potenciālos substrāta piebarošanas ātruma aprēķinus. Bez korekcijas šādus artefaktus varētu nepareizi interpretēt kā strauju dzīvotspējīgas biomasas pieaugumu, kas var izsaukt pārmērīgu substrāta padeves pieprasījumu, tālāk apdraudot procesa stabilitāti vai pat rezultējoties fermentācijas pārtraukšanā. Katrai konstatētajai anomālijai seko 15 minūšu validācijas logs, kura laikā tiek turpinātas korekcijas. Ja tiek novērots signāla pīķis, ko raksturo straujš pieaugums un atbilstošs kritums, abi notikumi tiek uzskatīti par vienu anomāliju, lai izvairītos no liekiem koriģējošiem pasākumiem, jo signāls atgriežas sākotnējā stāvoklī. Signāla anomāliju noteikšanas un korekcijas algoritma veiktspēja demonstrēta reāllaika fermentācijas procesa simulācijā (2.4. C att.).



2.4. att. Algoritma darbības vizuāls pārskats: (A) signāla kvalitātes uzlabošana ar priekšapstrādi; (B) anomāliju detektēšanas metožu sliekšņu salīdzinājums ar *DRA* transformētajam signālam; (C) algoritma darbība reāllaika fermentācijas simulācijā.

Izvēlētā trīs soļu pieeja tika veiksmīgi izmantota rekombinanta *P. pastoris* fermentāciju simulācijās, nodrošinot precīzu un stabilu sensora signālu, neraugoties uz traucējumiem. Izmantojot statisko slieksni 1,06 ($pF \cdot cm^{-1}$) un *DRA* metodi (logu izmēri $w_1 = 1$, $w_2 = 15$), pieeja sasniedza F1 testa vērtējumu 0,79 (faktiski nodrošinot 79 % precizitāti), demonstrējot spēcīgu signāla anomāliju detektēšanas veiktspēju. Tā vienkāršība, zemās skaitļošanas izmaksas un pielāgošanās spēja padara to labi piemērotu reāllaika procesu monitoringam un vadībai dažādās fermentācijās ar dažādiem sensoriem [60]. Tomēr, lai gan šāda signālu apstrāde ievērojami uzlabo procesu vadības uzticamību, vienlīdz svarīgi ir identificēt un risināt sensoru anomāliju pamatcēloņus, lai nodrošinātu mērījumu ilgtermiņa stabilitāti un procesa noturību.

3. *P. pastoris* fermentāciju modelēšana

Publikācijas

- **Bolmanis, E**.; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. pastoris* GS115 Mut+ Bioreactor Process under Controlled Residual Methanol Concentration. *Bioprocess Biosyst. Eng.* **2022**, *45*, 1447–1463 kā [4].
- **Bolmanis, E.**; Bogans, J.; Akopjana, I.; Suleiko, A.; Kazaka, T.; Kazaks, A. Production and Purification of Soy Leghemoglobin from *Pichia pastoris* Cultivated in Different Expression Media. *Processes* **2023**, *11*, 3215 kā [56].
- **Bolmanis, E.**; Galvanauskas, V.; Grigs, O.; Vanags, J.; Kazaks, A. Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development. *Fermentation* **2025**, *11*, 411 kā [57].

Procesa matemātiskā modelēšana ir būtiska *P. pastoris* fermentācijas labākai izpratnei, optimizēšanai un vadībai. Modelēšanas pieejas aptver gan mehānistiskus modeļus, kas apraksta bioloģiskos procesus, izmantojot bioķīmiskos un fizioloģiskos pamatprincipus, gan datos balstītus modeļus, piemēram, statistikas un mašīnmācīšanās metodes, kas empīriskas attiecības atvasina no procesa datiem. Mehānistiskie modeļi sniedz ieskatu sistēmas darbībā, bet tiem nepieciešamas padziļinātas zināšanas par pašu sistēmu un detalizēta parametru noteikšanas procedūra [23]. No otras puses, datos balstītie modeļi spēj apstrādāt un interpretēt sarežģītu, nelineāru procesa dinamiku ar minimālām iepriekšējām zināšanām, lai gan tie lielā mērā ir atkarīgi no datu kvalitātes un pēc savas būtības ir "melnā kaste" [48]. Lai novērstu šos ierobežojumus, arvien vairāk tiek lietoti hibrīdie modeļi, kuros mehānistiskā izpratne ir integrēta datos balstītā elastībā [29, 49, 50]. Efektīva modelēšana atbalsta procesu izstrādi, mērogošanu un reāllaika vadību, uzlabojot produktivitāti, produktu kvalitāti un atkārtojamību gan laboratorijas, gan rūpnieciskā mēroga fermentācijās.

3.1. Mehānistiskā modelēšana

Mehānistiskais bioreaktora modelis tika izstrādāts, izmantojot fermentācijas datu kopu *P. pastoris* celmam, kas ražo HBcAg. Šeit nav atspoguļoti glicerīna fāzes modelēšanas rezultāti, jo metanola indukcijas fāze ir svarīgāka rekombinanto proteīnu ražošanas kontekstā; pilnīgs pārskats par modelēšanas rezultātiem, tostarp glicerīna fāzi, sniegts oriģinālpublikācijā [4].

Tika izstrādāts makrokinētisks procesa modelis, kas aptver intracelulārās enerģijas un metabolītu līdzsvaru metanola metabolisma laikā, pamatojoties uz Ren *et al.* formulējumu [61]:

$$\begin{bmatrix} \frac{3}{1-\varphi} & \frac{3}{1-\varphi} - K_1 & 0 & 0 \\ \frac{5\varphi+1}{1-\varphi} & \frac{6\varphi}{1-\varphi} - K_1 - 4K_2 & 5 & -2 \\ -1 & -3K_1 - K_2 - \frac{1}{Y_{\mathrm{ATP}}} & 1 & 2P/O \\ 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} q_G \\ \mu \\ q_{Ac} \\ q_{O_2} \end{bmatrix} = \begin{bmatrix} q_{\mathrm{MeOH}} \\ 0 \\ mATP_{\mathrm{MeOH}} \\ 0 \end{bmatrix}, \tag{3.1.}$$

kur φ – formaldehīda frakcija, kas oksidēta līdz formiātam; K_1 un K_2 – modeļa parametri; Y_{ATP} – *ATP* ražības koeficients [g·mol^{-1}]; *P*/*O* – oksidatīvās fosforilēšanas efektivitātes koeficients; q_G – īpatnējais glikolīzes ātrums [mol·g^{-1}·h^{-1}]; μ – īpatnējais augšanas ātrums [h^{-1}]; q_{Ac} – īpatnējais acetil-CoA ražošanas ātrums [mol·g^{-1}·h^{-1}]; q_{O2} – īpatnējais skābekļa uzņemšanas ātrums [mol·g^{-1}·h^{-1}].

Biomasas īpatnējais augšanas ātrums (μ) tika iegūts, atrisinot vienādojumu sistēmu ar lineāras algebras metodēm. Īpatnējais metanola uzņemšanas ātrums (q_{MeOH}) tika aprēķināts, izmantojot nemonotoni pieaugošu funkciju, ko sākotnēji ierosināja *Jackson & Edwards* [62]:

$$q_{\mathrm{MeOH}} = \frac{q_{\max} \times S}{K_S + S + (S^2/K_i)} \times M, \tag{3.2.}$$

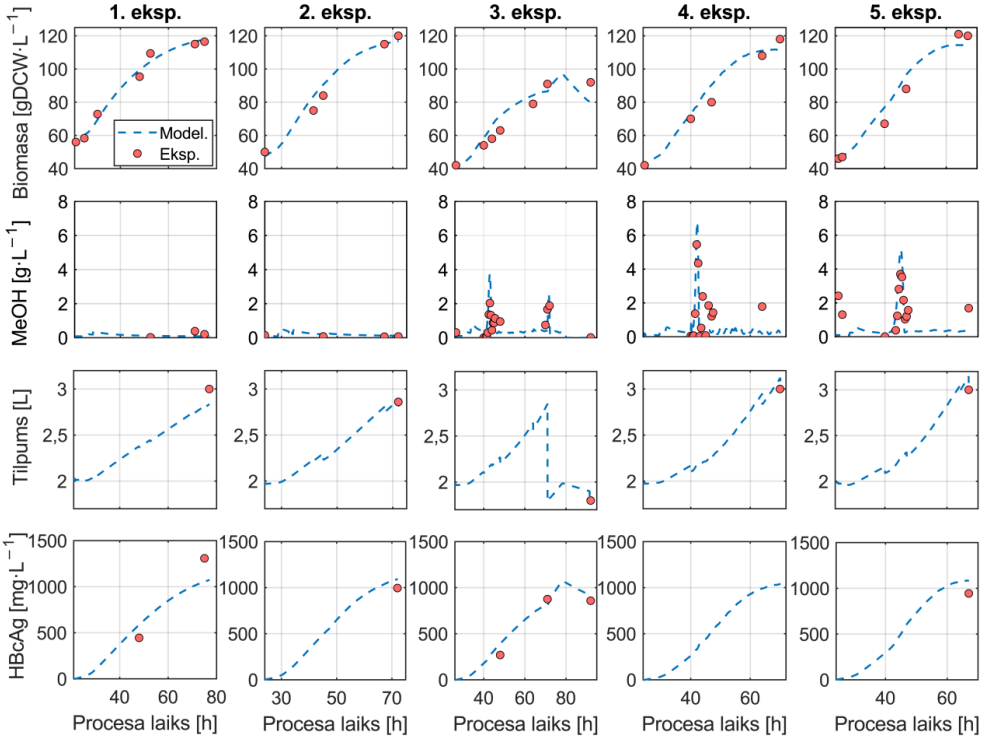
kur q_{max} – maksimālais īpatnējais metanola uzņemšanas ātrums [g·g^{-1}·h^{-1}]; *S* – metanola koncentrācija barotnē [g·L^{-1}]; K_S – metanola piesātinājuma konstante [g·L^{-1}]; K_I – metanola inhibīcijas konstante [g·L^{-1}]; *M* – metanola molmasa [g·mol^{-1}].

Produkta uzkrāšana (q_P) tika aprakstīta, izmantojot *Luedeking-Piret* modeli, kas saista produkta veidošanos gan ar augšanu saistītiem, gan ar augšanu nesaistītiem mehānismiem:

$$q_P = \mu \times Y_{PX}, \tag{3.3.}$$

kur μ – īpatnējais šūnu augšanas ātrums [h^{-1}]; Y_{PX} – īpatnējais produkta iznākuma koeficients [g·g^{-1}].

Aprēķinātais augšanas, substrāta uzņemšanas un produkta veidošanās ātrums tika iekļauts bioreaktora masas bilances diferenciālvienādojumos, lai simulētu šūnu biomasas (*X*), metanola koncentrācijas (*S*), kultūras tilpuma (*V*) un produkta uzkrāšanās (*P*) dinamiku fermentācijas procesā. Optimālie modeļa parametri tika noteikti, pamatojoties uz zinātniskajā literatūrā atspoguļotajām vērtībām, un precizēti, izmantojot fermentācijas datos balstītu parametru atlases procedūru. Izstrādātais modelis spēja veiksmīgi reproducēt galveno procesa parametru dinamiku *P. pastoris* fermentācijas procesa metanola indukcijas fāzē [4]. Attiecīgie simulācijas rezultāti parādīti 3.1. attēlā.



3.1. att. Mehānistiskās modelēšanas rezultāti, kuros atainota šūnu biomasas, metanola koncentrācijas, reaktora tilpuma un produkta (HBcAg) koncentrācijas dinamika piecās *P. pastoris* fermentācijās.

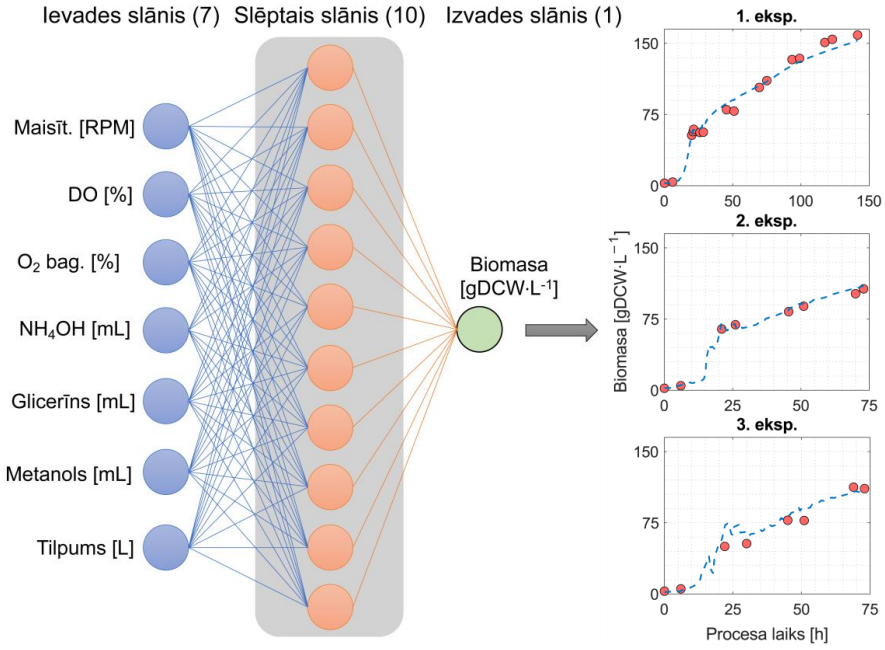
Modelis demonstrēja labu precizitāti attiecībā uz biomasu (5,05 % *NRMSE*), reaktora tilpumu (5,65 %) un produkta koncentrāciju (8,57 %). Tomēr tas bieži vien pārāk zemu novērtēja metanola koncentrāciju pēc šūnu adaptācijas, un tā rezultātā tika ietekmēta precizitāte (20,83 %) [4]. Neskatoties uz to, metanola uzkrāšanās paaugstināta substrāta piebarošanas ātruma laikā tika atspoguļota precīzi. Lai gan ir nepieciešama turpmāka pilnveide, šis modelis demonstrē retu mēģinājumu modelēt metanola koncentrācijas dinamiku *P. pastoris* fermentācijā.

Noslēgumā tika veikta modeļa parametru jutīguma analīze, lai novērtētu modeļa stabilitāti un noteiktu svarīgākos modeļa parametrus. Rezultāti liecināja, ka dažiem parametriem bija augsta jutība, kur pat nelielas novirzes būtiski ietekmēja modeļa precizitāti. Vislielākā jutība tika novērota parametriem glicerīna augšanas fāzē, kuru sākotnēji nelielās kļūdas vērtības visā turpmākajā fermentācijas procesā mēdza uzkrāties un pieaugt, kas radīja negatīvu ietekmi uz modeļa precizitāti [4]. Šie secinājumi liecina, ka modelis ir vispiemērotākais lietošanai tieši metanola indukcijas fāzē, kur tas demonstrē lielāku stabilitāti un augstāku precizitāti.

3.2. Datos balstīta modelēšana

Datos balstītā modelēšanā tiek izmantoti vēsturiskie procesa dati, lai atklātu empīriskas sakarības un prognozētu sistēmas uzvedību, nepaļaujoties uz detalizētām mehāniskām zināšanām [63]. Šī pieeja ir īpaši vērtīga sarežģītas dabas biotehnoloģiskos procesos, piemēram, *P. pastoris* fermentācijās, kur procesa kritisko parametru nelineārā dinamika un limitētā procesa izpratne ierobežo tīri mehānistisku modelēšanu. Izmantojot mašīnmācīšanās metodes, datos balstīti modeļi procesu datos spēj atrast sarežģītas ievades-izvades sakarības, atbalstīt fermentācijas reāllaika monitoringu (matemātiskie sensori) un uzlabot prognozēšanas precizitāti ar nosacījumu, ka ir pieejamas augstas kvalitātes reprezentatīvas datu kopas [56, 63].

Šūnu biomasas koncentrācijas novērtēšanai tika izveidots matemātiskais (*software*) sensors uz mākslīgo neironu tīklu (*ANN*) bāzes, izmantojot tikai standarta bioreaktora mērījumus no diviem fermentācijas procesiem [56]. Ievades dati ietvēra maisītāja ātrumu (*RPM*), *DO* (%), ieejas gaisa O_2 bagātināšanas pakāpi (%), pievadīto bāzes, glicerīna un metanola tilpumu (mL) un reaktora tilpumu (L). Savukārt reāllaika *in situ* optiskās biomasas zondes dati kalpoja kā mērķa datu kopa apmācībai. Datu kopa tika sadalīta, atvēlot 70 % datu apmācībai, 15 % testēšanai un 15 % validācijai. Lai samazinātu pēkšņas svārstības un troksni izstrādātajā biomasas matemātiskajā sensorā, tika izmantots *Savitzky-Golay* filtrs ar pirmās kārtas polinomu un loga izmēru 29. Modeļu apmācībai tika izmantots divu slāņu neironu tīkls ar 10 sigmoidālās funkcijas aktivētiem slēptiem neironiem un vienu lineāru izejas neironu, kā parādīts 3.2. attēlā.



3.2. att. *ANN* balstītā matemātiskā biomasas sensora struktūra un modelēšanas rezultāti.

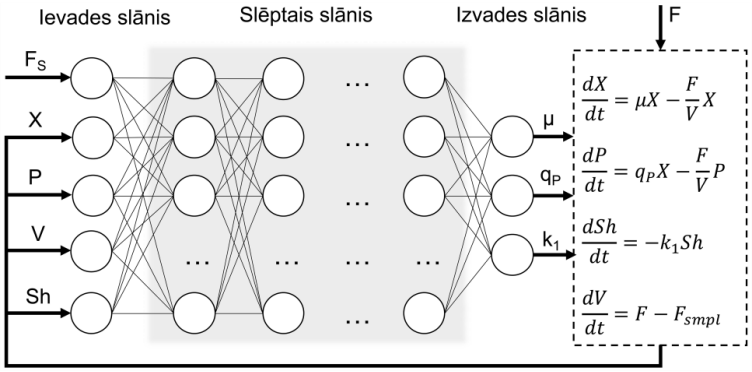
Izstrādātais datos balstītais modelis precīzi atspoguļoja šūnu biomasas dinamiku apskatītājās fermentācijās, demonstrējot labu atbilstību gandrīz visos eksperimentos (3.2. att.).

Kopējā precizitāte modeļa apmācībai izmantotajā datu kopā tika lēsta 3,72 % *NRMSE* apmērā; tomēr metodoloģiskas kļūdas dēļ netika novērtēta tās spēja vispārināt iepriekš neredzētus datus. Izstrādātais matemātiskais sensors paļaujas tikai uz standarta bioreaktora mērījumiem un neietver papildu signālus, piemēram, CO_2 koncentrāciju reaktora izejas gāzu plūsmā kā citos līdzīgos lietojumos [46, 64]. Lai gan tas var nedaudz ierobežot precizitāti, tas likvidē nepieciešamību pēc papildu sensoriem. Kopumā sasniegto veiktspēju var uzskatīt par pietiekamu izmantošanai rekombinantā *P. pastoris* kultivēšanā kā papildu mērījumu eksperimentālajām analīzēm.

3.3. Hibrīdā modelēšana

Bioloģisko sistēmu augstā sarežģītība un nelineārā augšanas dinamika biotehnoloģisko procesu monitoringa un vadības jomās joprojām rada būtiskas grūtības. Reaģējot uz to, hibrīdās modelēšanas pieejas, apvienojot mehānistisku un datos balstītu modeļu stiprās puses, ir kļuvušas par spēcīgiem rīkiem fermentāciju modelēšanā. Šiem modeļiem ir būtiska nozīme biotehnoloģiskās ražošanas digitālajā evolūcijā, jo īpaši tāpēc, ka mašīnmācīšanās kļūst arvien nozīmīgāka attiecībā uz spēju uztvert procesa dinamiku, neprasot pilnīgas sistēmas zināšanas [29, 49]. Optimālas neironu tīkla arhitektūras izvēle ir ļoti svarīga, lai panāktu augstu modeļa precizitāti un vispārināšanu dziļās mācīšanās lietojumos.

P. pastoris fermentācijai tika izstrādāts universāls hibrīdais procesa modelis, izmantojot vēsturisko procesu datu kopu, kas ietver 17 fermentācijas eksperimentus, kas tika veikti šī promocijas darba izstrādes gaitā [57]. Hibrīdā procesa modeļa struktūra redzama 3.3. attēlā.



3.3. att. Izveidotā hibrīdā fermentācijas procesa modeļa struktūras shematisks attēlojums.

Modeļa ievades slāni veido pieci ievades parametri: substrāta (metanola) piebarošanas ātrums (F_S, mL·min^{-1}); sausu šūnu biomasas koncentrācija (X, gDCW·L^{-1}); produkta koncentrācija (P, mg·L^{-1}); kultūras tilpums (V, L); empīrisks metanola šoka faktors (Sh). Šoka faktors, kas sākotnēji definēts kā $Sh(0) = 1$, raksturo metanola piebarošanas kumulatīvo citotoksisko ietekmi uz šūnām un var tikt uzskatīts par neizmērāmu iekšējā stāvokļa parametru. Modelis ģenerē trīs izvades parametrus – īpatnējo šūnu augšanas ātrumu (μ), īpatnējo produkta ražošanas ātrumu (qp) un šoka faktora izmaiņu tempu (k_1). Pēc tam šie rezultāti tiek nodoti modeļa parametriskajam komponentam, kas formulēts kā vienkāršu diferenciālvienādojumu

(*ODE*) sistēma, balstoties bioreaktora materiālajā bilancē un fizikālķīmiskos pieņēmumos. Šajā struktūrā vienīgie ārējās ievades parametri ir substrāta piebarošanas ātrums (F_S), paraugu ņemšanas ātrums (F_{smpl}) un kopējās tilpuma plūsmas ātrums (F).

Lai efektīvi optimizētu hibrīda modeļa slēptā slāņa arhitektūru, tika izmantota stratēģija, kas ietver trīs soļus. Pirmkārt, Beijesa optimizācija nodrošināja ātru galveno modeļa hiperparametru izpēti, proti, slāņa tipu (*LSTM* vai pilnībā savienotu (*FC*)), slāņu skaitu, aktivācijas funkciju un neironu skaitu. Hibrīdie modeļi tika apmācīti paralēli 10 ciklu garumā ar paaugstinātu mācīšanās ātrumu, un optimālās arhitektūras tika izvēlētas, pamatojoties uz validācijas kļūdas un Akaike informācijas kritērija (AICc) vērtībām, lai efektīvi līdzsvarotu modeļa precizitāti ar tā sarežģītību. Pēc tam iegūtās likumsakarības tika izmantotas, lai samazinātu arhitektūru kombināciju skaitu nākamajā solī, kur fokusētā veidā tika atlasīti labākie kandidāti, novērtējot 200 aktivācijas funkciju kombinācijas (*LeakyReLU*, *ReLU*, *Tanh* vai bez), *LSTM* slāņa (1–5) un pilnībā savienotu slāņu (1–10) neironu skaitu. No šiem kandidātiem tika atlasīti pieci labākie modeļi. Noslēgumā šie modeļi tika pilnībā apmācīti (20 000 iterāciju), un tika izvēlēts visefektīvākais modelis; tika novērtēta arī *dropout* slāņa izmantošana, lai uzlabotu noturību un vispārināšanu.

Beijesa optimizācija efektīvi identificēja daudzsološas hibrīda modeļu arhitektūras, koncentrējot meklēšanu uz augstas veiktspējas hiperparametru reģioniem, ievērojami samazinot izvērtēto modeļu skaitu. Vislabākās modeļu arhitektūras konsekventi sastāvēja no *LSTM* slāņa, kam sekoja viens *FC* slānis. Turpmākā visaptverošā meklēšana precizēja modeļa atlasi, novērtējot visas iespējamās slēpto neironu skaita un aktivācijas funkciju kombinācijas. Izmantojot gan validācijas kļūdas, gan AICc vērtības, no Pareto frontes tika atlasīti pieci līdzsvaroti modeļi, kas demonstrēja gan augstu prognozēšanas precizitāti, gan mērenu sarežģītību, tādējādi nodrošinot vispārināmību un skaitļošanas efektivitāti (3.1. tabula).

3.1. tabula

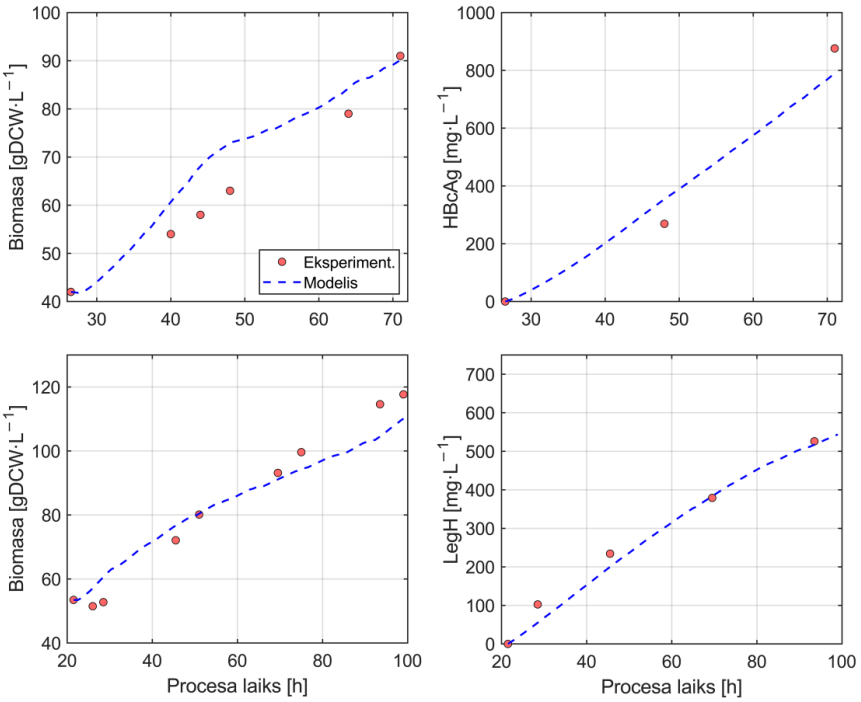
Labāko modeļa arhitektūru apkopojums

***LSTM* neironi**	***FC* neironi**	**Aktivācijas funkcija**	**Validācijas kļūda [%]**	**Parametru skaits**	***AICc***
3	5	*LeakyReLU*	7,28	146	1294
2	10	*LeakyReLU*	6,37	127	1155
2	9	*Tanh*	8,14	121	1236
2	**8**	***ReLU***	**4,93**	**115**	**998**
1	9	*Tanh*	8,27	76	1090

Pēdējā optimizācijas solī katra no piecām atlasītajām arhitektūrām tika apmācīta 10 reizes (20 000 iterācijās), lai novērtētu noturību un veiktspējas konsekvenci pie dažādām nejauša rakstura inicializācijām (3.1. tabula). Visefektīvākais modelis sastāvēja no diviem *LSTM* slāņiem un astoņiem neironiem *FC* slānī, kas aktivēti ar *ReLU* aktivizācijas funkciju. Šis modelis uzrādīja vismazāko validācijas kļūdu (4,93 %) un *AICc* (998) vērtību, nodrošinot optimālu kompromisu starp prognozēšanas precizitāti un vienkāršību (3.4. att.). *Dropout* slāņa

ieviešana (ar 0,1–0,5 varbūtību) konsekventi pasliktināja veiktspēju, liecinot, ka modelis jau bija pietiekami robusts.

Kā redzams 3.4. attēlā, modelis efektīvi atspoguļo gan šūnu biomasas, gan produktu koncentrāciju dinamiku fermentācijas procesa gaitā. Tomēr tas nespēj pilnībā izskaidrot nelielo biomasas koncentrācijas samazinājumu uzreiz pēc metanola indukcijas, kas atspoguļo šūnu pielāgošanos metanola metabolismam. Tas rezultējās nelielā biomasas koncentrācijas pārvērtēšanā agrīnajā metanola indukcijas fāzē. Sakarā ar minimālo šūnu biomasas pieaugumu metanola adaptācijas fāzē biomasas paraugu ņemšana netika veikta pietiekami bieži; tomēr šādi mērījumi ir būtiski, lai modelis varētu precīzi mācīties un atspoguļot šim posmam raksturīgo izaugsmes stagnāciju. Lai gan eksperimentālo produktu koncentrāciju mērījumu pieejamība ir limitēta, modelis demonstrē labu precizitāti arī produkta koncentrācijas novērtēšanā.



3.4. att. Ar atlasīto hibrīdo modeli noteiktās šūnu biomasas un produktu koncentrācijas HBcAg (augšējā rinda) un LegH (apakšējā rinda) fermentācijas procesiem.

3.4. Modeļu veidu salīdzinājums

Lai būtu iespējams adekvāti un jēgpilni salīdzināt mehānistiskās, datos balstītās un hibrīdās modelēšanas pieejas, visi trīs modeļu tipi tika novērtēti, izmantojot vienu un to pašu datu kopu. Lai gan iepriekšējās nodaļās katra modeļa spējas tika demonstrētas atsevišķi, izmantojot dažādas datu kopas, lai tieši varētu salīdzināt modelēšanas precizitāti, ir nepieciešami vienādi apmācības un testēšanas nosacījumi. Tāpēc tika izvēlēta vienota datu kopa (20 eksperimenti), lai nodrošinātu to, ka visi modeļi izmanto vienu un to pašu ievades informāciju. Šī standartizētā

pieeja ļauj objektīvi novērtēt modeļu prognozēšanas precizitāti, vispārināšanas spēju un modeļa sarežģītību dažādās modelēšanas paradigmās.

Ņemot vērā to, ka promocijas darba tapšanas gaitā tika pētīti trīs atšķirīgi rekombinanti produkti, katram produktam raksturīgie parametri tika optimizēti neatkarīgi katrā attiecīgā eksperimentālajā apakškopā, lai nodrošinātu precīzu un objektīvu produkta koncentrācijas novērtējumu. Visbeidzot, lai pilnībā izmantotu pieejamos datus, tika izmantota četrkārša savstarpējā validācija, nejauši sadalot datu kopu četrās apakškopās, no kurām katrai ir atsevišķi apmācības un testēšanas nodalījumi. Katra testēšanas kārta ietvēra divus eksperimentus ar HBcAg producentu, divus ar LegH producentu un vienu ar Qβ producentu, nodrošinot līdzsvarotu un reprezentatīvu novērtējumu visiem produktu veidiem. Modeļa veiktspēja tika novērtēta kā testēšanas rezultātu vidējā vērtība ± standartnovirze visās savstarpējās validācijas kopās. Visa eksperimentālo datu kopa apkopota 3.2. tabulā.

3.2. tabula

Eksperimentālo datu kopa modeļu salīdzināšanai

Eksp. Nr.	**Celms, produkts**	**Indukcijas ilgums [h]**	**Biomasa [gDCW·L^{-1}]**	**Piebarošanas ātr. [mL·min^{-1}]**	**V_{beigu} [L]**
1.	GS115, HBcAg	65	37,5–101,6	0,12–0,78	2,85
2.	GS115, HBcAg	45	40,6–113,5	0,12–1,00	3,09
3.	GS115, HBcAg	43	41,2–120,1	0,12–0,98	3,13
4.	GS115, HBcAg	50	59,2–120,1	0,12–0,36	2,54
5.	GS115, HBcAg	51	41,4–96,6	0,12–0,36	2,87
6.	GS115, HBcAg	48	49,1–120,0	0,12–0,50	2,88
7.	GS115, HBcAg	43	53,7–101,5	0,12–0,36	2,74
8.	GS115, HBcAg	54	44,1–84,0	0,12–0,56	2,75
9.	X-33, LegH	65	55,4–123,2	0,12–0,36	2,57
10.	X-33, LegH	46	49,5–95,4	0,12–0,60	2,98
11.	X-33, LegH	65	48,9–111,2	0,12–0,36	2,85
12.	X-33, LegH	50	45,3–101,3	0,12–0,36	2,61
13.	X-33, LegH	45	52,9–103,1	0,12–0,36	2,55
14.	X-33, LegH	46	45,1–101,3	0,12–0,36	2,52
15.	X-33, LegH	65	51,0–101,7	0,12–0,36	2,66
16.	X-33, LegH	46	50,6–92,4	0,12–0,60	3,00
17.	X-33, Qβ	65	52,5–117,6	0,12–0,49	3,23
18.	X-33, Qβ	48	49,3–117,2	0,12–1,00	3,40
19.	X-33, Qβ	55	50,1–107,7	0,12–0,36	2,84
20.	X-33, Qβ	52	52,9–112,6	0,12–0,87	3,45

Nav pārsteidzoši, ka modelēšanas sniegums dažādiem modeļu veidiem ievērojami atšķīrās. Katra modeļa tipa apmācības un testēšanas kļūdu dati apkopoti 3.3. tabulā.

3.3. tabula

Modeļu snieguma salīdzinājums, vidējā precizitāte (NRMSE)

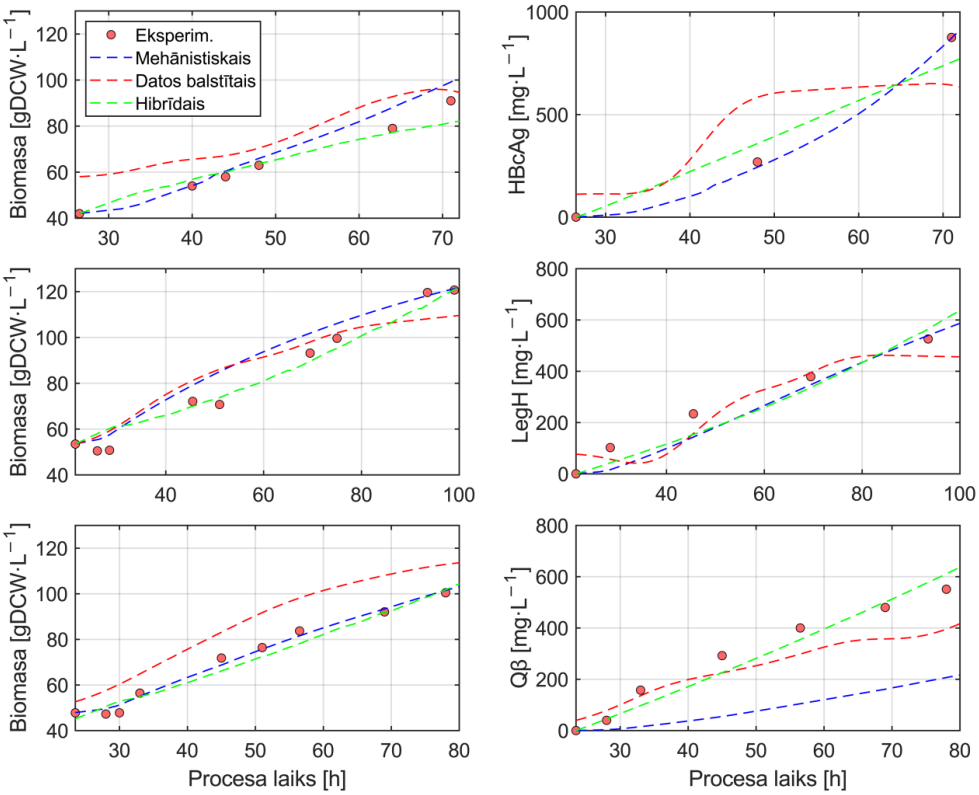
Modelis	Metrika	Biomasa [%]	Produkts [%]	Vidējais [%]
Mehānistisks	Apmācības	10,9 ± 2,0	19,6 ± 2,7	15,3 ± 2,2
	Testēšanas	13,1 ± 4,4	65,4 ± 33,8	**39,2 ± 18,9**
Datos balstīts	Apmācības	7,9 ± 1,3	7,7 ± 2,2	7,8 ± 0,5
	Testēšanas	14,8 ± 2,3	41,7 ± 9,2	**28,2 ± 5,4**
Hibrīds	Apmācības	9,1 ± 1,9	11,7 ± 3,1	10,4 ± 2,5
	Testēšanas	9,1 ± 2,0	13,1 ± 5,5	**11,1 ± 2,6**

3.3. tabulā novērojamas vairākas atšķirības starp trim modeļu veidiem. Visi modeļi uzrādīja pietiekamu precizitāti biomasas koncentrācijas prognozēšanā apmācības laikā, un *NRMSE* svārstījās no 7,9 % līdz 10,9 %. Tomēr hibrīda modelis sasniedza zemāko biomasas prognozēšanas testa kļūdu (9,1 %), kas liecina par labāku vispārināšanu attiecībā uz jauniem, iepriekš modelim neredzētiem datiem, salīdzinot ar mehānistiskajiem (13,1 %) un datos balstītajiem (14,8 %) modeļiem. Tas liecina, ka mehānisku sakarību integrēšana datos balstītā modelēšanas pieejā nodrošina lielāku robustumu, jo īpaši biomasas dinamikas atspoguļošanā dažādos procesa apstākļos.

Attiecībā uz produktu koncentrācijas aplēsēm atšķirības starp modeļu veidiem kļuva izteiktākas. Hibrīdais modelis ievērojami pārspēja pārējos modeļus ar 13,1 % testa kļūdu, kas bija salīdzinoši tuva tā apmācības laikā sasniegtajam rezultātam (11,7 %), liecinot par spēcīgu modeļa vispārināšanas spēju un uzticamu produkta biosintēzes dinamikas sakarību apguvi. Turpretī mehānistiskajam modelim bija ievērojama veiktspējas atšķirība ar salīdzinoši augstu apmācības kļūdu (19,6 %) un ļoti lielu testa kļūdu (65,4 %), kas atspoguļoja nepietiekamu spēju pielāgoties eksperimentālo datu mainīgumam un ierobežotu spēju uztvert rekombinantās proteīnu ekspresijas dinamiku. Datos balstītais modelis uzrādīja labākus rezultātus nekā mehāniskais modelis (41,7 %), taču tas joprojām būtiski atpalika no hibrīdā analoga, iespējams, tāpēc, ka tam nebija ar fermentācijas procesu saistītu iepriekšēju zināšanu.

Kopumā hibrīdais modelis uzrādīja vislabāko vispārējo veiktspēju ar vidējo apmācības kļūdu 10,4 % un testa kļūdu 11,1 %, kas liecina gan par augstu precizitāti, gan spēju sekmīgi vispārināt apgūtās sakarības. Mehānistiskais modelis demonstrēja ierobežotu elastību (39,2 % testa kļūda), savukārt datos balstītais modelis uzrādīja pārmērīgas pielāgošanas (*overfitting*) pazīmes ar zemu apmācības kļūdu (7,8 %), bet salīdzinoši augstu testa kļūdu (28,2 %). Šie rezultāti izceļ hibrīdā modeļa spēju efektīvi apvienot mehānistiskās zināšanas ar datu virzītu apmācību, padarot to par drošu un uzticamu izvēli tādu sarežģītu procesu modelēšanai kā *P. pastoris* fermentācijas.

3.5. attēlā redzama modeļu veiktspēja biomasas un produktu koncentrācijas prognozēšanā reprezentatīvam eksperimentam no katra rekombinantā proteīna producenta. Hibrīdais modelis pastāvīgi nodrošina visprecīzākos rezultātus, efektīvi atspoguļojot gan šūnu biomasas pieauguma, gan produktu uzkrāšanas tendences.



3.5. att. Modeļu veiktspējas salīdzinājums šūnu biomasas un produktu koncentrāciju novērtēšanai katra rekombinantā proteīna producenta vienā reprezentatīvā fermentācijā.

Lai šos rezultātus aplūkotu perspektīvā, jāņem vērā vairāki faktori, jo īpaši datu kopas kvalitāte, daudzveidība un pilnīgums. Dati aptvēra trīs dažādas *P. pastoris* konstrukcijas no diviem atšķirīgiem celmiem GS115 un savvaļas tipa X-33. Lai gan šūnu biomasas augšanas un produkta uzkrāšanās dinamika kopumā bija līdzīga, zināms bioloģiska rakstura mainīgums bija sagaidāms. Turklāt eksperimenti tika veikti laika posmā, kas aptver vairākus gadus, un, lai gan tika ievēroti standartizēti protokoli, pastāv nelielu darbību vai tehnisko neatbilstību iespējamība. Būtiska nozīme bija arī datu iztrūkumam, jo īpaši attiecībā uz rekombinanto proteīnu koncentrāciju. Produkta kvantitatīvā noteikšana ietvēra darbietilpīgas attīrīšanas procedūras, kurām bieži bija nepieciešamas vairākas hromatogrāfijas procedūras, kas būtiski ierobežoja mērījumu biežumu. Tā rezultātā modeļiem dažkārt bija problēmas precīzi atspoguļot produktu uzkrāšanās dinamiku. Kopumā šie faktori varēja veicināt lielāku modelēšanas kļūdu.

Neraugoties uz šīm problēmām, hibrīdais modelis demonstrēja spēcīgu sniegumu, precīzi atspoguļojot gan biomasas, gan produktu dinamiku dažādos apstākļos. Tā spēju vispārināt apstiprināja ciešā atbilstība starp apmācības un testēšanas kļūdām, kas liecināja, ka modelis pārmērīgi nepielāgojās apmācības datiem, bet gan spēja no tiem identificēt būtiskās sakarības un saglabāja prognozējošo noturību arī attiecībā uz iepriekš neredzētiem datiem.

3.5. Pārneses mācīšanās

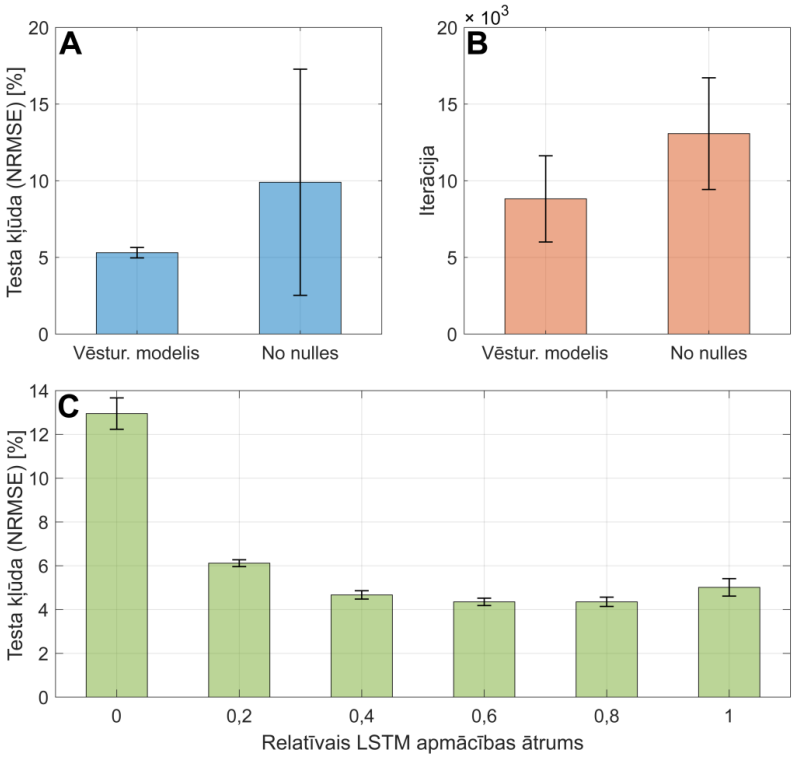
Pārneses mācīšanās ir mašīnmācīšanās tehnika, kas izmanto zināšanas, kas iegūtas, apmācot modeli vienam uzdevumam vai datu kopai, lai uzlabotu veiktspēju saistītā, bet atšķirīgā uzdevumā vai datu kopā [65]. Hibrīdas bioprocesu modelēšanas kontekstā tā ļauj pielāgot iepriekš apmācītu modeli jauniem produktiem, celmiem vai procesiem, paātrinot apmācību, samazinot nepieciešamību pēc jauniem lielapjoma datiem un uzlabojot prognozēšanas precizitāti. Tas ir īpaši vērtīgi bioprocesu izstrādē, kur augstas kvalitātes eksperimentālo datu ģenerēšana ir gan laikietilpīga, gan darbietilpīga, padarot pārneses mācīšanos par praktisku stratēģiju modeļa mērogojamības, efektivitātes un lietojamības uzlabošanai dažādos biotehnoloģiskās ražošanas scenārijos [57, 65].

Šajā pētījumā tika lietota pārneses mācīšanās, lai pielāgotu hibrīdo modeli, kas apmācīts ar datu kopu no 17 HBcAg un LegH fermentāciju eksperimentiem, Qβ ražošanas procesam, izmantojot datus tikai no diviem eksperimentiem (3.2. tab., 17. un 18. eksperiments). Galvenā ideja bija izmantot vēsturisko procesa modeli kā sākotnējo stāvokli Qβ modeļa apmācībai. Apmācot *LSTM* slāņa parametrus ar samazinātu mācīšanās ātrumu (0–1,0 attiecībā pret pārējo tīklu), modelis saglabāja iepriekš apgūtās parametru laika dinamikas, vienlaikus pielāgojoties jaunajai Qβ datu kopai. Šī stratēģija ļāva modelim paturēt atmiņā galvenās *P. pastoris* fermentācijas iezīmes no vēsturiskā modeļa, tādējādi uzlabojot vispārināšanu un mazinot ierobežojumus, kas saistīti ar apmācību uz tik neliela datu kopuma.

Vispirms tika novērtēta vēsturiskā hibrīdā modeļa pielāgošana Qβ datu kopai, salīdzinot tā testa kļūdas vērtību (3.2. tab., 19. eksperiments) un izkliedi ar sākuma modeli, kas apmācīts no nulles ar nejauši inicializētiem parametriem. Rezultāti rāda, ka Qβ modeļa apmācības inicializēšana ar jau iepriekš apmācītu vēsturisko procesa modeli, nevis apmācība no nulles, radīja ievērojami zemākas testa kļūdas vērtības (vidēji 5,31 % pret 9,90 %) ar būtiski samazinātu mainīgumu (standartnovirze ± 0,34 pret ± 7,38) (3.6. A att.). Turklāt apmācība tika pabeigta ātrāk, prasot mazāk iterāciju (vidēji 8820 pret 13070) (3.6. B att.). Šie atklājumi liecina, ka vispārīgie procesa modeļi, apmācīti uz līdzīgām datu kopām, var kalpot kā efektīvi sākuma punkti jaunu hibrīdo modeļu izstrādei *P. pastoris* fermentācijās [57, 64]. Saglabājot vēsturisko procesu zināšanas, šādi modeļi ne tikai samazina prognozēšanas kļūdu, bet arī saīsina apmācības laiku, kas ir īpaši vērtīgi, strādājot ar nelielām datu kopām.

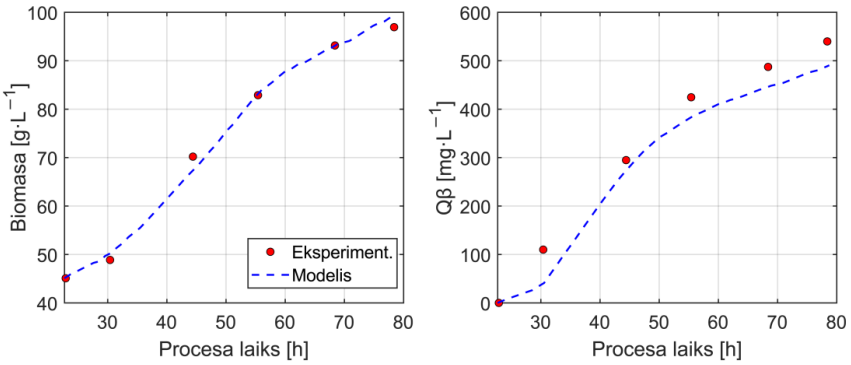
Lai efektīvi realizētu pārneses mācīšanos, optimālais *LSTM* slāņa relatīvais mācīšanās ātrums tika noteikts, sistemātiski pārbaudot visu tā diapazonu. Tā kā *LSTM* slānis "atceras" procesa dinamiku, tā mācīšanās ātruma pielāgošana ļauj modelim saglabāt daļu iepriekšējo zināšanu, vienlaikus pielāgojoties jaunam datu kopumam. Rezultāti liecina, ka *LSTM* slāņa apmācīšana ar 0,6–0,8 no pārējā modeļa mācīšanās ātruma nodrošina vislabāko veiktspēju ar vidējo prognozēšanas kļūdu 4,53 ± 0,20 % 10 atkārtojumos (3.6. C att.). Šie atklājumi liecina,

ka tikai daļēja atmiņas slāņu apmācība, nevis to pilnīga iesaldēšana, nodrošina efektīvu līdzsvaru starp iepriekšējā modeļa zināšanu saglabāšanu un jaunā procesa raksturīgo iezīmju uztveršanu [57]. Šajā konkrētajā gadījumā *LSTM* mācīšanās ātruma iestatīšana minētajā diapazonā ļauj efektīvi realizēt pārneses mācīšanos ar minimālu testēšanas kļūdu un izkliedi.



3.6. att. Vidējā prognozēšanas kļūda (A) un apmācības iterāciju skaits (B) hibrīda Qβ procesa modelim, kad tas ir pielāgots no vēsturiskā modeļa vai apmācīts no nulles; vidējais testa zudums kā *LSTM* slāņa relatīvā mācīšanās ātruma funkcija (C).

Galīgais hibrīda procesa modelis, pielāgots no vēsturiskā modeļa ar *LSTM* relatīvo mācīšanās ātrumu 0,6, uzrādīja spēcīgu prognozēšanas veiktspēju, precīzi atspoguļojot galveno mainīgo parametru dinamiku visā Qβ fermentācijas procesā. Kopumā tas sasniedza testēšanas *NRMSE* 4,35 %, ar 3,16 % biomasas koncentrācijai un 5,64 % produkta koncentrācijai, demonstrējot uzticamu abu procesa parametru prognozēšanas spēju (3.7. att.).



3.7. att. Šūnu biomasas un produkta koncentrāciju dinamika, kas prognozēta, izmantojot hibrīdo procesa modeli.

Tā kā biomasas augšanas dinamika ir līdzīga arī dažādiem *P. pastoris* celmiem, galvenais variācijas avots, modelējot fermentācijas dažādos producentos, rodas no atšķirībām produkta uzkrāšanās dinamikā. Tādējādi tieša modeļa pārnešana no viena producenta uz citu bez pielāgošanas var radīt neprecīzas prognozes, jo iegūtā produkta kinētikas reprezentācija var nebūt vispārīga. Lai to risinātu, papildu pārneses mācīšanai ar daļēju slāņa iesaldēšanu var izpētīt līdzīgas stratēģijas, piemēram, iekļaujot modelī papildu adaptācijas moduļus, hibrīda modeļa segmentēšanu atsevišķos apakštīklos katram parametram (piemēram, viens modulis biomasai, otrs – produktam utt.) vai produkta specifisko parametru pielāgošanu. Šī pieeja palīdz saglabāt vispārīgas procesa zināšanas, piemēram, biomasa augšanas dinamiku, vienlaikus ļaujot modelim uztvert katram producentam raksturīgas iezīmes, piemēram, rekombinantā produkta uzkrāšanos. Šī pieeja ļauj efektīvi pielāgoties jauniem producentiem pat ar ierobežotiem eksperimentāliem datiem, samazinot nepieciešamību modeļus apmācīt pilnībā no jauna.

Tā kā pieejami tikai divi eksperimenti, treniņu dati ir nepietiekami, lai pilnībā izpētītu procesa parametru telpu. Rezultātā modelis var labi darboties tikai apstākļos, kas ir līdzīgi tiem, kas sastapti apmācības laikā, savukārt prognozes neizpētītajā parametru telpā, visticamāk, būs neuzticamas ierobežotās ekstrapolācijas spējas dēļ. Apmācība uz jauna datu kopuma, izmantojot vēsturisko procesa modeli, ļauj modelim saglabāt vērtīgas zināšanas no daudzveidīgākā vēsturiskā datu kopuma, tādējādi uzlabojot veiktspēju procesa apstākļos, ko Qβ datu kopums neietver, kā tas demonstrēts šajā darbā.

Šis pētījums parāda, ka pārneses mācīšanās, izmantojot vēsturisku hibrīdo modeli, ir efektīva stratēģija, lai veiksmīgāk izstrādātu modeļus jauniem *P. pastoris* fermentācijas procesiem ar ierobežotām eksperimentālo datu kopām. Sākot Qβ modeļa apmācību, par pamatu ņemot iepriekš apmācītu vēsturisko modeli un daļēji pielāgojot *LSTM* slāni ar optimizētu relatīvo mācīšanās ātrumu, bija iespējams saglabāt vispārējās parametru dinamikas, vienlaikus pielāgojoties celma specifiskai produkta uzkrāšanai. Šī metode ne tikai samazināja testa kļūdu un mainīgumu, salīdzinot ar apmācību no nulles, bet arī paātrināja apmācību, nodrošinot galīgo modeli, kas spēj precīzi prognozēt gan biomasas, gan produkta koncentrācijas. Šie rezultāti uzsver iepriekšējo zināšanu izmantošanas vērtību hibrīdo modeļu pārneses apmācībā, lai uzlabotu modeļa vispārināšanu un efektivitāti bioprocesu lietojumos.

4. Fermentācijas procesu vadība

Publikācijas

- **Bolmanis, E**.; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. pastoris* GS115 Mut+ Bioreactor Process under Controlled Residual Methanol Concentration. *Bioprocess Biosyst. Eng.* **2022**, *45*, 1447–1463 kā [4].
- **Bolmanis, E.**; Galvanauskas, V.; Grigs, O.; Vanags, J.; Kazaks, A. Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development. *Fermentation* **2025**, *11*, 411 kā [57].

Substrāta piebarošanas ātruma kontrole ir kritisks aspekts fermentācijas procesu vadībā, tieši ietekmējot šūnu augšanu, produkta veidošanos un kopējo procesa produktivitāti. Precīza substrāta piebarošana, piemēram, glikozes, glicerīna vai metanola, nodrošina to, ka mikroorganismi saņem optimālu barības vielu piegādi, lai uzturētu vielmaiņas aktivitāti, vienlaikus izvairoties no paaugstināta substrāta koncentrācijas radītas inhibīcijas vai barības vielu trūkuma. Efektīvas piebarošanas ātruma stratēģijas palīdz uzturēt vēlamo augšanas ātrumu, novērst toksisku blakusproduktu uzkrāšanos šūnās un uzlabot procesa produktivitāti. Ir izstrādātas dažādas vadības pieejas – no vienkāršiem piebarošanas profiliem līdz uzlabotām, uz sarežģītiem procesa modeļiem balstītām un reāllaika atgriezeniskās saites kontroles sistēmām, kuru mērķis ir optimizēt substrāta piegādi un stabilizēt fermentācijas dinamiku [18, 33]. Tādējādi robusta substrāta piebarošanas kontrole ir būtiska, lai nodrošinātu reproducējamus un efektīvi mērogojamus fermentācijas procesus.

4.1. Metanola koncentrācijas kontrole

Metanola koncentrācijas kontrole *P. pastoris* fermentācijās ir būtiska, lai optimizētu rekombinanto proteīnu ražošanu, vienlaikus izvairoties no substrāta inhibīcijas vai toksicitātes [12, 66]. Precīza kontrole nodrošina to, ka metanols tiek uzturēts līmenī, kas atbalsta šūnu metabolismu un proteīnu ekspresiju, neradot šūnām papildu stresu vai pārmērīgu substrāta uzkrāšanos. Bieži izmantotās stratēģijas ietver tiešsaistes metanola sensoru izmantošanu kombinācijā ar vadības algoritmiem, piemēram, vienkāršiem PID cilpas kontrolieriem, lai dinamiski pielāgotu metanola piebarošanas ātrumu un uzturētu optimālu substrāta koncentrāciju visā indukcijas fāzē [33].

Lai izpētītu metanola koncentrācijas ietekmi uz rekombinantā proteīna biosintēzi, tika veikta virkne eksperimentu, kuros tika pētīta HBcAg produktivitāte, ja metanola līmenis ir 0,01 $g \cdot L^{-1}$, 1,0 $g \cdot L^{-1}$ un 2,0 $g \cdot L^{-1}$ [4]. Lai metanola koncentrāciju uzturētu nemainīgā līmenī, tika izstrādāts PI (proporcionālais-integrālais) kontrolieris, kas balstīts uz metanola sensora (*BCP-EtOH, BlueSens*) signālu. Metanola sensora tiešsaistes signāls tika iepriekš apstrādāts, izmantojot 2.2. apakšnodaļā aprakstīto slīdošā vidējā algoritmu. Kontroles algoritms regulēja metanola piebarošanas ātrumu ar mērķi stabilizēt sensora signālu ap vēlamo iestatījuma punktu (izvēlēto koncentrāciju), izmantojot atgriezeniskās saites kontroles vienādojumu, kas iekļauj PI algoritmu:

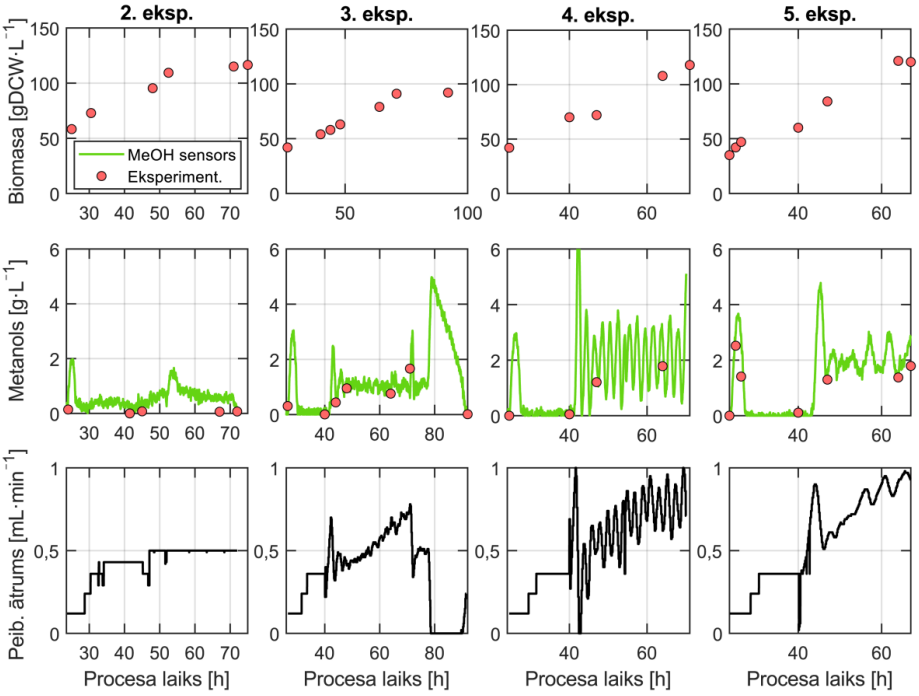
$$F_{t+\Delta t} = F_t - \frac{V}{(S_0 - S)} \times \frac{dS}{dt} + K_p \left[(\varepsilon_t - \varepsilon_{t-1}) + \frac{\Delta t}{\tau_I} \varepsilon_t \right], \quad (4.1.)$$

kur F_t – substrāta piebarošanas ātrums pašreizējā laikā t [mL·min^{-1}]; V – reaktora tilpums [L]; S_0 – substrāta koncentrācija piebarošanas šķīdumā [g·L^{-1}]; S – substrāta koncentrācija barotnē [g·L^{-1}]; K_p – proporcionālā pieauguma parametrs [L^2·g^{-1}·h^{-1}]; ε – starpība starp izmērīto un iestatīto koncentrāciju [g·L^{-1}]; Δt – laika intervāls starp soļiem [min]; τ_I – integrālā laika konstante [min].

Modeļa integrēšana ar PI vadības algoritmu ļāva saglabāt nemainīgus parametrus K_p un τ_I visā kultivēšanas procesā. Tomēr metanola kontroles precizitāte bija ļoti jutīga pret izvēlēto K_p skaitlisko vērtību. Šis parametrs, kura vērtība svārstījās no 0,02 līdz 0,05 [L^2·g^{-1}·h^{-1}], tika pielāgots, pamatojoties uz vēlamo metanola koncentrāciju. Pretstatā tam integrālās laika konstantes τ_I vērtība visos eksperimentos tika noteikta – 10 minūtes.

2. eksperimentā netika piemērota metanola kontrole. Metanola piebarošanas ātrums tika nedaudz palielināts, lai novērtētu tā ietekmi uz HBcAg produktivitāti, nodrošinot bāzes līnijas datus salīdzināšanai. 3. eksperimentā, sākot no 40. kultivācijas stundas, metanola koncentrācija tika kontrolēta 1,0 [g·L^{-1}] līmenī, izmantojot PI kontrolieri. Kontrolieris veiksmīgi uzturēja metanola koncentrāciju noteiktajā apmērā ar vidējo novirzi ± 0,28 [g·L^{-1}] (28 % *NRMSE*) līdz 72. kultivācijas stundai, kad no bioreaktora tika aizvadīts 1 [L] kultūras, izraisot īslaicīgu metanola pieaugumu. Lai gan kontrolieris sākotnēji pielāgojās, metanola līmenis drīz atkal pieauga, liecinot par iespējamu kultūras pārbarošanu.

4. eksperimentā metanola koncentrācijas uzstādījums tika palielināts līdz 2,0 [g·L^{-1}], bet tika saglabāta tā pati kontroles parametru vērtība (K_p = 0,05 [L^2·g^{-1}·h^{-1}]), kas izraisīja ievērojamu nestabilitāti, metanolam svārstoties starp 1,0 un 3,0 [g·L^{-1}] ar vidējo novirzi ± 1,26 [g·L^{-1}]. Samazinot K_p vērtību līdz 0,002 [L^2·g^{-1}·h^{-1}] 5. eksperimentā, ievērojami uzlabojās vadības kvalitāte un vidējā novirze bija ± 0,67 [g·L^{-1}] (63 % *NRMSE*). Šie rezultāti, kas redzami 4.1. attēlā, apliecina, ka PI kontrolieris metanola līmeni spēj kontrolēt ar apmierinošu precizitāti, ja tas ir pareizi noregulēts. Tomēr tā efektivitāte bija ļoti jutīga pret kontroles parametru izvēli, īpaši proporcionālā pieauguma parametru. Lai gan kontrolieris demonstrēja zināmu pielāgošanās spēju procesa traucējumiem, stabilitāte tika apdraudēta, ja parametru vērtības nebija optimālas vai procesa izmaiņas bija pēkšņas. Šie novērojumi ilustrē vajadzību pēc rūpīgas un adaptīvas kontroles iestatījumu pielāgošanas, lai nodrošinātu stabilu metanola regulēšanu fermentācijās.



4.1. att. Šūnu biomasas, metanola koncentrācijas un metanola piebarošanas ātruma dinamika *P. pastoris* fermentācijas eksperimentos ar metanola koncentrāciju PI-kontroli.

PI algoritmā balstītais metanola piebarošanas ātruma kontrolieris izrādījās efektīvs rekombinantās *P. pastoris* fermentācijās. Precīzi noregulēts, izmantojot izplūdes gāzu metanola sensora signālu kā atgriezeniskās saites elementu, tas ļāva kontrolēt metanola koncentrāciju izvēlētā līmenī ar vidējo novirzi ± 0,28–0,67 [g·L^{-1}], kas atbilst *NRMSE* 28–63 %. Tomēr, ņemot vērā tā augsto jutību pret kontroles parametru skaitlisko vērtību izvēli, ieteicams ieviest automatizētu regulēšanas procedūru, lai nodrošinātu stabilu veiktspēju dažādos metanola līmeņos. Kontroles precizitāti varētu vēl vairāk uzlabot, iekļaujot *in situ* metanola sensora zondi, kas nodrošina precīzāku mērījumu ar ātrāku atbildes laiku.

4.2. Uz modeli balstīta prognozējošā vadība

Hibrīdās *MPC* sistēmas integrē datos balstītus modeļus, piemēram, neironu tīklus, ar pamatprincipu zināšanām, lai nodrošinātu precīzu prognozēšanas spēju un reāllaika optimizāciju sarežģītos biotehnoloģiskos procesos. Šie kontrolieri izmanto gan mehānistiskās izpratnes, gan mašīnmācīšanās stiprās puses, lai risinātu nelineāru procesa dinamiku, neprognozētus procesa traucējumus un mērījumu troksni fermentācijās. Biotehnoloģiskajā ražošanā hibrīda *MPC* ir īpaši piemērota pieeja fermentācijās ar piebarošanu, kur fizioloģiskā mainība un substrātu-produktu mijiedarbības ir grūti atspoguļojamas tikai ar mehānistiskiem modeļiem [67, 68].

Pamatojoties uz hibrīdo procesa modeli, tika izstrādāts *MPC* ietvars, lai regulētu šūnu augšanu tuvu maksimālajam īpatnējam augšanas ātrumam [69]. *MPC* aprēķināja optimālo

substrāta piebarošanas ātrumu $F_S(t)$, kas nepieciešams, lai sekotu iepriekš noteiktai augšanas trajektorijai $\mu_{set}(t)$. Tā kā hibrīdais modelis, kur F_S ir ievade un μ ir izvade, nav tieši invertējams, katrā vadības solī tika lietota skaitliskā optimizācija, izmantojot *MATLAB fminbnd* funkciju noteiktajās robežās $F_S \in [0.36, 1.00]$ mL·min^{-1}:

$$\min_{Fs\in[0.36,1.00]\, mL\times\mathrm{min}^{-1}} \sum_{k=1}^{Np}[\mu(k) - \mu_{set}(k)]^2, \tag{4.2.}$$

ņemot vērā hibrīdā modeļa dinamiku:

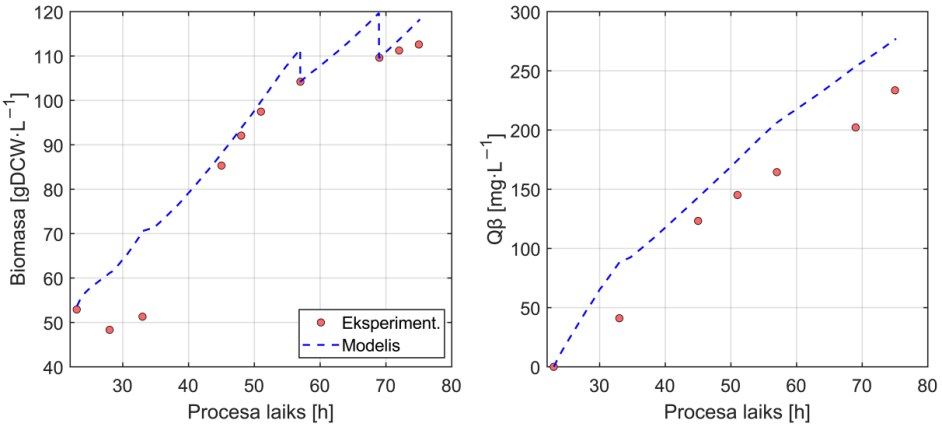
$$x(k+1) = f_{\mathrm{hybrid}}(x(k), F_S(k)), \tag{4.3.}$$

kur $x(k)$ apzīmē funkcijas stāvokļa vektoru un $\mu(k)$ ir prognozētais pieauguma temps laika solī k. Kontroles un prognozēšanas horizonti tika iestatīti uz $N_C = 1$ stundu un $N_P = 12$ stundām, attiecīgi. Hibrīdais modelis tika simulēts ar 1 minūtes paraugu ņemšanas intervālu, lai nodrošinātu precīzas prognozes.

Lai saglabātu augstu pielāgošanās spēju, modelis tika no jauna apmācīts aptuveni trīs reizes dienā pēc katras paraugu ņemšanas, izmantojot jaunos biomasas mērījumus $X_{meas}(t)$. Reāllaika procesa dati, tostarp substrāta piebarošanas ātrums, bāzes un putu dzēsēja padeve, tika integrēti *MATLAB* vidē, kas savienota ar bioreaktora *SCADA* sistēmu, izmantojot *OPC* serveri, iespējojot reāllaika slēgtās cilpas vadību.

MPC vadība tika sākta pēc metanola adaptācijas fāzes (8–10 h pēc indukcijas). Izvēlētais šūnu augšanas ātruma iestatījums $\mu_{set}(t)$ tika piemērots pakāpeniski: 0,04 h^{-1} (0–12 h), 0,02 h^{-1} (12–24 h) un 0,01 h^{-1} (24-36 h), līdzsvarojot produktivitāti un šūnu stresu.

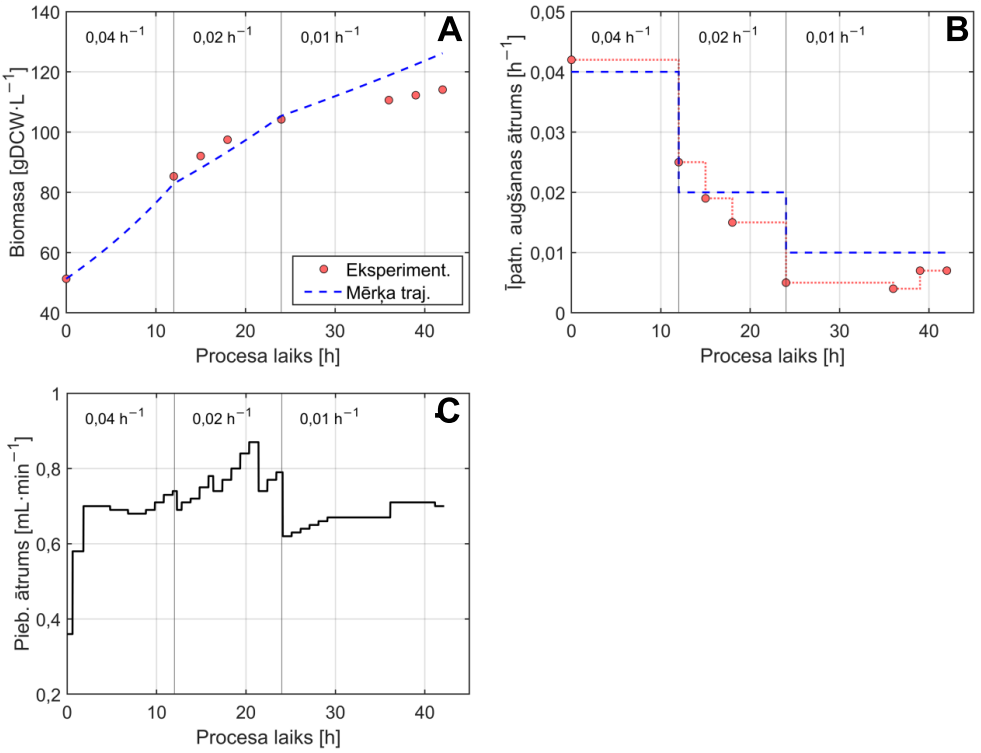
Lai novērtētu hibrīdā *MPC* ietvara praktisko lietojamību, tika veikta eksperimentāla validācija, kontrolējot metanola piebarošanas ātrumu reālā fermentācijas procesā. Tas ļāva novērtēt sistēmas spēju reāllaikā prognozēt un regulēt galvenos fermentācijas procesa parametrus (4.2. att.).



4.2. att. Hibrīdās *MPC* vadības prognozētās un eksperimentāli noteiktās biomasas un produkta koncentrācijas validācijas fermentācijā.

Biomasas prognozēšanas precizitāte var definēt kā labu, tomēr modelis pārvērtēja biomasas pieaugumu sākotnējās 8–12 stundās pēc indukcijas, kamēr šūnas adaptējas metanola uzņemšanai, un atkal fermentācijas beigās. Manuālas trajektorijas korekcijas izrādījās nepieciešamas fermentācijas beigu daļā un tika veiktas modeļa apmācības laikā, pamatojoties uz eksperimentālajiem datiem, un tā rezultātā kopējais biomasas *NRMSE* sasniedza 6,5 %. Līdzīgi arī produkta koncentrācija tika konsekventi pārvērtēta, uzrādot kļūdu 14,6 % apmērā [57].

Neskatoties uz šiem prognozēšanas ierobežojumiem, hibrīdā *MPC* vadības veiktspēja kopumā bija noturīga. Sistēma veiksmīgi ģenerēja substrāta piebarošanas profilus, kas uzturēja vēlamo īpatnējo augšanas ātrumu, demonstrējot efektīvu regulāciju pat modelēšanas neprecizitāšu klātbūtnē. Kontroliera izvēlētās šūnu biomasas izsekošanas veiktspēja redzama 4.3. attēlā.



4.3. att. Mērķa un eksperimentālās biomasa augšanas (A), īpatnējā augšanas ātruma (B) un substrāta piebarošanas ātruma (C) grafiki. Vertikālās līnijas norāda μ izmaiņas.

Hibrīdā *MPC* sistēma demonstrēja labu sniegumu, izsekojot īpatnējā augšanas ātruma iestatījumam visā fermentācijas procesā ar vidējo kļūdu 10,6 % *NRMSE*. Tomēr neliela novirze no mērķa biomasa trajektorijas tika novērota pēdējās 12 stundās, īpaši pēc tam, kad augšanas ātrums tika samazināts līdz 0,01 h^{-1}. Šī novirze, visticamāk, ir saistīta ar metanola citotoksisko efektu uz šūnām, kas var kavēt šūnu metabolismu un biomasas veidošanos fermentācijas beigu

fāzē. Šī tendence ir atspoguļota arī augšanas ātruma izsekošanas grafikā, kur mērītais īpatnējais augšanas ātrums šajā periodā pastāvīgi bija zem mērķa.

MPC sistēmas veiktspēju varētu uzlabot vēl vairāk, veicot visaptverošāku procesa parametru izpēti plašākā diapazonā. Sistemātiska plašāka kultivācijas apstākļu diapazona pārbaude, tostarp piebarošanas ātruma trajektorijas, indukcijas ilgums un īpatnējā augšanas ātruma iestatījumi, palīdzētu identificēt optimālas vadības stratēģijas. Šī papildu informācija uzlabotu kontroliera spēju tikt galā ar procesa mainīgumu un traucējumiem, tādējādi uzlabojot tā robustumu un pielāgojamību dažādos fermentācijas scenārijos. Turklāt visaptverošāks procesa modelis atbalstītu paplašinātas *MPC* lietojumprogrammas, piemēram, optimālo piebarošanas ātrumu novērtēšanu, lai maksimizētu procesa produktivitāti.

Kopumā, neskatoties uz nelielām neatbilstībām procesa beigās, *MPC* sistēma lielākajā daļā fermentācijas uzturēja precīzu augšanas regulāciju, izceļot tās robustumu, uzticamību un potenciālo lietojumu rekombinantu *P. pastoris* fermentāciju vadībā.

SECINĀJUMI

1. Reāllaika sensoru signālu priekšapstrāde būtiski uzlabo signāla kvalitāti – metanola sensora svārstības tika samazinātas par 63 %, bet dielektriskās spektroskopijas trokšņu līmenis par 33 %. Turklāt permitivitātes anomāliju noteikšanas un novēršanas algoritms uzrādīja 79 % precizitāti, stiprinot sensoru uzticamību monitoringa un vadības lietojumos.
2. Modelējot *P. pastoris* fermentāciju procesus, hibrīdie modeļi precizitātē pārspēja mehānistiskās un datos balstītās pieejas, sasniedzot zemāko testa *NRMSE* (11,1 ± 2,6 %), salīdzinot ar attiecīgi 39,2 ± 18,9 % un 28,2 ± 5,4 %.
3. Pielāgojot vēsturisko hibrīdo procesa modeli Qβ datu kopai ar 0,6 relatīvo mācīšanās ātrumu, tika uzrādīts labāks rezultāts nekā modeļa apmācībā no nulles, samazinot vidējo testa kļūdu un novirzi (no 9,90 ± 7,38 % līdz 5,31 ± 0,34 %) ar mazāku iterāciju skaitu (8820 pret 13 070) un demonstrējot testa *NRMSE* 4,35 %, pārneses mācīšanās rezultātā nodrošinot precīzas modeļa prognozes ar ierobežotu datu apjomu no tikai trīs eksperimentiem.
4. PI piebarošanas ātruma kontrolieris uzrādīja vidēju precizitāti metanola koncentrācijas regulācijā ar aprēķinātajām *NRMSE* vērtībām 28 % un 63 % pie 1 [$g \cdot L^{-1}$] un 2 [$g \cdot L^{-1}$], attiecīgi. Tā parametru jutīgums uzsver nepieciešamību pēc automatizētas parametru pielāgošanas procedūras un *in situ* sensora iekļaušanas, lai uzlabotu veiktspēju, paātrinot sensora reakcijas laiku un uzlabojot signāla kvalitāti dinamiskos fermentācijas apstākļos.
5. Hibrīdais *MPC* demonstrēja robustu *P. pastoris* fermentācijas vadību, saglabājot vēlamo īpatnējo augšanas ātrumu ar vidējo izsekošanas kļūdu 10,6 % *NRMSE*, neskatoties uz modelēšanas neprecizitātēm. Tā kā MPC veiktspēja lielā mērā bija atkarīga no procesa modeļa kvalitātes, šie rezultāti uzsver, ka augsta modeļa kvalitāte ir kritiska veiksmīgai *MPC* ieviešanai biotehnoloģiskos procesos un pamato turpmāku hibrīdā *MPC* izpēti *P. pastoris* fermentāciju kontekstā.

LITERATŪRAS SARAKSTS

1. Jayakrishnan, A.; Wan Rosli, W. R.; Tahir, A. R. M.; Razak, F. S. A.; Kee, P. E.; Ng, H. S.; Chew, Y. L.; Lee, S. K.; Ramasamy, M.; Tan, C. S.; et al. Evolving Paradigms of Recombinant Protein Production in Pharmaceutical Industry: A Rigorous Review. *Sci* **2024**, *6*.
2. Tripathi, N.K.; Shrivastava, A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Front Bioeng Biotechnol* **2019**, *7*, doi:10.3389/fbioe.2019.00420.
3. Mordor Intelligence. Biologics Market Size & Share Analysis – Growth Trends and Forecast (2026–2031). **2025**.
4. Bolmanis, E.; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. Pastoris* GS115 Mut+ Bioreactor Process under Controlled Residual Methanol Concentration. *Bioprocess Biosyst Eng* **2022**, 45, 1447–1463, doi:10.1007/s00449-022-02754-4.
5. Letourneur, O.; Watelet, B. HBcAg Expression and Diagnostic and Therapeutic Uses. EP Patent No. EP1256804A1, **2001**.
6. Freivalds, J.; Dislers, A.; Ose, V.; Skrastina, D.; Cielens, I.; Pumpens, P.; Sasnauskas, K.; Kazaks, A. Assembly of Bacteriophage Qβ Virus-like Particles in Yeast *Saccharomyces Cerevisiae* and *Pichia Pastoris*. *J Biotechnol* **2006**, *123*, 297–303, doi:10.1016/j.jbiotec.2005.11.013.
7. Spohn, G.; Keller, I.; Beck, M.; Grest, P.; Jennings, G. T.; Bachmann, M. F. Active Immunization with IL-1 Displayed on Virus-like Particles Protects from Autoimmune Arthritis. *Eur J Immunol* **2008**, *38*, 877–887, doi:10.1002/eji.200737989.
8. Vuitika, L.; Côrtes, N.; Malaquias, V. B.; Silva, J. D. Q.; Lira, A.; Prates-Syed, W. A.; Schimke, L. F.; Luz, D.; Durães-Carvalho, R.; Balan, A.; et al. A Self-Adjuvanted VLPs-Based Covid-19 Vaccine Proven Versatile, Safe, and Highly Protective. *Sci Rep* **2024**, *14*, doi:10.1038/s41598-024-76163-w.
9. Fraser, R. Z.; Shitut, M.; Agrawal, P.; Mendes, O.; Klapholz, S. Safety Evaluation of Soy Leghemoglobin Protein Preparation Derived From *Pichia Pastoris*, Intended for Use as a Flavor Catalyst in Plant-Based Meat. *Int J Toxicol* **2018**, *37*, 241–262, doi:10.1177/1091581818766318.
10. Reyes, T. F.; Chen, Y.; Fraser, R. Z.; Chan, T.; Li, X. Assessment of the Potential Allergenicity and Toxicity of *Pichia* Proteins in a Novel Leghemoglobin Preparation. *Regulatory Toxicology and Pharmacology* **2021**, *119*, 104817, doi:10.1016/j.yrtph.2020.104817.
11. Cregg, J. M.; Cereghino, J. L.; Shi, J.; Higgins, D. R. Recombinant Protein Expression in *Pichia Pastoris*. *Mol Biotechnol* **2000**, *16*, 23–52, doi:10.1385/MB:16:1:23.
12. Potvin, G.; Ahmad, A.; Zhang, Z. Bioprocess Engineering Aspects of Heterologous Protein Production in *Pichia Pastoris*: A Review. *Biochem Eng J* **2012**, *64*, 91–105, doi:10.1016/j.bej.2010.07.017.

13. Yang, Z.; Zhang, Z. Engineering Strategies for Enhanced Production of Protein and Bio-Products in *Pichia Pastoris*: A Review. *Biotechnol Adv* **2018**, *36*, 182–195, doi:10.1016/j.biotechadv.2017.11.002.
14. Karbalaei, M.; Rezaee, S. A.; Farsiani, H. Pichia Pastoris: A Highly Successful Expression System for Optimal Synthesis of Heterologous Proteins. *J Cell Physiol* **2020**, *235*, 5867–5881, doi:10.1002/jcp.29583.
15. Macauley-Patrick, S.; Fazenda, M. L.; McNeil, B.; Harvey, L. M. Heterologous Protein Production Using the *Pichia Pastoris* Expression System. *Yeast* **2005**, *22*, 249–270, doi:10.1002/yea.1208.
16. U. S. Food and Drug Administration *GRAS Notice No. 204. Phospholipase C Enzyme Preparation from Pichia Pastoris Expressing a Heterologous Phospholipase C Gene*; **2006**.
17. Moraes, L. M. P. de; Marques, H. F.; Reis, V. C. B.; Coelho, C. M.; Leitão, M. de C.; Galdino, A. S.; Porto de Souza, T. P.; Piva, L. C.; Perez, A. L. A.; Trichez, D.; et al. Applications of the Methylotrophic Yeast *Komagataella Phaffii* in the Context of Modern Biotechnology. *Journal of Fungi* **2024**, *10*.
18. Mears, L.; Stocks, S. M.; Sin, G.; Gernaey, K. V. A Review of Control Strategies for Manipulating the Feed Rate in Fed-Batch Fermentation Processes. *J Biotechnol* **2017**, *245*, 34–46, doi:10.1016/j.jbiotec.2017.01.008.
19. Lindskog, E. K. The Upstream Process: Principal Modes of Operation; Elsevier Ltd., **2018**.
20. Lim, H. C.; Shin, H.S. Introduction to Fed-Batch Cultures. In *Fed-Batch Cultures*; Cambridge University Press, **2018**.
21. Villadsen, J.; Patil, K. R. Optimal Fed-Batch Cultivation When Mass Transfer Becomes Limiting. *Biotechnol Bioeng* **2007**, *98*, 706–710, doi:10.1002/bit.21451.
22. Grigs, O.; Galvanauskas, V.; Dubencovs, K.; Vanags, J.; Suleiko, A.; Berzins, T.; Kunga, L. Model Predictive Feeding Rate Control in Conventional and Single-Use Lab-Scale Bioreactors: A Study on Practical Application. *Chem Biochem Engineer Quart* **2016**, *30*, 47–60, doi:10.15255/CABEQ.2015.2212.
23. Mears, L.; Stocks, S. M.; Albaek, M. O.; Sin, G.; Gernaey, K. V. Mechanistic Fermentation Models for Process Design, Monitoring, and Control. *Trends Biotechnol* **2017**, *35*, 914–924.
24. Cinar, A.; Parulekar, S. J.; Undey, C.; Birol, G. *Batch Fermentation*; CRC Press, **2003**.
25. Bailey, J. E. Mathematical Modeling and Analysis in Biochemical Engineering: Past Accomplishments and Future Opportunities. *Biotechnol Prog* **1998**, *14*, 8–20, doi:10.1021/bp9701269.
26. Noll, P.; Henkel, M. History and Evolution of Modeling in Biotechnology: Modeling & Simulation, Application and Hardware Performance. *Comput Struct Biotechnol J* **2020**, *18*, 3309–3323, doi:10.1016/j.csbj.2020.10.018.
27. Helleckes, L. M.; Hemmerich, J.; Wiechert, W.; von Lieres, E.; Grünberger, A. Machine Learning in Bioprocess Development: From Promise to Practice. *Trends Biotechnol* **2023**, *41*, 817–835, doi:10.1016/j.tibtech.2022.10.010
28. Pinto, J.; Mestre, M.; Ramos, J.; Costa, R. S.; Striedner, G.; Oliveira, R. A General Deep Hybrid Model for Bioreactor Systems: Combining First Principles with Deep Neural

Networks. *Comput Chem Eng* **2022**, *165*, 107952, doi:10.1016/j.compchemeng.2022.107952.
29. Agharafeie, R.; Ramos, J. R. C.; Mendes, J. M.; Oliveira, R. From Shallow to Deep Bioprocess Hybrid Modeling: Advances and Future Perspectives. *Fermentation* **2023**, *9*, 1–22, doi:10.3390/fermentation9100922.
30. Pinto, J.; Ramos, J. R. C.; Costa, R. S.; Oliveira, R. Hybrid Deep Modeling of a GS115 (Mut^+) *Pichia Pastoris* Culture with State–Space Reduction. *Fermentation* **2023**, *9*, doi:10.3390/fermentation9070643.
31. Pinto, J.; Ramos, J. R. C.; Costa, R. S.; Rossell, S.; Dumas, P.; Oliveira, R. Hybrid Deep Modeling of a CHO-K1 Fed-Batch Process: Combining First-Principles with Deep Neural Networks. *Front Bioeng Biotechnol* **2023**, *11*, doi:10.3389/fbioe.2023.1237963.
32. Ramos, J. R. C.; Pinto, J.; Poiares-Oliveira, G.; Peeters, L.; Dumas, P.; Oliveira, R. Deep Hybrid Modeling of a HEK293 Process: Combining Long Short-Term Memory Networks with First Principles Equations. *Biotechnol Bioeng* **2024**, *121*, 1554–1568, doi:10.1002/bit.28668.
33. Bolmanis, E.; Dubencovs, K.; Suleiko, A.; Vanags, J. Model Predictive Control – A Stand Out among Competitors for Fed-Batch Fermentation Improvement. *Fermentation* **2023**, *9*, 206, doi:10.3390/fermentation9030206.
34. Schwenzer, M.; Ay, M.; Bergs, T.; Abel, D. Review on Model Predictive Control: An Engineering Perspective. *J Adv Manuf Technol* **2021**, *117*, 1327–1349.
35. Jelsch, M.; Roggo, Y.; Kleinebudde, P.; Krumme, M. Model Predictive Control in Pharmaceutical Continuous Manufacturing: A Review from a User's Perspective. *Eur J Pharm* **2021**, *159*, 137–142, doi:10.1016/j.ejpb.2021.01.003.
36. Craven, S.; Whelan, J.; Glennon, B. Glucose Concentration Control of a Fed-Batch Mammalian Cell Bioprocess Using a Nonlinear Model Predictive Controller. *J Process Control* **2014**, *24*, 344–357, doi:10.1016/j.jprocont.2014.02.007.
37. Dubencovs, K.; Suleiko, A.; Sile, E.; Petrovskis, I.; Akopjana, I.; Suleiko, A.; Galvanauskas, V.; Tars, K.; Vanags, J. The Application of Adaptive Model Predictive Control for Fed-Batch *Escherichia Coli* BL21 (DE3) Cultivation and Biosynthesis of Recombinant Proteins. *Fermentation* **2023**, *9*, doi:10.3390/fermentation9121015.
38. Ashoori, A.; Moshiri, B.; Khaki-Sedigh, A.; Bakhtiari, M. R. Optimal Control of a Nonlinear Fed-Batch Fermentation Process Using Model Predictive Approach. *J Process Control* **2009**, *19*, 1162–1173, doi:10.1016/j.jprocont.2009.03.006.
39. Reyes, S. J.; Durocher, Y.; Pham, P. L.; Henry, O. Modern Sensor Tools and Techniques for Monitoring, Controlling, and Improving Cell Culture Processes. *Processes* **2022**, *10*, 189, doi:10.3390/pr10020189.
40. Niu, S. S.; Xiao, D. *Process Control*; Springer International Publishing, **2022**.
41. Vojinović, V.; Cabral, J. M. S.; Fonseca, L. P. Real-Time Bioprocess Monitoring: Part I: *In Situ* Sensors. *Sens Actuators B Chem* **2006**, *114*, 1083–1091, doi:10.1016/j.snb.2005.07.059.
42. Sundström, H.; Enfors, S. O. Software Sensors for Fermentation Processes. *Bioprocess Biosyst Eng* **2008**, *31*, 145–152, doi:10.1007/s00449-007-0157-5.

43. Panda, S.; Mehlawat, S.; Dhariwal, N.; Sanger, A.; Kumar, A. Comprehensive Review on Gas Sensors: Unveiling Recent Developments and Addressing Challenges. *Materials Science and Engineering: B* **2024**, *308*.
44. Biechele, P.; Busse, C.; Solle, D.; Scheper, T.; Reardon, K. Sensor Systems for Bioprocess Monitoring. *Eng Life Sci* **2015**, *15*, 469–488, doi:10.1002/elsc.201500014.
45. Grigs, O.; Bolmanis, E.; Galvanauskas, V. Application of *In-Situ* and Soft-Sensors for Estimation of Recombinant *P. Pastoris* GS115 Biomass Concentration: A Case Analysis of HBcAg (Mut^+) and HBsAg (Mut^S) Production Processes under Varying Conditions. *Sensors* **2021**, *21*, 1268, doi:10.3390/s21041268.
46. Beiroti, A.; Hosseini, S. N.; Norouzian, D.; Aghasadeghi, M. R.; Jajan, L. H. G. Development of Soft Sensors for Online Biomass Prediction in Production of Hepatitis B Vaccine. *Biointerface Res Appl Chem* **2023**, *13*, 1–13, doi:10.33263/BRIAC132.194.
47. Mandenius, C. F.; Gustavsson, R. Mini-Review: Soft Sensors as Means for PAT in the Manufacture of Bio-Therapeutics. *J Chem Technol Biotechnol* **2015**, *90*, 215–227, doi:10.1002/jctb.4477.
48. Tsopanoglou, A.; Jiménez del Val, I. Moving towards an Era of Hybrid Modelling: Advantages and Challenges of Coupling Mechanistic and Data-Driven Models for Upstream Pharmaceutical Bioprocesses. *Curr Opin Chem Eng* **2021**, *32*, doi:10.1016/j.coche.2021.100691.
49. von Stosch, M.; Glassey, J. *Hybrid Modeling in Process Industries*; CRC Press, **2018**.
50. Schweidtmann, A. M.; Zhang, D.; von Stosch, M. A Review and Perspective on Hybrid Modeling Methodologies. *Digit Chem Eng* **2024**, *10*, 100136, doi:10.1016/j.dche.2023.100136.
51. Roeva, O.; Slavov, T. PID Controller Tuning Based on Metaheuristic Algorithms for Bioprocess Control. *Biotechnol Biotechnol Equip* **2012**, *26*, 3267–3277, doi:10.5504/BBEQ.2012.0065.
52. Nagy, Z. K. Model Based Control of a Yeast Fermentation Bioreactor Using Optimally Designed Artificial Neural Networks. *Chem Eng J* **2007**, *127*, 95–109, doi:10.1016/j.cej.2006.10.015.
53. Eaton, J. W.; Rawlings, J. B. Model-Predictive Control of Chemical Processes. *Chem Eng Sci* **1992**, *47*, 705–720, doi:10.1016/0009-2509(92)80263-C.
54. Findeisen, R.; Allgöwer, F. An Introduction to Nonlinear Model Predictive Control. In Proceedings of the 21st Benelux Meeting on Systems and Control; Veldhoven, **2002**.
55. Horta, A. C. L.; Da Silva, A. J.; Sargo, C. R.; Cavalcanti-Montaño, I. D.; Galeano-Suarez, C. A.; Velez, A. M.; Santos, M. P.; Gonçalves, V. M.; Giordano, R. C.; Zangirolami, T. C. 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. *Braz J Chem Eng* **2015**, *32*, 821–829, doi:10.1590/0104-6632.20150324s00003534.
56. Bolmanis, E.; Bogans, J.; Akopjana, I.; Suleiko, A.; Kazaka, T.; Kazaks, A. Production and Purification of Soy Leghemoglobin from *Pichia Pastoris* Cultivated in Different Expression Media. *Processes* **2023**, *11*, 3215, doi:10.3390/pr11113215.

57. Bolmanis, E.; Galvanauskas, V.; Grigs, O.; Vanags, J.; Kazaks, A. Leveraging Historical Process Data for Recombinant *P. Pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development. *Fermentation* **2025**, *11*, 411, doi:10.3390/fermentation11070411.
58. Pentjuss, A.; Bolmanis, E.; Suleiko, A.; Didrihsone, E.; Suleiko, A.; Dubencovs, K.; Liepins, J.; Kazaks, A.; Vanags, J. Pichia Pastoris Growth – Coupled Heme Biosynthesis Analysis Using Metabolic Modelling. *Sci Rep* **2023**, *13*, 15816, doi:10.1038/s41598-023-42865-w.
59. Grigs, O.; Bolmanis, E.; Kazaks, A. HBsAg Production in Methanol Controlled *P. Pastoris* GS115 Mut^S Bioreactor Process. *Key Eng Mater* **2021**, *903*, 40–45, doi:10.4028/www.scientific.net/KEM.903.40.
60. Bolmanis, E.; Uhlendorff, S.; Pein-Hackelbusch, M.; Galvanauskas, V.; Grigs, O. Anomaly Detection and Removal Strategies for In-Line Permittivity Sensor Signal Used in Bioprocesses. *Front. Bioeng. Biotechnol.* **2025**, *13*, doi: 10.3389/fbioe.2025.1609369.
61. Ren, H. T.; Yuan, J. Q.; Bellgardt, K. H. Macrokinetic Model for Methylotrophic *Pichia Pastoris* Based on Stoichiometric Balance. *J Biotechnol* **2003**, *106*, 53–68, doi:10.1016/j.jbiotec.2003.08.003.
62. Jackson, J. V.; Edwards, V. H. Kinetics of Substrate Inhibition of Exponential Yeast Growth. *Biotechnol Bioeng* **1975**, *17*, 943–964, doi:10.1002/bit.260170702.
63. Zhu, X.; Rehman, K. U.; Wang, B.; Shahzad, M. Modern Soft-Sensing Modeling Methods for Fermentation Processes. *Sensors (Switzerland)* **2020**, *20*.
64. Wang, B.; Nie, Y.; Zhang, L.; Song, Y.; Zhu, Q. An Soft-Sensor Method for the Biochemical Reaction Process Based on LSTM and Transfer Learning. *Alex Eng J* **2023**, *81*, 170–177, doi:10.1016/j.aej.2023.09.007.
65. Rogers, A. W.; Vega-Ramon, F.; Yan, J.; del Río-Chanona, E. A.; Jing, K.; Zhang, D. A Transfer Learning Approach for Predictive Modeling of Bioprocesses Using Small Data. *Biotechnol Bioeng* **2022**, *119*, 411–422, doi:10.1002/bit.27980.
66. Gurramkonda, C.; Adnan, A.; Gäbel, T.; Lünsdorf, H.; Ross, A.; Nemani, S. K.; Swaminathan, S.; Khanna, N.; Rinas, U. Simple High-Cell Density Fed-Batch Technique for High-Level Recombinant Protein Production with *Pichia Pastoris*: Application to Intracellular Production of Hepatitis B Surface Antigen. *Microb Cell Fact* **2009**, *8*, 1–8, doi:10.1186/1475-2859-8-13.
67. Wu, Z.; Christofides, P. D.; Wu, W.; Wang, Y.; Abdullah, F.; Alnajdi, A.; Kadakia, Y. A Tutorial Review of Machine Learning-Based Model Predictive Control Methods. *Rev Chem Eng* **2024**.
68. Rathore, A. S.; Mishra, S.; Nikita, S.; Priyanka, P. Bioprocess Control: Current Progress and Future Perspectives. *Life* **2021**, *11*, 557, doi:10.3390/life11060557.

APPENDICES / PIELIKUMI

Model Predictive Control – A Stand Out among Competitors for Fed-Batch Fermentation Improvement

Emils Bolmanis, Konstantins Dubencovs, Arturs Suleiko, Juris Vanags
Fermentation **2023**, *9*, 3, DOI: 10.3390/fermentation9030206

E.B. input: Conceptualization, writing – original draft preparation, writing – review and editing, visualization, formal analysis, investigation.

fermentation

MDPI

Review

Model Predictive Control—A Stand Out among Competitors for Fed-Batch Fermentation Improvement

Emils Bolmanis [1,2,3], Konstantins Dubencovs [2,3,4], Arturs Suleiko [2,4] and Juris Vanags [4,*]

[1] Latvian Biomedical Research and Study Centre, Ratsupites Street 1-k1, LV-1067 Riga, Latvia
[2] Laboratory of Bioengineering, Latvian State Institute of Wood Chemistry, Dzerbenes Street 27, LV-1006 Riga, Latvia
[3] Institute of General Chemical Engineering, Faculty of Materials Science and Applied Chemistry, Riga Technical University, Paula Valdena Street 3, LV-1658 Riga, Latvia
[4] Bioreactors.net AS, Dzerbenes Street 27, LV-1006 Riga, Latvia
* Correspondence: juris.vanags@bioreactors.net

Abstract: The fed-batch cultivation is in many ways a benchmark for fermentation processes, and it has been an attractive choice for the biotechnological production of various products in the past decades. The majority of biopharmaceuticals that are presently undergoing clinical trials or are available on the market are manufactured through fed-batch fermentations. A crucial process parameter in fed-batch cultivations is the substrate feed rate, which directly influences the overall process productivity, product quality and process repeatability; henceforth, effective control of this parameter is imperative for a successful fed-batch fermentation process. Two distinct control strategies can be distinguished—open-loop and closed-loop (feedback) control. Each of these methods has its own set of benefits, limitations and suitability for specific bioprocesses. This article surveys and compares the most popular open- and closed-loop methods for substrate feed rate control in fed-batch fermentations. Emphasis is placed on model-predictive feed rate control (MPC)—a stand out among other methods that offers a promising application perspective. The authors also demonstrate a practical example of the implementation of a robust, flexible MPC solution that is suitable for various cultures and runs on standard computer hardware, thus overcoming one of the main reported MPC drawbacks—high computational requirements.

Keywords: fermentation; bioreactors; fed batch; feed rate control; model-based control; PID control; artificial neural network (ANN); fuzzy logic; model predictive control (MPC)

Citation: Bolmanis, E.; Dubencovs, K.; Suleiko, A.; Vanags, J. Model Predictive Control—A Stand Out among Competitors for Fed-Batch Fermentation Improvement. *Fermentation* **2023**, *9*, 206. https://doi.org/10.3390/fermentation9030206

Academic Editors: Francesca Raganati and Alessandra Procentese

Received: 21 December 2022
Revised: 21 February 2023
Accepted: 21 February 2023
Published: 22 February 2023

1. Introduction

Three main modes exist for operating a fermentation process, namely batch, fed batch and continuous cultivation. Each operational mode has its own set of advantages and disadvantages. To create a cost-effective process, bioengineers must take into account various factors, such as media and supplement costs, process duration, biomass growth and viability, product yield and product quality. Moreover, the concentrations of growth-associated nutrients and by-products in the culture medium are also important considerations. Therefore, during process development, bioengineers must carefully weigh the options of batch, fed batch or continuous cultivation approaches [1].

In a batch process, a fixed volume of a medium is introduced to a microbial culture (Figure 1, Batch). As the microorganisms multiply, they use up the nutrients in the medium and generate by-products, which gradually slows their growth and causes them to enter the stationary growth phase. In contrast, continuous fermentation involves a constant flow of fresh cultivation medium into the bioreactor, typically at a constant rate, while the spent medium and cells are simultaneously collected, thereby maintaining a constant medium level in the bioreactor (Figure 1, continuous). This process enables the replenishment of

consumed nutrients and removal of toxic metabolites from the culture while keeping the culture volume constant.

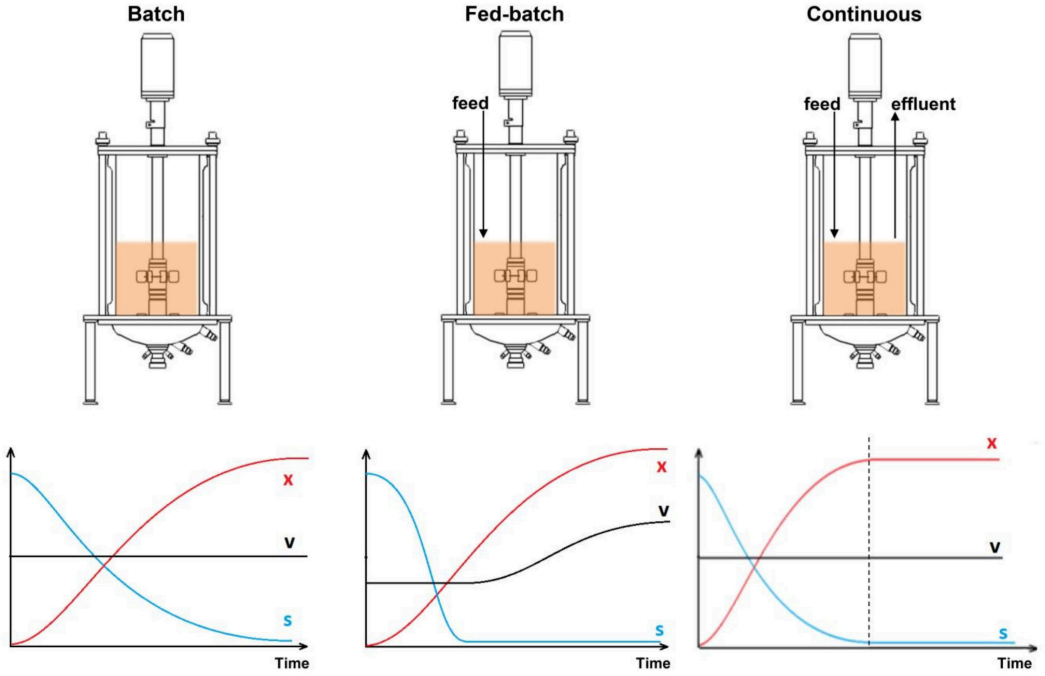


Figure 1. Modes of operation for cell cultivation. X—biomass, V—volume, S—substrate concentration. Vertical dashed line indicates the onset of the continuous process, as usually it starts with a batch or a fed-batch phase.

Fed-batch fermentation is comparable to the batch process, but after the microorganisms consume all the substrate and nutrients in the cultivation medium, additional substrate and nutrients are added to the bioreactor (Figure 1, fed batch), thereby extending the active growth phase. The feed rate can be adjusted to regulate both the microorganism growth or the product formation rate. When designing a feed profile, it is essential to consider the maximum specific substrate utilization rate for the employed microbial strain. If more substrate is provided to the bioreactor than the cells can utilize, it will eventually start to accumulate and, upon reaching high concentrations, could potentially impede biomass growth, either directly or by promoting the synthesis of inhibitory metabolites. A similar situation may arise when the product has an inhibitory effect on the cells; thus, it gradually slows growth as the product accumulates in the reactor. A fed-batch operation typically results in significant biomass accumulation, making it particularly beneficial for bioprocesses that seek to generate high cell densities and product titers, especially when the target product is positively linked with microbial growth. Additionally, because the substrate is not overfed during the process and by-product accumulation is limited, both product and biomass yields per fed substrate are elevated [2]. In general, a higher product titer can be obtained per fermentation using a fed-batch operation, which practically increases the product space time yields when compared to batch processes. To illustrate, Muradi et al. improved bioethanol productivity 1.8 fold in *S. cerevisiae* fermentation, in comparison of a 72 h long batch versus a 72 h long fed-batch fermentation, thus increasing the overall process productivity [3].

Fed batch has become a benchmark for bioprocesses, and it has been a popular choice for production for several decades due to its robustness and elevated product yields per run and per time. Currently, most biotherapeutics in the market or clinical phases are produced using fed-batch fermentations [4]. Moreover, a vast range of amino acids, antibiotics, enzymes, solvents, vitamins and other products are now produced using this approach [2]. Human insulin was historically the first "golden molecule" produced in the biotech industry using an *E. coli* fed-batch cultivation. In this process, additional substrate is fed into the reactor at a rate to maintain low growth rates, which in turn suppresses by-product (acetate) formation and allows the culture to grow to high cell densities [5]. The fed-batch strategy is also commonly used in large-scale cultivations to address limiting factors such as mass and heat transfer. For example, the substrate feed is adjusted to maintain sufficient mixing, oxygenation and cooling conditions. This helps to ensure the optimal growth and productivity of the microbial culture throughout the fermentation process [6]. Therefore, the fed-batch mode of operation is commonly used in the bioprocess industry since it can overcome most of the limitations of batch and continuous processes. However, achieving an efficient and consistent fed-batch operation can be challenging due to the need to maintain an optimal substrate feed rate.

The rate at which substrate is fed into the bioreactor has a significant impact on both the metabolic processes and the volume changes in the system. The concentration of substrate in the growth medium is a critical factor that directly influences various aspects of the fermentation process. Specifically, it affects the rate of cell growth, oxygen utilization and the formation rates of both product and by-products. Therefore, maintaining an optimal substrate concentration is essential to ensure the efficient use of nutrients, minimize the production of undesirable by-products and maximize the yield of desired products. In turn, changes in the volume dynamics affect the concentration of all the elements in the bioreactor, and the medium's viscosity can also impact both oxygen uptake and transfer rates. If the substrate feed rate is too low, the process will not reach its maximum productivity. If it is too high, it could result in an overflow metabolism and an increase in by-product formation. Additionally, during exponential growth, changes in metabolic processes, volume dynamics and other potential disturbances cause the maximum substrate uptake rate value to continuously change in a nonlinear manner, making the task of controlling the feed rate a challenging one [7].

Two distinct groups of feed rate control principles can be distinguished—open- and closed-loop control. Each of these approaches has its own strengths and limitations. However, the ultimate goal of any strategy is to achieve a balance between system implementation costs and control efficiency. This balance can affect the overall cost effectiveness of the process. Since microorganisms are fairly complex living systems, process productivity, product characteristics and the overall nature of each fermentation run can vary even when the same process parameters are used. As a result, feedback-based control methods are typically used in fed-batch fermentations to automatically adjust the substrate feed rate profile. This approach can provide significant benefits for overall process reproducibility by accommodating the dynamic and nonlinear changes in the system and ensuring that the nutrient demand is met consistently throughout the fermentation process.

This study examines some of the widely used substrate feed rate control strategies in fed-batch fermentations by comparing them based on the reports available in the literature. The analysis of practical examples from the authors' previous experience provides a more comprehensive understanding of various practical aspects when selecting a particular substrate feed rate control strategy. Additionally, the paper presents a practical example of developing a robust and flexible model predictive control (MPC) solution, which runs on standard hardware. The study highlights the strengths of the MPC method and demonstrates how it can overcome one of its primary drawbacks—high computational demands. The presented example showcases the possibility of achieving an efficient and effective substrate feed rate control while also reducing computational demands, making the process more practical and cost effective.

2. Open-Loop Feed Rate Control

In an open-loop feed rate controller (also known as a nonfeedback controller), the control action taken is not dependent on the current process output. This type of control does not use feedback to assess whether the output has achieved the desired input or process set point. In fermentation systems, this approach is typically implemented in the form of a precalculated substrate feed rate time profile. The controller relies solely on the predetermined feed rate profile, which is based on the expected nutrient demand of the microorganisms and does not take into account any deviations in the actual nutrient demand or fluctuations in the system. Open-loop control can be a simple and cost-effective solution for certain systems, but it may not be as accurate or reliable as the feedback-based methods in more complex systems. The system is schematically illustrated in Figure 2.

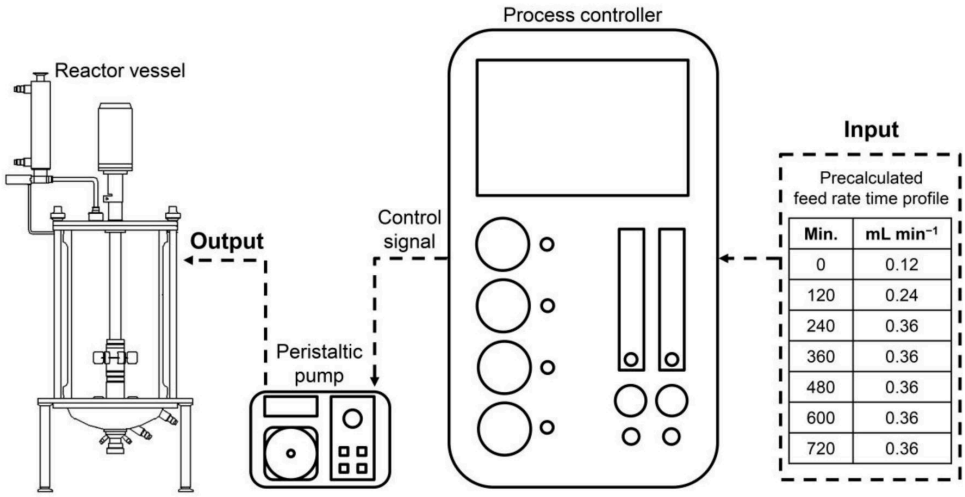


Figure 2. Open-loop feed rate control in a bioprocess.

Many industrial substrate-limited fermentation processes are operated in an open loop as this is the easiest way to organize a fed-batch process [8]. The feed rate time profile is calculated based on the initial conditions of the process. The first step is usually to calculate the initial feed rate at the start of the fed-batch phase. The equation for the initial feed rate (F_0) calculation is presented in Equation (1):

$$F_0 = \left(\frac{\mu}{Y_{X/S,max}} + m_S \right) \times \frac{V_0 X_0}{F_S} \text{ [L/h]} \quad (1)$$

where μ—specific cell growth rate (h^{-1}), $Y_{X/S,max}$—maximum cell biomass yield from substrate (g g^{-1}), m_S—specific cell maintenance coefficient (g g^{-1} h^{-1}), V_0—initial culture volume (L), X_0—initial cell biomass concentration (g L^{-1}) and F_S—substrate mass fraction in the feed solution (g g^{-1}) [9].

Once the initial feed rate has been determined, the remaining substrate feed rate time profile can be calculated using open-loop control methods. The specific approach used to calculate the feed rate profile will depend on the defined control goal of the fermentation process. There are several strategies that can be employed to determine the open-loop feed rate time profile for fed-batch fermentation. The choice of the appropriate strategy will depend on the specific goals of the fermentation process and the characteristics of the microorganisms being used.

A linear feed rate increase is the most commonly used feed rate profile, where the feed rate is gradually increased at a constant rate over time. A linear feed rate increase

can be implemented by determining the initial feed rate, the final feed rate and the length of the fermentation process, and then calculating the rate of increase in the feed rate. The mathematical expression of this method is presented in Equation (2).

The exponential feed rate increase involves an exponential increase in the feed rate over time. The exponential feed rate increase strategy can be more effective than the linear feed rate increase strategy in achieving a faster increase in product concentration, but it may not be suitable for all types of processes or for achieving certain control goals. In fed-batch fermentation processes, using an exponential feed rate is a simple, predetermined method to maintain a steady specific growth rate below a critical level in order to prevent the formation of undesired by-products. This approach allows for better control of the growth rate and the formation of the end product [10]. The mathematical expression of this method is presented in Equation (3):

$$F_{lin}(t) = F_0 + (F_{end} - F_0) \times (t/T) \text{ [L/h]} \quad (2)$$

$$F_{exp}(t) = F_0 \times e^{\mu t} \text{ [L/h]} \quad (3)$$

where F_0—initial feed rate [L/h], F_{end}—feed rate at the end of substrate feeding (L/h), t—time after starting feed (h), T—process end time (h), μ—specific cell growth rate (h^{-1}).

Some other strategies include step-wise feed rate change (a sudden change in the feed rate at a specific point in time), pulse feeding (adding a pulse of substrate to the reactor at certain intervals), feed-on-demand (adding substrate to the reactor only when it is needed, based on the current state of the process) and gradual feed rate decrease (gradually decreasing the feed rate over time). It is important to note that the best substrate feed rate time profile will depend on the specific process, the control goal and the available resources. Some strategies may be more appropriate for certain types of processes or for achieving certain control goals.

In addition to basic strategies, more advanced methods can be used to calculate substrate feed rate time profiles, such as following a selected specific growth rate time profile [11,12]. The specific growth rate (μ) is a crucial variable in biotechnological processes as it greatly impacts the physiological state of the microbial culture, the production of cell biomass and the quantity and quality of desired products [13,14]. It is a key factor that shapes the outcome of the fermentation process; hence, efficiently controlling this parameter may prove beneficial for process productivity and repeatability.

Open-loop feed rate profiles can also be calculated based on a process model (also called model-based control). Understandably, this method requires a robust process model, and its accuracy is dependent on the accuracy of the model itself. Several examples of the successful use of this method have been reported [15–17].

Although easy to implement, such systems are not able to predict or react to unexpected disturbances during cultivation. For example, some inhibitory factors can cause changes in the specific substrate utilization rate, and the selected cell growth rate can no longer be attained. This can lead to lowered product yields, higher batch-to-batch variability and potentially even to whole-batch discard. In some cases, unexpected deviations from the expected performance of a fed-batch bioprocess can be addressed by an operator who manually adjusts the feed rate profile. However, this approach depends heavily on the operator's experience and introduces a significant "human factor" into the process. In industries such as biopharmaceutical production, relying on manual adjustments may not be sustainable or compliant with current Good Manufacturing Practice guidelines. Therefore, it is important to develop and implement reliable feed rate control strategies that minimize the need for operator intervention and ensure consistent process performance.

In practice, the selected feed rate control strategy is very much dependent on the characteristics of the cultivation process in question. Sometimes high process productivity can be achieved with relatively simple feed rate control strategies, thus excluding the need to develop sophisticated and labor-intensive control schemes. One example, where an open-loop substrate feed profile can provide sufficient productivity, is the classic Invitrogen

corporations' *Pichia pastoris* cultivation protocol for recombinant protein production [18]. The methanol feed profile in this particular protocol suggests a step-wise feed rate increase, starting with a feed rate of 3.6 mL/h/L for the first 3–5 h, which is then doubled to 7.3 mL/h/L for 2 h and, finally, increased to 10.9 mL/h/L for the rest of the cultivation. Such a feed profile may not be entirely optimal from a maximum-productivity standpoint, for example, because with a constant feed rate, the amount of substrate per cells decreases during cultivation (cell biomass increases, feed rate remains constant), causing a decrease in the specific growth rate with time [9].

Regardless, this cultivation strategy is still able to produce high recombinant protein yields. We have reported a case in which this particular feed profile produced the highest yet reported recombinant hepatitis B core antigen virus-like particle (HBcAg VLPs) yield [19]. In this particular research, the effect of residual methanol in the concentration range of 0.1–2.0 g/L (usually associated with an increased recombinant protein productivity) on HBcAg VLP production was screened. The residual methanol concentration was controlled using a methanol sensor and by manipulating the methanol feed rate using a feedback PI-based feed rate control algorithm. Despite employing this closed-loop feed rate control method, the highest HBcAg VLP yield was produced in the cultivation conducted according to the Invitrogen cultivation guidelines with an open-loop feed profile, where no significant accumulation of methanol occurred and the methanol concentration was limiting for most of the induction phase (see Figure 3).

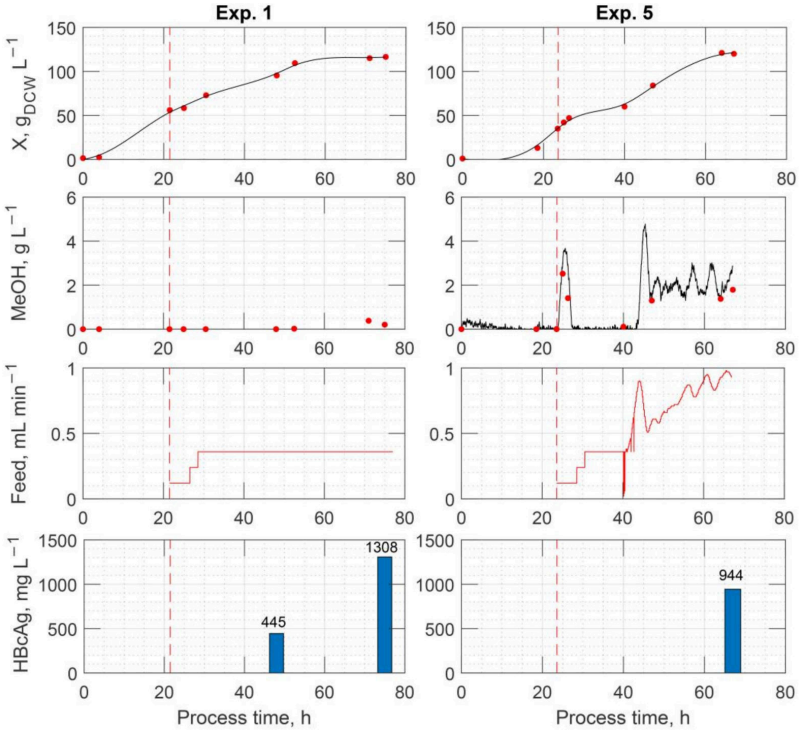


Figure 3. A comparison of two recombinant *P. pastoris* fermentation processes for HBcAg production employing open-loop and closed-loop (PI) methanol feed profiles. X—dry cell biomass, MeOH—methanol concentration, Feed—substrate feed rate, HBcAg—product concentration. Vertical dashed line indicates the onset of methanol induction phase. A significantly higher process productivity was achieved using a simple open-loop feed profile. Reprinted/adapted with permission from Ref. [19]. 2023, Springer Nature.

There are plenty of reports where the optimization of the Invitrogen protocol leads to elevated recombinant protein yields. The cultivation media composition, induction temperature, dissolved oxygen concentration (DO) and residual methanol concentration are most often the targets for product yield optimization in *P. pastoris* [9,20,21]. However, the substrate feed rate profile can also play a significant role in maximizing process productivity. For example, the length of the glycerol fed-batch phase and reached cell biomass concentration can have a significant effect on recombinant protein biosynthesis [22–24]. A similar case is true for the methanol induction phase, as the selection of an optimal feed rate profile can be crucial to the product concentration at the end of the fermentation. The open-loop profiles most often employed are either linear or exponential (to maintain a constant specific growth rate or induction pressure) [22,25].

Garcia et al. compared two open-loop substrate feed systems—offline optimizing control and exponential feeding—in a *C. necator* fermentation producing a poly(3-hydroxybutyrate-co-3-hydroxyvalerate) (PHA) copolymer. The offline optimizing control approach employed an unstructured process model to calculate the feed profile, whereas the exponential feed profile was calculated to maintain a constant specific growth rate of 0.15 h^{-1}. The fed-batch optimizing control strategy resulted in the highest reaction yield, highest PHA content and highest productivity. However, the attained product yield obtained (mass of polymer produced per mass of substrate consumed by the microorganism) was not satisfactory. Furthermore, the fed-batch-exponential strategy resulted in the lowest reaction yield, showing that this strategy is not at all suitable. To further improve the process productivity, the authors implemented a closed-loop control strategy, which uses biomass and substrate information predicted online by soft sensors to maintain a preset substrate concentration in a 500 L bioreactor process. However, the produced polymer concentration was even lower than in the previous 5 L open-loop experiments [26]. Vanichsriratana et al. compared two optimal open-loop and closed-loop feed rate control strategies in a simulated fermentation process. The authors found that the closed-loop optimal feed rate control provided a better result mainly due to the feedback property allowing the system to compensate for the modeling error [27]. In a different article, the authors compared an open-loop and a closed-loop control system and concluded that when the process model was an exact representation of the plant, both methods gave similar results in a fermentation producing primary and secondary metabolites [28].

In general, open-loop control may not always lead to the highest product yields and does not account for system disturbances. However, depending on the specifics of a given bioprocess, it can still be a valid approach for feed rate control. As demonstrated in the previous examples, open-loop control can sometimes even outperform closed-loop systems. Despite its limitations, the main benefit of open-loop control is its ease of implementation and the fact that it does not rely on measured variables. It is important to carefully evaluate the specific requirements of each bioprocess to determine whether open-loop control is an appropriate choice for feed rate control.

3. Closed-Loop Feed Rate Control

In a closed-loop feed rate controller, the input is adjusted based on the output of the system, which is typically a measured process variable (PV) signal. This allows the system to compensate for any disturbances that may affect the process and ensure that the output remains within the desired range. The system is schematically illustrated in Figure 4.

Closed-loop feed rate control strategies cover a wide range of methods, but the common denominator is the variation in the control action in order to respond to nonlinear dynamics or uncertainties during fed batch processes. The type of controller is determined by the control action calculation method employed. Closed-loop strategies may also be adaptive and thus able to adapt to the unexpected variance in the system parameters. An adaptive control system is one in which the controller parameters are adjusted automatically to compensate for changing process conditions [29]. Due to the complex and nonlinear dynamics of a biological system, the use of adaptive strategies is very com-

pelling for fermentation processes. The most popular closed-loop strategies are briefly presented further.

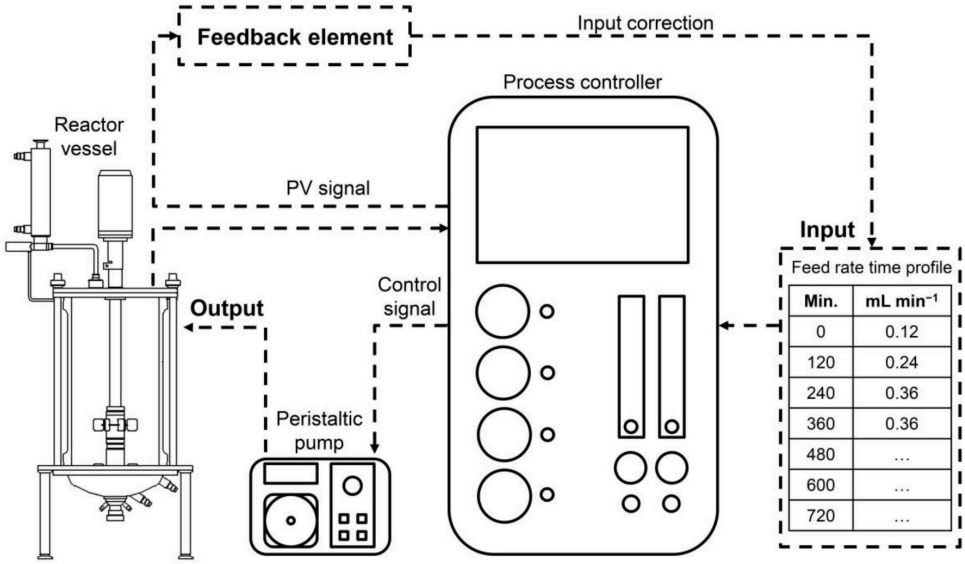


Figure 4. An example of closed-loop (feedback) feed rate control in a bioprocess.

3.1. PID Control

The most common closed-loop feed rate control method is the proportional–integral–derivative or PID controller. It works by continuously calculating the error signal value as the difference between the desired set point (SP) and the measured process variable (PV) and applies a correction to the input based on proportional, integral and derivative terms. Of these three terms, the first is proportional to the deviation between the targeted value and the currently observed value, the second is proportional to the integral of the deviation and the third term is proportional to the derivative of the deviation. Namely, the proportional term responds to the current error, the integral term responds to the accumulated error over time and the derivative term responds to the rate of change of the error. The overall control function for the substrate feed rate control in its classical mathematical form is presented in Equation (4):

$$F(t) = K_p e(t) + K_i \int_0^t e(t)dt + K_d \frac{de(t)}{dt} \quad (4)$$

where $F(t)$—feed rate correction signal value; K_p, K_i and K_d—the coefficients denoting the proportional, integral and derivative terms, respectively; $e(t)$—deviation from set-point value (SP-PV).

To implement digital controllers, which is the case for most biotechnical applications, discrete sampling periods are used, requiring the use of a discrete form of the PID equation to approximate the integral of the error and derivative. This discrete modification involves replacing the continuous integral with a summation of the error, with T representing the time between sampling instances and n representing the number of sampling instances [30]. The discrete three-term controller allows for the independent tuning of each term, which can be an advantage for the control of nonlinear processes. The discrete (sampled) form of

the algorithm has many forms based on the technique used for discretization [31,32], but can, for example, be expressed as:

$$F(t) = F_{t-1} + K_p(e_t - e_{t-1}) + K_i e_t T + K_d \frac{(e_t + e_{t-2} - 2e_{t-1})}{T} \quad (5)$$

where $F(t)$—the control output at time step t; e—the difference between the setpoint and measured output (error) at current (t) or previous ($t-1$, $t-2$) sampling times; T—the time between sampling instances (sampling time) [33].

In fed-batch processes, PID control is usually implemented in the form of indirect feedback control schemes that couple the substrate feed rate with measurements of pH (pH-stat) and/or dissolved oxygen concentration (DO-stat) [34,35]. The pH-stat method is based on the phenomena that pH usually decreases when the substrate is consumed (due to the consumption of NH_4^+ ions) and rises along with the excretion of ammonium ions, usually due to the depletion of the carbon source or cell lysis [34,36]. Similarly, the DO-stat is based on the fact that DO increases sharply when a key substrate is depleted [35,37].

Several other PID control schemes are also widely used in the biotechnological industry. In *Saccharomyces cerevisiae* (baker's yeast) cultivations, the ethanol concentration in the medium can be maintained at a constant low level. In this way, the specific growth rate is kept at maximum while avoiding growth-inhibitory ethanol formation [38]. The PID-based control shows favorable results as the overall performance of the process is better than in the case of open-loop control and the overall implementation complexity is not that high. Despite this, the use of this method is limited in some cases due to the lack of reliable on-line or express at-line measurement systems for determining the values of control parameters and due to the general nonlinearity of biological processes [39]. For example, it is difficult to accurately estimate the concentrations of cell biomass or most substrates on line.

The signal quality of optical or capacitive biomass sensors are often influenced by such disturbances as mixing, aeration, foam and the state of the culture, while the enzymatic sensors used to measure the substrate concentrations often do not provide sufficient accuracy and reliability [40]. The availability of accurate substrate concentration measurement systems, however, depend on the substrate in question as, for example, there are several systems to precisely estimate the concentration of methanol in recombinant *Picha pastoris* cultivations [41,42]. More accurate measurements of the mentioned control parameters can be obtained by in-line chromatography or flow cytometry; nevertheless, the said approaches do not guarantee that the state of the system will be correctly evaluated. For example, the lack of culture growth does not necessarily indicate malnutrition but may be a result of inhibitory metabolite formation. Additionally, maintaining a certain critical substrate concentration (S_{crit}) does not guarantee that there is no overflow metabolism in the system, since S_{crit} can change from process to process [43].

The critical factor in ensuring high-quality bioreactor operation is the effective setup and tuning of the controller parameters to address the sources of process variability [44]. Due to substantial fluctuations in the process dynamics during operation, standard PID controllers with fixed tuning parameters are insufficient for precise process control. As a result, several methods have been suggested for tuning the PID controller parameters in biotechnological processes under conditions that change over time. A continuous cultivation process has an equilibrium for which the controller can be tuned. Fed-batch processes, however, have no such point, as the system constantly changes and thus cannot be linearized. Therefore, it is necessary to use nonclassical tuning methods to achieve the best overall PID control for the entire operating range of the given system [31]. Numerous PID parameter tuning methods are available—both manual and automatic (online) [45,46]. Some of the most commonly used tuning methods are gain scheduling [12,47–49], first-principle models [50,51], fuzzy rule-based systems [52], metaheuristic models [31,45] and other techniques [30,46,53]. By introducing an automatic tuning procedure, the closed-loop system is periodically tested, and the test characteristics automatically determine new controller settings and thus introduce adaptive control [29].

To present an actual example of the role that PID parameter tuning plays in fed-batch cultivations, we can look to our previous experience, controlling the residual methanol concentration via a PI-based substrate feed rate control algorithm in *P. pastoris* cultivations. The cell biomass growth, methanol concentration and feed rate time profiles of three experiments are presented in Figure 5. In Exp. 3, methanol was successfully controlled at 1.0 g/L; however, in both Exps. 4 and 5, a residual methanol concentration of 2.0 g/L was selected as the set-point value. In Exp. 4, the same control parameters were used as in Exp. 3 and, consequently, the system was unable to successfully maintain a methanol concentration of 2.0 g/L, illustrated by the wild fluctuations (±1.0 g/L) around the set-point value. After manually tuning the PI control algorithm parameters in a process simulation, the experiment was repeated (Exp. 5). Now, the residual methanol concentration control was much more accurate and the deviation around the set-point value much lower (±0.39 g/L) [19].

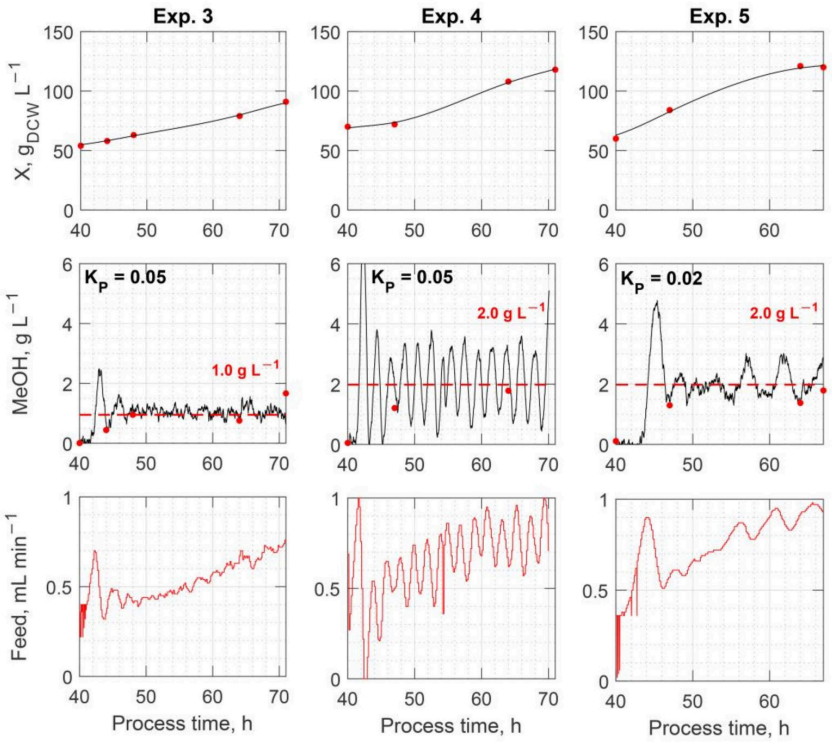


Figure 5. Three *P. pastoris* cultivations at increased residual methanol levels, illustrating the effect PI parameter (K_p) tuning has on control accuracy. Horizontal dashed line indicates the methanol concentration set point. Reprinted/adapted with permission from Ref. [19]. 2023, Springer Nature.

Despite the difficulties posed by the necessary parameter tuning procedures, PID, PD and PI closed-loop feed rate control can efficiently control the substrate feed in a way to achieve the set process goals. For example, we reported a case of recombinant hepatitis B surface antigen virus-like particle (HBsAg VLPs) production in the yeast *P. pastoris*, where sufficient productivity could not be achieved by open-loop feed rate control methods [54]. The product would be mainly in nonsoluble form (not suitable for VLP formation) and disappear during the initial purification steps. However, cultivation in an elevated residual methanol concentration of 6.0 g/L would prove to be beneficial to the product, and 186 mg/L of purified HBsAg VLPs were successfully produced. Case in point, the residual methanol concentration was controlled using a PI-based methanol feed rate control algo-

rithm and adjusted according to the real-time methanol concentration measurement of a methanol sensor probe.

Priyanka et al. compared the performance of an augmented decoupled adaptive control strategy that uses the substrate and the DO as controlled variables vs. a PID controller in a *P. pastoris* X-33 Mut$^+$ fermentation to produce rHSA. The augmented controller consisted of a decoupled adaptive control (DAC) and a decoupled input–output linearizing control (DIOLC) in a combined architecture, allowing for the simultaneous control of both the substrate and dissolved oxygen concentration. The performance of the augmented DAC-DIOLC controller showed deviations of only 0.021 g/L from the substrate set point compared to 0.09 g/L in PID, and a 1.12% deviation from the DO set point compared to 11.86% in PID. Overall, the use of the augmented controller resulted in a 1.5-fold increase in rHSA production [55]. Brignoli et al. implemented and optimized a novel proportional-integral (PI) feedforward–feedback controller to maintain a desired specific growth rate of a *K. marxianus* culture. The proposed control logic provided robust set-point tracking of an exponentially evolving fed-batch culture while ensuring improved noise and oscillation management. It was demonstrated that a strain of *K. marxianus* could be grown successfully in fed-batch mode under substrate-limited, aerobic conditions at different set points ranging from 0.1 h^{-1} to 0.4 h^{-1} [12]. Kager et al. implemented a nonlinear model predictive controller (MPC) in a *P. chrysogenum* fed-batch process and compared it to a PI(D) and an open loop feedback control scheme. The authors found that the implemented PI(D) controller showed some instabilities due to changing process dynamics, and the open-loop feedback inverse control (MBC) was limited in its linearized control space. The authors concluded that to optimally control the nonlinearities and discontinuities of the process, MPC is needed to satisfy multiple objectives and to predict process events in order to prevent the formation of gluconic acid and mannitol during the first phase of the experiment. This enabled an ongoing product formation, which led to 14% higher product concentrations and improved product and biomass yields [56].

To summarize, a PID controller can enhance the overall stability and efficiency of the process by regulating the substrate feed rate and compensating for unexpected disturbances. However, for the controller to be effective, it must be properly tuned, which can be challenging due to the nonlinearity of bioprocesses. Controllers with fixed tuning parameters may not be accurate throughout the entire process, thus requiring careful attention to tuning during operation [13].

3.2. Artificial Neural Networks (ANN)

An artificial neural network (ANN) is a computational algorithm that mimics the structure of a biological brain network. ANNs are used to estimate and predict bioprocess variables based on environmental and physiological information available from on-line sensors. They can describe complex nonlinear systems without the need for complex model equations, but require significant amounts of past process data to train the network accurately and predict the relationship between system inputs and outputs [7].

ANNs have layers of nodes, also called artificial neurons, which are typically grouped into subsets or layers, including an input layer, one or more hidden layers and an output layer. Each node is connected to others and has a weight and a threshold. A node is activated and sends data to the next layer of the network if its output exceeds the threshold value. If not, no data are passed to the next layer. The outputs of the neurons from one layer represent the inputs for the next layer. A neuron can have any number of inputs, but only one output that is usually related to the inputs by a transfer function. The most common transfer functions are the sigmoid, hyperbolic tangent, sine, linear and saturated linear functions. The argument for these functions is calculated by adding the inputs to a neuron, where each input is multiplied by its corresponding weight, indicating the connection strength between neurons [39]. Figure 6 shows a schematic representation of an artificial neural network.

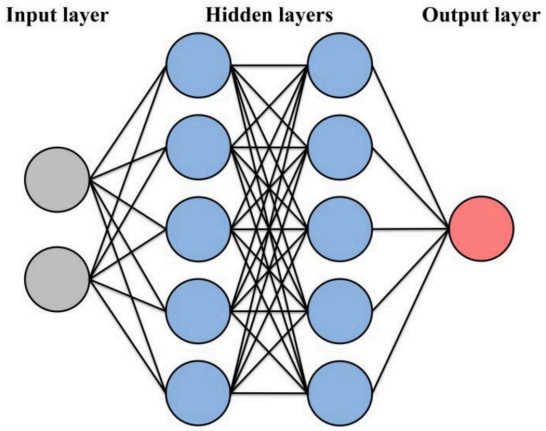


Figure 6. A schematic representation of an artificial neural network. The circles represent interconnected nodes (artificial neurons).

Although ANNs are generally quick and straightforward to implement and perform well, they lack interpretability; thus, limited process knowledge is gained. Additionally, the network is trained for a certain scale and, therefore, will need to be retrained during the process scale up [57]. Furthermore, ANNs have limited ability to extrapolate beyond the range of data used for training, in contrast to physically based mathematical models or rule-based systems, which can be more flexible and widely applicable [58].

It is possible either to apply the ANN as a predictor for a variable of interest and then incorporate this into a feedback control algorithm or to utilize the ANN directly in an optimization algorithm to solve for optimal control solutions [59,60]. ANNs have previously been shown to be successful at predicting the behavior of various fermentation systems and, consequently, have been employed in bioprocess control applications. The current reported applications include the use of ANN models to control reactor temperature [39], specific growth rate [61], maximize cell biomass [60,62,63], product concentration [63–67] and to estimate [63,67] or follow a preselected substrate concentration trajectory [59].

Chaudhuri et al. demonstrated the applicability of an ANN to calculate optimal substrate feed profiles in two simulations in comparison to reported kinetic models. The simulation results for the secreted protein and invertase production demonstrated that the ANN model managed to capture the essential features of the process kinetics, and therefore, the model can be used for the dynamic simulation of the process. The optimal feeding policies obtained with the model agrees reasonably well with the previously reported results, which allows the authors to conclude that the main advantage of the approach lies in the fact that optimization can be achieved without knowledge of the detailed kinetic model of the process [68]. Galvanauskas et al. compared the specific growth rate control performance of an adaptive PI controller based on the gain scheduling technique and a model-free adaptive controller based on an artificial neural network. Both controllers delivered a comparable control performance and were suitable for application when using the substrate limitation approach and substrate feeding rate manipulation. However, the authors concluded that considering the efforts for controller design and tuning (including the development of the adaptation/learning algorithms), the model-free adaptive control algorithm proved to be more attractive for industrial applications, especially when only limited knowledge of the process and its mathematical model is available. Additionally, the model-free adaptive controller also tended to deliver a better control quality under low specific growth rate conditions during the recombinant protein production phase [49]. Zelic et al. investigated three different models: the unstructured mechanistic black-box model, the input–output neural-network-based model and the externally recurrent neural

network model to describe the pyruvate production process from glucose and acetate using a genetically modified *E. coli* YYC202 strain. The unstructured mechanistic model could not adequately fit the experimental data obtained in the different processes; however, the results obtained using the neural networks showed satisfactory compliance with experimental values. The performance of the neural networks conclusively showed that they have the potential to be implemented as an online state estimator, facilitating the control of pyruvate production, and can also be used in process optimization [69]. Rashid et al. conducted a study where neural network and fuzzy logic methods are used in the design of a controller for baker's yeast production. After the first iteration, the ANN-based controller showed unsatisfactory results; hence, a set of fuzzy rules were introduced to characterize the biomass concentration. To obtain a reliable assessment, the results were then compared with those of the PI controller. Unlike the PI controller, the developed controller proved capable to follow the time-varying characteristics of the process and deal with the nonlinearity of the process [70]. Pantano et al. successfully implemented a closed-loop controller integrated with a neural network state estimator to track three optimal profiles of an important nonlinear biological process by manipulating the nutrient and inductor feed rates. The proposed controller was compared with a PI controller, showing a better control response. Moreover, through several closed-loop simulation tests, the proposed controller was shown to be not only simple and efficient with the neural state estimation, but also sufficiently robust to compensate for the mismatches in the model parameters as well as the internal and external perturbations of the system [71].

In summary, ANNs have demonstrated the successful prediction of fermentation system behavior based on measured variables and have been applied to various control applications. Although they can be relatively easy to implement, building an accurate ANN requires significant historical process data, and limited process knowledge is gained from the network. Additionally, ANNs are trained on a single scale and are limited in their ability to extrapolate data, which restricts their application.

3.3. Fuzzy Control

Fuzzy logical control is a control strategy intended to address uncertainties in nonlinear systems, such as batch, fed-batch and continuous fermentations. It leverages the knowledge and expertise of the process operator to control the fermentation. The fundamental principle of fuzzy control is to convert quantitative data into qualitative parameters [7].

A set of fuzzy if–then rules and an inference mechanism are used to determine the input–output mapping of the system. First, the input data are fuzzified—converted from numerical data into 'membership functions' based on the degree to which they fit in a fuzzy set (given a value from 0 to 1). To perform fuzzification, a range of possible values is defined for each input variable. This range serves as a scale for comparing the data values. The variable is then described using a fuzzy set and is assigned a degree of membership. For instance, a substrate concentration reading of 2.0 g/L might be compared to the range of 0–10 g/L and assigned to the fuzzy set "Low" with a degree of membership of 0.8 and to the fuzzy set "OK" with a degree of membership of 0.2. It is important to note that the total degree of membership for all fuzzy sets for a given variable must equal one. A fuzzy set is described by a matrix that includes the original value of the variable and the degree to which it belongs to a category on a scale from 0 to 1. The current state of the system is interpreted using the fuzzy sets. Fuzzy control is based on a set of "fuzzy rules" that describe the conditions of the system and are derived from the operators' experience with the process in the form of conditional statements using such terms as 'if' and 'then'. For example, If S (substrate concentration) is "Low", Then F (substrate feed rate) is "High". In this case, the controller assesses the substrate concentration as low and, consequently, increases substrate feed rate. The operator's experience is incorporated into the controller via a process called defuzzification—the process of transforming a fuzzy output into a numerical value that represents the best estimate of the actual output that can be used to make decisions [7,72]. A fuzzy logic controller is schematically illustrated in Figure 7.

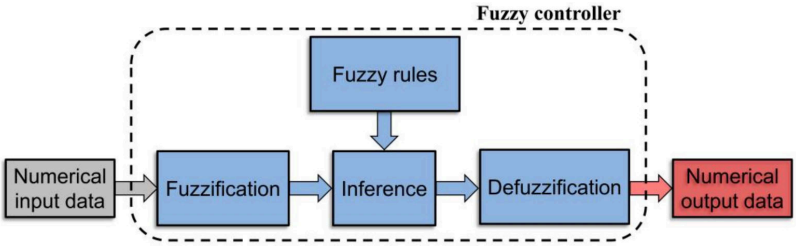


Figure 7. A schematic representation of a fuzzy logic controller.

In contrast to ANNs, the outcome of fuzzy control is to identify the current state of the system, and it therefore provides some information to the operator. The main advantage of the fuzzy logic approach is the greater flexibility to capture incomplete and imperfect aspects of the process.

Fuzzy controllers have also previously been successfully employed in bioprocess applications; for example, to estimate substrate concentrations [73–76], cell biomass [74], fermentation temperature [77] and adjust the substrate feed rate [75].

Ye et al. developed a five-layer neural–fuzzy network to control the fed-batch cultivation conditions of a recombinant *E. coli* strain for maximum β-galactosidase production [78]. The authors applied a fuzzy network to control the glucose feed rate based on deviations from set points of the pH value, which can be used to indirectly measure the glucose concentration in the broth, and the specific cell-growth rate. The specific growth rate is the most important variable, influencing both cell growth and recombinant product expression. The biomass concentration of recombinant *E. coli* improved from 20 g dry cell weight/L under conventional control to 84 g dry cell weight/L under neural–fuzzy control. However, the relative activity of β-galactosidase was not high due to a high residual glucose concentration after induction. To overcome the low relative activity, the researchers created two fuzzy networks—one for the period before induction and one for after. Adding the second fuzzy network increased the relative activity of β-galactosidase four fold while a biomass of 50 g (dry cell weight)/L was reached, which still represents a significant improvement over conventional techniques [78,79].

Fonseca et al. proposed an adaptive fuzzy feedforward–feedback control structure (A4FB) in which a Mamdani fuzzy system was used to adapt the feedforward control law parameters. The A4FB was tested in a CSTR reactor, in which an enzymatic reaction of starch hydrolysis by *Aspergillus niger* glucoamylase takes place. A comparison between the A4FB and the classical feedforward–feedback control was also provided. The results indicated that the adaptive fuzzy feedforward control performed better in terms of disturbance rejection since its performance index was 0.645 while the classical feedforward–feedback control achieved only 0.406. In conclusion, the authors list the main advantages of the developed adaptive fuzzy control system: easy deployment of a regulatory control action without prior nonlinear process modeling; ease to adapt from the fuzzy logic to other systems; and better rejection of the process disturbances [75]. Choi et al. investigated the effect of controlling the substrate feed rate with an ANN and fuzzy logic controllers to enhance flavonoid production in a *Scutellaria baicalensis* Georgi plant cell fed-batch fermentation. The experimental results showed that the ANN controller with a genetic algorithm improved the flavonoid production compared with a simple fuzzy logic control system. Furthermore, the specific production yield and flavonoid productivity also increased. Therefore, the authors concluded that an artificial neural network is a more suitable controller than a fuzzy logic controller in the fed-batch cultivation of *S. baicalensis* G. [80]. Yuan et al. investigated an approach in which a fuzzy logic controller was developed based on an artificial neural network model. The controller was then tested online with an objective to obtain a high concentration of cell mass in the shortest fed-batch culture time. The experimental

results revealed that the fuzzy neural network controller on-line feeding of the substrate performed, as the authors described, "well" [81].

In summary, fuzzy control methods are useful for dealing with nonlinear systems such as fermentations and are considered more intuitive to users. However, there is a lack of recent examples of their use in substrate feed rate control applications. Fuzzy logic controllers require a good understanding of the bioprocess to establish a comprehensive rule base, which is dependent on human expertise. While they provide insight into the dynamics of complex systems, their effectiveness is limited by the need for human input and the lack of recent examples of their application in this specific area.

3.4. Model Predictive Control

One of the most promising methods for closed-loop control in fermentation systems is currently model predictive control (MPC). This method evaluates the difference between the predicted and reference values of the controlled variable in order to determine an appropriate control action. A robust prediction process model is required to simulate the fermentation process up to a future time horizon, which allows for predictions of both the current output and future system states. These predictions are then assessed based on an optimization of a cost function over the entire process time, which ultimately determines the appropriate control action (i.e., feed rate correction) required at the current time. An MPC controller is schematically illustrated in Figure 8.

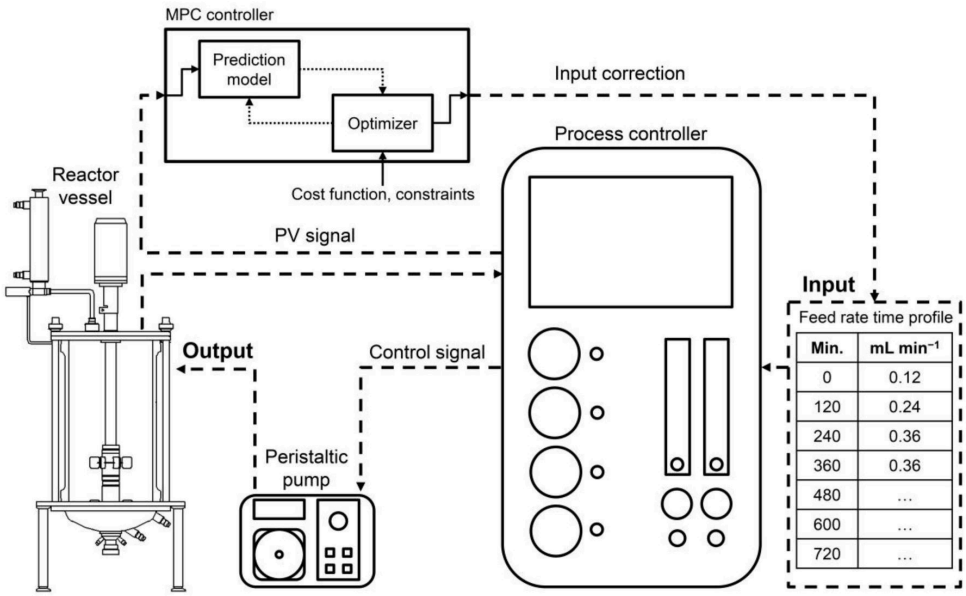


Figure 8. An example of an MPC controller.

The basic operating principle of model predictive control (MPC) involves several steps. First, the initial process parameter values are defined, and the desired process reference trajectories are established. Then, using historical process data, a model is constructed to simulate the behavior of the fermentation process over a defined time period (the prediction horizon), allowing for the prediction of the system state. The predicted trajectories are then compared to the reference trajectories, and if the deviation exceeds a defined threshold, the MPC algorithm modifies the feed rate profile within a predetermined range over a specific time period (the control horizon) until new data are supplied. The operational principle of

an MPC controller is presented in Figure 9, where T_k is the actual point in time, T_{k-1} and T_{k-2} are past (measured) values and T_{k+1} and T_{k+2} are future (predicted) values.

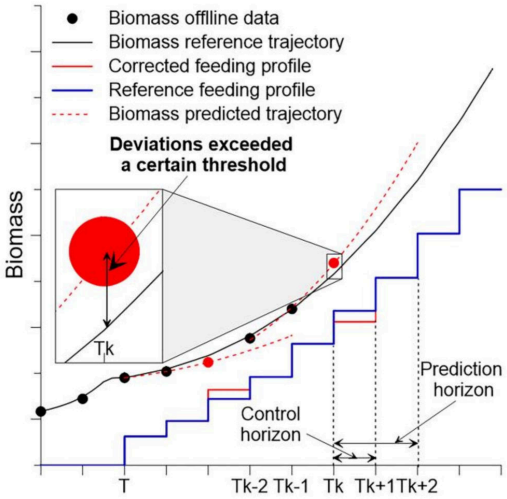


Figure 9. MPC operation principle [82]. The model estimates the values of cell biomass in the prediction horizon and adjusts the substrate feeding profile to minimize the deviation from the biomass reference trajectory.

To control the feed rate using MPC, choosing an appropriate optimization function is a crucial aspect. Two commonly used approaches are following a reference trajectory for either cell biomass or substrate concentration during cultivation. The reference trajectories can be obtained from previously reported processes [83], calculated using mathematical process models [84] or determined experimentally, for example, by employing the probing method [85].

Both classic mechanistic [86,87] and empirical models (ANN [88,89], partial least squares [90]) can be used to develop predictions. Empirical models can be quite accurate within the boundaries of the process parameters they are originally derived from, but they are poorly scalable and present less predictive and analytical insight, in contrast to mechanistic models [90]. Additionally, the employment of a model predicting the behavior of the system, along with a desired reference trajectory, permits the prediction of possible process parameter deviations in advance, which is not possible using other methods of direct control.

Many examples for the application of MPC for the control of yeast [88,89], bacteria [84,87,91] and mammalian cell cultivations [83,86,92] can be found in the literature. Some examples of use in alcohol biosynthesis [88,89], recombinant proteins [83], hormones [84], flavonoids [93] and antibiotic [90] production have been also reported. However, half of these studies were carried out on various kinds of simulators [88–92] and not dedicated towards direct operation in real systems.

Kuprijanov et al. demonstrated how an MPC controller can be easily implemented in industrial bioreactor automation systems for fed-batch fermentation processes. Using only standard sensors (pH, DO) and at-line biomass and glucose measurements, and by varying the substrate feed rate, the authors were able to demonstrate the ability of the MPC controller to follow a preset biomass growth profile and thus improved process repeatability and safety [87].

Lupenza et al. developed an Event-Triggered Feed-Forward Control (ET-FFC) scheme for the K12 *E. coli* fed-batch fermentation process. Based on the data, a Proportional Integral (PI) and a model predictive (MPC) controllers were designed to control the cell biomass

concentration by manipulating the substrate feed rate. The closed-loop performances of the proposed controllers were evaluated and analyzed through a simulation of an *E. coli* fermentation process. The results showed that the MPC-based ET-FFC scheme provided better performances with a minimum integral square error over the PI-based ET-FFC scheme [94]. Ulonska et al. compared the performance of two different model-based control strategies with respect to the simultaneous set-point control of two individual substrate uptake rates by two substrate feeds in an *E. coli* fed-batch process. The compared controllers were an elemental balance controller (EBC) and a model predictive controller (MPC) based on a mechanistic model. Both controllers showed comparable behavior and were generally capable of fulfilling their tasks. The MPC was based on a better and more flexible description of the system, whereas the EBC was easier and showed more stable behavior. The authors concluded that for the investigated application, the EBC was preferable due to its simplicity. However, the potential of the MPC was clearly in its prediction power and flexibility towards objective functions. Therefore, it would be the controller of choice in the case of product-related objective functions [95]. Schneider et al. compared the performance of two model predictive controllers—a nonlinear open-loop feedback optimal (OLFO) and a linear dynamic matrix control (DMC)—in *Cyathus striatus* fermentation. The authors found that the DMC controller performed unexpectedly well in this particular process as it is more suited for fast, discrete and linear systems. However, the authors concluded that better results can be achieved with the OLFO controller with more frequent model parameter identification and if the time horizon for the identification process is the past 24 h and not the entire process time [96]. Aehle et al. employed an MPC system to ensure a high batch-to-batch reproducibility in an animal cell (CHO-cell) culture for recombinant therapeutic protein (EPO) production. The control objective was to identify and control an optimal specific growth rate by controlling the oxygen mass consumed by the cells by manipulating the glutamine feed rate. The authors judged the performance of the controller to be fairly good, which could best be judged by high batch-to-batch reproducibility obtained in cultures that are operated with this controller [83].

To summarize, model predictive control (MPC) is a powerful closed-loop control method for substrate feed rate control that optimizes the control actions for the full process time, not just at the current time instant. MPC also models the impact of disturbances as part of the optimization problem. However, the effectiveness of this method heavily depends on the accuracy of the process model and its ability to handle unexpected disturbances. Some of the main drawbacks of this method include the requirement for robust process models, which may not always be available, and its computational expense.

4. A Practical Example of MPC Implementation

In practice, MPC is a powerful and prospective feed rate control approach in a wide variety of fed-batch bioprocesses. There are also some commercial MPC solutions available; however, they are highly process specific and are designed to be employed mainly in the biopharma industry. For example, ABB offers a nonlinear MPC system for mammalian cell (CHO) cultivation for biopharmaceutical production [86,97]. A similar system was also announced to be in development by Bilfinger [98]. The market currently lacks universal, sufficiently robust, affordable and easy-to-use MPC options for a wide range of advanced bioprocess control applications; therefore, it may be necessary to tailor the MPC system to one's specific process.

However, implementing this approach demands proficiency in bioprocess modeling and programming. In the case of adapting the MPC system to a different bioprocess, most of the development process has to be repeated. To showcase the feasibility of this approach and to provide a clearer picture of the development process, we have previously designed an MPC system, which was constructed using a standard hardware PC and a Siemens Simatic S7-1200/1500 series programmable logic controller (PLC). The module communication scheme and hardware implementation of the proposed system are illustrated in Figures 10 and 11, respectively. SCADA, installed on a standard hardware personal com-

puter, was connected to the Siemens Simatic S7-1500 PLC through the router via an Ethernet link. The implementation of the proposed MPC algorithm was realized using a PC-based program, developed in a Matlab (Mathworks, USA) environment. The data exchange of the process and control variables between the MPC control algorithm and SCADA, programmed in a PcVue (PcVue Solutions, France) environment, was implemented every 1 s through an OPC server. The feeding was controlled gravimetrically using the scale module (Bioreactors.net, Latvia) or volumetrically by means of an analogue peristaltic pump.

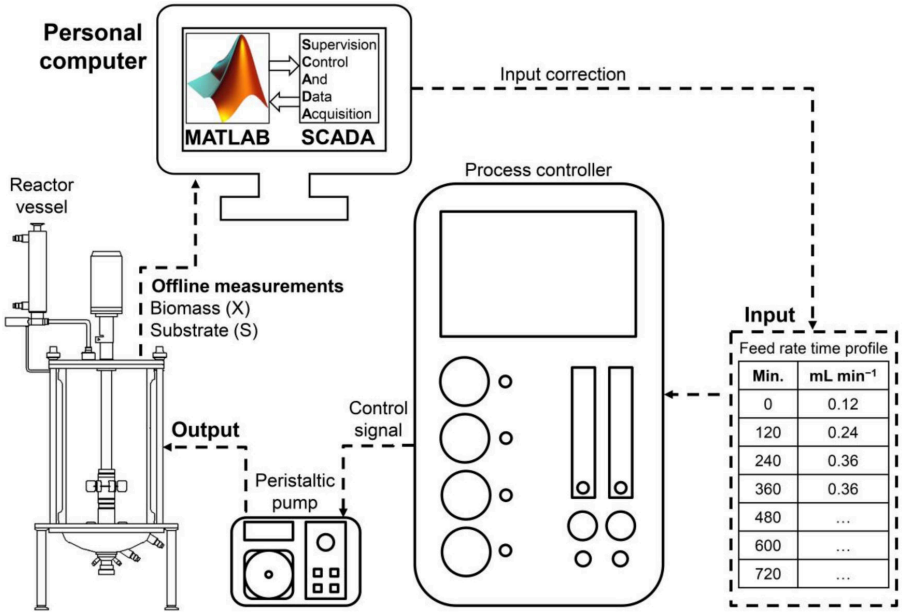


Figure 10. The communication scheme of a MATLAB-based MPC solution.

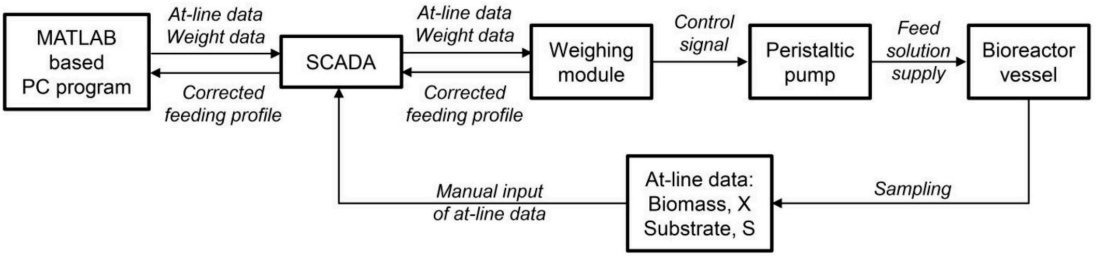


Figure 11. The information exchange diagram of the proposed MPC system [82].

To employ this MPC system, it is first necessary to establish the optimal trajectories for either biomass growth or substrate concentration (so-called "golden batch") [99]. Both the empirical (modeling) and experimental methods mentioned in the previous paragraphs can be employed to construct said reference trajectories. Before starting the process, the initial X_0, S_0 and V_0 values must be defined, and the program then calculates a possible feeding profile and models the reference trajectory of $X(t)$, $S(t)$ and $V(t)$ for the entire cultivation. After starting the MPC control program, sampling must be carried out. After each sampling procedure, the biomass and sugar measurements are entered in the system, a prediction trajectory is calculated and the deviation from the reference trajectory is estimated. If it is within preset boundaries, no action is carried out; however, if the deviation is too great,

a corrective feed action is implemented in order to minimize the difference between the measured process and reference values. This MPC algorithm is schematically illustrated in Figure 12.

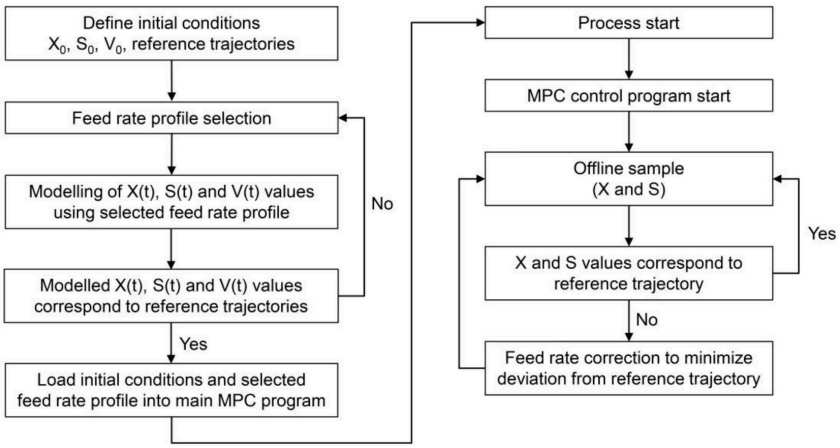


Figure 12. The schematic algorithm of the MPC program.

We have previously reported several cases where this MPC approach has been successfully employed [82,99].

For example, this MPC solution was used in a recombinant *E. coli* BL21 process to successfully follow a predefined cell biomass reference trajectory in a glucose-limited fed-batch process. The proposed approach was implemented and tested in a lab-scale bioreactor system EDF-5.4/BIO-4 (Bioreactors.net, Riga, Latvia). An additional test run using the proposed approach was performed on a system consisting of a single-use bioreactor (SUB) CellVessel 5.7 (CerCell, Herlev, Denmark) connected to a BIO-4 controller. Fermentations were started as batch cultures and continued as fed-batch when the MPC controller activated feeding to follow the predefined reference growth trajectory.

The time profiles of the cell biomass growth, substrate (glucose) concentration and feed rate are presented in Figure 13. A cell biomass concentration of 93.6 g/L was reached in the most productive run (Exp-3); however, significant glucose accumulation was noted towards the end of the process due to culture overfeeding, and significant corrective actions were necessary. In a more stable and safer yet still very productive run (Exp-2), a cell biomass concentration of 79.8 g/L, in contrast to 65.8 g/L, produced in the initial experiment (Exp-1) was reached after 24 h, employing the proposed system. The estimated and experimentally measured cell biomass mean deviations from the preset reference value at the end of the processes were 4.6% and 3.8%, respectively. The developed and implemented MPC system was used to track the selected biomass reference trajectories. The MPC system demonstrated a good control performance and reduced process variability as compared to the system with an open-loop feeding profile control. Importantly, it was also proven that the results produced in the best run (Exp-2) were reproducible in a different system—in this case, a single-use bioreactor vessel [99].

Another successful example of this system's use was reported in a yeast *Kluyveromyces marxianus* fed-batch cultivation process [82]. Laboratory-scale cultivations were performed in a 5.4 L working volume bioreactor EDF-5.4_1 (Bioreactors.net, Latvia). During cultivations, offline glucose and biomass measurements were carried out every 30–120 min, and the obtained results were uploaded into SCADA, which passed the updated information to the MPC software. The reference trajectories X_{ref}, S_{ref}, V_{ref} and F_{ref} were calculated using the model described by Grigs et al. [99].

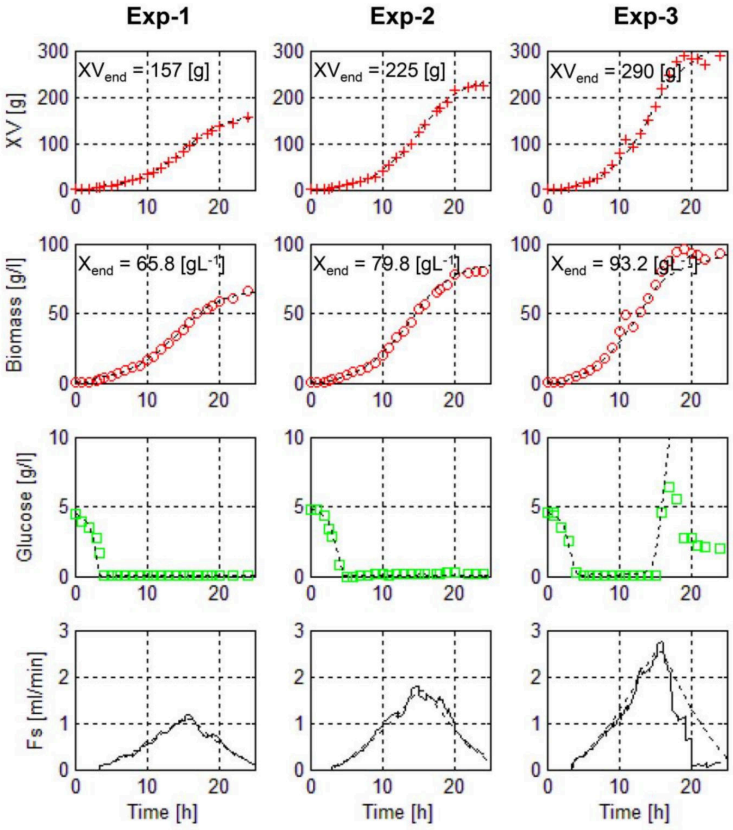


Figure 13. Cell biomass growth, substrate concentration and feed rate profiles during 3 experimental runs employing MPC feed rate control to estimate optimal MPC control parameters [99]. Dashed lines represent reference trajectories. Process productivity increases with each experiment and optimal MPC control parameters are identified after 3 runs. Figure adapted with permission from the author.

The MPC algorithm was successfully applied for substrate feed rate control during cultivation. Mathematical model parameters were fine tuned during the experimental work to successfully estimate the yeast biomass growth and substrate consumption kinetics, which in turn allowed for the successful construction of desirable reference growth trajectories to maintain a constant specific growth rate. Along with a novel optimized synthetic cultivation medium, the employed MPC feed rate control system facilitated the production of 70 g/L yeast dry biomass concentration and a 2-Phenylethanol biosynthesis rate of 0.372 g/L/h (74% conversion from 2-phenylalanine) (Figure 14) [82].

Overall, the presented MPC system is flexible enough to potentially be employed for a wide range of microorganisms, while at the same time providing good control accuracy and being relatively easy to use. It has the potential to produce comparable results to other, much more sophisticated MPC systems with lower resource requirements, while still being user friendly and widely applicable.

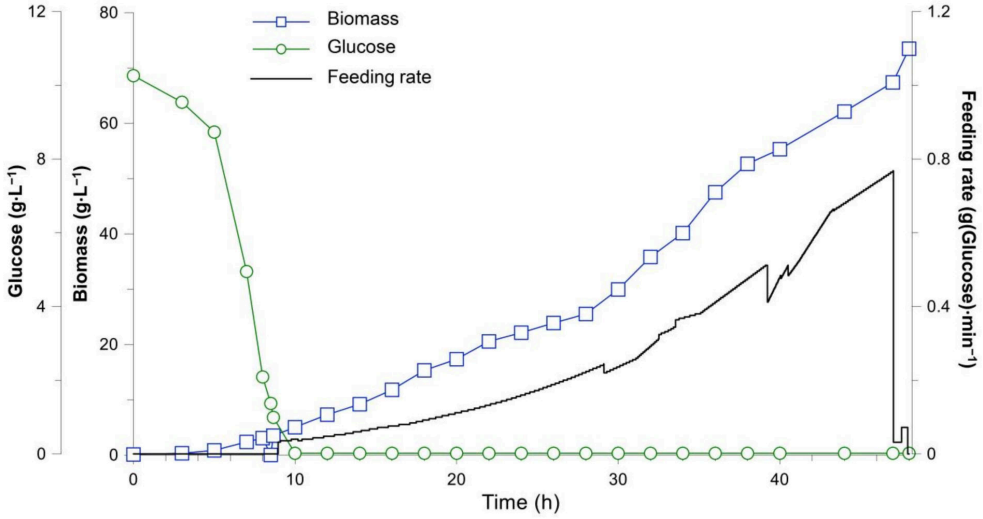


Figure 14. Cell growth and substrate feeding dynamics during yeast *K. marxianus* bioreactor cultivation. The MPC system maintains the cell growth trajectory close to the reference growth profile with limited glucose accumulation in the medium, improving process yield and repeatability.

5. Discussion

With the rise in popularity and accessibility of computers, advanced process engineering tools for biotechnology have also emerged, including modeling, process control and optimization. However, applying these tools to cultivation processes remains challenging due to the limited availability of experimental data for model building, as well as uncertainties associated with bioprocess dynamics. In addition, the lack of reliable and cost-effective online sensors to monitor certain key process variables poses a further challenge [100]. The industry is increasingly adopting the Quality-by-Design (QbD) approach, which involves designing quality into the product. To achieve this, advanced bioprocess control and insight are essential, which can only be attained by utilizing advanced process engineering tools, such as comprehensive data analysis and process modeling [101].

To reduce implementation costs, many industrial fed-batch cultivation processes still employ open-loop feeding strategies. With this approach, there is no need for online measurements or prior user experience. However, it may not be the most optimal control strategy and may result in reduced process efficiency and product quality [100,102]. However, this approach has limitations in reacting to unexpected system deviations and accounting for nonlinear behavior, potentially decreasing overall batch-to-batch reproducibility and resulting in poor performance. Furthermore, it also provides limited process insight. Despite these limitations, the open-loop approach may still be appropriate for certain bioprocesses and product requirements and should not be entirely dismissed.

Closed-loop strategies are an alternative to open-loop feeding strategies, which allow for the calculation of the substrate feed rate based on measured online data or their estimates (soft sensors), and can react to system disturbances [100]. While the use of soft sensors may seem appealing, they are often difficult to implement and require extensive data from previous processes. Most closed-loop solutions use a PID controller, which requires extensive parameter tuning and may not account for the dynamic and nonlinear behavior of living systems. More advanced controllers, such as ANN and fuzzy logic controllers, can handle the dynamic and nonlinear behavior of living cells, but they require historical data to create empirical process models, which are difficult to interpret and provide little insight into the nature of the system. Fuzzy logic controllers also require previous operator experience to evaluate the state of the process. Despite their drawbacks, closed-loop

strategies are superior to open-loop strategies as they can react to disturbances and provide better batch-to-batch reproducibility.

Model predictive control (MPC) is a promising approach for substrate feed rate control in fed-batch cultivations, provided a robust process model is available. This method can effectively account for disturbances and the nonlinear dynamics of living organisms, while also offering important insights into the process. MPC also allows for the use of at-line measurements to track deviations from the reference growth trajectory and make automatic corrections. Based on our experience, MPC is particularly suitable for controlling the feed rate in fast-growing microorganism fermentations such as yeasts and bacteria.

MPC has some limitations that should be considered. One of the main requirements for the successful implementation of MPC is a robust process model, which can be a challenge in cases where the system has not been well characterized or there is limited information available. Furthermore, the accuracy of the model must be validated through experimental data, which may require several runs. The development of bioprocess models also necessitates a specialized skill set, including a solid understanding of bioengineering and basic programming. There are also reports of the use of ANN [88,103,104] and fuzzy logic [105] models in MPC systems. The second limitation of MPC is that it involves computationally intensive calculations, which can be time consuming. In addition, it has high hardware and software requirements, similar to ANNs. However, it is possible to overcome these limitations, as demonstrated in the previous section, by using a standard personal computer and replacing expensive online sensors with standard at-line measurements. This makes MPC more appealing to a wider audience and promotes its use in fed-batch fermentations.

There are several reports where the performance of MPC solutions have been compared in fed-batch fermentations. For example, Jabarivelisdeh et al. implemented and applied a closed-loop bilevel problem based on MPC in order to handle uncertainties in modeling parameters in an *E. coli* fermentation for ethanol production. To evaluate the results, the authors compared the maximum ethanol productivity with an open-loop controller. The MPC solution showed a better performance and allowed the system to achieve both higher cell biomass concentrations and ethanol yields [106]. Joynes and Zhang designed a feedback control loop by using MPC and a conventional proportional and integral (PI) controller to control the glucose concentration at 15 min sampling intervals. The authors found that incorporating a linear ARX model in both controllers improved the control performance in both cases. The adaptive MPC had the lowest integral absolute error (IAE) on average, outperforming PI control, though the results varied run to run [107]. Eaton and Rawlings compared an MPC system with a conventional feedback control and found that MPC provided a superior performance in the case for nonminimum phase systems with input constraints when future set points were known [108]. Finally, Karra et al. developed a novel adaptive MPC formulation for multivariable time-varying systems. The proposed adaptive MPC (AMPC) scheme was able to achieve tight control of time-varying semi-batch processes and was capable of managing large transitions in the operating point of a continuously operated complex multivariable process. The authors also demonstrated that the AMPC scheme performed considerably better than the conventional nonadaptive MPC scheme for the servo control problems [109].

The main takeaway is that there are currently no universal substrate feed rate control methods that are ideal for every possible application. Each method has its strengths and weaknesses and varying applicability in bioprocesses. A comparison of both the benefits and drawbacks of the mentioned substrate feed rate control methods can be seen in Table 1.

An excellent review of the future perspectives of bioprocess control was presented by Rathore et al. [110]. Despite the significant progress in this field, it is evident that the future holds immense potential for further advancements. More advanced closed-loop substrate feed rate control methods, such as MPC, ANNs and fuzzy logic controllers, provide promising opportunities for the successful automation of fed-batch fermentations, addressing some of the challenges faced by commercial-scale production. Enhanced knowledge in process modeling will lead to a better understanding of the bioprocess, even facilitating

the creation of process digital twins—a virtual model capable of accurately simulating the fermentation process. The improvements in process monitoring with more robust at- and in-line sensors will drive an increase in the overall productivity of the fermentation process. Lastly, the use of sophisticated and improved bioprocess data analysis techniques will enable the analysis of the ever-increasing amount of data generated, which is becoming increasingly complex and voluminous.

Table 1. A comparison of substrate feed rate control methods in fed-batch bioprocesses.

Feed Rate Control Method	Benefits	Drawbacks
Open loop	Easy implementation	Inability to react to disturbances Reduced batch-to-batch repeatability
PID	Ability to react to disturbances Simple implementation	Requires accurate parameter tuning Unable to account for nonlinear system dynamics
ANN	Able to account for nonlinear system dynamics Relatively easy and fast to implement	Historical process data required for training Provides little insight into the process
Fuzzy logic	Provides some insight into the process More intuitive to the user	Requires human interpretation and a high degree of understanding of the bioprocess
MPC	Provides high process repeatability	Reliant on process model accuracy Requires historical process data for model parameter fine tuning

6. Conclusions

Accurately controlling the substrate feed rate in fed-batch bioprocesses is critical to achieving high overall process productivity, and thus requires careful optimization. Several substrate feed rate control strategies are available, each with their own strengths and weaknesses. However, there is no universally perfect control method that can be applied to all bioprocesses. As such, selecting the control method that is best suited for the specific application is essential.

Model predictive control has emerged as a promising substrate feed rate control method, despite being described as computationally expensive, process specific and difficult to implement. However, as demonstrated in this article, our past experience has shown that implementing a simple, robust and accurate MPC system may not always be challenging. We showcased that a standard hardware personal computer can be used to implement an MPC system that provides accurate microbial biomass control along a predefined growth trajectory in bacteria and yeast fed-batch fermentations. This example indicates that the main drawbacks of MPC can be surmounted, and hopefully, it will encourage the wider adoption of model predictive control in various bioprocess applications.

Author Contributions: Conceptualization, J.V., K.D., A.S. and E.B.; writing—original draft preparation, E.B.; writing—review and editing, E.B., K.D., J.V. and A.S.; visualization, E.B. and K.D.; supervision, J.V. and K.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Latvian ERDF project No. 1.1.1.1/20/A/137.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yang, Y.; Sha, M. *A Beginner's Guide to Bioprocess Modes-Batch, Fed-Batch, and Continuous Fermentation*; Eppendorf Inc.: Enfield, CT, USA, 2017; pp. 1–16.
2. Lim, H.C.; Shin, H.S. Introduction to Fed-Batch Cultures. In *Fed-Batch Cultures*; Cambridge University Press: Cambridge, UK, 2018; pp. 1–18. ISBN 9781139018777.
3. Muradi, N.A.; Adeni, D.S.A.; Suhaili, N. Enhancement of Very High Gravity Bioethanol Production via Fed-Batch Fermentation Using Sago Hampas as a Substrate. *Asia Pac. J. Mol. Biol. Biotechnol.* **2020**, *28*, 44–51. [CrossRef]
4. Lindskog, E.K. *The Upstream Process: Principal Modes of Operation*; Elsevier Ltd.: Amsterdam, The Netherlands, 2018; ISBN 9780128125526.
5. Posten, C. *Integrated Bioprocess Engineering*; De Gruyter: Berlin, Germany, 2018; ISBN 9783110315394.
6. Villadsen, J.; Patil, K.R. Optimal Fed-Batch Cultivation When Mass Transfer Becomes Limiting. *Biotechnol. Bioeng.* **2007**, *98*, 706–710. [CrossRef] [PubMed]
7. Mears, L.; Stocks, S.M.; Sin, G.; Gernaey, K.V. A Review of Control Strategies for Manipulating the Feed Rate in Fed-Batch Fermentation Processes. *J. Biotechnol.* **2017**, *245*, 34–46. [CrossRef] [PubMed]
8. Oliveira, R.; Simutis, R.; De Azevedo, S.F. Design of a Stable Adaptive Controller for Driving Aerobic Fermentation Processes near Maximum Oxygen Transfer Capacity. *J. Process Control* **2004**, *14*, 617–626. [CrossRef]
9. Looser, V.; Bruhlmann, B.; Bumbak, F.; Stenger, C.; Costa, M.; Camattari, A.; Fotiadis, D.; Kovar, K. Cultivation Strategies to Enhance Productivity of Pichia pastoris: A Review. *Biotechnol. Adv.* **2014**, *33*, 1177–1193. [CrossRef]
10. Chai, W.Y.; Teo, K.T.K.; Tan, M.K.; Tham, H.J. Fermentation Process Control and Optimization. *Chem. Eng. Technol.* **2022**, *45*, 1731–1747. [CrossRef]
11. Urniezius, R.; Galvanauskas, V.; Survyla, A.; Simutis, R.; Levisauskas, D. From Physics to Bioengineering: Microbial Cultivation Process Design and Feeding Rate Control Based on Relative Entropy Using Nuisance Time. *Entropy* **2018**, *20*, 779. [CrossRef]
12. Brignoli, Y.; Freeland, B.; Cunningham, D.; Dabros, M. Control of Specific Growth Rate in Fed-Batch Bioprocesses: Novel Controller Design for Improved Noise Management. *Processes* **2020**, *8*, 679. [CrossRef]
13. Galvanauskas, V.; Simutis, R.; Levišauskas, D.; Urniežius, R. Practical Solutions for Specific Growth Rate Control Systems in Industrial Bioreactors. *Processes* **2019**, *7*, 693. [CrossRef]
14. Schuler, M.M.; Marison, I.W. Real-Time Monitoring and Control of Microbial Bioprocesses with Focus on the Specific Growth Rate: Current State and Perspectives. *Appl. Microbiol. Biotechnol.* **2012**, *94*, 1469–1482. [CrossRef]
15. Levisauskas, D.; Galvanauskas, V.; Henrich, S.; Wilhelm, K.; Volk, N.; Lübbert, A. Model-Based Optimization of Viral Capsid Protein Production in Fed-Batch Culture of Recombinant Escherichia coli. *Bioprocess Biosyst. Eng.* **2003**, *25*, 255–262. [CrossRef]
16. Lee, J.; Ramirez, W.F. Optimal Fed-Batch Control of Induced Foreign Protein Production by Recombinant Bacteria. *AIChE J.* **1994**, *40*, 899–907. [CrossRef]
17. Krämer, D.; Wilms, T.; King, R. Model-Based Process Optimization for the Production of Macrolactin D by Paenibacillus Polymyxa. *Processes* **2020**, *8*, 752. [CrossRef]
18. Invitrogen Corporation Pichia Fermentation Process Guidelines. Available online: https://tools.thermofisher.com/content/sfs/manuals/pichiaferm_prot.pdf (accessed on 20 February 2023).
19. Bolmanis, E.; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled P. pastoris GS115 Mut+ Bioreactor Process under Controlled Residual Methanol Concentration. *Bioprocess Biosyst. Eng.* **2022**, *45*, 1447–1463. [CrossRef]
20. Potvin, G.; Ahmad, A.; Zhang, Z. Bioprocess Engineering Aspects of Heterologous Protein Production in Pichia pastoris: A Review. *Biochem. Eng. J.* **2012**, *64*, 91–105. [CrossRef]
21. Yang, Z.; Zhang, Z. Engineering Strategies for Enhanced Production of Protein and Bio-Products in Pichia Pastoris: A Review. *Biotechnol. Adv.* **2018**, *36*, 182–195. [CrossRef]
22. Zhang, W.; Inan, M.; Meagher, M.M. Fermentation Strategies for Recombinant Protein Expression in the Methylotrophic Yeast Pichia Pastoris. *Biotechnol. Bioprocess Eng.* **2000**, *5*, 275–287. [CrossRef]
23. Duarte, D.R.; Barroca-Ferreira, J.; Gonçalves, A.M.; Santos, F.M.; Rocha, S.M.; Pedro, A.Q.; Maia, C.J.; Passarinha, L.A. Impact of Glycerol Feeding Profiles on STEAP1 Biosynthesis by Komagataella pastoris Using a Methanol-Inducible Promoter. *Appl. Microbiol. Biotechnol.* **2021**, *105*, 4635–4648. [CrossRef]
24. Bahrami, A.; Shojaosadati, S.A.; Khalilzadeh, R.; Farahani, E.V. Two-Stage Glycerol Feeding for Enhancement of Recombinant HG-CSF Production in a Fed-Batch Culture of Pichia pastoris. *Biotechnol. Lett.* **2008**, *30*, 1081–1085. [CrossRef]
25. Dagar, V.K.; Khasa, Y.P. Combined Effect of Gene Dosage and Process Optimization Strategies on High-Level Production of Recombinant Human Interleukin-3 (HIL-3) in Pichia Pastoris Fed-Batch Culture. *Int. J. Biol. Macromol.* **2018**, *108*, 999–1009. [CrossRef]
26. García, C.; Alcaraz, W.; Acosta-Cárdenas, A.; Ochoa, S. Application of Process System Engineering Tools to the Fed-Batch Production of Poly(3-Hydroxybutyrate-Co-3-Hydroxyvalerate) from a Vinasses–Molasses Mixture. *Bioprocess Biosyst. Eng.* **2019**, *42*, 1023–1037. [CrossRef] [PubMed]
27. Vanichsriratana, W. Comparison of Open Loop Optimal Control and Closed Loop Optimal Control of a Fermentation Process. In Proceedings of the UKACC International Conference on Control (Control '96), Exeter, UK, 2–5 September 1996; Volume 1996, pp. 258–263.

28. Vanichsriratana, W.; Zhang, B.S.; Leigh, J.R. Optimal Control of a Fed-Batch Fermentation Process. *Trans. Inst. Meas. Control* **1997**, *19*, 240–251. [CrossRef]
29. Hahn, J.; Edgar, T.F. Process Control Systems. In *Encyclopedia of Physical Science and Technology*; Elsevier: Amsterdam, The Netherlands, 2003; pp. 111–126.
30. Astrom, K.J. *PID Controllers; Theory, Design, and Tuning*, 2nd ed.; Instrument Society of America: Pittsburg, PA, USA, 1995; ISBN 1556175167.
31. Roeva, O.; Slavov, T. PID Controller Tuning Based on Metaheuristic Algorithms for Bioprocess Control. *Biotechnol. Biotechnol. Equip.* **2012**, *26*, 3267–3277. [CrossRef]
32. Schaepe, S.; Kuprijanov, A.; Aehle, M.; Simutis, R.; Lübbert, A. Simple Control of Fed-Batch Processes for Recombinant Protein Production with E. *Coli. Biotechnol. Lett.* **2011**, *33*, 1781–1788. [CrossRef] [PubMed]
33. Chopda, V.R.; Rathore, A.S.; Gomes, J. Maximizing Biomass Concentration in Baker's Yeast Process by Using a Decoupled Geometric Controller for Substrate and Dissolved Oxygen. *Bioresour. Technol.* **2015**, *196*, 160–168. [CrossRef]
34. Yeo, Y.K.; Kwon, T.I. Control of pH Processes Based on the Genetic Algorithm. *Korean J. Chem. Eng.* **2004**, *21*, 6–13. [CrossRef]
35. Akesson, M.; Hagander, P. *Control of Dissolved Oxygen in Stirred Bioreactors*; Technical Reports TFRT-7571; LUND Universit: Lund, Sweden, 1998; pp. 1–16.
36. Suzuki, T.; Yamane, T.; Shimizu, S. Phenomenological Background and Some Preliminary Trials of Automated Substrate Supply in pH-Stat Modal Fed-Batch Culture Using a Setpoint of High Limit. *J. Ferment. Bioeng.* **1990**, *69*, 292–297. [CrossRef]
37. Cutayar, J.M.; Poillon, D. High Cell Density Culture of E. Coli in a Fed-Batch System with Dissolved Oxygen as Substrate Feed Indicator. *Biotechnol. Lett.* **1989**, *11*, 155–160. [CrossRef]
38. Dairaku, K.; Izumoto, E.; Morikawa, H.; Shioya, S.; Takamatsu, T. An Advanced Micro-Computer Coupled Control System in a Baker's Yeast Fed-Batch Culture Using a Tubing Method. *J. Ferment. Technol.* **1983**, *61*, 189–196.
39. Nagy, Z.K. Model Based Control of a Yeast Fermentation Bioreactor Using Optimally Designed Artificial Neural Networks. *Chem. Eng. J.* **2007**, *127*, 95–109. [CrossRef]
40. Grigs, O.; Bolmanis, E.; Galvanauskas, V. Application of In-Situ and Soft-Sensors for Estimation of Recombinant P. pastoris GS115 Biomass Concentration: A Case Analysis of HBcAg (Mut+) and HBsAg (MutS) Production Processes under Varying Conditions. *Sensors* **2021**, *21*, 1268. [CrossRef]
41. Guarna, M.M.; Lesnicki, G.J.; Tam, B.; Robinson, J.; Radziminski, C.Z.; Hasenwinkle, D.; Boraston, A.; Jervis, E.; MacGillivray, R.T.A.; Turner, R.F.B.; et al. On-Line Control of Methanol Concentration in Pichia Pastoris Cultures. *Biotechnol. Bioeng.* **1997**, *56*, 279–286. [CrossRef]
42. Katakura, Y.; Zhang, W.; Zhuang, G.; Omasa, T.; Kishimoto, M.; Goto, Y.; Suga, K.I. Effect of Methanol Concentration on the Production of Human B2- Glycoprotein I Domain V by a Recombinant Pichia Pastoris: A Simple System for the Control of Methanol Concentration Using a Semiconductor Gas Sensor. *J. Ferment. Bioeng.* **1998**, *86*, 482–487. [CrossRef]
43. Valentinotti, S.; Srinivasan, B.; Holmberg, U.; Bonvin, D.; Cannizzaro, C.; Rhiel, M.; von Stockar, U. Optimal Operation of Fed-Batch Fermentations via Adaptive Control of Overflow Metabolite. *Control Eng. Pract.* **2003**, *11*, 665–674. [CrossRef]
44. Simutis, R.; Lübbert, A. Bioreactor Control Improves Bioprocess Performance. *Biotechnol. J.* **2015**, *10*, 1115–1130. [CrossRef]
45. Roeva, O.; Slavov, T. Fed-Batch Cultivation Control Based on Genetic Algorithm PID Controller Tuning. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Dimov, I., Dimova, S., Kolkovska, N., Eds.; Lecture Notes in Computer Science; Springer Berlin Heidelberg: Berlin/Heidelberg, Germany, 2011; Volume 6046, pp. 289–296. ISBN 978-3-642-18465-9.
46. Soons, Z.I.T.A.; Van Straten, G.; Van Der Pol, L.A.; Van Boxtel, A.J.B. Online Automatic Tuning and Control for Fed-Batch Cultivation. *Bioprocess Biosyst. Eng.* **2008**, *31*, 453–467. [CrossRef]
47. Kuprijanov, A.; Gnoth, S.; Simutis, R.; Lübbert, A. Advanced Control of Dissolved Oxygen Concentration in Fed Batch Cultures during Recombinant Protein Production. *Appl. Microbiol. Biotechnol.* **2009**, *82*, 221–229. [CrossRef]
48. Butkus, M.; Levišauskas, D.; Galvanauskas, V. Simple Gain-Scheduled Control System for Dissolved Oxygen Control in Bioreactors. *Processes* **2021**, *9*, 1493. [CrossRef]
49. Galvanauskas, V.; Simutis, R.; Vaitkus, V. Adaptive Control of Biomass Specific Growth Rate in Fed-Batch Biotechnological Processes. A Comparative Study. *Processes* **2019**, *7*, 810. [CrossRef]
50. Rajinikanth, V.; Latha, K. Modeling, Analysis, and Intelligent Controller Tuning for a Bioreactor: A Simulation Study. *ISRN Chem. Eng.* **2012**, *2012*, 413657. [CrossRef]
51. Levišauskas, D. An Algorithm for Adaptive Control of Dissolved Oxygen Concentration in Batch Culture. *Biotechnol. Tech.* **1995**, *9*, 85–90. [CrossRef]
52. Babuška, R.; Damen, M.R.; Hellinga, C.; Maarleveld, H. Intelligent Adaptive Control of Bioreactors. *J. Intell. Manuf.* **2003**, *14*, 255–265. [CrossRef]
53. Levišauskas, D.; Simutis, R.; Galvanauskas, V. Adaptive Set-Point Control System for Microbial Cultivation Processes. *Nonlinear Anal. Model. Control* **2015**, *21*, 153–165. [CrossRef]
54. Grigs, O.; Bolmanis, E.; Kazaks, A. HBsAg Production in Methanol Controlled *P. Pastoris* GS115 MutS Bioreactor Process. *Key Eng. Mater.* **2021**, *903*, 40–45. [CrossRef]
55. Priyanka; Roy, S.; Chopda, V.; Gomes, J.; Rathore, A.S. Comparison and Implementation of Different Control Strategies for Improving Production of RHSA Using Pichia Pastoris. *J. Biotechnol.* **2019**, *290*, 33–43. [CrossRef]

56. Kager, J.; Tuveri, A.; Ulonska, S.; Kroll, P.; Herwig, C. Experimental Verification and Comparison of Model Predictive, PID and Model Inversion Control in a Penicillium Chrysogenum Fed-Batch Process. *Process Biochem.* **2020**, *90*, 1–11. [CrossRef]
57. Lübbert, A.; Simutis, R. Using Measurement Data in Bioprocess Modelling and Control. *Trends Biotechnol.* **1994**, *12*, 304–311. [CrossRef]
58. Schubert, J.; Simutis, R.; Dors, M.; Havlik, I.; Lübbert, A. Bioprocess Optimization and Control: Application of Hybrid Modelling. *J. Biotechnol.* **1994**, *35*, 51–68. [CrossRef]
59. Ferreira, L.S.; De Souza, M.B.; Folly, R.O.M. Development of an Alcohol Fermentation Control System Based on Biosensor Measurements Interpreted by Neural Networks. *Sens. Actuators B Chem.* **2001**, *75*, 166–171. [CrossRef]
60. Chen, L.Z.; Nguang, S.K.; Chen, X.D.; Li, X.M. Modelling and Optimization of Fed-Batch Fermentation Processes Using Dynamic Neural Networks and Genetic Algorithms. *Biochem. Eng. J.* **2004**, *22*, 51–61. [CrossRef]
61. Beiroti, A.; Hosseini, S.N.; Aghasadeghi, M.R.; Norouzian, D. Comparative Study of μ-stat Methanol Feeding Control in Fed-batch Fermentation of Pichia pastoris Producing HBsAg: An Open-loop Control versus Recurrent Artificial Neural Network-based Feedback Control. *J. Chem. Technol. Biotechnol.* **2019**, *94*, 3924–3931. [CrossRef]
62. Jenzsch, M.; Simutis, R.; Eisbrenner, G.; Stückrath, I.; Lübbert, A. Estimation of Biomass Concentrations in Fermentation Processes for Recombinant Protein Production. *Bioprocess Biosyst. Eng.* **2006**, *29*, 19–27. [CrossRef]
63. Zhang, A.H.; Zhu, K.Y.; Zhuang, X.Y.; Liao, L.X.; Huang, S.Y.; Yao, C.Y.; Fang, B.S. A Robust Soft Sensor to Monitor 1,3-Propanediol Fermentation Process by Clostridium butyricum Based on Artificial Neural Network. *Biotechnol. Bioeng.* **2020**, *117*, 3345–3355. [CrossRef]
64. Peng, J.; Meng, F.; Ai, Y. Time-Dependent Fermentation Control Strategies for Enhancing Synthesis of Marine Bacteriocin 1701 Using Artificial Neural Network and Genetic Algorithm. *Bioresour. Technol.* **2013**, *138*, 345–352. [CrossRef]
65. Peng, W.; Zhong, J.; Yang, J.; Ren, Y.; Xu, T.; Xiao, S.; Zhou, J.; Tan, H. The Artificial Neural Network Approach Based on Uniform Design to Optimize the Fed-Batch Fermentation Condition: Application to the Production of Iturin A. *Microb. Cell Fact.* **2014**, *13*, 1–10. [CrossRef]
66. Chen, F.; Li, H.; Xu, Z.; Hou, S.; Yang, D. User-Friendly Optimization Approach of Fed-Batch Fermentation Conditions for the Production of Iturin A Using Artificial Neural Networks and Support Vector Machine. *Electron. J. Biotechnol.* **2015**, *18*, 273–280. [CrossRef]
67. Vijaya Krishna, V.V.S.; Pappa, N.; Joy, P.; Vasantharani, S. Realization of Deep Learning Based Embedded Soft Sensor for Bioprocess Application. *Intell. Autom. Soft Comput.* **2022**, *32*, 781–794. [CrossRef]
68. Chaudhuri, B.; Modak, J.M. Optimization of Fed-Batch Bioreactor Using Neural Network Model. *Bioprocess Eng.* **1998**, *19*, 71–79. [CrossRef]
69. Zelić, B.; Bolf, N.; Vasić-Rački, D. Modeling of the Pyruvate Production with Escherichia Coli: Comparison of Mechanistic and Neural Networks-Based Models. *Bioprocess Biosyst. Eng.* **2006**, *29*, 39–47. [CrossRef]
70. Rashid, M.M.; Hizbullah; Mohammad, N.; Khan, M.J.H. Advanced Control Technique for Substrate Feed Rate Regulation of a Fed Batch Fermentation. *Asian J. Biochem.* **2012**, *7*, 1–15. [CrossRef]
71. Pantano, M.N.; Serrano, M.E.; Fernández, M.C.; Rossomando, F.G.; Ortiz, O.A.; Scaglia, G.J.E. Multivariable Control for Tracking Optimal Profiles in a Nonlinear Fed-Batch Bioprocess Integrated with State Estimation. *Ind. Eng. Chem. Res.* **2017**, *56*, 6043–6056. [CrossRef]
72. Babuška, R.; Verbruggen, H.B. An Overview of Fuzzy Modeling for Control. *Control Eng. Pract.* **1996**, *4*, 1593–1606. [CrossRef]
73. Abyad, M.; Karama, A.; Khallouq, A. Fuzzy Takagi-Sugeno Based Modelling and Control for an Alcoholic Fermentation Process. In Proceedings of the 2017 International Conference on Electrical and Information Technologies (ICEIT), Rabat, Morocco, 15–18 November 2018; Volume 2018, pp. 1–6. [CrossRef]
74. Abyad, M.; Karama, A.; Khallouq, A. Takagi-Sugeno Tracking Control of a Fermentation Process with Respect to Asymmetric Constraints. *Int. J. Adapt. Control Signal Process.* **2020**, *34*, 266–282. [CrossRef]
75. Fonseca, R.R.; Sencio, R.R.; Franco, I.C.; Da Silva, F.V. An Adaptive Fuzzy Feedforward-Feedback Control System Applied to a Saccharification Process. *Chem. Prod. Process Model.* **2018**, *13*, 1–12. [CrossRef]
76. Escalante-Sánchez, A.; Barrera-Cortés, J.; Poggi-Varaldo, H.M.; Ponce-Noyola, T.; Baruch, I.S. A Soft Sensor Based on Online Biomass Measurements for the Glucose Estimation and Control of Fed-Batch Cultures of Bacillus Thuringiensis. *Bioprocess Biosyst. Eng.* **2018**, *41*, 1471–1484. [CrossRef]
77. Fonseca, R.R.; Franco, I.C.; Da Silva, F.V. Bioreactor Temperature Control Using a Generic Fuzzy Feedforward Controlsystem. In Proceedings of the 15th IASTED International Conference Intelligent Systems and Control (ISC 2016), Campinas, Brazil, 16–18 August 2016; pp. 275–280. [CrossRef]
78. Ye, K.; Jin, S.; Shimizu, K. Fuzzy Neural Network for the Control of High Cell Density Cultivation of Recombinant Escherichia Coli. *J. Ferment. Bioeng.* **1994**, *77*, 663–673. [CrossRef]
79. Baughman, D.R.; Liu, Y.A. Process Forecasting, Modeling, and Control of Time-Dependent Systems. In *Neural Networks in Bioprocessing and Chemical Engineering*; Academic Press: Cambridge, MA, USA, 1995; pp. 228–364. [CrossRef]
80. Choi, J.W.; Lee, W.; Cho, J.M.; Kim, Y.K.; Park, S.Y.; Lee, W.H. Control of Feed Rate Using Neurocontroller Incorporated with Genetic Algorithm in Fed-Batch Cultivation of Scutellaria baicalensis Georgi. *J. Microbiol. Biotechnol.* **2002**, *12*, 687–691.
81. Yuan, Y.J.; Miao, Z.Q.; Li, S.Y. The Fuzzy Neural Network Controller in Yeast Fed-Batch Fermentation. *Chem. Eng. Commun.* **1999**, *174*, 167–183. [CrossRef]

82. Dubencovs, K.; Liepins, J.; Suleiko, A.; Suleiko, A.; Vangravs, R.; Kassaliete, J.; Scerbaka, R.; Grigs, O. Optimization of Synthetic Media Composition for Kluyveromyces marxianus Fed-Batch Cultivation. *Fermentation* **2021**, *7*, 62. [CrossRef]
83. Aehle, M.; Bork, K.; Schaepe, S.; Kuprijanov, A.; Horstkorte, R.; Simutis, R.; Lübbert, A. Increasing Batch-to-Batch Reproducibility of CHO-Cell Cultures Using a Model Predictive Control Approach. *Cytotechnology* **2012**, *64*, 623–634. [CrossRef]
84. Hjersted, J.L.; Henson, M.A. Optimization of Fed-Batch Saccharomyces Cerevisiae Fermentation Using Dynamic Flux Balance Models. *Biotechnol. Prog.* **2008**, *22*, 1239–1248. [CrossRef]
85. Åkesson, M.; Hagander, P.; Axelsson, J.P. A Probing Feeding Strategy for Escherichia Coli Cultures. *Biotechnol. Tech.* **1999**, *13*, 523–528. [CrossRef]
86. Craven, S.; Whelan, J.; Glennon, B. Glucose Concentration Control of a Fed-Batch Mammalian Cell Bioprocess Using a Nonlinear Model Predictive Controller. *J. Process Control* **2014**, *24*, 344–357. [CrossRef]
87. Kuprijanov, A.; Schaepe, S.; Simutis, R.; Lübbert, A. Model Predictive Control Made Accessible to Professional Automation Systems in Fermentation Technology. *Biosyst. Inf. Technol.* **2013**, *2*, 26–31. [CrossRef]
88. Ławryńczuk, M. Modelling and Nonlinear Predictive Control of a Yeast Fermentation Biochemical Reactor Using Neural Networks. *Chem. Eng. J.* **2008**, *145*, 290–307. [CrossRef]
89. Abdulrahman, A. Control of a Yeast Fermentation Bioreactor Using Model Predictive Control Based on Radial Basis Function Network Modeling. *J. Sci. Technol.* **2014**, *19*, 24–45.
90. Zhang, H.; Lennox, B. Integrated Condition Monitoring and Control of Fed-Batch Fermentation Processes. *J. Process Control* **2004**, *14*, 41–50. [CrossRef]
91. Santos, L.O.; Dewasme, L.; Coutinho, D.; Wouwer, A. Vande Nonlinear Model Predictive Control of Fed-Batch Cultures of Micro-Organisms Exhibiting Overflow Metabolism: Assessment and Robustness. *Comput. Chem. Eng.* **2012**, *39*, 143–151. [CrossRef]
92. Gorrini, F.; Biagiola, S.; Figueroa, J.L.; Wouwer, A.V. Reaction Rate Estimation and Model Predictive Control of Hybridoma Cell Cultures. *IFAC-PapersOnLine* **2019**, *52*, 715–720. [CrossRef]
93. Ionkova, I.; Sasheva, P.; Ionkov, T. Enhanced Production of Flavonoids in Astragalus Missouriensis, Using Bioreactor by Model Based Control of the Bioprocess. In Proceedings of the 7th Conference on Medicinal and Aromatic Plants of Southeast European Countries, Subotica, Serbia, 12 May 2012; pp. 332–337.
94. Lupenza, L.B.; Subbian, S.; Murugan, C. Development of Event Triggered Feed Forward Control Scheme for Fed-Batch E. Coli Fermentation Process. In Proceedings of the 2021 IEEE 18th India Council International Conference (INDICON), Guwahati, India, 19–21 December 2021; pp. 1–6.
95. Ulonska, S.; Waldschitz, D.; Kager, J.; Herwig, C. Model Predictive Control in Comparison to Elemental Balance Control in an E. Coli Fed-Batch. *Chem. Eng. Sci.* **2018**, *191*, 459–467. [CrossRef]
96. Schneider, R.; Jalel, N.A.; Munack, A.; Leigh, J.R. Adaptive Predictive Control for the Fed Batch Fermentation Process. In Proceedings of the International Conference on Control '94, Coventry, UK, 21–24 March 1994; pp. 249–254.
97. Controlling the Living Biopharmaceutical Factory. Available online: https://new.abb.com/control-systems/industry-specific-solutions/pharmaceutical-and-life-sciences/controlling-the-living-biopharmaceutical-factory (accessed on 20 February 2023).
98. Sommeregger, W.; Sissolak, B.; Kandra, K.; von Stosch, M.; Mayer, M.; Striedner, G. Quality by Control: Towards Model Predictive Control of Mammalian Cell Culture Bioprocesses. *Biotechnol. J.* **2017**, *12*, 1600546. [CrossRef]
99. Grigs, O.; Galvanauskas, V.; Dubencovs, K.; Vanags, J.; Suleiko, A.; Berzins, T.; Kunga, L. Model Predictive Feeding Rate Control in Conventional and Single-Use Lab-Scale Bioreactors: A Study on Practical Application. *Chem. Biochem. Eng. Q. J.* **2016**, *30*, 47–60. [CrossRef]
100. Ochoa, S. Fed-Batch Fermentation—Design Strategies. In *Comprehensive Biotechnology*; Elsevier: Amsterdam, The Netherlands, 2019; Volume 2, pp. 586–600. ISBN 9780444640475.
101. Yu, L.X.; Amidon, G.; Khan, M.A.; Hoag, S.W.; Polli, J.; Raju, G.K.; Woodcock, J. Understanding Pharmaceutical Quality by Design. *AAPS J.* **2014**, *16*, 771–783. [CrossRef] [PubMed]
102. Meyer, H.-P.; Minas, W.; Schmidhalter, D. Industrial-Scale Fermentation. In *Industrial Biotechnology*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2016; pp. 1–53.
103. Kiran, A.U.M.; Jana, A.K. Control of Continuous Fed-Batch Fermentation Process Using Neural Network Based Model Predictive Controller. *Bioprocess Biosyst. Eng.* **2009**, *32*, 801–808. [CrossRef] [PubMed]
104. Eeranna, B.C.; Rao, P.S.; Reddy, G.P. Artificial Neural Network Based Modeling and Control of Bioreactor. *Int. J. Eng. Res. Technol.* **2014**, *3*, 117–122.
105. Zulkeflee, S.A.; Aziz, N.; Campus, E.; Ampangan, S.; Tebal, N. Control Implementation in Bioprocess System: A Review. *Int. Conf. Control. Instrum. Mechatron. Eng.* **2007**, 798–804.
106. Jabarivelisdeh, B.; Findeisen, R.; Waldherr, S. Model Predictive Control of a Fed-Batch Bioreactor Based on Dynamic Metabolic-Genetic Network Models. *IFAC PapersOnLine* **2018**, *51*, 34–37. [CrossRef]
107. Joynes, L.; Zhang, J. A Feedback Control Strategy for a Fed-Batch Monoclonal Antibody Production Process Utilising Infrequent and Irregular Sampled Measurements. *Processes* **2022**, *10*, 1448. [CrossRef]
108. Eaton, J.W.; Rawlings, J.B. Model-Predictive Control of Chemical Processes. *Chem. Eng. Sci.* **1992**, *47*, 705–720. [CrossRef]

109. Karra, S.; Shaw, R.; Patwardhan, S.C.; Noronha, S. Adaptive Model Predictive Control of Multivariable Time-Varying Systems. *Ind. Eng. Chem. Res.* **2008**, *47*, 2708–2720. [CrossRef]
110. Rathore, A.S.; Mishra, S.; Nikita, S.; Priyanka, P. Bioprocess Control: Current Progress and Future Perspectives. *Life* **2021**, *11*, 557. [CrossRef]

Appendix II
2. Pielikums

Anomaly Detection and Removal Strategies for In-Line Permittivity Sensor Signal Used in Bioprocesses

Emils Bolmanis*, Selina Uhlendorff*, Miriam Pein-Hackelbusch, Vytautas Galvanauskas, Oskars Grigs

Frontiers in Bioengineering and Biotechnology **2025**, *13*, DOI: 10.3389/fbioe.2025.1609369

* These authors have contributed equally to this work and share first authorship.

E.B. input: Conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing – original draft, writing – review and editing.

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 30 July 2025
DOI 10.3389/fbioe.2025.1609369

OPEN ACCESS

EDITED BY
Krist V. Gernaey,
Technical University of Denmark, Denmark

REVIEWED BY
Ezhaveni Sathiyamoorthi,
Yeungnam University, Republic of Korea
Oscar Andrés Prado-Rubio,
National University of Colombia, Manizales, Colombia

*CORRESPONDENCE
Oskars Grigs,
oskars.grigs@kki.lv

†These authors have contributed equally to this work and share first authorship

RECEIVED 10 April 2025
ACCEPTED 21 July 2025
PUBLISHED 30 July 2025

CITATION
Bolmanis E, Uhlendorff S, Pein-Hackelbusch M, Galvanauskas V and Grigs O (2025) Anomaly detection and removal strategies for in-line permittivity sensor signal used in bioprocesses. *Front. Bioeng. Biotechnol.* 13:1609369. doi: 10.3389/fbioe.2025.1609369

Anomaly detection and removal strategies for in-line permittivity sensor signal used in bioprocesses

Emils Bolmanis[1,2,3†], Selina Uhlendorff[4†], Miriam Pein-Hackelbusch[4], Vytautas Galvanauskas[5] and Oskars Grigs[1*]

[1]Laboratory of Bioengineering, Latvian State Institute of Wood Chemistry, Riga, Latvia, [2]K. Tars Lab, Latvian Biomedical Research and Study Centre, Riga, Latvia, [3]Institute of Biomaterials and Bioengineering, Riga Technical University, Riga, Latvia, [4]Institute for Life Science Technologies ILT.NRW, OWL University of Applied Sciences and Arts, Lemgo, Germany, [5]Department of Automation, Kaunas University of Technology, Kaunas, Lithuania

Introduction: In-line sensors, which are crucial for real-time (bio-) process monitoring, can suffer from anomalies. These signal spikes and shifts compromise process control. Due to the dynamic and non-stationary nature of bioprocess signals, addressing these issues requires specialized preprocessing. However, existing anomaly detection methods often fail for real-time applications.

Methods: This study addresses a common yet critical issue: developing a robust and easy-to-implement algorithm for real-time anomaly detection and removal for in-line permittivity sensor measurement. Recombinant *Pichia pastoris* cultivations served as a case study. Trivial approaches, such as moving average filtering, do not adequately capture the complexity of the problem. However, our method provides a structured solution through three consecutive steps: 1) Signal preprocessing to reduce noise and eliminate context dependency; 2) Anomaly detection using threshold-based identification; 3) Validation and removal of identified anomalies.

Results and discussion: We demonstrate that our approach effectively detects and removes anomalies by compensating signal shift value, while remaining computationally efficient and practical for real-time use. It achieves an F1-score of 0.79 with a static threshold of 1.06 pF/cm and a double rolling aggregate transformer using window sizes $w1 = 1$ and $w2 = 15$. This flexible and scalable algorithm has the potential to bridge a crucial gap in process real-time analytics and control.

KEYWORDS

in-situ, permittivity, dielectric spectroscopy, signal preprocessing, dynamic threshold, static threshold, anomaly validation, *Pichia pastoris*

1 Introduction

The quest for efficiency, safety and sustainability is driving new developments in the bioprocess industry. This includes monitoring, controlling and predicting cell cultivation processes as continuously as possible and in real-time. Achieving this requires comprehensive process knowledge as well as the analysis of various process parameters, which are recorded using modern sensor technology (Mandenius and Gustavsson, 2015). In-line sensors, which do not influence the process or the product and continuously supply process data in real-time, are particularly important here. They can ensure early detection of deviations, such as nutrient limitations, and are thus able to optimally determine feeding profiles or harvesting times, for example. Since a single sensor signal can rarely provide information about such a complex process as cell cultivation, it is worthwhile to use mathematical models to fuse the signals of different sensors into a so-called soft sensor. In the development of such soft sensors, the quality of the data is of crucial importance so that the mathematical model to be created on the basis of the data is not negatively influenced (Warne et al., 2004; Brunner et al., 2021).

According to the International Standard ISO/IEC 25012:2008, data quality is defined as 'degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions', where data quality characteristics are defined as 'category of data quality attributes that bears on data quality' (International Organization for Standardization, 2008). For a detailed list of the 15 characteristics defined there, we refer to this standard. Among other important characteristics, the most important one to consider is credibility. Credibility refers to the extent to which data possesses attributes that are considered authentic and trustworthy by users within a given context of application (International Organization for Standardization, 2008). At this point, the aforementioned process knowledge comes into play, helping to assess whether or not the captured sensor data is true and believable.

An example for this is the recorded trend of the viable biomass (more precisely, the signal which is correlated with it), which, depending on the process control strategy, should correspond to classical growth kinetics. If irregularities such as spikes or signal shifts are detected, this indicates in most cases an anomaly of the sensor signal and not of the true viable biomass value. To record such a signal in-line and in real-time, permittivity probes that can infer viable cell density are suitable, for example. These probes polarize cells with intact cell membranes through an alternating electric field, while dead cells with damaged membranes are not polarized and thus not measured (Metze et al., 2020). However, it is important to correlate the probe signals with off-line reference analytics to make a qualitative statement about the viable cell density (Ramm et al., 2023).

Depending on the sensor and the underlying measurement technology, signal anomalies can be caused by external process changes, such as a change in agitator speed, the addition of an antifoam agent (Grigs et al., 2021a) or movement of bubbles near the sensor tip (Fehrenbach et al., 1992; Konstantinov et al., 1992; Münzberg et al., 2017; Katla et al., 2019; Brignoli et al., 2020). In such cases, the recorded sensor signal does not reflect the true viable biomass value. If this erroneous, unreliable data were fed into a mathematical model without any preprocessing, this would lead to a flawed model. Therefore, it is of utmost importance to detect and filter out signal anomalies in the preprocessing step (Kadlec et al., 2009; Jiang et al., 2021).

However, there are both process-inherent and application-dependent aspects that must be considered in such a preprocessing task. One of the process-inherent aspects is that the recorded viable cell density signal is time dependent, as bioprocesses generally are, which is why the signal is non-stationary. This means that changes in the mean (increasing signal due to increasing cell biomass) and variance (e.g., increased signal noise at low cell densities) can be observed. Another obvious but important aspect to consider is the fact that both data preprocessing and modeling must be possible in real-time, i.e., during the ongoing process. The acceptable latency between the time when an event occurs in the process and the time it is detected depends on the application's goal. If the intended use of the developed soft sensor is solely for monitoring purposes, for example, there are lower demands placed on latency compared to when it is intended for control, where rapid responsiveness is of great importance. Dependent on this, filters and algorithms are used in data preprocessing and mathematical model building, which may entail a time delay or high computational power. Further requirements for streaming algorithms can be found in (Ahmad et al., 2017; Blázquez-García et al., 2022).

The existing anomaly detection techniques can be categorized based on input dimensionality, learning type category, and method family. Input dimensionality differentiates between univariate and multivariate data types and describes the extent to which algorithms can handle inter-variable. In terms of learning types, techniques can be classified as unsupervised, semi-supervised, or supervised. The method families can be broadly divided into six categories: forecasting, reconstruction, distance, encoding, distribution, and tree methods (Schmidl et al., 2022). However, none of these categorizations provide insight into whether the respective algorithms are fundamentally suitable for real-time application in non-stationary processes.

Regarding the categorization of different anomalies, a common distinction is made between point anomalies and sequential anomalies (Schmidl et al., 2022), with the former often appearing as contextual anomalies in time series data (Chandola et al., 2009). This is because a signal value recorded at a specific time point may represent an anomaly due to its context but would not be classified as an anomaly if it occurred at a different time. Since this context dependency complicates anomaly detection, it is beneficial to transform the sensor signal data in such a way that the contextual information is removed, leaving point anomalies without context dependency. For detecting those, Chandola et al. propose classifying anomaly detection techniques into six categories, including, for example, classification based techniques, nearest neighbor based techniques, clustering based techniques and statistical techniques (parametric methods such as gaussian model-based and non-parametric methods such as histogram-based) (Chandola et al., 2009).

Both, traditional, manual anomaly detection and modern machine learning methods have the disadvantage of rarely working in real-time (Hill and Minsker, 2010; Ahmad et al., 2017). Even algorithms that could theoretically be applied in

real-time do not inherently guarantee that their implementation as a streaming algorithm will work in practice. In this regard, the data acquisition rate and the computation time of the algorithm must always be taken into account, which, depending on the process, may make real-time integration impossible (Blázquez-García et al., 2022).

Assuming that our transformed, context-free data follows a Gaussian distribution, we have chosen Gaussian Model-Based techniques for this study. These methods offer the advantage that they can be applied as streaming algorithms due to their low computational power and are also relatively easy to understand and implement. The latter was important to us so that a broad readership can apply our algorithm to their own sensor signal data.

To the best of our knowledge, the topic of anomaly detection and removal in sensor signal data for recording in-line viable cell biomass in bioprocesses remains largely unexplored in the existing literature. The only known contribution in this area is the work by Grigs et al. (2021a), which forms the basis of the present study. Building on this groundwork, our study addresses a critical gap and pioneers further exploration into this underdeveloped yet essential field.

Using permittivity measurements from recombinant *P. pastoris* fermentations, this study aimed to develop an algorithm for detecting and removing signal anomalies in real-time. To achieve this goal, three main questions needed to be addressed.

1. How to overcome the non-stationarity of the signal?
2. How to detect anomalies?
3. How to remove anomalies?

Based on the three questions above, our approach can be divided into three consecutive steps, into which both, this study and the algorithm, are divided. Step 1) is the signal preprocessing, which includes the reduction of noise and the transformation of the smoothed signal to remove context dependency. Step 2) is the anomaly detection and the associated selection of an appropriate threshold, based on which an anomaly is classified as such. Step 3) is the validation of anomalies and their removal.

Our requirements for the algorithm included the possibility of real-time in-line application and minimal complexity in terms of mathematical and computational aspects.

2 Theory

2.1 Signal preprocessing

With regard to the variance of the signal over time, it becomes apparent that data smoothing is necessary to reduce noise. To minimize signal noise before the actual signal anomaly detection and removal, various smoothing methods seem suitable, which will be discussed in more detail below. For all methods, the window size w is a freely selectable and optimizable parameter. However, it should be noted that this choice involves a trade-off when implementing the filter in real-time: the larger the chosen window size w, the stronger the noise reduction, but also the greater the time delay between input (raw signal) and output (smoothed signal), which can be described by $(w\text{-}1)/2$ (Harju et al., 1996).

The moving mean smooths signals by calculating the mean of data points over a specific window size, which is typically centered on the point being analyzed. The moving median works the same way, except that the median is used as the aggregation function instead of the mean.

In the Gaussian filter, a weight is calculated for each data point within the selected window based on an underlying Gaussian function, and the value of the data point is multiplied by the respective weight. To obtain the smoothed value for the central point of the window, the sum of the weighted data is divided by the sum of the weights. In addition to the window size, the standard deviation σ of the Gaussian function is a freely selectable parameter.

The local linear/quadratic regression (lowess/loess) smooths values by fitting a linear or quadratic function to the data points within a window using weighted least squares. Tricubic weighting is typically used, giving more weight to the nearest and less weight to the furthest points. The robust variant is more resistant to outliers but more computationally expensive as the regression is adjusted not just by simple, but by iterated weighted least squares (Cleveland, 1979).

The Savitzky-Golay filter fits a polynomial of degree n to the data points within a window. The window size must be at least $n + 1$ points, and it is typically centered on the point being analyzed. The result of the filter is a smoothed value for the center point within the window (Savitzky and Golay, 1964). The degree n of the polynomial function is a freely selectable parameter.

In addition to the variance inhomogeneity of the signal over time, the change of the signal mean is another factor of non-stationarity. To address this issue, it is advisable to transform the signal in such a way that the mean of the transformed signal remains constant over time. To achieve this, a double rolling aggregate (DRA) can be used. This transformer consists of two windows, which can be freely sized, moving in parallel along the time axis over the data series. These windows can move side by side or overlap, and within each window, the data are aggregated according to the chosen aggregation function. The DRA compares the aggregated metrics of the two windows by subtracting the metric of one window from the metric of the other and saves those differences as the transformed signal. So if there is a sudden increase in signal, it is first reflected in the metric of the right window. Consequently, the difference between both window metrics, i.e., the transformed signal, also increases significantly.

2.2 Anomaly detection

To assess whether the difference between the two metrics of the rolling windows, referred to as transformed signal, is significant enough to be considered an anomaly, a threshold value is required above which the corresponding signal is classified as an anomaly. However, this threshold must be chosen wisely to avoid classifying too many values as false positives if the threshold is too low, and to ensure that anomalies are still recognized as such if the threshold is too high. When choosing an appropriate threshold, there are generally two different approaches. Either a threshold is set manually based on experience and visual assessment of the transformed signal, or the threshold is set based on the location and scale estimators of the respective data. The latter approach can

be applied both off-line to the entire dataset and in process simulations, when only the past and present data is available at any given time point, to the local areas defined by a predefined window. When implemented in process simulations, unlike the manual method, the threshold is not static but dynamic, adapting to the continuously provided new data. For this dynamic determination of the threshold value, various approaches are available (Jones, 2019; Berger and Kiefer, 2021), which can be expressed in the form of Equation 1. The following sections will detail three methods applied in this study.

$$threshold = location\,estimator \pm threshold\ factor * scale\,estimator \quad (1)$$

The probably best-known and most frequently used method is the 3-sigma rule (Pearson, 2001; 2002; Chiang et al., 2003; Lin et al., 2007; Zhu et al., 2018; Jones, 2019). The 3-sigma rule states that, in a normal distribution, approximately 99.73% of the data points will fall within three standard deviations of the mean (Zhao et al., 2013). This means that the probability of a data point lying outside of this range is very low, making it a useful rule of thumb for identifying outliers. It is important to note that the anomaly detection result depends on the relationship between the threshold factor and the window size (Shiffler, 1988). For example, Berger et al. describe that when using the 3-sigma rule, the sample size must be > 10 in order to possibly detect any outliers (Berger and Kiefer, 2021). In general, the maximum threshold factor can be calculated by $(w-1)/\sqrt{w}$ with the window size w. As the name implies, in this method the threshold factor is set to 3, and the standard deviation from the mean (location estimator) is used as the scale estimator (Equation 2).

$$threshold = \bar{x} \pm 3 * \sigma \quad (2)$$

However, since both the mean and especially the standard deviation are very outlier-sensitive and can be overestimated by outliers, the masking effect occurs, leading to false negatives. In addition, the 3-sigma rule assumes symmetry, which can lead to false positives if this assumption is violated (Jones, 2019). Therefore, it is advisable to use more robust methods, where the mean and standard deviation are replaced by more robust location and scale estimators. The mean can be replaced by the median, and there are two options for replacing the standard deviation. If the standard deviation is replaced by the median of absolute deviation (MAD) scale estimate, the resulting method is called the Hampel identifier (Pearson, 2005). The drawback of this more robust method is that more values tend to be identified as false positives, known as swamping (Davies and Gather, 1993; Pearson, 2005), which is the opposite of the masking effect. The MAD scale estimate is the product of the constant b and the MAD. The value of the constant b depends on the underlying distribution and can be calculated as the reciprocal value of the 75th percentile (Huber, 2011; Leys et al., 2013). For a normal distribution, b is 1.4826 (Davies and Gather, 1993; Rousseeuw and Croux, 1993; Chiang et al., 2003; Pearson, 2005; Lin et al., 2007). The threshold factor is usually set at 2.0, 2.5 or 3.0 (Miller, 1991). The general notation is shown in Equation 3.

$$threshold = \tilde{x} \pm threshold\ factor * (b \times median\{|x_i - \tilde{x}|\}) \quad (3)$$

Within a signal window, where the signal values change mainly due to the noise and not due to a process trend, normal distribution is most probable.

The other option for replacing the standard deviation with a more robust scale estimator is the interquartile range (IQR) scale estimate. It is based on the range between the 75th percentile (Q_3) and the 25th percentile (Q_1) and is less sensitive to outliers than the standard deviation but more sensitive than the MAD scale estimate. Similar to the MAD scale estimate, a correction factor is introduced for the IQR scale estimate, which is 1.35 for a threshold factor of 2 (Equation 4).

$$\begin{aligned} threshold &= \tilde{x} \pm threshold\ factor * 1.35 * \left(\frac{Q_3 - Q_1}{1.35}\right) \\ &= \tilde{x} \pm threshold\ factor * 1.35 * \sigma \end{aligned} \quad (4)$$

These outlier detection limits correspond to approximately ± 2.7*σ, as the IQR divided by this correction factor leads to an unbiased estimate of the standard deviation σ for normally distributed data (Venables and Ripley, 2002; Pearson, 2005; Higgins and Green, 2008).

While there are more advanced methods for anomaly detection, such as machine learning-based and deep learning-based approaches (e.g., autoencoders), these techniques fall outside the scope of this study. The primary reason is their complexity and the requirement for sufficiently large datasets to ensure good model performance (Darban et al., 2024; Iqbal and Amin, 2024). In the context of bioprocesses, obtaining such large datasets is often challenging, as data collection is typically expensive and time-consuming. Moreover, the effective implementation of these advanced methods demands interdisciplinary expertise in data science, statistics, and bioprocess engineering, which can limit their accessibility and practical adoption in many industrial and academic settings. Therefore, we focus on threshold-based methods that are more practical given the constraints of bioprocess monitoring.

3 Materials and methods

Since this study refers to the data from Grigs et al., only the key aspects of the cultivation and data acquisition are described below. For details, we refer to the original study where the experimental data were recorded (Grigs et al., 2021a). HBcAg (Mut^+) and HBsAg (Mut^S) recombinant *P. pastoris* GS115 strains (obtained from Latvian Biomedical research and study centre) were cultivated in 5 L fully automated bench-top bioreactor systems EDF-5.4 (Biotehniskais Centrs, Riga, Latvia). Residual methanol levels varied between 0.01–7 g/L during the protein production phase, process temperature was 30°C ± 0.1 C (or 24°C ± 0.1 °C for Exp. 2) and the aeration rate was set at 3.0 slpm. The dissolved oxygen level varied between 3%–40% and the set-point of 30% ± 5% was controlled by automatically adjusting the stirrer rotational speed (200–1000 rpm) or additional inlet air enrichment with oxygen. The permittivity signal (Hamilton, Bonaduz, Switzerland, Incyte) was recorded every 60 s. According to the manufacturer, the permittivity probe has an accuracy of ± 1 pF/cm or ± 1%, whichever is greater across the full measurement range. Zero calibration was conducted before inoculation using cell-free culture media under process conditions. The duration of the time series and the scale of permittivity values for each experiment are summarized in

Supplementary Table S1. Out of a dataset of 13 experiments, only eight contained permittivity sensor data and were selected for this study. The names of the eight experiments considered from the original study (1s–4s; 3c–6c) correspond to experiment numbers 1–8 in this work. The suffixes s and c in the original study refer to the particular *P. pastoris* producer strain employed, with s denoting the hepatitis B surface antigen (HBsAg) producer and c the hepatitis B core antigen (HBcAg) producer (Grigs et al., 2021b; Bolmanis et al., 2022).

MATLAB version R2021b (Mathworks, Natick, MA, USA) with the Statistics and Machine Learning Toolbox was used for algorithm code and figure creation. The algorithm was visualized using draw.io (https://drawio.com).

Algorithm implementation and calculations were performed on a desktop computer with an Intel i5-6600 (3.90 GHz) processor and 16 GB RAM.

In the development of the algorithm, we followed three steps: signal preprocessing, anomaly detection, and anomaly validation and removal. Accordingly, this chapter will go through these steps in sequence.

3.1 Signal preprocessing

To evaluate the performance of the smoothing method, a reference signal with no noise is required. The permittivity signals were first smoothed off-line, utilizing the complete dataset to obtain a 'noiseless' signal as a reference. Several methods, namely, the moving mean and median, Gaussian filter, Savitzky-Golay filter, local linear and quadratic regression, and their robust equivalents were applied and the results were visually compared in terms of preserving the original signal pattern and removing most signal fluctuations.

Once the optimal off-line data smoothing method was identified, the resulting 'noiseless' signal served as a reference. The noise in the smoothed signals was then assessed by calculating the average normalized root mean square error (NRMSE) for each experimental dataset (Equation 5).

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2}{n}}}{y_{max} - y_{min}} 100\% \quad (5)$$

Where y_i is the ith reference noiseless permittivity signal value, y_i^* is the smoothed permittivity signal value, $y_{\min}$ and $y_{\max}$ are the minimum and maximum values of reference y_i.

Higher signal fluctuations (noise) directly correspond to an increased NRMSE value. In contrast to the previous step, this smoothing was performed in fermentation process simulations, where only past and present (not future) data is available to the model at any given time. Several different signal filtration methods and parameters were investigated using the *smoothdata* function (MATLAB). Namely, moving mean and median, Gaussian filter, Savitzky-Golay filter, local linear and quadratic regression, and their robust equivalents. For each of these methods, the optimal smoothing window sizes were identified, achieving the best signal noise reduction performance, as indicated by the lowest NRMSE. Since signal filtration in real-time often introduces a signal delay, this delay must also be taken into account. The smoothed signal delays in process simulations were estimated using signal cross-correlation function *xcorr* (MATLAB) comparing the transformed raw and smoothed permittivity signals. Cross-correlation is a widely used technique for estimating signal delay by measuring the similarity between two signals as a function of time lag. In this approach, one signal is systematically shifted relative to the other, and their correlation is computed at each shift. The time lag corresponding to the maximum correlation value indicates the estimated delay between the signals (Müller et al., 2003).

For all methods, we investigated window sizes of 2–180 data points, corresponding to 2–180 min. For the Gaussian filter, the standard deviation was fixed to be 1/5th of the total window width. A lowess/loess smoothing technique was applied using weighted linear or quadratic least squares with a first- or second-degree polynomial model. The degree n of the polynomial function for the Savitzky-Golay filter was set to two.

For the transformation of the smoothed signal to remove context dependency, we used the DRA with the mean as aggregate function. The window sizes *w1* and *w2* varied from one to 20 with increments of one; the windows did overlap entirely.

3.2 Anomaly detection

The anomaly detection is based on the selection of an appropriate threshold, based on which an anomaly is classified as such. For this, the approaches outlined in the theory section were compared. On the one hand, this includes the manual method, where the optimal threshold value was determined from a range from 0.10 to 1.45 in increments of 0.01. The other approaches are based on the real-time implementation of a dynamic threshold. In addition to the 3-sigma rule, we applied the Hampel identifier where we set the factor b to 1.4826, since we assume normal distribution. In accordance with the 3-sigma rule, the threshold factor was set to three. Furthermore, we exerted the threshold determination using the IQR scale estimate with ± 2.7*σ. The performance of the method was compared as described below.

First, the signal anomalies for each experiment were manually annotated (Supplementary Table S1). The decrease in the permittivity signal after approx. 20–30 h was not considered anomalous as it was due to switching the feed substrate from glycerol to methanol, which is usually followed by an adaptation period corresponding to a reduced cell viability and little to no growth for approx. 1–2 h (Ferreira et al., 2014). Then, the threshold methods were applied to the transformed signals of each experiment. Various window *w3* sizes were chosen between 120 and 180 for the 3-sigma method and from two to 30 for the MAD and IQR method. In both cases, *w3* was varied in increments of two. To determine which window size produced the best results, the true positives (TP), false positives (FP) and false negatives (FN) were analyzed by comparing the detected anomalies with the annotated anomalies. Using these measures, the precision (TP/(TP + FP)), recall (TP/(TP + FN)) and F1-score (2 * precision * recall/(precision + recall)) were calculated. The best results are indicated by the highest average F1-score across all experiments. In this way, the window size for each method that achieves the best results (i.e., the highest average F1-scores) can be determined. Finally, the highest F1-scores of the different methods can be

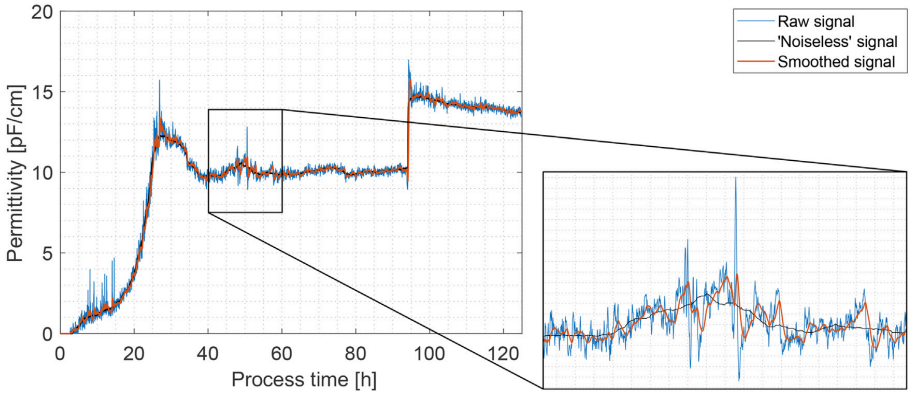


FIGURE 1
Comparison of raw (colored in blue), 'noiseless' (off-line 'loess' filter) (colored in black) and real-time smoothed (Gaussian filter, w = 70) (colored in orange) permittivity signals in Exp. 1 for signal noise filtering method performance estimation.

compared to identify the best method. Exp. 6 was omitted from this calculation as no significant signal anomalies were detected for this experiment.

3.3 Anomaly removal

If a signal anomaly is detected, the permittivity signal will be corrected by replacing the anomalous value with the mean of the previous 15 values. Once the anomaly has passed, a 15-min validation window begins to estimate the new signal baseline. The baseline level before the anomaly is then subtracted from the baseline level after the anomaly to determine a correction term, which is applied to the permittivity signal.

4 Results and discussion

To address the in-line permittivity sensor signal anomalies during recombinant yeast *P. pastoris* fermentations, we developed an algorithm for real-time detection and removal of these anomalies. The algorithm consists of three consecutive steps: 1) signal preprocessing, 2) anomaly detection, 3) anomaly validation and removal. Each step is detailed in the following sections.

4.1 Signal preprocessing

By analyzing the experimental dataset, we found that the permittivity signal contains significant noise and the overall signal quality should be subject to improvement. The signal noise was also even more prominent in two experiments (3 and 4), which were conducted in a different cultivation medium, indicating that the permittivity signal noise could be affected, for example, by medium conductivity. Henceforth, we found it essential to include a prerequisite signal noise filtering step to improve overall signal quality prior to anomaly detection.

'Noiseless' reference signals were obtained by smoothing the raw permittivity signals off-line utilizing the whole dataset with different methods. The best results with high preservation of the original signal pattern and removing most signal fluctuations were achieved by using a local quadratic regression smoothing filter with a smoothing factor of 0.03 (see Figure 1). Other smoothing methods failed to fully encapsulate the underlying signal characteristics by either cutting off distinctive signal peaks or misrepresenting anomalous signal jumps and spikes.

The smoothing performance of various methods was evaluated by calculating the NRMSE between the 'noiseless' reference and the smoothed signals in fermentation process simulations utilizing only past and present data at any given time point, as well as estimating the signal delay between the raw and smoothed signals. The best results were achieved, using a Gaussian smoothing filter with a window size of 70. In this case, an average NRMSE of 4.56% with a standard deviation of ± 1.40% was achieved (in comparison to 6.76% ± 1.93% for the raw signal) with an average estimated signal delay over all experiments of 6.4 min. Similar performance was noted by the moving mean filter (4.89% ± 1.55%), however, significantly higher signal delays of an average of 10.1 min were noted. The performance of other methods were deemed unsatisfactory either due to higher NRMSE values or prolonged signal delays. In the case of robust local linear/quadratic regressions (rloess/rlowess), a significantly higher computational burden was noted and thus these methods were excluded from consideration for real-time signal smoothing implementation. For the extended results, we refer to Supplementary Table S2.

As a result of this step, a higher quality permittivity signal was produced for signal anomaly detection in the next step. Much of the signal noise was removed and, although slight signal delays were introduced (which is to be expected), we estimate that they are not significant enough not to warrant using the filtered signal for real-time substrate feed rate adjustment in yeast or mammalian cell fed-batch bioprocesses, for example. Of course, the acceptable signal delay is highly process-specific and should be considered with every application. The average specific growth rate for *P. pastoris* Mut$^+$

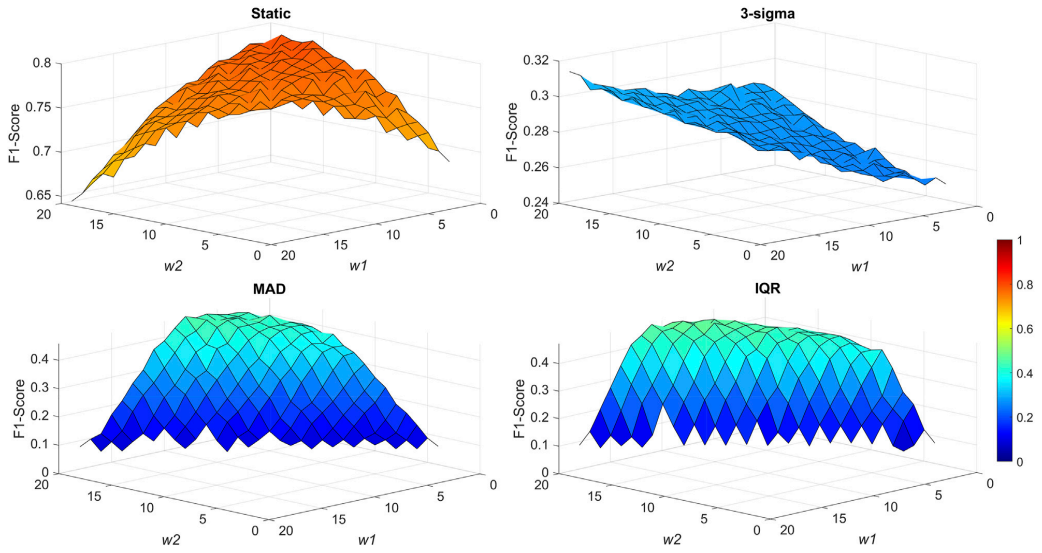


FIGURE 2
Highest F-scores achieved for respective types of thresholds at selected DRA transformer window sizes *w1* and *w2* with optimized threshold values (for static threshold type) and window sizes *w3* (for dynamic threshold types).

phenotype on methanol varies between 0.02 and 0.15 h^{-1} (Looser et al., 2015). This represents an average biomass increase by 2.0%–15.0% every hour. Hence, a signal delay of 5–10 min can be considered insignificant. The results for signal real-time smoothing in Exp. 1, using a Gaussian filter with a window size of 70, are shown in Figure 1. Permittivity signal preprocessing is used quite often when employing an in-line sensor probe (Ramm et al., 2023), however, necessary signal quality is often determined by the way the signal is to be used. For example, some authors have used the permittivity signal only for monitoring purposes, thus choosing not to apply any additional signal processing steps (Sarrafzadeh et al., 2005; Meitz et al., 2016; Pentjuss et al., 2023; Sakiyo and Németh, 2023). On the other hand, when choosing (or by necessity) to filter the permittivity signal, a moving average filter or a variation of it is often employed with window sizes varying from 15–110 samples (Da Silva et al., 2013; Downey et al., 2014; Horta et al., 2015). Horta et al. thereby developed a smoothed moving average filter, which performed better in permittivity sensor signal noise reduction than a classical moving average filter (Horta et al., 2012). The filtered signal was then used to estimate the cell growth rate (μ) and control the substrate feed rate in *E. coli* cultivations. The authors also emphasize that an efficient noise filter was essential for a good performance of the control system. A similar control strategy was also employed by Da Silva et al. (2013).

4.2 Anomaly detection

For signal anomaly detection using the DRA transformer, we investigated four different strategies for anomaly threshold determination. In addition to the manually selected static threshold, we tested three different variants of a dynamic threshold based on the 3-sigma rule, MAD and IQR scale estimates. The respective resulting F1-scores are shown in Figure 2. We also examined consecutive (once per selected period, *w4*) dynamic threshold calculations, however, in all cases, a lower F-score value was achieved and, thus, this method was discarded.

As can be seen in Figure 2, the best anomaly detection performance was demonstrated using the static threshold. An F1-score of 0.7935 was achieved with window sizes of $w1 = 1$, $w2 = 15$, and a threshold value of 1.06 pF/cm. An F1-score of 0.8 is usually considered a good result (Fränti and Mariescu-Istodor, 2023) as the algorithm demonstrates good anomaly prediction performance. The static threshold method was also the least computationally expensive and easy to implement, in comparison to the dynamic threshold methods, thus promoting its use in real-time process implementation.

The dynamic threshold methods produced significantly lower F1-scores, all of which were below 0.5 and can be considered as not good enough. The MAD and IQR approaches produced similar results, as the methods are quite similar themselves. In both cases, the F1-score was significantly impacted by the detection of false positive and false negative anomalies. Regarding the 3-sigma threshold, the performance was similar to the static threshold in detecting true positive and false negative anomalies, however, a very high number of false positive anomalies were detected, impacting the overall F1-score. For extended results, refer to Supplementary Table S3.

With the static threshold approach, the F1-score criterion, referred to as 'precision', was 1.0 across all processes, indicating that every anomaly detected was in fact an anomaly. The other criterion, 'recall', demonstrating how many of all signal anomalies were correctly identified, varied from 0.47 to 0.91. On average,

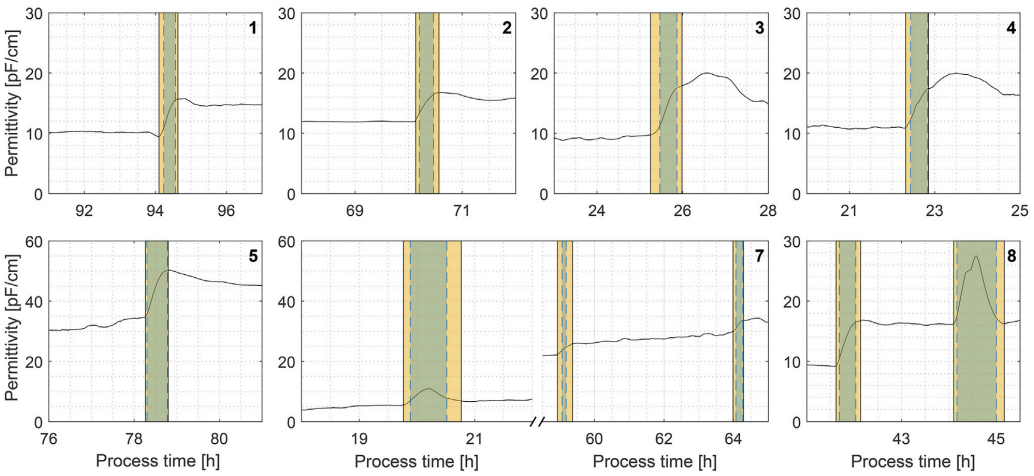


FIGURE 3
Comparison of manually annotated permittivity signal anomalies (orange shading) vs. successfully detected anomalies (blue shading) in fermentation process simulations. Exp. 6 was omitted because it contained no detected anomalies.

20 anomalous signal data points were not detected in each process (false negatives). Although that may seem significant initially, this count mainly arises from undetected anomalous signal values just prior and after detected signal anomalies (see Figure 3). The orange shading indicates manual anomaly annotations, and the blue shading shows the algorithm-detected anomalies. Overlap indicates good performance (e.g., Exp. 5). Most false negatives in the F-test result from slight delays in detection or early cutoffs as the signal flattens after an anomaly. It is in part caused by signal smoothing, as the signal change before and after anomalies is not so sudden and prominent anymore, hence the spike in the DRA transformed signal is also slightly delayed. In this case, it can be envisioned as a tradeoff between signal anomaly detection time and overall detection robustness.

Additionally, it can be noted that in most cases (excluding the 3-sigma threshold), *w1* size was quite low (1 or 2). This corresponds to the swiftness of anomaly detection, as with a smaller *w1* size, the anomalies are detected more quickly due to the mean of the window increasing more rapidly due to sudden signal jumps. With greater window sizes, the increase is slower, however, the detection can be seen as more robust.

In the case with all of the dynamic thresholds, the results were worse than expected. The dynamic threshold calculations are carried out, based on past signal values, hence, if the signal volatility suddenly increases, the dynamic threshold value increase is delayed by design. For example, if signal volatility has been low, the dynamic threshold is also low, but, if the volatility suddenly increases, the signal threshold is still low, thus, signal anomalies are detected. Assuredly, this may not be a problem when implementing such algorithms off-line (using the whole dataset), but in a real-time implementation this phenomenon could only be overcome by introducing some sort of signal volatility prediction parameter, which is beyond the scope of this article.

4.3 Anomaly validation and removal

In the final step, the detected permittivity signal anomalies are removed by introducing an alternative (corrected) permittivity sensor signal. When an anomaly is detected, the signal is corrected by replacing the current permittivity value with a mean of 15 past values prior to anomaly detection. Thus, the sudden nature of signal anomalies does not interfere, for example, with substrate feed rate calculations. On the other hand, the sudden increase in permittivity signal value would be estimated as a sudden increase in viable cell concentration by the substrate feeding algorithm and, thus, a drastic corrective action of the feed rate would follow. Such severe alterations to the substrate feeding profile would certainly lead to profound negative effects on process productivity and even result in batch discard.

This phase is initiated just after the detection of a signal anomaly and is characterized by an anomaly validation period of 15 min. During this transition period, the permittivity signal correction continues, estimating the corrected signal as a mean of past 15 values. In the case of a signal spike, anomaly detection is triggered by a sudden signal jump upwards. However, it is always followed by a sudden signal drop of similar magnitude. In such cases, the second anomaly often falls within the validation period. If so, then both anomalies are grouped into one and, due to the nature of these signal spikes, the corrective action is often minor as, after the anomaly, the signal returns to its previous level.

If a signal shift occurs, it is detected as a single signal anomaly. In this case, during the validation period, the new signal level is estimated. If additional anomalies are detected within the initial 15-point window, the window is dynamically extended to ensure that at least 15 min of valid data follow the last detected anomaly. A correction factor (Fc) is then introduced to compensate for the signal shift that has occurred. Fc is estimated as the difference

10.3389/fbioe.2025.1609369

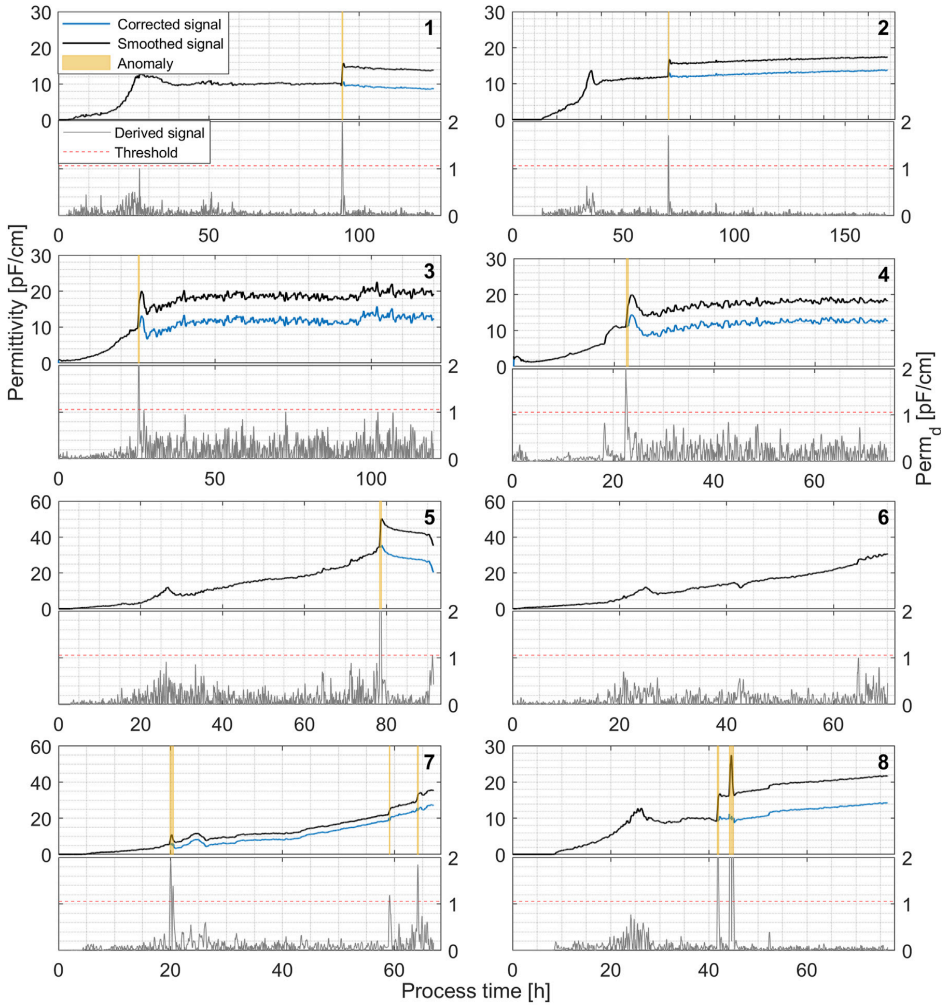


FIGURE 4
Permittivity signal anomaly detection and removal algorithm performance in real-time simulated recombinant *P. pastoris* fermentation processes. Raw signal is filtered in real-time using a Gaussian smoothing filter (w = 70) and a DRA-transformed (lower plot) Perm$_d$ signal ($w1$ = 1, $w2$ = 15) is used for anomaly detection with a static threshold of 1.06 pF/cm.

between 15 mean signal values before and after (validation period) anomaly detection. This value is then used to compensate for the previous signal values and introduce the corrected signal (see Figure 4).

4.4 Anomaly detection and removal algorithm

The combination of the aforementioned steps resulted in the creation of a novel permittivity signal anomaly removal algorithm. The algorithm is implemented in real-time simulations of recombinant *P. pastoris* fed-batch bioprocesses and effectively detects and removes in-line permittivity sensor signal anomalies. The schematic representation of the algorithm can be seen in Figure 5.

The bioreactor data processing program operates in a loop, loading data every minute (Step 1). The program then preprocesses the signal using a Gaussian smoothing filter and applies the signal correction factor Fc (initially set to zero) (Steps 2–3). A derived signal value (Perm$_d$) is then calculated using the DRA transformer ($w1$ = 1, $w2$ = 15) and compared to a threshold of 1.06 pF/cm (Step 5). If Perm$_d$ exceeds the threshold, the time point is logged as an anomaly (Step 6), and the signal value is replaced by the mean of the past 15 values (Step 7) and the program returns to Step 1.

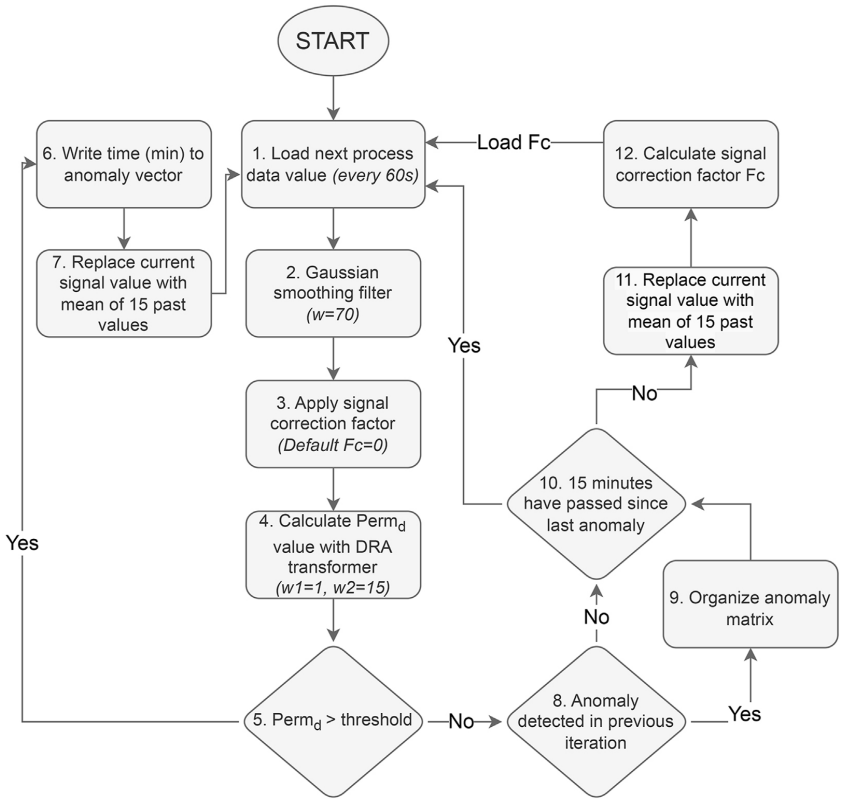


FIGURE 5
Schematic representation of the implemented permittivity signal anomaly detection and removal algorithm.

If the Perm$_d$ value does not exceed the threshold, then the algorithm evaluates whether an anomaly was previously detected in the previous iteration (Step 8). In all cases, the anomalies were registered as strings of consecutive time series, thus, if an anomaly was detected in the previous iteration and is not detected anymore in the next step, it indicates that the anomaly has passed. Furthermore, the anomaly can now be organized into the anomaly matrix, registering the anomaly start time in column 1 and end time in column 2 (Step 9).

Signal spikes are often detected as two separate anomalies, as both the initial signal jump and subsequent drop are detected by the DRA transformer. To avoid unnecessary signal overcorrection, a 15-min anomaly validation period was implemented (Step 10). Multiple anomalies within these 15 min are merged together as a single anomaly mainly to filter out signal spikes. When a signal spike occurs, the signal tends to return to the previous level after the spike has passed, thus, minimal or no corrective action is usually necessary.

During the validation period, permittivity signal values are replaced with means of 15 past values to compensate for the signal level after the shift, and a signal correction factor Fc is calculated from signal values before and after the anomaly (Steps 11–12). The anomaly validation period is also crucial for Fc calculation as the 15-min window provides a chance to estimate the extent of the permittivity signal shift. The correction factor Fc is calculated by subtracting the mean of 15 permittivity signal values prior to a detected anomaly from the mean of 15 values after an anomaly. It is then used in subsequent iterations to compensate for the permittivity signal shift that occurred during each detected signal anomaly.

The particular algorithm, when implemented in MATLAB, managed to successfully detect and remove permittivity signal anomalies for the selected dataset in real-time process simulations with an average computation time per iteration loop of 0.32 milliseconds, greatly improving overall permittivity signal quality. Thus, proving to be a rather straightforward and easy to implement tool for real-time permittivity signal anomaly removal, promoting the use of viable cell concentration measurement for substrate feed rate calculation in fed-batch bioprocesses.

The exact cause of these permittivity signal anomalies remains unclear, however, a significant correlation can be established with antifoam solution addition and changes in agitation, which often precede said anomalies. Studies have demonstrated that introducing small quantities of antifoam leads to a reduction in gas hold-up and an increase in average bubble diameter. This enlargement of bubble

size results in a decreased specific surface area and medium surface tension (Al-Masry et al., 2006; Routledge, 2012). We presume that antifoam addition increases culture medium density primarily by reducing entrapped air bubbles, facilitating the formation of larger bubbles that rise and escape more easily, thereby decreasing gas hold-up. This reduction in gas volume results in a denser liquid phase, which accounts for the consistent upward shifts in permittivity signals. Previous studies have reported that antifoam addition correlates with increased culture density (Routledge et al., 2011). This theory is also supported by visual assessment of *P. pastoris* cultivation media volume prior and after antifoam addition. This suggests that incorporating small amounts of antifoam at the start of fermentation could be beneficial, provided it is compatible with the selected microorganism and bioprocess.

5 Conclusion

This study tackles a key challenge in *in-situ* measurement related to viable biomass concentration: the development of a robust and easily implementable algorithm for real-time anomaly detection and removal in permittivity sensor data. Unlike simplistic methods, which fail to capture the complexity of the issue, our approach offers a structured three-step solution: (1) Signal preprocessing to minimize noise and remove context dependency; (2) Anomaly detection through threshold-based identification; and (3) Validation and removal of detected anomalies.

As a result, we present a general workflow with defined steps for in-line permittivity sensor signal anomaly detection and removal. This approach enabled reliable real-time anomaly detection and removal in permittivity sensor data from recombinant *P. pastoris* fermentations while maintaining computational efficiency, making it practical for real-time applications. With a static threshold of 1.06 and a double rolling aggregate transformer using window sizes $w1 = 1$ and $w2 = 15$, it achieves an F1-score of 0.79. This flexible algorithm has the potential to bridge a critical gap in process analytics and control for real-time bioprocess monitoring, while its ease of implementation promotes the use of in-line permittivity measurements in monitoring and control applications in other cultivations.

Data availability statement

Online records of bioreactor parameters and raw in-line permittivity sensor data analyzed for this study can be found under https://dx.doi.org/10.5281/zenodo.14264619.

Author contributions

EB: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing. SU: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Validation, Writing – original draft, Writing – review and editing. MP-H: Conceptualization, Formal Analysis, Project administration, Resources, Supervision, Writing – review and editing. VG: Conceptualization, Formal Analysis, Methodology, Validation, Writing – review and editing. OG: Conceptualization, Data curation, Formal Analysis, Project administration, Resources, Software, Supervision, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The current research was supported by LSIWC grant No. 07-24. Furthermore, we acknowledge the support of the German Federal Ministry of Research, Technology and Space through the funding program Forschung an Fachhochschulen under contract number 13FH045KX2. The work was also partially supported by Riga Technical University through doctoral grant No. 1094, as part of the project funded by the European Union Recovery and Resilience Facility (Project No. 5.2.1.1.i.0/2/24/I/CFLA/003), titled 'Implementation of Consolidation and Management Changes at Riga Technical University, Liepaja University, Rezekne Academy of Technology, Latvian Maritime Academy, and Liepaja Maritime College for Advancing Excellence in Higher Education, Science, and Innovation'.

Acknowledgments

The study utilized bioprocess data obtained within the framework of a postdoctoral project (ERDF and Latvian state-funded research application grant No. 1.1.1.2/VIAA/1/16/186).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2025.1609369/full#supplementary-material

References

Ahmad, S., Lavin, A., Purdy, S., and Agha, Z. (2017). Unsupervised real-time anomaly detection for streaming data. *Neurocomputing* 262, 134–147. doi:10.1016/j.neucom.2017.04.070

Al-Masry, W. A., Ali, E. M., and Aqeel, Y. M. (2006). Effect of antifoam agents on bubble characteristics in bubble columns based on acoustic sound measurements. *Chem. Eng. Sci.* 61, 3610–3622. doi:10.1016/j.ces.2006.01.009

Berger, A., and Kiefer, M. (2021). Comparison of different response time outlier exclusion methods: a simulation study. *Front. Psychol.* 12, 675558. doi:10.3389/fpsyg.2021.675558

Blázquez-García, A., Conde, A., Mori, U., and Lozano, J. A. (2022). A review on outlier/anomaly detection in time series data. *ACM Comput. Surv.* 54, 1–33. doi:10.1145/3444690

Bolmanis, E., Grigs, O., Kazaks, A., and Galvanauskas, V. (2022). High-level production of recombinant HBcAg virus-like particles in a mathematically modelled P. pastoris GS115 Mut+ bioreactor process under controlled residual methanol concentration. *Bioprocess Biosyst. Eng.* 45, 1447–1463. doi:10.1007/s00449-022-02754-4

Brignoli, Y., Freeland, B., Cunningham, D., and Dabros, M. (2020). Control of specific growth rate in fed-batch bioprocesses: novel controller design for improved noise management. *Processes* 8, 679. doi:10.3390/pr8060679

Brunner, V., Siegl, M., Geier, D., and Becker, T. (2021). Challenges in the development of soft sensors for bioprocesses: a critical review. *Front. Bioeng. Biotechnol.* 9, 722202. doi:10.3389/fbioe.2021.722202

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: a survey. *ACM Comput. Surv.* 41, 1–58. doi:10.1145/1541880.1541882

Chiang, L. H., Pell, R. J., and Seasholtz, M. B. (2003). Exploring process data with the use of robust outlier detection algorithms. *J. Process Control* 13, 437–449. doi:10.1016/S0959-1524(02)00068-9

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* 74, 829–836. doi:10.2307/2286407

Darban, Z. Z., Webb, G. I., Pan, S., Aggarwal, C. C., and Salehi, M. (2024). Deep learning for time series anomaly detection. *A Surv.* doi:10.48550/arXiv.2211.05244

Da Silva, A. J., Horta, A. C. L., Velez, A. M., Iemma, M. R. C., Sargo, C. R., Giordano, R. L., et al. (2013). Non-conventional induction strategies for production of subunit swine erysipelas vaccine antigen in rE. coli fed-batch cultures. *Springerplus* 2, 322. doi:10.1186/2193-1801-2-322

Davies, L., and Gather, U. (1993). The identification of multiple outliers. *J. Am. Stat. Assoc.* 88, 782–792. doi:10.1080/01621459.1993.10476339

Downey, B. J., Graham, L. J., Breit, J. F., and 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. doi:10.1002/btpr.1845

Fehrenbach, R., Comberbach, M., and Pêtre, J. O. (1992). On-line biomass monitoring by capacitance measurement. *J. Biotechnol.* 23, 303–314. doi:10.1016/0168-1656(92)90077-m

Ferreira, A. R., Dias, J. M. L., Stosch, M. von, Clemente, J., Cunha, A. E., and Oliveira, R. (2014). Fast development of Pichia pastoris GS115 Mut(+) cultures employing batch-to-batch control and hybrid semi-parametric modeling. *Bioprocess Eng.* 37, 629–639. doi:10.1007/s00449-013-1029-9

Fränti, P., and Mariescu-Istodor, R. (2023). Soft precision and recall. *Pattern Recognit. Lett.* 167, 115–121. doi:10.1016/j.patrec.2023.02.005

Grigs, O., Bolmanis, E., and Galvanauskas, V. (2021a). Application of *in-situ* and soft-sensors for estimation of recombinant P. pastoris GS115 biomass concentration: a case analysis of HBcAg (Mut+) and HBsAg (MutS) production processes under varying conditions. *Sensors (Basel)* 21, 1268. doi:10.3390/s21041268

Grigs, O., Bolmanis, E., and Kazaks, A. (2021b). HBsAg production in methanol controlled P. pastoris GS115 Mut^S bioreactor process. *KEM* 903, 40–45. doi:10.4028/www.scientific.net/KEM.903.40

Harju, P. T., Ovaska, S. J., and Valimaki, V. (1996). "Delayless signal smoothing using a median and predictive filter hybrid," in *Proceedings of third international conference on signal processing (ICSP'96)* (IEEE), 87–90.

Higgins, J., and Green, S. (2008). *Cochrane handbook for systematic reviews of interventions* (Chichester: Wiley).

Hill, D. J., and Minsker, B. S. (2010). Anomaly detection in streaming environmental sensor data: a data-driven modeling approach. *Environ. Model. & Softw.* 25, 1014–1022. doi:10.1016/j.envsoft.2009.08.010

Horta, A. C. L., Da Silva, A. J., Sargo, C. R., Cavalcanti-Montaño, I. D., Galeano-Suarez, I. D., Velez, A. M., et al. (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. *Braz. J. Chem. Eng.* 32, 821–829. doi:10.1590/0104-6632.20150324s00003534

Horta, A. C. L., Sargo, C. R., Da Silva, A. J., Carvalho Gonzaga, M. de, Dos Santos, M. P., Gonçalves, V. M., et al. (2012). Intensification of high cell-density cultivations of rE. coli for production of S. pneumoniae antigenic surface protein, PspA3, using model-based adaptive control. *Bioprocess Eng.* 35, 1269–1280. doi:10.1007/s00449-012-0714-4

Huber, P. J. (2011). "Robust statistics," in *International encyclopedia of statistical science*. Editor M. Lovric (Berlin, Heidelberg: Springer Berlin Heidelberg), 1248–1251.

International Organization for Standardization (2008). *Software engineering — software product quality requirements and evaluation (SQuaRE) — data quality model*, 2008.*ISO/IEC 25012:2008*

Iqbal, A., and Amin, R. (2024). Time series forecasting and anomaly detection using deep learning. *Comput. & Chem. Eng.* 182, 108560. doi:10.1016/j.compchemeng.2023.108560

Jiang, Y., Yin, S., Dong, J., and Kaynak, O. (2021). A review on soft sensors for monitoring, control, and optimization of industrial processes. *IEEE Sensors J.* 21, 12868–12881. doi:10.1109/JSEN.2020.3033153

Jones, P. R. (2019). A note on detecting statistical outliers in psychophysical data. *Atten. Percept. Psychophys.* 81, 1189–1196. doi:10.3758/s13414-019-01726-3

Kadlec, P., Gabrys, B., and Strandt, S. (2009). Data-driven soft sensors in the process industry. *Comput. & Chem. Eng.* 33, 795–814. doi:10.1016/j.compchemeng.2008.12.012

Katla, S., Mohan, N., Pavan, S. S., Pal, U., and 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. doi:10.1002/jctb.6118

Konstantinov, K. B., Pambayun, R., Matanguihan, R., Yoshida, T., Perusicn, C. M., and Hu, W. S. (1992). On-line monitoring of hybridoma cell growth using a laser turbidity sensor. *Biotechnol. Bioeng.* 40, 1337–1342. doi:10.1002/bit.260401107

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766. doi:10.1016/j.jesp.2013.03.013

Lin, B., Recke, B., Knudsen, J. K., and Jørgensen, S. B. (2007). A systematic approach for soft sensor development. *Comput. & Chem. Eng.* 31, 419–425. doi:10.1016/j.compchemeng.2006.05.030

Looser, V., Bruhlmann, B., Bumbak, F., Stenger, C., Costa, M., Camattari, A., et al. (2015). Cultivation strategies to enhance productivity of pichia pastoris: a review. *Biotechnol. Adv.* 33, 1177–1193. doi:10.1016/j.biotechadv.2015.05.008

Mandenius, C.-F., and Gustavsson, R. (2015). Mini-review: soft sensors as means for PAT in the manufacture of bio-therapeutics. *J. Chem. Technol. & Biotechnol.* 90, 215–227. doi:10.1002/jctb.4477

Meitz, A., Sagmeister, P., Lubitz, W., Herwig, C., and Langemann, T. (2016). Fed-batch production of bacterial ghosts using dielectric spectroscopy for dynamic process control. *Microorganisms* 4, 18. doi:10.3390/microorganisms4020018

Metze, S., Ruhl, S., Greller, G., Grimm, C., and Scholz, J. (2020). 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. doi:10.1007/s00449-019-02216-4

Miller, J. (1991). Short report: reaction time analysis with outlier exclusion: Bias varies with sample size. *Q. J. Exp. Psychol. A* 43, 907–912. doi:10.1080/14640749108400962

Müller, T., Lauk, M., Reinhard, M., Hetzel, A., Lücking, C. H., and Timmer, J. (2003). Estimation of delay times in biological systems. *Ann. Biomed. Eng.* 31, 1423–1439. doi:10.1114/1.1617984

Münzberg, M., Hass, R., Dinh Duc Khanh, N., and Reich, O. (2017). Limitations of turbidity process probes and formazine as their calibration standard. *Anal. Bioanal. Chem.* 409, 719–728. doi:10.1007/s00216-016-9893-1

Pearson, R. K. (2001). Exploring process data. *J. Process Control* 11, 179–194. doi:10.1016/S0959-1524(00)00046-9

Pearson, R. K. (2002). Outliers in process modeling and identification. *IEEE Trans. Contr. Syst. Technol.* 10, 55–63. doi:10.1109/87.974338

Pearson, R. K. (2005). *Mining imperfect data: dealing with contamination and incomplete records*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Pentjuss, A., Bolmanis, E., Suleiko, A., Didrihsone, E., Suleiko, A., Dubencovs, K., et al. (2023). Pichia pastoris growth-coupled heme biosynthesis analysis using metabolic modelling. *Sci. Rep.* 13, 15816. doi:10.1038/s41598-023-42865-w

Ramm, S., Rodríguez, T. H., Frahm, B., and Pein-Hackelbusch, M. (2023). "Systematic preprocessing of dielectric spectroscopy data and estimating viable cell densities," in *2023 IEEE 21st international conference on industrial informatics (INDIN)* (IEEE), 1–6.

Rousseeuw, P. J., and Croux, C. (1993). Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* 88, 1273–1283. doi:10.2307/2291267

Routledge, S. J. (2012). Beyond de-foaming: the effects of antifoams on bioprocess productivity. *Comput. Struct. Biotechnol. J.* 3, e2012100001. doi:10.5936/csbj.201210014

Routledge, S. J., Hewitt, C. J., Bora, N., and Bill, R. M. (2011). Antifoam addition to shake flask cultures of recombinant Pichia pastoris increases yield. *Microb. Cell Fact.* 10, 17. doi:10.1186/1475-2859-10-17

Sakiyo, J. J., and Németh, Á. (2023). The potential of bacilli-derived biosurfactants as an additive for biocontrol against Alternaria alternata plant pathogenic fungi. *Microorganisms* 11, 707. doi:10.3390/microorganisms11030707

Sarrafzadeh, M. H., Belloy, L., Esteban, G., Navarro, J. M., and Ghommidh, C. (2005). Dielectric monitoring of growth and sporulation of Bacillus thuringiensis. *Biotechnol. Lett.* 27, 511–517. doi:10.1007/s10529-005-2543-x

Savitzky, A., and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36, 1627–1639. doi:10.1021/ac60214a047

Schmidl, S., Wenig, P., and Papenbrock, T. (2022). Anomaly detection in time series: a comprehensive evaluation. *Proc. VLDB Endow.* 15, 1779–1797. doi:10.14778/3538598.3538602

Shiffler, R. E. (1988). Maximum Z scores and outliers. *Am. Statistician* 42, 79–80. doi:10.1080/00031305.1988.10475530

Venables, W. N., and Ripley, B. D. (2002). *Modern applied statistics with S*. New York, NY: Springer.

Warne, K., Prasad, G., Rezvani, S., and Maguire, L. (2004). Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion. *Eng. Appl. Artif. Intell.* 17, 871–885. doi:10.1016/j.engappai.2004.08.020

Zhao, Y., Lehman, B., Ball, R., Mosesian, J., and Palma, J.-F. de (2013). "Outlier detection rules for fault detection in solar photovoltaic arrays," in 2013 twenty-eighth annual IEEE applied power electronics conference and exposition (APEC) *(IEEE)*, 2913–2920.

Zhu, J., Ge, Z., Song, Z., and Gao, F. (2018). Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data. *Annu. Rev. Control* 46, 107–133. doi:10.1016/j.arcontrol.2018.09.003

Appendix III
3. Pielikums

High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. pastoris* GS115 Mut$^+$ Bioreactor Process under Controlled Residual Methanol Concentration

Emils Bolmanis, Oskars Grigs, Andris Kazaks, Vytautas Galvanauskas
Bioprocess and Biosystems Engineering **2022**, *45*, DOI: 10.1007/s00449-022-02754-4

E.B. input: Conceptualization, formal analysis, investigation, methodology, software, writing – original draft preparation, writing – review and editing, visualization.

Bioprocess and Biosystems Engineering
https://doi.org/10.1007/s00449-022-02754-4

RESEARCH PAPER

High-level production of recombinant HBcAg virus-like particles in a mathematically modelled *P. pastoris* GS115 Mut$^+$ bioreactor process under controlled residual methanol concentration

Emils Bolmanis[1,2] · Oskars Grigs[1] · Andris Kazaks[2] · Vytautas Galvanauskas[3]

Received: 3 May 2022 / Accepted: 4 July 2022

Abstract
Recombinant hepatitis B core antigen (HBcAg) molecules, produced in heterologous expression systems, self-assemble into highly homogenous and non-infectious virus-like particles (VLPs) that are under extensive research for biomedical applications. HBcAg production in the methylotrophic yeast *P. pastoris* has been well documented; however, productivity screening under various residual methanol levels has not been reported for bioreactor processes. HBcAg production under various excess methanol levels of 0.1, 1.0 and 2.0 g L^{-1} was investigated in this research. Results indicate that, under these particular conditions, the total process and specific protein yields of 876–1308 mg L^{-1} and 7.9–11.2 mg g_{DCW}^{-1}, respectively, were achieved after 67–75 h of cultivation. Produced HBcAg molecules were efficiently purified and the presence of highly immunogenic, correctly formed and homogenous HBcAg-VLPs with an estimated purity of 90% was confirmed by electron microscopy. The highest reported HBcAg yield of 1308 mg L^{-1} and 11.2 mg g_{DCW}^{-1} was achieved under limiting residual methanol concentration, which is about 2.5 times higher than the next highest reported result. A PI-algorithm-based residual methanol concentration feed rate controller was employed to maintain a set residual methanol concentration. Finally, mathematical process models to characterise the vegetative, dead and total cell biomass (X_v, X_d and X), substrate (Glycerol and Methanol) concentration, reactor volume (V), and product (HBcAg) dynamics during cultivation, were identified. A rare attempt to model the residual methanol concentration during induction is also presented.

Keywords Hepatitis B core antigen (HBcAg) · *Pichia pastoris* · *Fed-batch* bioreactor · Cultivation process modelling · Residual methanol PI control

Abbreviations

HBcAg	Hepatitis B core Antigen
DO	Dissolved Oxygen
OD	Optical Density at 590 nm
DCW	Dry Cell Weight
WCW	Wet Cell Weight
GMP	Good Manufacturing Practice
PID	Proportional, Integral and Derivative control parameters
PLC	Programmable Logic Controller
HPLC	High-Performance Liquid Chromatography
GC	Gas Chromatography
SCADA	Supervisory Control and Data Acquisition interface
Gly	Glycerol
MeOH	Methanol

List of symbols

X_v	Vegetative cell biomass concentration [g L^{-1}]
X_d	Dead (non-vegetative) cell biomass concentration [g L^{-1}]
X	Total cell biomass concentration [g L^{-1}]
S	Substrate concentration [g L^{-1}]
P	Product concentration [g L^{-1}]
V	Cultivation medium volume [L]
τ_I	Integral time constant [min.]
K_p	Proportional gain parameter [$L^2\ g^{-1}\ h^{-1}$]

✉ Emils Bolmanis
e.bolmanis@hotmail.com

✉ Andris Kazaks
andris@biomed.lu.lv

[1] Latvian State Institute of Wood Chemistry, Riga, Latvia

[2] Latvian Biomedical Research and Study Centre, Riga, Latvia

[3] Department of Automation, Kaunas University of Technology, Kaunas, Lithuania

Published online: 08 August 2022

$K1, K2, k_1, k_2$	Model parameters
Y_{ATP}	ATP yield coefficient [g mol^{-1}]
Y_{PX}	Specific product yield coefficient [g g^{-1}]
P/O	Oxidative phosphorylation effectiveness coefficient
φ	Fraction of formaldehyde oxidised to formate
mATP	ATP maintenance coefficient [mol g^{-1} h^{-1}]
q_G	Specific glycolysis rate [mol g^{-1} h^{-1}]
q_{Ac}	Specific acetyl-CoA production rate [mol g^{-1} h^{-1}]
q_{O2}	Specific oxygen uptake rate [mol g^{-1} h^{-1}]
q_{Gly}	Specific glycerol uptake rate [g g^{-1} h^{-1}]
$q_{Gly,R}$	Specific glycerol uptake rate according to the Regulator model [g g^{-1} h^{-1}]
$q_{Gly,M}$	Specific glycerol uptake rate according to the Monod model [g g^{-1} h^{-1}]
$q_{min,Gly}$	Minimum specific glycerol uptake rate [g g^{-1} h^{-1}]
$q_{max,Gly}$	Maximum specific glycerol uptake rate [g g^{-1} h^{-1}]
q_{MeOH}	Specific methanol uptake rate [g g^{-1} h^{-1}]
$q_{MeOH,max}$	Maximum specific methanol uptake rate [g g^{-1} h^{-1}]
q_P	Specific product formation rate [g g^{-1} h^{-1}]
S_{Gly}	Glycerol concentration in cultivation medium [g L^{-1}]
S_{MeOH}	Methanol concentration in cultivation medium [g L^{-1}]
S_F	Substrate concentration in feed [g L^{-1}]
M	Substrate molar mass [g mol^{-1}]
K_{Gly}	Glycerol saturation constant [g L^{-1}]
K_{MeOH}	Methanol saturation constant [g L^{-1}]
K_i	Methanol inhibition constant [g L^{-1}]
K_d	Cell-specific death rate [h^{-1}]
μ	Specific biomass growth rate [h^{-1}]
F_{evp}	Evaporation rate [L h^{-1}]
F_s	Substrate feed rate [L h^{-1}]
F_{NH3}	Base addition rate [L h^{-1}]
F_{AF}	Antifoam addition rate [L h^{-1}]
F_{smpl}	Sampling rate [L h^{-1}]

Introduction

Hepatitis B virus core antigen (HBcAg) is a protein subunit of the hepatitis B virus icosahedral core shell. Intracellularly expressed in heterologous production systems such as bacteria, yeasts, plants or mammalian cell cultures, and the recombinant HBcAg molecules self-assemble in to non-infectious virus-like particles (VLPs) [1]. These particles, being highly immunogenic and known to induce significant T-cell responses and high antibody titres in the exposed host organism [2], have been extensively studied and used in vaccine development [3]. Consequently, the efficient production of purified HBcAg-VLPs is a key aspect for both research and commercial applications.

HBcAg-VLPs have been recognised as a promising VLP carrier platform since the 1980s; naturally, they have also been extensively studied [4, 5]. Recent reviews illustrate that HBcAg-VLPs have successfully been employed as epitope carriers in different vaccine candidates against a myriad of diseases, contrary to its fraternal HBV surface antigen (HBsAg), still lack successful commercial examples [3, 6]. A promising HBcAg-VLP-based commercial vaccine candidate against malaria (MalariVax) was in development; however, it was discontinued after failing the sporozoite challenge [7]. On the contrary, a malaria vaccine candidate (Mosquirix™) based on HBsAg-VLPs was recently approved by the WHO for general use in Africa [8]. Several HBcAg-VLP-based vaccine patent applications can also be found against various diseases: melioidosis [9], hepatocarcinoma, and hepatitis B [10, 11], and HPV [12]. Despite the lack of successful commercial HBcAg-VLP-based vaccines, this platform still stands out as one of the frontrunners for future vaccine design.

A wide range of expression systems are available for recombinant protein production and each of them has their own advantages. The choice of a production platform is in large part dictated by the properties of the target protein—its structure and biological activity [13]. For the past 20–30 years, the *P. pastoris* expression system has been successfully employed for the production of various recombinant proteins for both research and industrial applications. Some of the main characteristics that make this methylotrophic yeast so suited for foreign protein expression include easy genetic manipulation, high-frequency DNA transformation, cloning by functional complementation, high levels of intra- and extracellularly expressed protein, and the ability to perform higher eukaryotic protein modifications (glycosylation, disulphide bond formation, and proteolytic processing) [14]. Also, relatively low levels of native secreted proteins allow for the use of simple purification strategies of the secreted recombinant proteins. When the economic aspects (high cell growth in minimal media and high product stability in prolonged processes) are taken into consideration, together with the powerful genetic techniques available, *P. pastoris* is clearly the system of choice for heterologous protein expression [15].

The two-stage cultivation on glycerol and methanol associated with Invitrogen's "*Pichia* Fermentation Process Guidelines" is well documented and present in most AOX1-promoter-based *P. pastoris* cultivation strategies. However, recent trends advocate for a move away from standard

protocols towards a more conceptual approach, which allows for the development of process-specific strategies tailored both to the specific combination of product/genetic construct and the characteristics of the bioreactor equipment [16]. The standardized approach employs a constant substrate feed rate during methanol induction. Since cell biomass concentration keeps increasing, but the substrate feed rate remains constant, the amount of substrate per biomass (g DCW^{-1} h^{-1}) gradually decreases, resulting in a decreasing specific growth rate and, consequently, productivity [16].

A myriad of different methanol feeding strategies have been reported with excellent recombinant product yields [16–18]. Two distinct groups can be identified—co-feeding and -*stat* strategies. Co-feeding strategies aim to supplement the methanol feed with additional substrates (glycerol, sorbitol, mannitol, etc.), whereas -*stat* strategies attempt to control the methanol feed in a way that a cultivation parameter (μ, DO, methanol concentration) remains at a pre-set value. One of the more popular approaches is the *methanol-stat* strategy, in which the residual methanol concentration is controlled at a set-point value during the methanol induction phase [17]. There are reports of this strategy leading to increased recombinant protein yields for several products; however, no reports can be found on the effect this strategy has on HBcAg production in *P. pastoris*.

The residual methanol concentration in AOX1-promoter-regulated *P. pastoris* systems is a critical factor to control, as it directly influences the production and proteolytic degradation of recombinant proteins, cell growth, lysis, and oxygen transfer; thus, an optimal methanol feeding strategy is crucial to maximize product yields [19]. Residual methanol concentrations between 2.0 and 3.5 g L^{-1} are considered optimal for protein production with *P. pastoris* [19]; however, some authors report success with concentrations in 4.0–6.0 g L^{-1} range [20, 21]. Monitoring and controlling the residual methanol concentration in the cultivation broth are crucial; therefore, a suitable methanol sensor and a methanol feed rate control algorithm should be used to ensure adequate methanol concentration control at a selected set-point.

Two types of control systems for residual methanol concentration control can be distinguished—open loop and closed loop. The open-loop systems do not require any on-line measurements (feedback); however, closed loop systems do. In this particular case, the methanol sensor on-line signal can be used as feedback for methanol feed rate control. Additionally, a control algorithm is necessary to successfully control the residual methanol concentration around set-point value with minimal deviation. For certain residual methanol level control, PI- [22] or PID- [23] control algorithms have been successfully employed.

Stringent requirements for producing biopharmaceutical products and their marketing in the corresponding regions, in accordance with good manufacturing practice (GMP) guidelines, are imposed by the EU [24], USA [25], and other developed countries. Since biotechnological microorganism cultivation processes are highly complex and their productivity is sensitive to optimal process parameter control, the above-mentioned directives require that, to ensure a high level of product quality and process repeatability (safety), advanced microorganism culture process control techniques must be employed. It is also mentioned that model-based methods facilitate an in-depth understanding of the process and expedite the performance of the procedures of cultivation process validation and revalidation [26]. Mathematical modelling of recombinant *P. pastoris* cultivation processes is well described; however, there are few examples modelling residual methanol concentrations during induction [27–31].

In the present work, we establish a *P. pastoris* cultivation procedure for high-level expression of HBcAg in a defined medium under various residual methanol levels controlled by the PI-feed rate control algorithm. A previously reported purification method [32] was employed with minor modifications, leading to superior results and producing HBcAg-VLPs with an estimated purity of 90%. Mathematical modelling was used to tune the K_p parameter for residual methanol PI-control algorithm. Suitable mathematical models and model parameters to model biomass growth, substrate (glycerol and methanol) consumption, and HBcAg formation were successfully identified.

Materials and methods

Construction of an expression vector and selection of clones

The construction of the expression vector pPIC-HBc and the selection of an HBc producer clone was described in [32].

Experimental conditions

Cultivation processes were carried out using a recombinant *Pichia pastoris* GS115 Mut^+ strain. Batch cultivation and feeding media solutions used in this study were prepared according to the "*Pichia fermentation process guidelines*" by Invitrogen Corporation [33]: 1.9 l of Basal Salts Medium [26.7 ml L^{-1} H_3PO_4 85%, 0.93 g L^{-1} $CaSO_4$, 18.2 g L^{-1} K_2SO_4, 14.9 g L^{-1} $MgSO_4{\cdot}7H_2O$, 4.13 g L^{-1} KOH, 40.0 g L^{-1} glycerol and 4.35 ml L^{-1} PTM_1 trace-element solution (0.02 g L^{-1} H_3BO_4, 5 ml L^{-1} H_2SO_4 98%, 6.0 g L^{-1} $CuSO_4{\cdot}5H_2O$, 0.08 g L^{-1} NaI, 3.0 g L^{-1} $MnSO_4{\cdot}H_2O$, 0.2 g L^{-1} $Na_2MoO_4{\cdot}2H_2O$, 0.5 g L^{-1} $Ca_2SO_4{\cdot}2H_2O$, 20.0 g L^{-1} $ZnCl_2$, 65.0 g L^{-1} $FeSO_4{\cdot}7H_2O$, 0.2 g L^{-1} biotin)], was inoculated with 100 mL of inoculum grown in BMGY medium (10.0 g L^{-1} yeast extract, 20.0 g L^{-1} peptone, 100 mM

potassium phosphate buffer, pH 6.0, 13.4 g L^{-1} yeast nitrogen base, 10.0 g L^{-1} glycerol, and 0.0004 g L^{-1} biotin) at 30 °C for 18–22 h in a shake flask at 250 RPM. Two feeding solutions were used—glycerol fed-batch solution (50% glycerol, 12 mL L^{-1} PTM_1) and methanol fed-batch solution (100% methanol, 12 mL L^{-1} PTM_1).

The bioreactor vessel was filled with distilled water and sterilized at 121 °C for 30 min. BSM, BMGY, and glycerol fed-batch solutions were autoclaved separately at 121 °C for 30 min. PTM_1 trace element and methanol fed-batch solutions were sterilized by filtration through a 0.2 µm filter.

Fermentations were carried out in a 5 l (2–4 l working volume) fully automated bench-top fermenter (*Biotehniskais centrs*, EDF-5.4/BIO-4, Latvia) schematically illustrated in Fig. 1. A calibrated pH sensor probe (*Hamilton*, EasyFerm Bio, Switzerland) was used to measure the medium pH during cultivation. Before the start of the cultivation process, the fermentation medium pH was adjusted at 5.0 ± 0.1 using a 28% NH_4OH solution, which was also used to maintain the set pH value during fermentation. The temperature was controlled at 30.0 ± 0.1 °C. The dissolved oxygen level was measured with a calibrated DO probe (*Hamilton*, OxyFerm, Switzerland) and kept above 30 ± 5% during fermentation by increasing the stirrer speed (200–1000 RPM) or inlet air enrichment with pure O_2. Constant air or air/oxygen mixture at a flow rate of 3.0 slpm was maintained during all processes. A condenser was used to condense the moisture from outlet gasses and excessive foam formation was controlled by adding antifoam 204 (*Sigma*) when necessary. Substrate feed solutions were pumped using a pre-calibrated high-accuracy peristaltic pump (*Longer-Pump,* BT100–2 J, China).

The cultivations were initiated with a glycerol batch phase. After 18–24 h, all batch glycerol is consumed and glycerol fed-batch solution is fed into the reactor at a rate of 0.61 mL min^{-1} for 4 h and until an optical density of

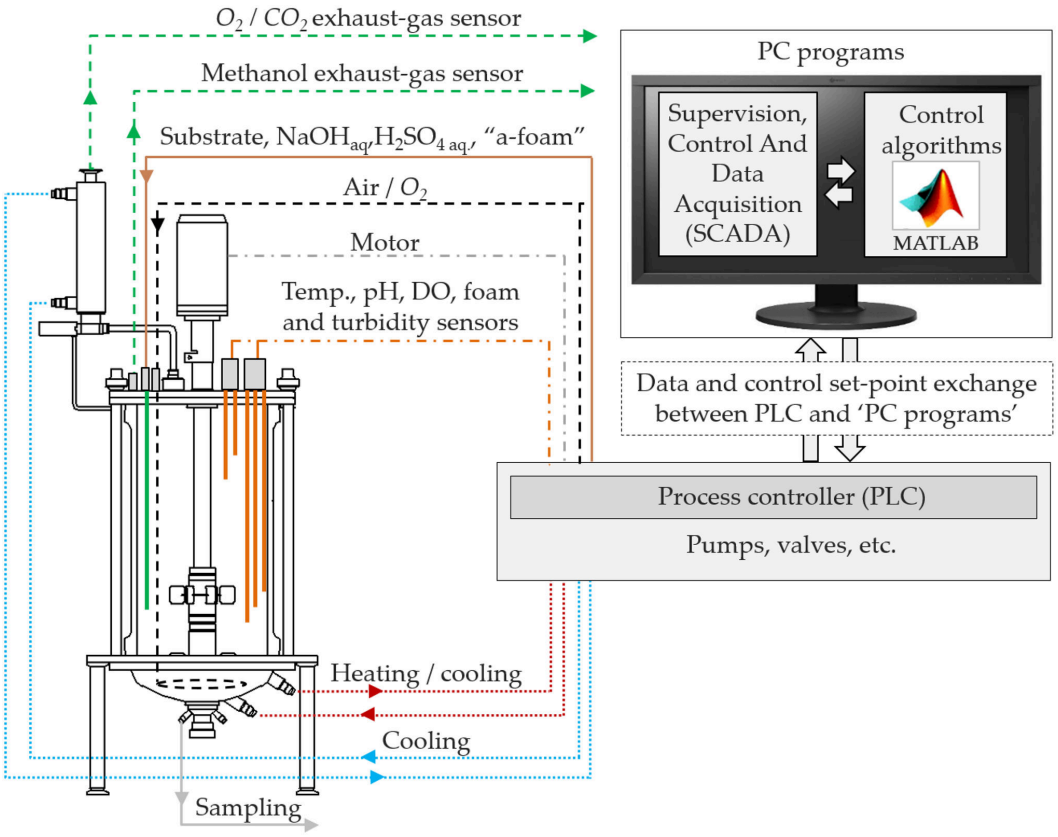


Fig. 1 Schematic diagram of the bioreactor and control architecture

100–120 is reached. Then, the feed substrate is switched to methanol and the solution is fed into the reactor at a rate of 0.12 mL min^{-1} for 5 h, then at a rate of 0.24 mL min^{-1} for 2 h, and finally at a rate of 0.36 mL min^{-1} up until 40 h of cultivation. At 40 h, a direct residual methanol concentration control using a PI-algorithm-based feed rate control is initiated at a selected set-point and carried out until the end of the process.

Analytical measurements

Cell growth was observed by off-line measurements of the culture optical density (OD) at a wavelength of 590 nm (GRANAT, KFK-2, St. Petersburg, Russia). Wet cell weight (WCW) and dry cell weight (DCW) measurements were determined gravimetrically. Biomass samples were placed in pre-weighted Eppendorf® tubes and centrifuged at 15,500*g* for 5 min. Afterwards, the supernatant was discarded, and the cells resuspended in distilled water and centrifuged once more. The liquid phase was discarded and the remaining wet cell biomass weighted. Afterwards, samples were dried at 105 °C until constant weight was reached and the dry cell biomass weighted. The DCW measurement was selected as the most accurate cell biomass assessment with an average error of ± 1.0 g L^{-1} [34].

Glycerol and methanol concentrations were measured off-line using, respectively, HPLC (Agilent, 1100 HPLC) and GC (Agilent, 6890 N GC).

Protein samples taken during cultivation were analysed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), with a 5% stacking and 15% separating polyacrylamide gel (PAAG), according to standard protocols. To visualize the separated protein bands, the gels were stained with 0.4% Coomassie Brilliant Blue (CBB) G-250 dye.

The level of HBcAg production was also estimated by disrupting 20 optical units of yeast cells by glass beads in 200 µL of 20 mM Tris–HCl, pH 8.0, eight times for 0.5 min. Debris was separated by low-speed centrifugation, and the supernatant was serially diluted for an immunodiffusion assay [35] using a polyclonal rabbit anti-HBcAg antibody.

For electron microscopy, the protein samples were adsorbed on carbon–formvar-coated copper grids and negatively stained with 1% uranyl acetate aqueous solution. The grids were examined with a JEM-1230 electron microscope (JEOL Ltd., Tokyo, Japan) at 100 kV.

Methanol concentration control

Residual methanol concentration was monitored in the exhaust gas and correlated to the concentration in the fermentation medium (*Bluesens*, BCP-EtOH, Germany). A moving average algorithm (step = 10) was used to decrease signal noise and increase overall quality.

For residual methanol concentration control, a direct substrate feed rate PI-control approach proposed by Cos et al. was implemented [22]. Methanol concentration mass balance equation

$$F \bullet M_0 - r_M \bullet V = V \bullet \frac{\mathrm{d}M}{\mathrm{d}t} + M\frac{\mathrm{d}V}{\mathrm{d}t}. \tag{1}$$

Methanol concentration change dynamics

$$\frac{\mathrm{d}M}{\mathrm{d}t} = -r_M + \frac{F\left(M_0 - M\right)}{V} \tag{2}$$

$$\frac{\mathrm{d}M'}{\mathrm{d}t} = -r_M + \frac{F'\left(M_0 - M'\right)}{V'}; \tag{3}$$

': deviation variables.

If enough short time intervals are taken into account, methanol concentration and methanol consumption rate variation can be considered non-significant (Eq. 4). The smoothing spline technique is applied for methanol concentration derivative calculation

$$F_{t+\Delta t} = F_t - \frac{V}{\left(M_0 - M\right)} \bullet \frac{dM}{dt}. \tag{4}$$

To maintain a constant methanol (inducer) concentration, control Eq. 4 should be updated with the PI feedback term (Eq. 5), to regulate methanol addition aiming at stabilizing the signal around the set methanol concentration value

$$F_{t+\Delta t} = F_t - \frac{V}{\left(M_0 - M\right)} \bullet \frac{\mathrm{d}M}{\mathrm{d}t} + K_p\left(\left(\varepsilon_t - \varepsilon_{t-1}\right) + \frac{\Delta t}{\tau_I}\varepsilon_t\right). \tag{5}$$

The combination between the model and the PI algorithm makes it possible for the control parameters, K_p and τ_I, to remain constant throughout the cultivation. The K_p used was within the range of 0.02–0.05 [$L^2\ g^{-1}\ h^{-1}$], and, as it will be demonstrated further, was dependent on selected residual methanol concentration; the constant $\tau_I = 10$ min was used in the experiments.

Mathematical modelling

The macrokinetic model presented by Ren et al. [28] was employed to model cell biomass (X_v, X_d and X) and glycerol concentrations (S_{Gly}) during HBcAg-producing *P. pastoris* cultivations. A non-monotonically increasing function equation was chosen to model the methanol concentration (S_{MeOH}); and the product accumulation dynamics (P_{HBcAg}) were estimated using a Luedeking–Piret model equation. MATLAB® 2018a was used for mathematical modelling.

Macrokinetic models describing the balances of energy and intracellular substances for both glycerol and methanol uptake phases were constructed in matrix form according to Ren et al. [28].

For the glycerol uptake phase

$$\begin{bmatrix} 1 & K1_{\mathrm{Gly}} & 0 & 0 \\ 2 & -4K2_{\mathrm{Gly}} & 5 & -2 \\ 1 & -K1_{\mathrm{Gly}} - K2_{\mathrm{Gly}} - \frac{1}{Y_{\mathrm{ATP}}} & 1 & 2P/O_{\mathrm{Gly}} \\ 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} q_G \\ \mu \\ q_{Ac} \\ q_{O_2} \end{bmatrix} = \begin{bmatrix} q_{\mathrm{Gly}} \\ 0 \\ m\mathrm{ATP}_{\mathrm{Gly}} \\ 0 \end{bmatrix}. \tag{6}$$

The respective biomass growth rate (μ) was determined using Eq. 6. To account for the characteristic lag phase in *P. pastoris* cultivations, the specific glycerol uptake rate (q_{Glyc}) was first described using the metabolic regulator model and afterwards with the well-known Monod model. The switch between the two models was initiated by the minimum function

$$\frac{\mathrm{d}q_{Gly,R}}{\mathrm{d}t} = k_1\left(q_{\mathrm{Gly}} + q_{\mathrm{min,Gly}}\right) + \left(-k_2 - \mu\right)q_{\mathrm{Gly},R} \tag{7}$$

$$q_{\mathrm{Gly},M} = \frac{q_{\mathrm{max,Gly}} * S_{\mathrm{Gly}}}{K_{S,\mathrm{Gly}} + S_{\mathrm{Gly}}} \tag{8}$$

$$q_{Gly} = \min\{q_{\mathrm{Gly},R}, q_{\mathrm{Gly},M}\} * M_{\mathrm{Gly}}. \tag{9}$$

For the methanol uptake phase

$$\begin{bmatrix} \frac{3}{1-\varphi} & \frac{3}{1-\varphi} - K1_{\mathrm{MeOH}} & 0 & 0 \\ \frac{5\varphi+1}{1-\varphi} & \frac{6\varphi}{1-\varphi} - K1_{\mathrm{MeOH}} - 4K2_{\mathrm{MeOH}} & 5 & -2 \\ -1 & -3K1_{\mathrm{MeOH}} - K2_{\mathrm{MeOH}} - \frac{1}{Y_{\mathrm{ATP}}} & 1 & 2P/O_{\mathrm{MeOH}} \\ 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} q_G \\ \mu \\ q_{Ac} \\ q_{O_2} \end{bmatrix} = \begin{bmatrix} q_{\mathrm{MeOH}} \\ 0 \\ m\mathrm{ATP}_{\mathrm{MeOH}} \\ 0 \end{bmatrix}. \tag{10}$$

The respective biomass growth rate (μ) values were determined by solving the equation. The specific methanol uptake rate (q_{MeOH}) was calculated according to the non-monotonically increasing function first described by Jackson and Edwards [36]

$$q_{\mathrm{MeOH}} = \frac{q_{\mathrm{max,MeOH}} * S_{\mathrm{MeOH}}}{K_{S,\mathrm{MeOH}} + S_{\mathrm{MeOH}} + (S^2_{\mathrm{MeOH}}/K_i)} * M_{\mathrm{MeOH}}. \tag{11}$$

As recombinant product biosynthesis under the AOX1 promoter takes place only during the methanol induction phase, product accumulation was modelled only during this phase and is described by the following equation:

$$q_P = \mu_{\mathrm{MeOH}} * Y_{PX}. \tag{12}$$

Finally, a bioreactor model was constructed to estimate the vegetative, dead and total cell biomass (X_v, X_d, and X, respectively), and substrate (S) and product (P) concentrations. A model describing the liquid media volume (V) is also included, as the reactor volume directly influences the concentration of all the aforementioned parameters

$$\frac{\mathrm{d}X_v}{\mathrm{d}t} = \mu X_v - K_d X_v - \frac{F_S + F_{\mathrm{NH3}} + F_{\mathrm{AF}} - F_{\mathrm{evp}}}{V} X_v \tag{13}$$

$$\frac{\mathrm{d}X_d}{\mathrm{d}t} = K_d X_v - \frac{F_S + F_{\mathrm{NH3}} + F_{\mathrm{AF}} - F_{\mathrm{evp}}}{V} X_d \tag{14}$$

$$X = X_v + X_d \tag{15}$$

$$\frac{\mathrm{d}S_{\mathrm{Gly}}}{\mathrm{d}t} = -q_{\mathrm{Gly}} X_v - \frac{F_S + F_{\mathrm{NH3}} + F_{\mathrm{AF}} - F_{\mathrm{evp}}}{V} S_{\mathrm{Gly}} + \frac{F_S}{V} S_{F,\mathrm{Gly}} \tag{16}$$

$$\frac{\mathrm{d}S_{MeOH}}{\mathrm{d}t} = -q_{\mathrm{MeOH}} X_v - \frac{F_S + F_{\mathrm{NH3}} + F_{AF} - F_{\mathrm{evp}}}{V} S_{\mathrm{MeOH}} + \frac{F_S}{V} S_{F,\mathrm{MeOH}} \tag{17}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = q_P X_v - \frac{F_S + F_{\mathrm{NH3}} + F_{\mathrm{AF}} - F_{\mathrm{evp}}}{V} P \tag{18}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = F_S + F_{\mathrm{NH3}} + F_{\mathrm{AF}} - F_{\mathrm{smpl}} - F_{\mathrm{evp}}. \tag{19}$$

Model parameter sensitivity analysis

To evaluate the model parameters that have the most impact on modelling results, a model parameter sensitivity analysis procedure was carried out. The value of one selected parameter was increased by 5%, while others remained constant, and the effect on modelled variables was estimated by evaluating the difference of the new and reference root-mean-square errors ($\mathrm{RMSE}_{5\%}$ and RMSE_0, respectively) using the following formula:

$$\mathrm{Err}_{5\%} = \frac{\mathrm{RMSE}_{5\%} - \mathrm{RMSE}_0}{\mathrm{RMSE}_0} * 100\%. \tag{20}$$

Downstream processing of HBcAg

To compare the outcome of recombinant HBcAg-VLPs produced under different cultivation conditions, a

well-established protocol from Freivalds et al. with some modifications was used [32].

Briefly, 4.0 g of frozen yeast cells were first suspended in 16 mL of lysis buffer (20 mM Tris–HCl, 100 mM NaCl, 0.1% Triton X-100, pH 8.0) and disrupted with a French press (4 cycles, 10′000 psi). The supernatant was separated by centrifugation (30 min, 15,500*g*, at 15 °C) and pH was adjusted to 8.0 using 0.5 M NaOH. The supernatant was then incubated at 65 °C for 1 h and subsequently centrifuged (15 min, 15,500*g*). Solid ammonium sulfate was added to the solution up to a saturation of 40%, incubated at 4 °C for 30 min, and centrifuged once more (15 min, 15,500*g*). The precipitate was then dissolved in 2 mL of 20 mM Tris–HCl buffer (pH 8.0) and loaded onto a gel filtration Sepharose 4 Fast Flow (50 mL bed volume) column in 20 mM Tris–HCl pH 8.0, 100 mM NaCl connected to an ÄKTA chromatography system (*Amersham Biosciences*, UK). The fractions containing HBcAg-VLPs were isolated, pooled, and loaded onto an anion-exchange Fractogel DEAE column (20 mL bed volume) connected to an ÄKTA chromatography system (*Amersham Biosciences*, UK). The column was preequilibrated with 20 mM Tris–HCl pH 8.0, 100 mM NaCl, and run at 5.0 mL min^{-1}. Column-bound proteins were eluted by a linear gradient with the same buffer containing 1 M NaCl.

Purity and quality of HBc VLPs were assessed by SDS-PAGE and electron microscopy and the concentration established by the Bradford assay method.

Results

HBcAg production

Scientific literature lacks successful examples of bioreactor cultivations at increased residual methanol concentrations for HBcAg production with *P. pastoris*; however, it is standard practice when producing HBcAg in flasks. Whether induction at increased residual methanol concentrations can improve recombinant HBcAg biosynthesis yields in *P. pastoris* was the main hypothesis investigated during this research. A series of five experiments was performed and three residual methanol levels during the induction phase were investigated – 0.1 (limiting), 1.0 and 2.0 g L^{-1}. The experiments can be summarized as follows:

- Exp. 1—according to the Invitrogen protocol; after adaptation to growth on methanol (process 30th hour), constant methanol feed solution addition with the rate of 0.36 mL min^{-1}.
- Exp. 2—after adaptation to growth on methanol (process 32th hour), constant methanol feed solution addition with the rate of 0.4–0.5 mL min^{-1}.
- Exp. 3—after adaptation to growth on methanol (process 40th hour), methanol feed rate control according to the PI algorithm, methanol set-point – 1.0 g L^{-1}; $K_p = 0.05$ L^2 g^{-1} h^{-1}, $\tau_I = 10$ min.
- Exp. 4—after adaptation to growth on methanol (process 40th hour), methanol feed rate control according to the PI algorithm, methanol set-point – 2.0 g L^{-1}; $K_p = 0.05$ L^2 g^{-1} h^{-1}, $\tau_I = 10$ min.
- Exp. 5—after adaptation to growth on methanol (process 40th hour), methanol feed rate control according to the PI algorithm, methanol set-point – 2.0 g L^{-1}; $K_p = 0.02$ L^2 g^{-1} h^{-1}, $\tau_I = 10$ min.

The biomass growth, methanol concentration, methanol feed rate, and HBcAg accumulation dynamics during these experiments are presented in Fig. 2.

Under various residual methanol levels, differing amounts of HBcAg (858–1308 mg L^{-1}) were recovered after 67–75 h of cultivation (40–45 h of induction). Interestingly, opposite to the reports of increased residual methanol levels promoting other recombinant product biosynthesis, the highest HBcAg concentration of 1308 mg L^{-1} was produced in Exp. 1, which was performed in accordance to the Invitrogen guidelines and without any residual methanol concentration control. Although the higher cell concentration reached in this experiment attributed to an increased overall process productivity, the specific HBcAg yields per DCW were also higher than in other experiments – 11.2 mg g^{-1}. The higher methanol feed rate (0.4–0.5 mL min^{-1}) and a slightly higher reached cell biomass in Exp. 2 did not amount to an increased final HBcAg concentration or specific yield – 996 mg L^{-1} and 8.3 mg g^{-1}, respectively.

Cultivation in Exp. 3 was carried out under a controlled residual methanol concentration of 1.0 g L^{-1}. The PI-algorithm-based methanol concentration control performed admirably and was able to maintain the residual methanol concentration at 1.0 g L^{-1} with an average deviation of ± 0.3 g L^{-1}. After 72 h of cultivation, 1 L of culture was removed to assess how the control would adapt to the new conditions. After a sharp methanol concentration increase, the feed rate values were decreased and increased again when the excess residual methanol was consumed. At approx. 78 h, a sudden methanol concentration spike was noted. It is unclear what exactly caused the cell methanol uptake to suddenly drop as such. Although the methanol feed was automatically stopped, it took about 12 h for the cells to consume the excess methanol. The results from this experiment also indicate that the optimal cultivation length for maximum HBcAg production is around 70–75 h, as an overall yield decrease can be noted the next day (92 h). Yet, a lower biomass concentration (92.0 g L^{-1}) and a lower both total and specific productivity were achieved in this

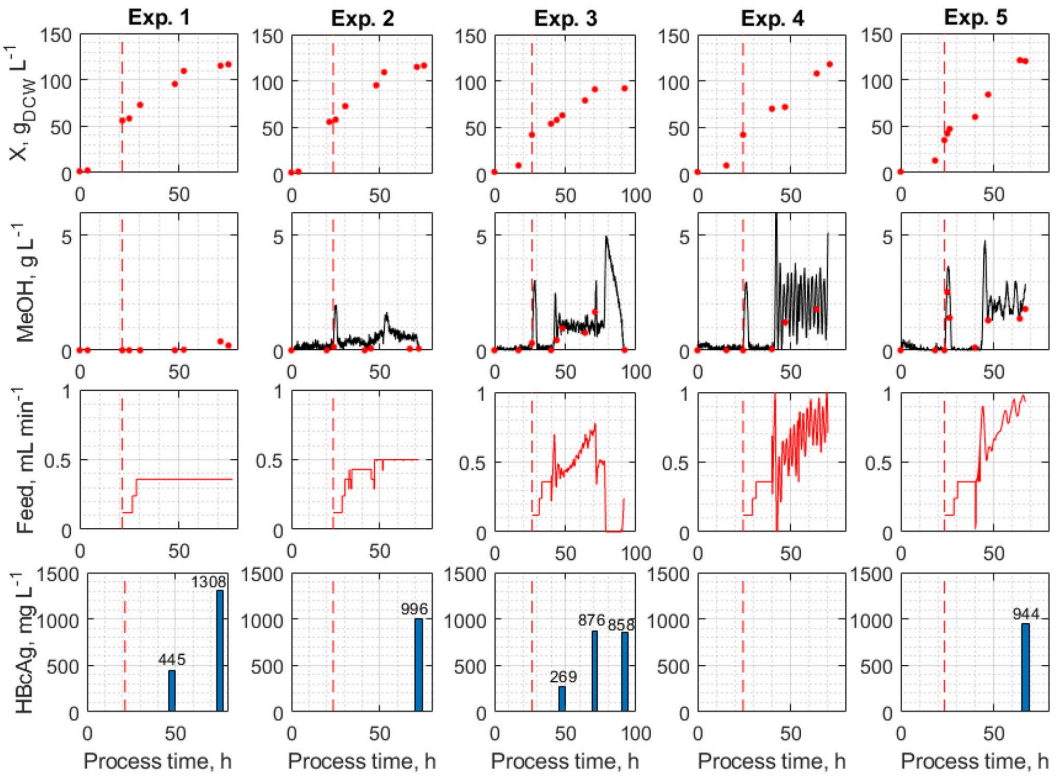


Fig. 2 Cell biomass growth (*X*), residual methanol concentration (MeOH), methanol feed rate, and recombinant HBcAg accumulation dynamics during cultivation experiments 1–5. Vertical dashed line indicates the start of the methanol induction phase

experiment in comparison to Exp. 1 − 876 mg L^{-1} and 9.6 mg g^{-1}, respectively.

Exp. 4 was performed at a residual methanol concentration of 2.0 g L^{-1}; however, the same control parameter values were used as in Exp. 3 and this led to a very poor methanol concentration control with an average deviation of approx. ± 1.0 g L^{-1}. This illustrates the importance of process-specific control parameter tuning. Consequently, a very poor production was shown by SDS-PAGE analysis and a decision was made not to purify the end culture. This experiment was repeated with more suitable PI-control parameter values ($K_P = 0.02$ L^2 g^{-1} h^{-1}) in Exp. 5 and a much better residual methanol concentration control was achieved − 2.0 ± 0.39 g L^{-1}. Consequently, a much better productivity was achieved − 944 mg L^{-1} and 7.9 mg g^{-1}. Although this productivity is comparable to the one achieved in Exp. 1 and was achieved in 67 h (instead of 75 h), around 20% more methanol was used, which would impact process overall economic feasibility. Taking this into consideration, Exp. 1 still seemingly has the upper hand production-wise.

The gel electrophoresis analysis from Exp. 1 in Fig. 3 illustrates the accumulation dynamics of HBcAg during methanol induction. Significant HBcAg accumulation can already be noted in the sample taken just 3.5 h after the start of methanol induction. Despite the thick HBcAg band of the 48th h sample, the specific yield of recombinant protein appears to be approx. 2–3 times higher for the end culture (75th process hour).

Previously processed samples were loaded onto a size-exclusion chromatography column packed with Sepharose 4 *Fast flow* resin. In laboratory scale, we used a 10 mm Tricorn column, packed at 550 mm with a flow rate of 0.30 mL min^{-1}. HBcAg elution profile is presented in Fig. 4, where the second peak represents the HBcAg VLP fractions (3–7).

SDS-PAGE gel showing HBcAg bands and immunodiffusion titres after each purification step is shown in Fig. 5.

Finally, the collected fractions from the size-exclusion chromatography are merged and loaded onto an anion-exchange column packed with a Fractogel DEAE resin. In

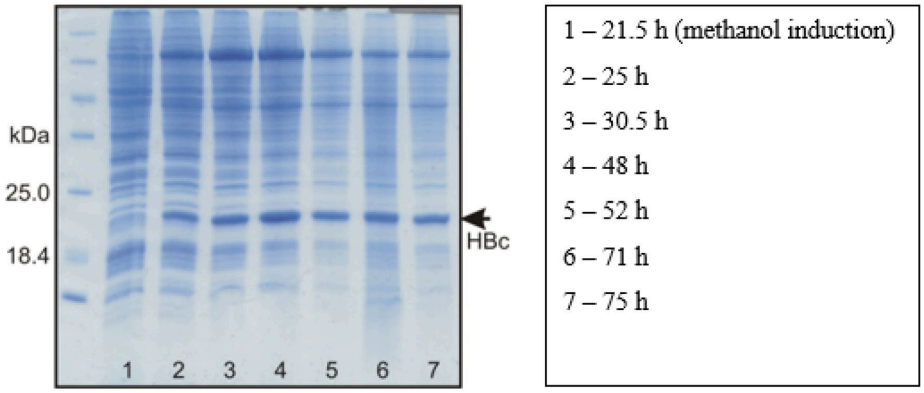


Fig. 3 SDS-PAGE analysis of HBcAg accumulation dynamics from Exp. 1

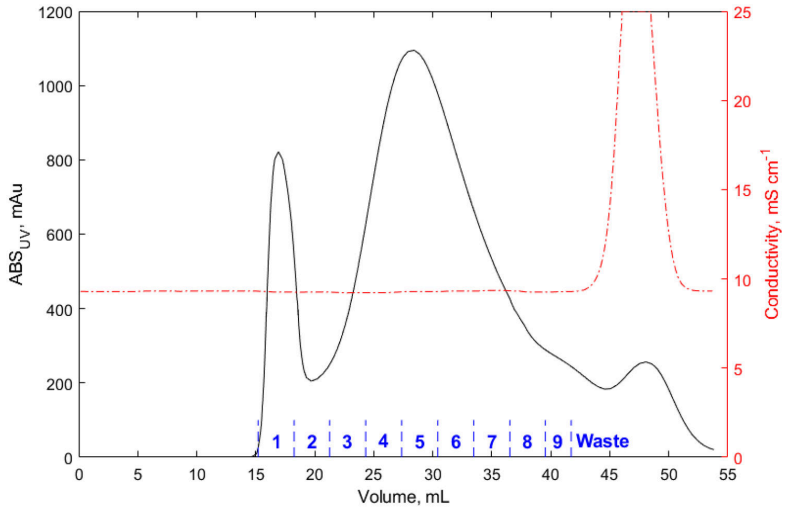


Fig. 4 Sepharose 4*FF* size-exclusion chromatography profile of HBcAg elution (Exp. 1, 75 h). Black line represents protein elution; red line—conductivity

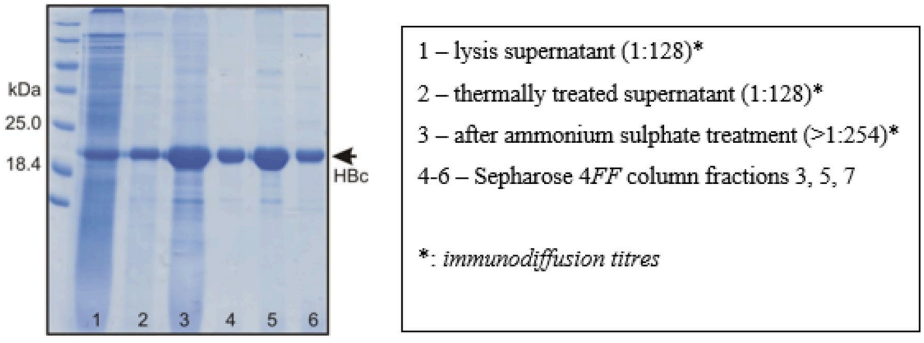


Fig. 5 SDS-PAGE analysis of HBcAg purification steps (Exp. 1, 75 h)

laboratory scale, we used a column with a bed volume of 20.0 mL and a flow rate of 4.0 mL min^{-1}. The HBcAg is bound to the column and eluted by a linear salt gradient. The HBcAg elution profile is shown in Fig. 6. Fractions 2 and 3 are collected and pooled for final HBcAg concentration determination and electron microscopy imaging.

The HBcAg concentration in the final fraction was analysed using the universal Bradford assay. It was estimated that the specific HBcAg yield reached 11.2 mg g^{-1} and the total yield was 1308 mg L^{-1}. The estimated product purity is roughly 90%. The presence of purified, correctly formed, homogenous HBcAg-VLPs is confirmed by the electron microscopy images in Fig. 7. As is apparent from the electron microscopy images, the purified HBcAg-VLPs from samples after 48 h and 75 h of cultivation indicate similar quality.

Mathematical modelling

Along with the Process Analytical Technology (PAT) and Good Manufacturing Practice (GMP) directive-driven requirements for knowledge-based in-depth fermentation process understanding, some typical large-scale culture problems can also be mediated by successfully employing

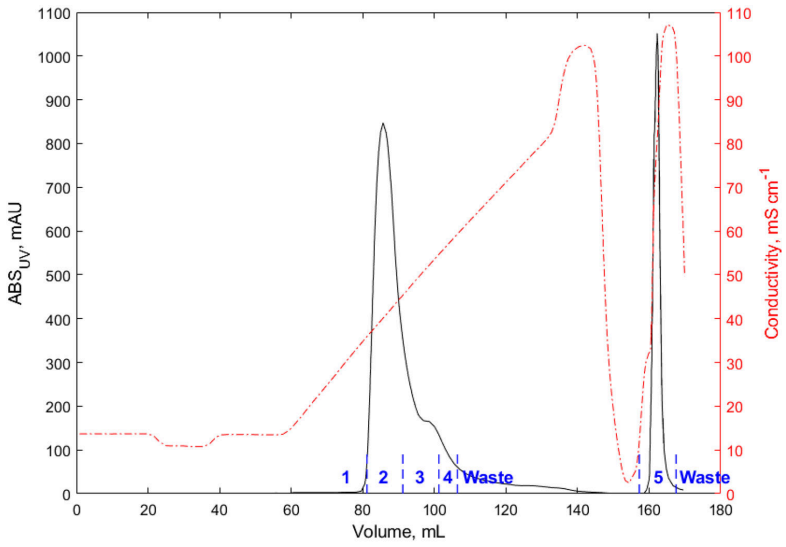


Fig. 6 HBcAg VLP purification with a Fractogel DEAE column. Black line represents protein elution, red line represents salt gradient (Exp. 1, 75 h)

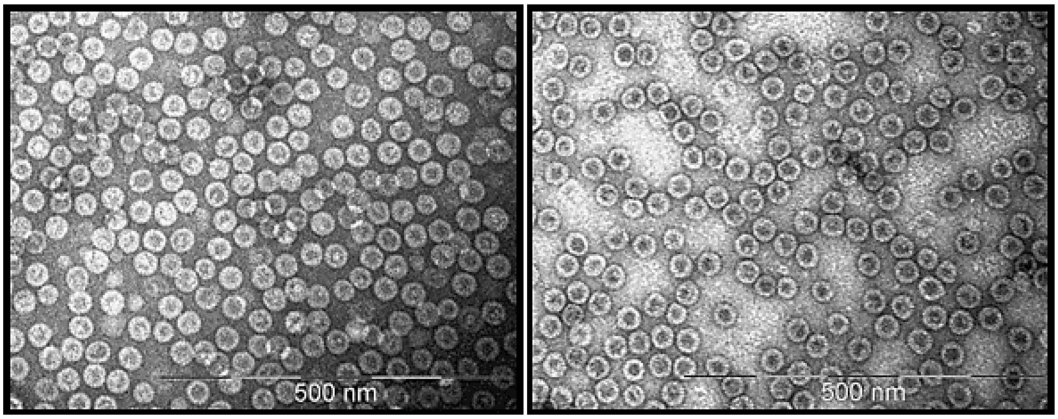


Fig. 7 Electron microscopy images of purified HBcAg-VLPs (Exp. 1, (left): 48 h; (right): 75 h)

fermentation process mathematical modelling. Namely, to select an appropriate substrate feeding rate profile that provides optimal cell biomass growth and avoids substrate-induced catabolite repression, while also mitigating the very high oxygen demand and metabolic heat production during the methanol induction phase.

When modelling *P. pastoris* cultivations, the goal should be to refrain from using batch-dependant parameters, and, if possible, to identify comprehensive parameter values that provide suitable accuracy, while remaining constant for all batches. The former approach also sometimes requires time-intensive parameter-identification procedures for every batch, whereas the latter does not.

Here, we present mathematical models to portray cell biomass (X_v, X_d, and X), glycerol (Gly) and methanol (MeOH) concentrations, HBcAg accumulation (P), and reactor volume (V) change dynamics during five HBcAg-producing *P. pastoris* cultivations. Experimental data were used to evaluate model precision.

To our knowledge, this is the first reported attempt to model HBcAg accumulation during a *P. pastoris* bioreactor cultivation and the residual methanol concentrations in this particular range. The modelled cultivation experiments are presented in Fig. 8. Identified model parameters and their respective reference values are compiled in Table 1.

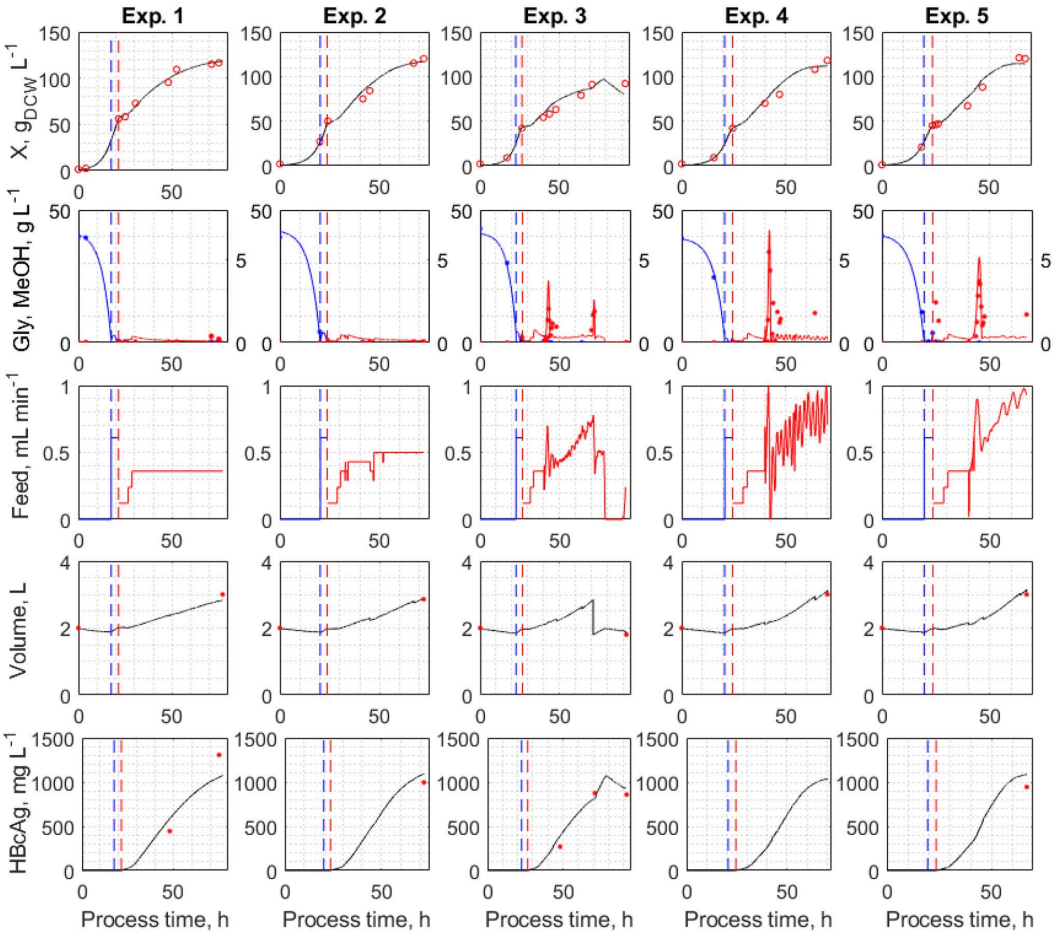


Fig. 8 Modelling results from 5 HBcAg-producing *P. pastoris* cultivations. Vertical blue dashed line indicates the onset of the glycerol fed-batch phase, and red line indicates the start of methanol induction

Table 1 Identified model parameters

Symbol	Unit	Glycerol phase		Methanol phase	
		Identified value	Reference	Identified value	Reference
K1	mol g^{-1}	0.001	0.001[b]	0.0015	0.0015[b]
*K2**	mol g^{-1}	0.0127	0.009–0.015[b]	0.017	0.012–0.017[b]
Y_{ATP}	g mol^{-1}	10.50	10.50[b]	10.50	10.50[b]
φ	–	–	–	0.25	0.25[b]
P/O	mol mol^{-1}	1.50	1.50[b] 1.53[c]	2.13	1.76–2.50[c]
K_S	g L^{-1}	0.30	0.30[b] 0.06–0.14[c] 0.10[a]	0.30	0.30[b] 0.10[a]
mATP	mol g^{-1} h^{-1}	0.0015	0.0015[b] 0.0015[c]	0.0001	0.0001[b] 0.0017–0.069[c]
q_{min}	mol g^{-1} h^{-1}	0.0006	0.0006[b]	–	–
q_{max}*	mol g^{-1} h^{-1}	0.067	0.053–0.062[b] 0.016–0.021[c] 0.004[a]	0.008	0.038–0.051[b] 0.776–1.095[c] 0.018[a]
K_i*	g L^{-1}	–	–	11.0	9.1–19.3[c] 7.6–8.9[d] 8.9[e]
k1	h^{-1}	0.40	0.40[b]	–	–
*k2**	h^{-1}	0.255	0.20–0.29[b]	–	–
F_{evp}*	L h^{-1}	0.0066	–	0.0066	–
S_F	g L^{-1}	500	–	790	–
K_d*	h^{-1}	0.0012	–	0.0012	–
Y_{PX}*	g g^{-1}	–	–	0.091	–
M	g mol^{-1}	92.09	92.09[f]	32.04	32.04[f]

*Determined experimentally; Jahic *et al.* [27]; bRen *et al.* [28]; cNiu *et al.* [29]; dBarrigon *et al.* [30]; eZhang *et al.* [31]; [f]Pubchem database [37]

Table 2 Model accuracy

Modelled variable	RMSE	NRMSE
X, g L^{-1}	6.05	5.05%
Gly., g L^{-1}	1.65	3.85%
MeOH, g L^{-1}	1.26	20.83%
Volume, L	0.07	5.65%
HBcAg, mg L^{-1}	112.1	8.57%

Model quality was estimated by means of root-mean-square error (RMSE) and normalized root-mean-square error (NRMSE) values. Calculated values are presented in Table 2.

The model parameter sensitivity analysis results are presented in Table 3. Table values indicate the deviation in each respective modelled variable RMSE, caused by a 5% increase of said model parameter.

The results from the model parameter sensitivity analysis procedure indicate that both glycerol and methanol modelled concentrations are very sensitive to most model parameter values. This can be explained by the relatively low residual concentrations of the substrates in the medium. As for the parameters that have the most impact on modelling results, constants *k1* and *k2* jump out. Such high sensitivity can be explained, first of all, because they directly influence the biomass (X) concentration, which in turn has a major effect on most other modelled variables. Second, because minor deviations from the biomass growth profile in the early cultivation hours can have a snowball effect in the late process hours.

Discussion

Historically, HBcAg-VLPs have played a key role in vaccine development and, consequently, the efficient production and purification of these particles can play a major role in commercial vaccine production. There are many reports of successfully producing these particles in various recombinant expression platforms; however, *P. pastoris* stands out due to its ability to reach high cell densities on minimal media, perform post-translational modifications, and drive high-level expression of recombinant proteins. The substrate feeding strategy during the methanol induction phase for recombinant product biosynthesis is recognised to have a significant

Table 3 Model parameter sensitivity analysis results

Param	Glycerol phase					Methanol phase				
	X	S_{Gly}	S_{MeOH}	V	P_{HBcAg}	X	S_{Gly}	S_{MeOH}	V	P_{HBcAg}
K1	1.3%	7.2%	− 1.6%	0%	0.1%	− 0.4%	0%	0.9%	0%	− 0.7%
K2	− 3.5%	− 52.8%	9.1%	0%	− 0.2%	0.2%	0%	10.0%	0%	0.1%
Y_{ATP}	13.2%	25.9%	− 11.2%	0%	6.3%	13.2%	25.9%	− 11.2%	0%	6.3%
φ	–					0%	0%	0%	0%	0%
P/O	7.0%	29.1%	2.7%	0%	1.1%	3.9%	0%	− 1.4%	0%	6.3%
K_S	3.7%	0.5%	− 10.5%	0%	0.7%	− 0.1%	0%	2.2%	0%	0%
mATP	3.8%	1.6%	− 8.3%	0%	− 0.7%	0%	0%	0.1%	0%	0%
q_{min}	3.7%	47.2%	− 5.8%	0%	− 0.1%	–				
q_{max}	− 1.1%	− 13.7%	5.8%	0%	− 0.1%	0%	0%	2.4%	0%	0%
K_i	–					0%	0%	− 1.3%	0%	0%
k1	14.6%	84.5%	41.8%	0%	− 1.9%	–				
k2	85.6%	88.4%	98.5%	0%	79.3%	–				
F_{evp}	5.1%	1.6%	− 9.5%	3.4%	3.3%	5.1%	1.6%	− 9.5%	3.4%	3.3%
K_d	0.8%	0%	0.9%	0%	− 0.9%	0.8%	0%	0.9%	0%	− 0.9%
Y_{PX}	–					0%	0%	0%	0%	8.8%

influence on the product yield and process productivity. Despite this, the effect of residual methanol concentration on HBcAg production with *P. pastoris* in bioreactor cultivations has not been extensively studied and reported. In this particular research, three residual methanol levels (0.1, 1.0, and 2.0 g L^{-1}) during the HBcAg-producing *P. pastoris* cultivation methanol induction phase were investigated.

Excellent recombinant HBcAg yields were achieved in all cultivations; however, the best yield was produced in Exp. 1—according to the Invitrogen guidelines and without residual methanol concentration control during induction (1308 mg L^{-1} and 11.2 mg g_{DCW}^{-1}). When the residual methanol concentration was controlled at 1.0 or 2.0 g L^{-1}, lower HBcAg yields were registered – 876 mg L^{-1} and 9.6 mg g_{DCW}^{-1} in Exp. 3, and 944 mg L^{-1} and 7.8 mg g_{DCW}^{-1} in Exp. 5, respectively. Additionally, approx. 10–20% more methanol was used in Exps. 3 and 5, which should also be taken into account when evaluating the economic feasibility of a production process. Also, residual methanol concentration control requires both the technical possibility and the implementation and tuning of a suitable methanol feed rate control algorithm. Overall, it is evident that in this particular case, residual methanol concentration control is not feasible. A review of reported *P. pastoris*-produced HBcAg yields is presented in Table 4.

Unfortunately, experimental screening at higher residual methanol levels (4.0–6.0 g L^{-1}) was not possible, due to the insufficiently low temperature of the tap water used for cooling. Consequently, the bioreactor's cooling system was unable to successfully maintain the temperature below 32 °C. To avoid this setback, an additional cooling system should be employed, for example, a circulation chiller. Another possible option to avoid cooling issues would be to grow the cells to lower densities; however, this would also lead to lower HBcAg yields. An example can be seen in Exp. 3—where lower cell densities lead to lower overall HBcAg yields, despite the second highest specific HBcAg yield.

The overall HBcAg yield in Exp. 1 is the highest reported in both scientific and patent literature to our knowledge. The second highest reported yield by Freivalds et al. [32] utilizes the same HBc producer clone and a similar purification procedure. There are two main differences. First, BMGY was selected as the cultivation medium. In general, production in minimal (salt) medium, such as BSM, is preferred over rich (complex) BMGY medium, as it eases subsequent product purification and improves overall process economic feasibility. Also, the ability to reach high cell densities and high-level production in minimal medium is one of the main advantages the yeast *P. pastoris* presents. Second, the authors did not employ a glycerol fed-batch phase, but added a glycerol–methanol mixture to facilitate faster adaptation to methanol consumption. While this may indeed facilitate better cell adaptation to methanol consumption, the absence of a glycerol fed-batch phase leads to lower cell biomass concentrations and, consequently, overall process yields. Although the specific HBcAg yields are comparable (11.2 vs 10.0 mg g_{DCW}^{-1}), we reached an almost 2.5 times higher biomass concentration and, consequently, a 2.5 times higher overall HBcAg yield.

The residual methanol concentration in Exps. 3–5 was controlled by altering the methanol feed rate according to the values calculated by the PI algorithm. This control solution proved to be able to control the residual methanol concentration at set-point value with sufficient accuracy. The main advantage being relatively easy implementation. This controller, however, is quite sensitive to the selected parameter

Table 4 Reported purified HBcAg yields

HBcAg yield		Host	Strain	Media[a]	MeOH (g L^{-1})	Cultivation length (h)	Scale	Refs.
mg L^{-1}	mg g_{DCW}^{-1}							
1308.0	11.2	*P. pastoris*	GS115 Mut$^+$	BSM	Limiting	75	5L bioreactor	This work
944.0	7.9	*P. pastoris*	GS115 Mut$^+$	BSM	0.2%	67	5L bioreactor	This work
876.0	9.6	*P. pastoris*	GS115 Mut$^+$	BSM	0.1%	71	5L bioreactor	This work
700.0	10.0*	*P. pastoris*	GS115 Mut$^+$	BMGY	Limiting	92	10L bioreactor	[32]
398.0	54.6	*E. coli*	BL21	M9Cas	–	16	Shake flasks	[38]
120.0	-	*P. pastoris*	GS115 Mut$^+$	BMGY/BMMY	0.5%	92	Shake flasks	[39]
80.0	-	*P. pastoris*	GS115	BMGY/BMMY	0.6%	114	Shake flasks	[40]
79.0	-	*E. coli*	BL21	2YT	–	24*	Shake flasks	[41]
69.0	-	*P. pastoris*	GS115 Mut$^+$	BMGY	Limiting	72	2L bioreactor	[42]
64.0	-	*P. pastoris*	GS115 Mut$^+$	BMGY	0.5%	144	Shake flasks	[41]
58.1	8.9*	*E. coli*	BL21	2YT	–	24*	Shake flasks	[43]
55.2	12.0*	*E. coli*	BL21	LB	–	24*	Shake flasks	[44]
50.0	-	*P. pastoris*	GS115 Mut$^+$	BMGY	Limiting	96	2L reactor	[45]
24.7	1.8*	*E. coli*	JM109	LB	–	–	Shake flasks	[43]
8.7	0.05*	*P. pastoris*	GS115	BMGY/ BMMY	0.5%	114*	Shake flasks	[43]
2.5	-	*E. coli*	BL21	2YT	–	24*	Shake flasks	[45]

*Estimated; [a]*BSM* Basal Salts Medium, *BMGY* Buffered Glycerol Complex Medium, *M9Cas* Trp-deficient M9 minimal medium supplemented with casamino acids, *BMMY* Buffered Methanol Complex Medium, *2YT* Yeast extract and Tryptone medium, *LB* Luria Bertani medium.

values, as can be seen in Exps. 4 and 5; therefore, a procedure for optimum control parameter value identification is necessary. The control was able to adapt to unexpected situations—for example, when approx. 1 L of culture was removed during Exp. 3 (71 h)—although a brief spike in the residual methanol concentration was noted, it was corrected shortly after. At times, however, the control was prone to increasing oscillations and operator involvement was necessary to stabilize the control signal. Overall, the control proved capable to control the methanol concentration at the set-point value with sufficient accuracy and can be employed in *P. pastoris* cultivation processes.

The purification procedure used in this research allowed for a rapid and efficient HBcAg VLP purification. When purifying recombinant proteins, an important aspect to consider is the time required for the purification procedure. In this case, the whole procedure can be carried out in 1 day. By following this purification protocol, highly immunogenic, correctly formed and homogenous HBcAg-VLPs with an estimated purity of 90% can be achieved.

As indicated earlier, process mathematical modelling can play an important role in recombinant product development and manufacture and, although there are several reports of successful *P. pastoris* cultivation process models, there is little information on modelling the residual methanol concentration during the induction phase. In this particular research, we present a process model that characterises cell biomass (X_v, X_d and X), substrate (Gly. and MeOH) concentration, reactor volume (V), and product (HBcAg) accumulation dynamics during cultivation. The experimental data accumulated during these cultivations were used to estimate the model accuracy.

Good results were achieved when modelling the cell biomass (± 6.05 g L^{-1}) and glycerol (± 1.65 g L^{-1}) concentrations. In prior research, using a dielectric spectroscopy probe and an off-gas analyser, we have observed that, during the start of the methanol induction phase—when significant methanol accumulation in the reactor takes place—a reduction in overall cell viability can be noted [34]. The experimental dry cell biomass measurements account also for dead (non-vegetative) cells; therefore, separate models were introduced to account for both vegetative cell (used in substrate and product calculations) and total cell (used for correlation with DCW measurements) biomass concentration. To estimate the dead cell fraction and offset the effect of these non-viable cells present on modelled cell biomass values, an empirical parameter K_d was introduced. This allowed to significantly improve the accuracy of cell biomass estimation, especially in the late induction phase, when the dead cell fraction composes a significant part of total cell biomass. However, some underestimation can still be noted. This is likely because the term K_d does not remain constant during the whole cultivation and is influenced by both the residual methanol concentration and cell exposure (induction) time.

The residual methanol concentration, however, proved difficult to model. Although the model was able to successfully

account for methanol concentration peaks at the start of residual methanol concentration control, the overall accuracy could still use more improvement (± 1.26 g L^{-1}). The model also somewhat underestimated the residual methanol concentration during cultivations. Looking at other residual methanol concentration modelling results, the model presented by Ren et al. was unable to account for a fast accumulation of methanol up to 10.0 g L^{-1} [28]. The model, however, was able to accurately estimate the methanol concentration in a methanol-limited cultivation. Although the model presented by Jahic et al. included methanol concentration estimation, they opted to represent the methanol concentration using experimental measurements, perhaps, thus, indicating that the model was not able to accurately estimate the accumulated methanol concentration up to nearly 2.0 g L^{-1} during methanol adaptation [27]. A more sophisticated approach is presented by Niu et al., as their model was able to account for methanol accumulation during initial adaptation (approx. 6.4 g L^{-1}) and at the end of the cultivation (approx. 16 g L^{-1}) [29]. Although the model is successful at estimating the residual methanol concentration at the start and the end of the methanol induction, it seemingly fails to account for a small methanol spike in the middle of cultivation. While this presented approach seems promising, it is also complex, as it incorporates real-time exhaust gas analysis and a dynamic parameter-identification procedure, which lessens the potential for practical application. Even though the accuracy of our presented model could still be improved, it demonstrates a rare attempt to model residual methanol concentration with sufficient accuracy and low complexity. The ability of the model to accurately predict increases (spikes) in methanol concentration can be very useful during both laboratory and industrial cultivations, as rapid detection of such spikes may help avoid deviating from the optimal growth trajectory or whole batch discard in general.

Recombinant HBcAg accumulation dynamics were also modelled (± 112.1 mg L^{-1}) and follow the experimentally determined yield values. However, the model could not account for the high productivity in Exp. 1, which could indicate that some other unknown factor could possibly have influenced the high yield in this particular experiment. A significant factor to note is also the fact that this measurement is very dependent on the estimated cell biomass concentration. Ensuring a higher cell biomass estimation precision would also positively impact this measurement in this case.

Overall, we have presented a relatively simple and robust bioreactor model to successfully portray cell biomass (X_v, X_d, and X) concentration, substrate (Gly. and MeOH), reactor volume (V), and product (HBcAg) dynamics during a *P. pastoris* GS115 cultivation under both limiting (0.1 g L^{-1}) and residual (1.0 and 2.0 g L^{-1}) methanol concentrations. Relatively easy implementation and parameter identification enables this approach to also be suitable for practical applications. The presented model could also be further used for model-based optimization and software-sensor development of the investigated process.

Conclusions

In this particular research, three residual methanol levels (0.1, 1.0, and 2.0 g L^{-1}) during the HBcAg-producing *P. pastoris* cultivation methanol induction phase were investigated. Increased residual methanol concentration did not have a profound effect on HBcAg accumulation dynamics. Considering the economic aspects of the production process, bioreactor cultivation under limiting methanol concentration, according to the Invitrogen cultivation guidelines, is preferable. The highest reported HBcAg VLP yield of 1308 mg L^{-1} and 11.2 mg g_{DCW}^{-1} is presented, which is about 2.5 times higher than the next highest result (489 mg L^{-1}) reported by Freivalds et al. [32].

We also demonstrate that the PI-algorithm-based residual methanol concentration control was able to maintain the selected set-point concentrations in the explored range; however, control quality was sensitive to the values of selected control parameters K_P and τ_I.

The produced recombinant HBcAg molecules were successfully purified using a modified purification protocol, and the presence of highly immunogenic, correctly formed and homogenous HBcAg-VLPs with an estimated purity of 90% was confirmed using electron microscopy.

A relatively simple and robust bioreactor model to portray cell biomass (X_v, X_d and X), substrate (Gly. and MeOH), reactor volume (V), and product (HBcAg) dynamics during a *P. pastoris* GS115 cultivation at both limiting (0.1 g L^{-1}) and residual (1.0 and 2.0 g L^{-1}) methanol concentrations is presented. A rare attempt to model residual methanol concentration during the induction phase is also presented. Relatively easy implementation and parameter identification enables this approach to also be suitable for practical applications. The presented model could be further used also for model-based optimization and software-sensor development of the investigated process.

Acknowledgements The authors would like to acknowledge the contribution of Inara Akopjana for preparing seed inoculation cultures and performing SDS-PAGE analysis, Janis Bogans for performing chromatography runs, Dr. Juris Jansons for providing the electron microscopy images, and Rita Skerbaka for the off-line glycerol and methanol analysis. This research was funded by the European Regional Development Fund (ERDF) and Latvian State Project Agreement No. 1.1.1.2/16/I/001 research application (Grant) No. 1.1.1.2/VIAA/1/16/186.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

1. Roldão A, Mellado MCM, Castilho LR et al (2010) Virus-like particles in vaccine development. Expert Rev Vaccines 9:1149–1176. https://doi.org/10.1586/erv.10.115
2. Milich DR, McLachlan A, Moriarty A, Thornton GB (1987) Immune response to hepatitis B virus core antigen (HBcAg): localization of T cell recognition sites within HBcAg/HBeAg. J Immunol 139:1223–1231
3. Moradi Vahdat M, Hemmati F, Ghorbani A et al (2021) Hepatitis B core-based virus-like particles: A platform for vaccine development in plants. Biotechnol Reports 29:e00605. https://doi.org/10.1016/j.btre.2021.e00605
4. Pumpens P (2008) Construction of Novel Vaccines on the Basis of Virus-Like Particles. In: Khudyakov YE (ed) Medicinal Protein Engineering. CRC Press, Boca Raton, pp 205–248
5. Pumpens P, Grens E (2016) The true story and advantages of the famous hepatitis B virus core particles: Outlook 2016. Mol Biol 50:558–576. https://doi.org/10.7868/S0026898416040091
6. Nooraei S, Bahrulolum H, Hoseini ZS et al (2021) Virus-like particles: preparation, immunogenicity and their roles as nanovaccines and drug nanocarriers. J Nanobiotechnology 19:1–27. https://doi.org/10.1186/s12951-021-00806-7
7. Schwartz L, Brown GV, Genton B, Moorthy VS (2012) A review of malaria vaccine clinical projects based on the WHO rainbow table. Malar J 11:1–22. https://doi.org/10.1186/1475-2875-11-11
8. Laurens MB (2019) RTS, S/AS01 vaccine (Mosquirix^{TM}): an overview. Hum Vaccines Immunother 16:480–489. https://doi.org/10.1080/21645515.2019.1669415
9. Whelan MA, Field RA (2016) Vaccines based on Hepatitis B core antigens
10. Xunlong S (2016) Chimeric hepatitis B virus core antigen therapeutic vaccine of a kind of targeting and uses thereof
11. Wang L (2013) HLA-DR9 restrictive regulatory T cell epitope of Hepatitis B virus core antigen and e antigen and application thereof
12. Ma Y (2016) Application of HBcAg (hepatitis B core antigen) virus-like particle serving as cancer therapeutic vaccine carrier
13. Owczarek B, Gerszberg A, Hnatuszko-Konka K (2019) A Brief Reminder of Systems of Production and Chromatography-Based Recovery of Recombinant Protein Biopharmaceuticals. Biomed Res Int 2019:4216060. https://doi.org/10.1155/2019/4216060
14. Cregg JM (2000) Recombinant protein expression in Pichia pastoris. Appl Biochem Biotechnol - Part B Mol Biotechnol 16:23–52. https://doi.org/10.1385/MB:16:1:23
15. Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM (2005) Heterologous protein production using the Pichia pastoris expression system. Yeast 22:249–270. https://doi.org/10.1002/yea.1208
16. Looser V, Bruhlmann B, Bumbak F et al (2015) Cultivation strategies to enhance productivity of Pichia pastoris: A review. Biotechnol Adv 33:1177–1193. https://doi.org/10.1016/j.biotechadv.2015.05.008
17. Liu WC, Inwood S, Gong T et al (2019) Fed-batch high-cell-density fermentation strategies for Pichia pastoris growth and production. Crit Rev Biotechnol 39:258–271. https://doi.org/10.1080/07388551.2018.1554620
18. Yang Z, Zhang Z (2018) Engineering strategies for enhanced production of protein and bio-products in Pichia pastoris: A review. Biotechnol Adv 36:182–195. https://doi.org/10.1016/j.biotechadv.2017.11.002
19. Potvin G, Ahmad A, Zhang Z (2012) Bioprocess engineering aspects of heterologous protein production in Pichia pastoris: A review. Biochem Eng J 64:91–105. https://doi.org/10.1016/j.bej.2010.07.017
20. Gurramkonda C, Adnan A, Gäbel T et al (2009) Simple high-cell density fed-batch technique for high-level recombinant protein production with Pichia pastoris: Application to intracellular production of Hepatitis B surface antigen. Microb Cell Fact 8:1–8. https://doi.org/10.1186/1475-2859-8-13
21. Hong F, Meinander NQ, Jönsson LJ (2002) Fermentation strategies for improved heterologous expression of laccase in Pichia pastoris. Biotechnol Bioeng 79:438–449. https://doi.org/10.1002/bit.10297
22. Cos O, Ramon R, Montesinos JL, Valero F (2006) A simple model-based control forPichia pastoris allows a more efficient heterologous protein production bioprocess. Biotechnol Bioeng 95:145–154. https://doi.org/10.1002/bit.21005
23. Pla IA, Damasceno LM, Vannelli T et al (2006) Evaluation of Mut+ and MutS Pichia pastoris phenotypes for high level extracellular scFv expression under feedback control of the methanol concentration. Biotechnol Prog 22:881–888. https://doi.org/10.1021/bp060012+
24. Parliament European (2003) Commission Directive 2003/94/EC of 8 October 2003 laying down the principles and guidelines of good manufacturing practice in respect of medicinal products for human use and investigational medicinal products for human use
25. Food and Drug Administration (2004) Guidance for Industry, PAT-A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance
26. Grigs O, Galvanauskas V, Dubencovs K, et al (2016) Model Predictive Feeding Rate Control in Conventional and Single-use Lab-scale Bioreactors: A Study on Practical Application. Chem Biochem Eng Q J 30:47–60. https://doi.org/10.15255/CABEQ.2015.2212
27. Jahic M, Rotticci-Mulder J, Martinelle M et al (2002) Modeling of growth and energy metabolism of Pichia pastoris producing a fusion protein. Bioprocess Biosyst Eng 24:385–393. https://doi.org/10.1007/s00449-001-0274-5
28. Ren HT, Yuan JQ, Bellgardt KH (2003) Macrokinetic model for methylotrophic Pichia pastoris based on stoichiometric balance. J Biotechnol 106:53–68. https://doi.org/10.1016/j.jbiotec.2003.08.003
29. Niu H, Daukandt M, Rodriguez C et al (2013) Dynamic modeling of methylotrophic Pichia pastoris culture with exhaust gas analysis: From cellular metabolism to process simulation. Chem Eng Sci 87:381–392. https://doi.org/10.1016/j.ces.2012.11.006
30. Barrigon JM, Valero F, Montesinos JL (2015) A macrokinetic model-based comparative meta-analysis of recombinant protein production by Pichia pastoris under AOX1 promoter. Biotechnol Bioeng 112:1132–1145. https://doi.org/10.1002/bit.25518
31. Zhang W, Bevins MA, Plantz BA et al (2000) ModelingPichia pastoris growth on methanol and optimizing the production of a recombinant protein, the heavy-chain fragment C of botulinum neurotoxin, serotype A. Biotechnol Bioeng 70:1–8. https://doi.org/10.1002/1097-0290(20001005)70:1%3c1::AID-BIT1%3e3.0.CO;2-Y
32. Freivalds J, Dislers A, Ose V et al (2011) Highly efficient production of phosphorylated hepatitis B core particles in yeast Pichia pastoris. Protein Expr Purif 75:218–224. https://doi.org/10.1016/j.pep.2010.09.010
33. Invitrogen Corporation (2002) Pichia Fermentation Process Guidelines. https://tools.thermofisher.com/content/sfs/manuals/pichiaferm_prot.pdf

34. Grigs O, Bolmanis E, Galvanauskas V (2021) Application of In-Situ and Soft-Sensors for Estimation of Recombinant P. pastoris GS115 Biomass Concentration: A Case Analysis of HBcAg (Mut+) and HBsAg (MutS) Production Processes under Varying Conditions. Sensors 21:1268. https://doi.org/10.3390/s21041268
35. Ouchterlony O (1966) The antigenic pattern of immunoglobulins. G Mal Infett Parassit 18:942–948
36. Jackson JV, Edwards VH (1975) Kinetics of substrate inhibition of exponential yeast growth. Biotechnol Bioeng 17:943–964. https://doi.org/10.1002/bit.260170702
37. National Institutes of Health PubChem database. https://pubchem.ncbi.nlm.nih.gov/
38. Berza I, Dishlers A, Petrovskis I et al (2013) Plasmid dimerization increases the production of hepatitis B core particles in E. coli. Biotechnol Bioprocess Eng 18:850–857. https://doi.org/10.1007/s12257-013-0188-5
39. Peng M, Chen M, Ling N et al (2006) Novel vaccines for the treatment of chronic HBV infection based on mycobacterial heat shock protein 70. Vaccine 24:887–896. https://doi.org/10.1016/j.vaccine.2005.12.050
40. Jing XD, Wei H, Li H (2018) A kind of expression and purification method of HBcAg
41. Li ZX, Hong GQ, Hu B et al (2007) Suitability of yeast- and Escherichia coli-expressed hepatitis B virus core antigen derivatives for detection of anti-HBc antibodies in human sera. Protein Expr Purif 56:293–300. https://doi.org/10.1016/j.pep.2007.08.004
42. Watelet B, Quibriac M, Rolland D et al (2002) Characterization and diagnostic potential of hepatitis B virus nucleocapsid expressed in E. coli and P. pastoris. J Virol Methods 99:99–114. https://doi.org/10.1016/S0166-0934(01)00385-8
43. Rolland D, Gauthier M, Dugua JM et al (2001) Purification of recombinant HBc antigen expressed in Escherichia coli and Pichia pastoris: Comparison of size-exclusion chromatography and ultracentrifugation. J Chromatogr B Biomed Sci Appl 753:51–65. https://doi.org/10.1016/S0378-4347(00)00538-7
44. Zhang Y, Liu Y, Zhang B et al (2021) In vitro preparation of uniform and nucleic acid free hepatitis B core particles through an optimized disassembly-purification-reassembly process. Protein Expr Purif 178:105747. https://doi.org/10.1016/j.pep.2020.105747
45. Letourneur O, Watelet B (2001) HBcAg expression and diagnostic and therapeutic uses

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com

Appendix IV
4. Pielikums

Production and Purification of Soy Leghemoglobin from *Pichia pastoris* Cultivated in Different Expression Media

Emils Bolmanis, Janis Bogans, Inara Akopjana, Arturs Suleiko, Tatjana Kazaka, Andris Kazaks

Processes **2023**, *11*, DOI: 10.3390/pr11113215

E.B. input: Conceptualization, formal analysis, investigation, methodology, software, writing – original draft preparation, writing – review and editing, visualization.

processes

MDPI

Article

Production and Purification of Soy Leghemoglobin from *Pichia pastoris* Cultivated in Different Expression Media

Emils Bolmanis [1], Janis Bogans [1], Inara Akopjana [1], Arturs Suleiko [2], Tatjana Kazaka [1] and Andris Kazaks [1,*]

[1] Latvian Biomedical Research and Study Centre, Ratsupites Street 1 k1, LV-1067 Riga, Latvia; emils.bolmanis@biomed.lu.lv (E.B.); janis.bogans@biomed.lu.lv (J.B.); inara@biomed.lu.lv (I.A.); tatyanav@biomed.lu.lv (T.K.)
[2] Laboratory of Bioengineering, Latvian State Institute of Wood Chemistry, Dzerbenes Street 27, LV-1006 Riga, Latvia; arturs.suleiko@bioreactors.net
* Correspondence: andris@biomed.lu.lv

Abstract: Plant-based meat alternatives, exemplified by Impossible Foods' Impossible Burger, offer a sustainable, ethical substitute for traditional meat, closely mimicking the taste and appearance of meat by utilizing soy leghemoglobin (LegH), a 16 kDa holoprotein found in soy plants structurally similar to heme in animal meat. Cultivation medium plays an important role in bioprocess development; however, medium development or optimization can be labor intensive, and thus the use of previously reported media can be enticing. In this study, we explored the expression of recombinant LegH in *Pichia pastoris* in various reported cultivation media (BSM, BMGY, FM22, D'Anjou, BSM/2, and RDM) and using different feeding approaches (μ-stat and mixed feed with sorbitol). Our findings indicate that optimization techniques tailored to the specific process did not increase LegH yields, highlighting the need to investigate strain-specific strategies. We also utilized the collected process data to create and train a novel artificial neural network-based soft sensor for estimating cell biomass, relying solely on standard bioreactor measurements (such as stirrer speed, dissolved oxygen, O_2 enrichment, base feed, glycerol feed, methanol feed, and reactor volume). This soft sensor proved to be robust and exhibited a strong correlation (3.72% WCW) with experimental data.

Keywords: soy leghemoglobin (LegH); *Pichia pastoris*; cultivation media; process optimization; recombinant proteins; protein purification; artificial neural networks; biomass soft sensor

Citation: Bolmanis, E.; Bogans, J.; Akopjana, I.; Suleiko, A.; Kazaka, T.; Kazaks, A. Production and Purification of Soy Leghemoglobin from *Pichia pastoris* Cultivated in Different Expression Media. *Processes* **2023**, *11*, 3215. https://doi.org/10.3390/pr11113215

Academic Editor: Enrico Marsili

Received: 26 October 2023
Revised: 9 November 2023
Accepted: 10 November 2023
Published: 12 November 2023

1. Introduction

Plant-based meat analogues are designed to provide an alternative to traditional meat products and mimic the taste, texture, and appearance of meat. These products have gained popularity due to concerns about the environmental impact of meat production, as well as health and ethical considerations. Soy leghemoglobin (LegH) is a small 16 kDa holoprotein (i.e., a protein plus a heme cofactor) found in soy plants that has a similar structure to heme in animal meat (a molecule found in animal blood and muscle tissue, responsible for the red color of meat and carrying oxygen in the blood). It is sometimes used as a meat flavoring and colorant in plant-based meat alternatives to create a more authentic meat-like taste, texture, and a reddish hue [1,2].

The Impossible Burger is a plant-based burger made by Impossible Foods Inc. (Redwood City, CA, USA), a company that specializes in developing and producing meat alternatives. The burger is made using a combination of plant-based ingredients, including soy and potato protein, coconut and sunflower oil, cellulose-based culinary binder, water, and the secret ingredient—soy leghemoglobin [1]. The LegH molecule, expressed in recombinant yeast *Pichia pastoris*, is what gives the burger its meat-like taste and aroma [3,4]. The Impossible Burger has gained popularity due to its close resemblance to traditional beef

burgers in terms of taste, texture, and appearance. It is also marketed as a more sustainable and ethical alternative to beef, as it requires less land, water, and other resources to produce.

LegH is a naturally occurring molecule found in animal tissue and certain microorganisms. However, assessing its safety becomes crucial when produced through genetic engineering for use in food products. Extensive studies on LegH as a food ingredient indicate minimal toxicological or allergic concerns [4–6], leading regulatory agencies in the United States and other countries to deem it safe for consumption [7,8]. However, approval in the EU and UK is still pending.

LegH, a single-unit hemoprotein present in leguminous plant root nodules, shares a three-dimensional structure akin to myoglobin, a hemoprotein in mammalian muscles. Its protein structure, mainly comprising alpha helices that form a stable framework, features eight helices creating a distinct pocket for heme binding [9]. In contrast to mammalian globin with four subunits, plant LegH comprises a single monomeric unit [10]. See Figure 1 for the structure of a soybean LegH molecule.

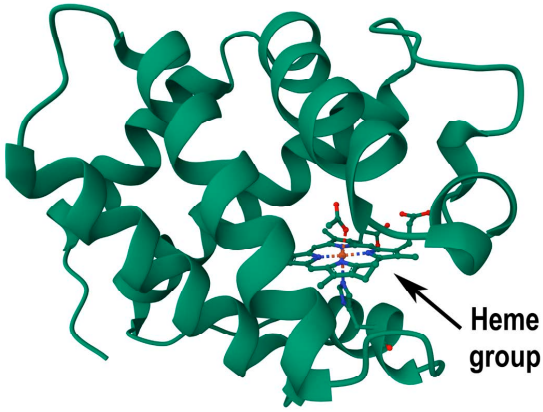


Figure 1. Three-dimensional structure of soy LegH A (PDB ID: 1BIN) [11,12].

Several microorganisms have successfully produced recombinant soy LegH. Recently, Shao et al. developed a high-yield secretion system for functional LegH expression using a *P. pastoris* yeast strain, through gene dosage optimization and heme pathway consolidation. These strategies increased LegH secretion by more than 83-fold, resulting in a maximum titer of 3.5 g/L, which is the highest ever reported for a secretory production of not only LegH, but also all heme-containing proteins [13]. Xue et al. reported the production of several heme proteins in the yeast *Saccharomyces cerevisiae*. An engineered *S. cerevisiae* strain produced a titer of 108.2 mg/L soybean LegH [14]. Recombinant LegH production has also been reported in bacterial *Escherichia coli* cells [11,15]. Jones et al. managed to achieve a yield of 20 mg/L pure product, after the optimization of the growth conditions in shake flasks [16].

Over the last two to three decades, the *P. pastoris* expression system has proven its efficacy in generating a diverse range of recombinant proteins for both research and industrial purposes. This methylotrophic yeast stands out as an excellent choice for expressing foreign proteins, thanks to its straightforward genetic manipulation, high-frequency DNA transformation, functional complementation cloning, robust intra- and extracellular protein expression capabilities, and proficiency in executing complex higher eukaryotic protein modifications such as glycosylation, disulfide bond formation, and proteolytic processing. Additionally, the low levels of native secreted proteins simplify the purification of the expressed recombinant proteins. When considering economic factors like high cell growth in minimal medium, prolonged process stability, and the availability of potent genetic techniques, *P. pastoris* undeniably emerges as the preferred system for heterologous protein expression [17,18].

The typical *P. pastoris* two-stage cultivation process is described in Invitrogen's "*Pichia Fermentation Process Guidelines*" [19]. This procedure comprises growing *P. pastoris* cells in a minimal medium, first using glycerol as a growth substrate until a suitable biomass concentration is reached, then inducing product biosynthesis by switching the substrate feed to methanol. Recent developments indicate a shift away from conventional protocols, embracing a more conceptual approach that enables the customization of process-specific strategies based on the unique attributes of the product/genetic construct and the equipment in the bioreactor [20].

Invitrogen Co.'s basal salt medium (BSM) stands out as the frequently employed minimal medium for achieving high cell density fermentation of the methylotrophic yeast P. pastoris. Despite its status as a standard medium, it may not be universally optimal and is known to exhibit certain drawbacks, including an unbalanced composition, precipitate formation, and issues related to ionic strength [21]. To circumvent the aforementioned problems, optimization of the BSM medium components is often carried out; however, this can be time and labor intensive. Therefore, opting for a previously developed medium may be preferable, as several formulations have been reported. For example, the FM22 medium by Stratton et al. [22] or the D'Anjou medium [23]. More recent minimal medium formulations include the rich defined medium (RDM) by Matthews et al. [24] and the MBSM medium by Pais-Chanfrau et al. [25]. Several authors have also reported that, reducing the salt concentrations of BSM may prove beneficial for recombinant product synthesis, while having little to no effect on cell growth [26,27].

The use of complex cultivation medium (a nutrient-rich medium that contains a variety of undefined components such as yeast extract, peptone), can sometimes produce better results than the minimal (defined) medium. However, the use of complex medium can make it difficult to control and optimize the growth conditions, result in batch-to-batch variability, and is generally much more expensive. The buffered glycerol complex medium (BMGY) is often employed in *P. pastoris* cultivations and is the go-to complex medium [24,26].

When analyzing the formulation of a *P. pastoris* cultivation medium, the study conducted by Wegner often serves as a benchmark [28]. In this study, the optimal ranges of important elements for cell growth (P, K, Mg, Ca, S, Fe, Zn, Cu, Mn) were determined experimentally in a continuous fermentation.

Artificial neural networks (ANNs) are computational algorithms designed to emulate the structure of biological brain networks, enabling the estimation and prediction of bioprocess variables using real-time sensor data, offering the capability to model intricate nonlinear systems without intricate model equations, although they necessitate substantial historical process data for precise network training and establishing the connections between input and output parameters [29]. While ANNs are typically efficient and easy to deploy with strong performance, their drawback lies in their lack of interpretability, leading to a restricted acquisition of process knowledge. Despite this limitation, ANNs have demonstrated success in predicting the behavior of diverse fermentation systems, prompting their utilization in bioprocess control applications. Recent applications of ANN models in cell biomass estimation, encompass regulating specific growth rate [30], optimizing cell biomass [31–33], and estimating [33,34] or tracking a predefined substrate concentration trajectory [35].

In this study, we explored the expression of recombinant LegH in *P. pastoris* using various documented cultivation media (BSM, BMGY, FM22, D'Anjou, BSM/2, RDM) and employed different feeding strategies (μ-stat and mixed feed with sorbitol). Generated process data were used to establish and train a novel artificial neural network-based soft sensor for cell biomass estimation, utilizing only standard bioreactor measurements (stirrer speed, dissolved oxygen, O_2 enrichment, base feed, glycerol feed, methanol feed, and reactor volume).

2. Materials and Methods

2.1. Construction of an Expression Vector and Selection of Clones

An artificial gene with *P. pastoris* optimized codons encoding LegH sequence (GenBank Acc. NP_001235248.2) was designed by GenScript and synthesized by BioCat GmbH (Heidelberg, Germany). This gene was subsequently incorporated into the pPICZC vector (Invitrogen) through *Eco*RI and *Not*I restriction sites. The resulting plasmid underwent linearization with *Pme*I and was introduced into the *P. pastoris* X-33 strain through electroporation. Mut^+ transformants were successfully obtained on agarized YPD plates containing 800 μg/mL zeocin, and the selected clones were further cultivated analytically in flasks using the rich BMGY medium with methanol induction over three days to identify the most efficient producer.

2.2. Experimental Conditions

A recombinant *P. pastoris* X-33 Mut^+ strain was used for the cultivation processes. The bioreactor vessel was filled with distilled water and subjected to sterilization at 121 °C for 30 min. Simultaneously, the cultivation media and glycerol fed-batch solutions underwent separate autoclaving under identical conditions. The trace element, vitamin, and methanol fed-batch solutions were sterilized through filtration using a 0.2 μm filter.

The fermentations were carried out in a 5 L bench-top fermenter (Bioreactors.net, EDF-5.4/BIO-4, Riga, Latvia) with a working volume ranging from 2 to 4 L, as illustrated in Figure 2. The pH levels were monitored using a calibrated pH sensor probe (Hamilton, EasyFerm Bio, Bonaduz, Switzerland) and adjusted to 5.0 ± 0.1 before initiating cultivation, maintaining the set value throughout fermentation using a 28% NH_4OH solution. Temperature control was set at 30.0 ± 0.1 °C, regulated by a temperature sensor and adjustments to the vessel jacket temperature. Dissolved oxygen (DO) levels were measured with a DO probe (Hamilton, Oxyferm Bio, Bonaduz, Switzerland) and kept above 30 ± 5% by modulating stirrer speed (200–1000 RPM) in Cascade 1 or enriching the inlet air with pure O_2 in Cascade 2. A consistent flow of air or an air/oxygen mixture at 3.0 slpm was maintained throughout the process. A condenser was employed to condense moisture from outlet gases, and antifoam 204 (Sigma, St. Louis, MO, USA) was added when needed to manage excessive foam formation. Substrate feed solutions were pumped using a high-precision peristaltic pump (Longer-Pump, BT100–2J, Baoding, China). A turbidity probe (Optek, ASD19- EB-01, Kitzingen, Germany) was employed in Experiments 1 and 9. Sensor signal was converted to wet cell biomass (g/L), according to a previously established correlation [36].

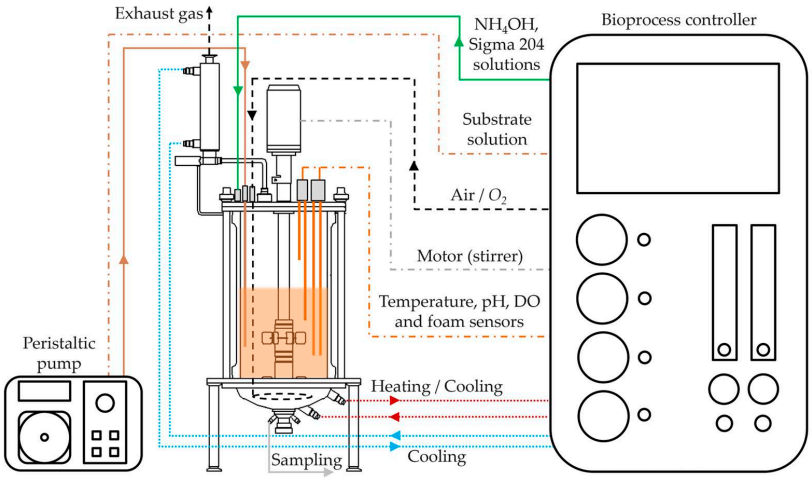


Figure 2. Schematic diagram of the bioreactor and control architecture.

The cultivation commenced with a glycerol batch phase. After 18–24 h, once the batch glycerol was exhausted, indicated by a sudden DO spike, a glycerol fed-batch solution was introduced into the reactor at a rate of 0.61 mL/min for 4 h or until reaching an optical density of 100–120. Subsequently, a brief feeding pause of 10–30 min allowed cells to consume any residual glycerol. Following this, the substrate feed transitioned to methanol, supplied to the reactor at a rate of 0.12 mL/min for 5 h, followed by 0.24 mL/min for 2 h, and finally 0.36 mL/min for the remainder of the cultivation.

In the mixed feed cultivation (Experiment 10), a mixture of methanol/sorbitol at a ratio of 0.5 C-mol/0.5 C-mol was used, according to Niu et al. [37]. The feed rate profile was the same as in previous cultivations; however, after the stirrer (Cascade 1) reached 1000 RPM, DO-stat feeding (Cascade 2) was activated instead of O_2 enrichment.

2.3. Cultivation Media

In order to evaluate the effect that the cultivation medium has on recombinant LegH biosynthesis in yeast *P. pastoris*, several reported minimal media formulations were selected. Namely, Invitrogen's BSM [19], FM22 medium reported by Stratton et al. [22], D'Anjou medium [23], BSM with the salt concentration reduced by half (denoted as BSM/2) [26,27], and the RDM without the addition of lipids reported by Matthews et al. [24]. In order to compare the performance of minimal and complex media, one experiment was carried out in BMGY medium. The compositions of the previously mentioned media are shown in Table A1 (Appendix A).

2.4. Downstream Processing of LegH

A total of 7.0 g of wet cells were resuspended in 35 mL of lysis buffer (20 mM Tris 8.0, 100 mM NaCl) and disrupted by French press (3 × 10,000 psi). The suspension was then centrifuged for 30 min at 18,500× *g* and the supernatant was buffer exchanged to 20 mM Tris 8.0 on XK26/20 column packed with 60 mL of Sephadex G-25 at 5 mL/min. Proteins were then loaded onto XK16/20 column packed with 20 mL of Sepharose Q HP equilibrated with 20 mM Tris 8.0. Bound proteins were eluted with a salt gradient using 20 mM Tris 8.0, 1 M NaCl at 3 mL/min. Finally, fractions containing the target protein were loaded onto XK16/70 column packed with 120 mL of Superdex 200 in PBS at 1 mL/min. All the columns were purchased from Cytiva. The first two processes were operated by Akta Pure 25, while the third was processed by Akta Prime Plus.

2.5. Analytical Measurements

Cell growth was monitored through offline measurements of wet cell weight (WCW), determined gravimetrically. Biomass samples were placed in pre-weighed Eppendorf® tubes and centrifuged at 15,500× *g* for 3 min. Subsequently, the supernatant was discarded, and the cells were resuspended in distilled water before undergoing another round of centrifugation. The liquid phase was discarded, and the remaining wet cell biomass was then weighed.

Protein samples collected during cultivation underwent analysis through sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), employing a 5% stacking and 15% separating polyacrylamide gel (PAAG), following established protocols. To visualize the distinct protein bands, the gels were stained with 0.4% Coomassie Brilliant Blue G-250 dye.

LegH amount was estimated by Coomassie-stained PAAG, using protein concentration standards. Relative proportions of target protein outcome after purification were calculated by measuring peak squares after size-exclusion chromatography.

2.6. ANN-Based Cell Biomass Soft Sensor Development

The ANN-based cell biomass soft sensor was developed in MATLAB R2021b, using the Neural Net Fitting toolbox. The cell biomass dataset (12,631 entries) generated from turbidity sensor measurements in Experiments 1 and 9, were used as the response variables.

Corresponding recorded process data of stirrer speed (RPM), dissolved oxygen (%), O_2 enrichment (%), pumped base (mL), glycerol feed (mL), methanol feed (mL), and reactor volume (L) were used as the predictor variables. Then, 70% of data were used for neural network training, 15% for testing, and 15% for validation. A two-layer feedforward network with 10 sigmoid hidden neurons and 1 linear output neuron, schematically illustrated in Figure 3, was selected for training.

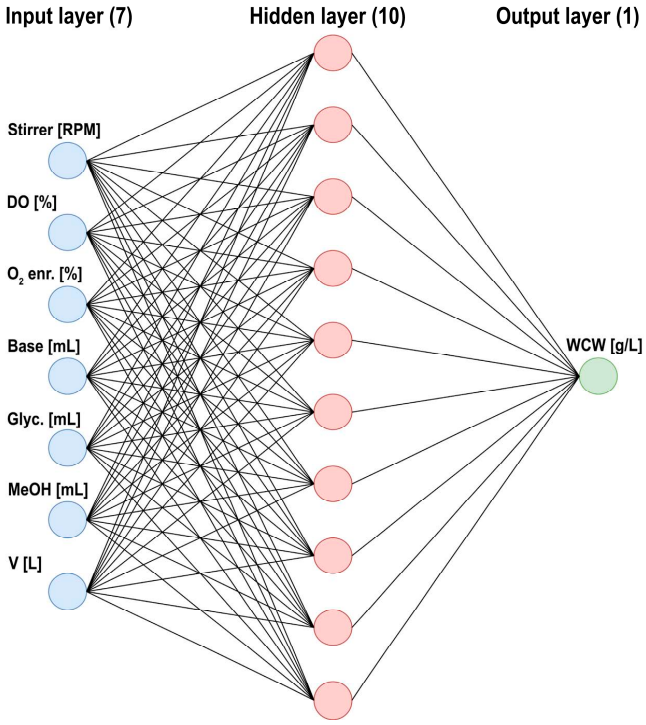


Figure 3. Structure of the developed ANN-based biomass soft sensor.

The network was trained using the Levenberg–Marquard training algorithm until 6 consecutive validation checks were failed. The model was then exported to MATLAB workspace and used for cell biomass concentration estimation.

Cultivation parameters are often influenced by external factors or signal noise; hence, signal filtering methods are popular in bioengineering. To reduce the sudden jumps and noise of the developed biomass soft sensor, a Savitzky–Golay filter was used with an order of 1 and frame length of 29. This filter significantly reduced sudden signal jumps, noise, and improved the overall performance of the sensor.

3. Results

3.1. Clone Selection

To select the best producer cells, eight zeocin-resistant clones were cultivated in flasks with rich BMGY medium, and LegH synthesis level was assessed three days post methanol induction. A product of predicted molecular mass was detectable for all clones compared (Figure 4A). Although LegH synthesis level is well detectable in all cases, it varies from clone to clone; however, the best production can be noted in clone No 3. Henceforth, this producer strain was selected for further investigation.

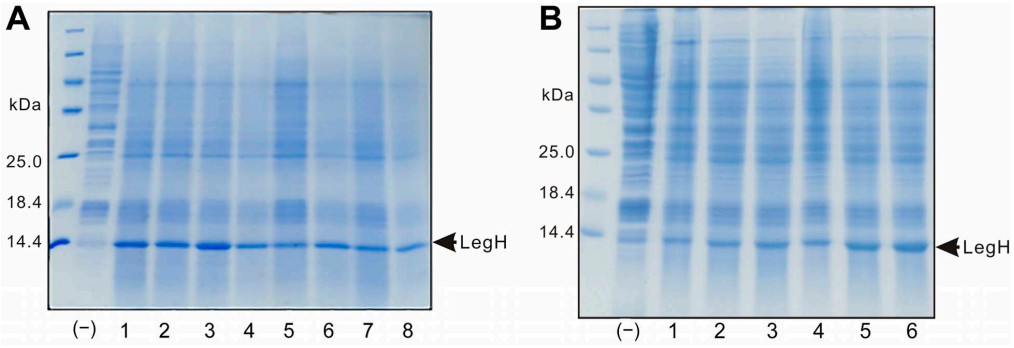


Figure 4. LegH synthesis level of eight selected zeocin-resistant clones (**A**) and clone No. 3 cultures induced at 24 °C (1–3) and 30 °C (4–6) (**B**) visualized by Coomassie-stained PAAG.

Flask experiments were also used to investigate the optimal LegH expression temperature. Inducing biosynthesis at a lowered temperature is a popular strategy to improve protein yield in some cases; therefore, two induction temperatures—24 and 30 °C were investigated. The Coomassie-stained PAAG from these experiments can be seen in (Figure 4B). Thicker LegH bands can be noted at an expression temperature of 30 °C; hence, this temperature was used during induction in consequent bioreactor experiments.

3.2. Cultivation Experiments

In order to establish a standardized bioreactor process, we cultured the chosen producer cells in Invitrogen's classical BSM medium five days after induction, and then monitored LegH synthesis levels at various time intervals using SDS–PAGE. For comparison, cultivation in complex BMGY medium was also carried out, according to the same protocol, in order to compare the productivity between minimal and complex media. The cultivation process parameters during these experiments are presented in Figure 5 and the LegH accumulation dynamics in post induction samples, visualized by Coomassie-stained PAAG, are shown in Figure 6.

According to the PAAG from Experiment 1 in Figure 6, it can be noted that the thickness of LegH band increases in the 7 h, 24 h, and 48 h samples post methanol induction. In the remaining samples, the increase is insignificant and difficult to observe. After performing sample purification, a LegH concentration of 1.56 mg/g wet cells is reached after 48 h on methanol. Although, the maxima of synthesis (1.62 mg/g) was reached on the fifth day of post methanol induction (120 h sample), the increase in specific product yield was only gradual. Therefore, the time point of 48 h after methanol induction was used in further experiments to compare the efficiency of LegH biosynthesis in different reported cultivation media.

The cultivation in rich BMGY medium was carried out, according to the same Invitrogen cultivation protocol and continued for 48 h after methanol induction. In this cultivation, an even higher yield of LegH—1.77 mg/g wet cells was achieved, indicating that the cultivation medium might have a significant effect on LegH productivity. Yet, employing complex (rich) medium in cultivations does present notable drawbacks, particularly on an industrial scale. These include diminished batch-to-batch repeatability stemming from variations in component composition, increased costs, and challenges in product purification. Additionally, the inclusion of meat peptone in the medium formulation raises ethical concerns, given LegH's primary use in vegan nutrition. Given the well-documented cultivation of *P. pastoris* on minimal media, our focus shifted exclusively to investigating minimal media.

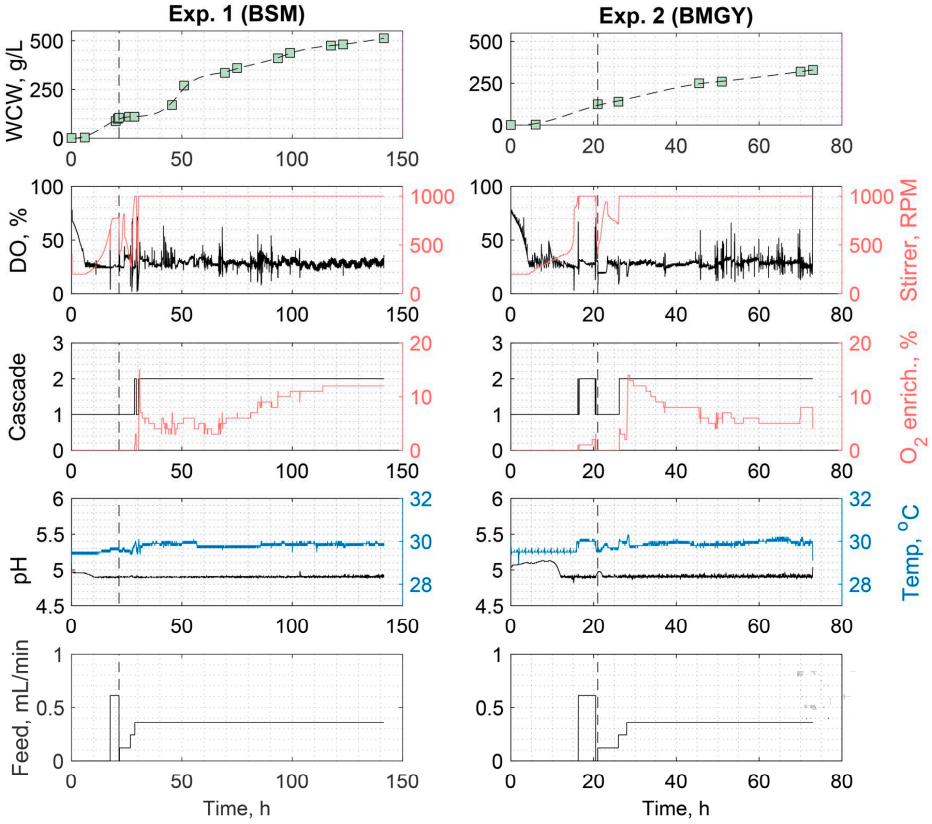


Figure 5. Cultivation parameters during *P. pastoris* X-33 LegH production processes in BSM (Experiment 1) and BMGY (Experiment 2) media. Green squares indicate experimental WCW measurements. Black/colored lines correspond to their respective axes. Vertical dashed line indicates the onset of methanol induction.

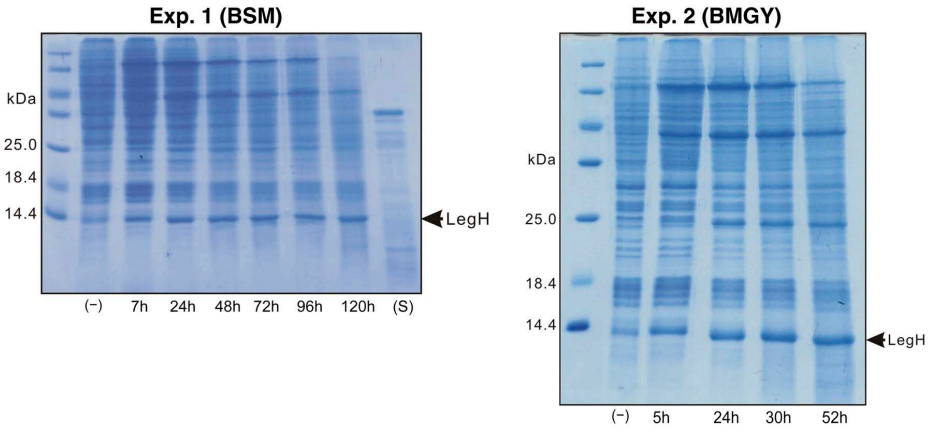


Figure 6. LegH accumulation dynamics in samples taken post methanol induction, visualized by Coomassie-stained PAAG. Negative control samples taken shortly before methanol induction. Supernatant (S) sample taken at 120 h.

3.3. Purification

Purification of LegH was processed in three steps. In the first step, excess salt was removed and a pH of 8.0 was established. In the second step, protein was attached to an anion-exchange matrix and eluted by increasing the amount of salt, resulting in removal of major contaminants (Figure 7A). Then, 1 mL of four major fractions, corresponding to the LegH peak, were taken for further purification and analysis. For final polishing, the four fractions were merged and the protein was passed through a size-exclusion chromatography column, which indicated that the majority of the protein is eluted in a monomeric state according to its molecular weight (Figure 7B). Moreover, Superdex column peak fractions, in contrast to the anion Q HP fractions, represented at least 90% pure LegH protein (Figure 7C), which allowed to further use the square of this peak for quantification of target protein and comparison of different cultivation processes. Attachment of the heme group to LegH is proven by the characteristic reddish color of peak protein fractions from the Superdex column (Figure 7D).

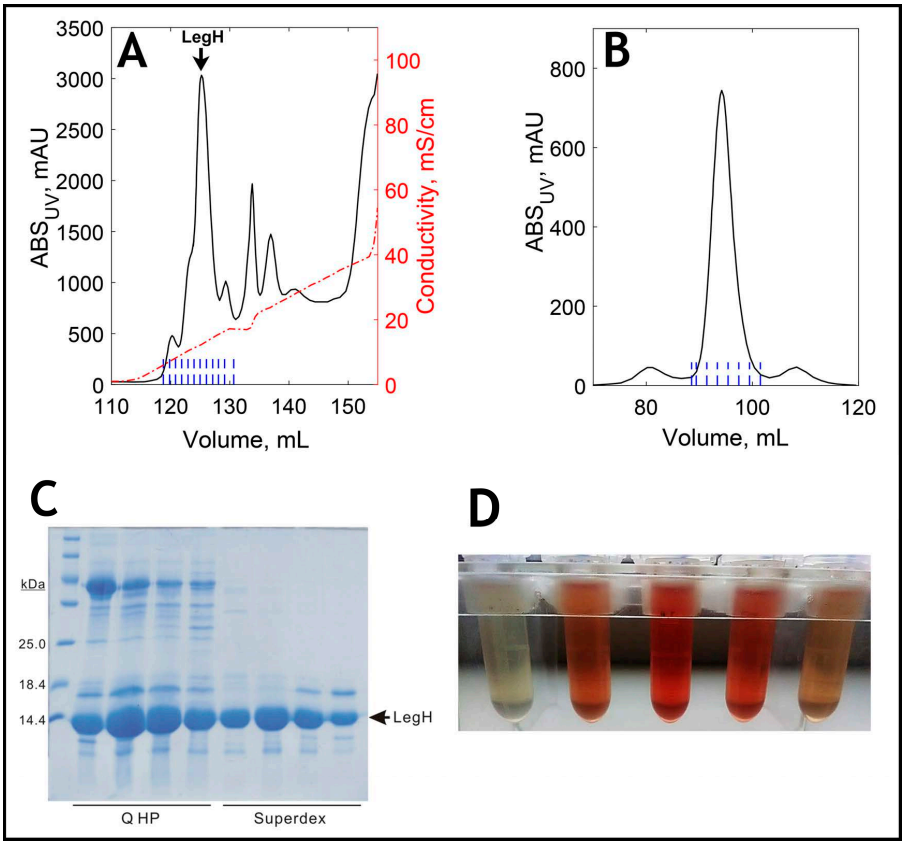


Figure 7. Purification of LegH. (**A**) Anion Q HP chromatography profile. (**B**) Size-exclusion Superdex 200 chromatography profile. (**C**) Coomassie-stained PAAG with both chromatography peak protein fractions containing LegH. (**D**) Peak fractions from size-exclusion chromatography column, demonstrating the characteristic reddish color of LegH. Blue vertical dashed lines indicate sampled fractions.

3.4. Reported Cultivation Medium Evaluation

The choice of cultivation medium holds considerable importance in bioprocess development. To explore whether the yield of LegH is impacted by the cultivation media

employed, cultivations were conducted in reported FM22, D'Anjou, BSM/2, and RDM media under uniform conditions, following the Invitrogen protocol. The cultivation parameter dynamics during these cultivation experiments are shown in Figure A1 (Appendix A).

The purified LegH results from the six cultivation experiments are compiled in Table 1 and visualized in Figure 8. The results indicate that the highest LegH productivity was achieved in BMGY medium; however, a slightly lower, but similar yield was noted in BSM. Out of the reported media formulations, the best performance was shown by BSM/2 medium (BSM salt concentrations reduced two times). However, the yields in the reported media were almost two times lower than in BSM or BMGY.

Table 1. LegH yields 2 days post methanol induction.

Exp.	Media	LegH Peak Square, u	LegH Yield, mg/g WCW	WCW, g/L	LegH Yield, mg/L
1	BSM	3500	1.56	345	537.2
2	BMGY	3991	1.77	339	600.5
3	FM22	1824	0.81	370	301.3
4	D'Anjou	1705	0.76	269	203.7
5	BSM/2	2170	0.96	382	365.6
6	RDM	1800	0.80	375	300.0

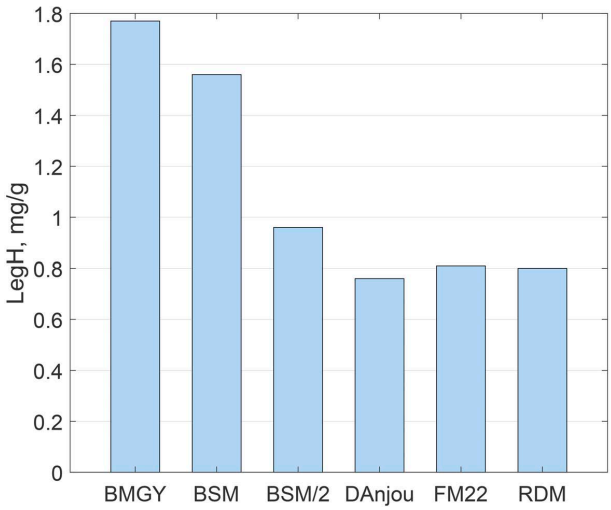


Figure 8. Relative amounts of LegH obtained in different cultivation processes 2 days post methanol induction.

3.5. Experiments to Improve LegH Expression

Since experiments with different cultivation media did not achieve an increased LegH yield, we decided to investigate, whether supplementing the BSM medium with 1 g/L glycine (Experiment 7) or a vitamin solution (Experiment 8) would have an effect on product yield. Additionally, a μ-stat feeding profile (Experiment 9) and mixed substrate (0.5 C-mol methanol/0.5 C-mol sorbitol) feed (Experiment 10) were investigated.

Glycine is an amino acid involved in the heme biosynthesis pathway. In a recent paper, we also hypothesized that upregulating the C1 metabolism pathway in mitochondria to increase glycine synthesis is necessary for improved heme biosynthesis [38]. Vitamin addition to cultivation media has been shown to improve recombinant product yield in

some cases [24,39]. To investigate the effect of vitamin addition, we supplemented the BSM medium with the vitamin solution used in the RDM medium formulation.

Inducing recombinant product biosynthesis by methanol mixed feed induction with sorbitol, is a popular strategy to improve recombinant protein yields. To investigate the effect that a mixed substrate feed has on LegH production, a cultivation process was carried out.

According to the popular Luedeking–Piret model, the protein production rate has an empirical relationship with the cell growth rate. There are many reports in the literature of a positive correlation between the specific cell growth rate (μ) and specific target protein production rate (q). To investigate, whether this correlation is also true for LegH, we conducted an experiment, in which we attempted to control the specific cell growth rate of *P. pastoris*, by manipulating the substrate (methanol) feed during the induction phase.

The cultivation parameters from the aforementioned experiments are presented in Figure A2 (Appendix A) and LegH yields in Table 2.

Table 2. LegH yields 2 days post methanol induction for Experiments 7-10.

Exp.	Media	LegH Peak Square, u	LegH Yield, mg/g WCW	WCW, g/L	LegH Yield, mg/L
7	BSM + Gly	2400	1.05	390	409.5
8	BSM + Vit.	2800	1.23	370	455.1
9	BSM (μ-stat)	2580	1.13	342	387.9
10	BSM (Sorb)	2160	0.95	370	350.8

The BSM medium supplementation with 1 g/L glycine did not yield a positive effect on the LegH synthesis level. The achieved LegH yield of 1.05 mg/g WCW was slightly lower than in Experiment 1; however, a higher cell concentration was achieved.

In the experiment where BSM was enriched with vitamins, we observed a reduced lag phase and a quicker adjustment to methanol uptake. These effects can be attributed to the presence of crucial vitamins that facilitate yeast metabolism. However, no increase in LegH productivity could be noted in this experiment.

In Experiment 9, we attempted to control the specific cell growth rate (μ) at 0.02 h^{-1} post methanol induction, by varying the methanol feed rate using a PID algorithm-based controller. Soon after initiating μ-stat control, it was noted that the increased feed rate caused a significant increase in metabolic heat production, as the fermentation temperature began to rise. The bioreactor cooling system was unable to maintain the temperature at 30 °C; therefore, the maximum feed rate was adjusted so that the temperature would not exceed 32 °C, which can be detrimental to recombinant protein biosynthesis. Although this restriction led to a lower average specific growth rate of 0.015 h^{-1}, it was still higher than in the typical BSM process (approx. 0.006–0.008 h^{-1}). Based on the findings, no enhancement in LegH productivity was observed in this experiment.

Finally, an experiment (Experiment 10) with mixed substrate induction was carried out. Methanol solution was supplemented with sorbitol at a ratio of 0.5 C-mol methanol/0.5 C-mol sorbitol. Induction was carried out, according to the Invitrogen protocol. This experiment revealed a rapid adjustment to methanol uptake, a well-known occurrence in mixed substrate induction with sorbitol. Although, a higher cell biomass concentration could be noted at the end of the process, the specific LegH productivity was reduced, perhaps due to the lower fraction of methanol in the feed solution.

3.6. ANN-Based Cell Biomass Soft Sensor

In Experiments 1 and 9, we employed an *in situ* turbidity probe to monitor the real-time growth of *P. pastoris* cell biomass. This monitoring process produced a substantial dataset, comprising 12,631 entries. This dataset served as the foundation for developing a neural network-based soft sensor for estimating cell biomass. To ensure that no additional

expensive sensors are necessary, we exclusively utilized parameters directly measured by the bioreactor system, which included stirrer speed, dissolved oxygen, oxygen enrichment, base pump, feed pump, and reactor volume.

The datasets from previous experiments were used to test and validate the created ANN model. The model was used to calculate WCW values, based on the input data recorded in the experiments. These values were then compared to their corresponding experimentally measured WCW values to determine the model accuracy. The dataset generated by the ANN-based soft sensor and experimental measurements is illustrated in Figure 9.

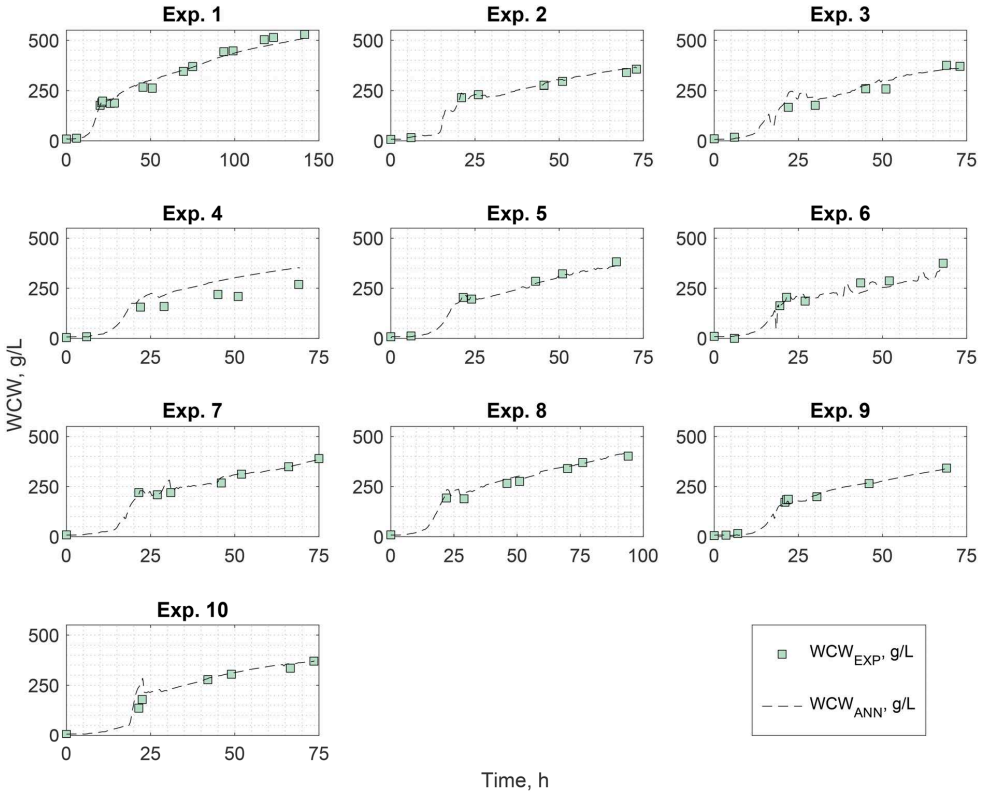


Figure 9. *P. pastoris* cell biomass during cultivation experiments, modeled by the ANN-based soft sensor.

As we can see, the developed soft sensor is able to accurately describe cell biomass dynamics in the selected cultivations. A good fit can be noted in almost all experiments.

In Experiment 4, the sensor fails to follow the biomass trajectory, as it overestimates the biomass concentration post induction. This most likely can be explained by the D'Anjou medium used in this experiment, as it significantly differs from other cultivation media. Also, significantly lower cell biomass measurements were registered in the particular cultivation experiment. Taking this into consideration, we can surmise that the developed soft sensor is not entirely applicable for cultivations in these media.

In Experiment 10, it is observed that at the beginning of methanol induction, the soft sensor tends to overestimate the cell biomass concentration. This could be attributed to the influence of sorbitol when co-fed with methanol. The transition phase to methanol uptake

is expedited, and cell metabolism resumes more rapidly than usual. Consequently, the sensor overestimates the presumed adapting cell biomass concentration.

To investigate the performance of the ANN-based soft sensor, RMSE and NRMSE values for each experiment and overall accuracy are compiled in Table 3.

Table 3. ANN-based cell biomass soft sensor accuracy.

Exp.	RMSE	NRMSE
1	14.78	2.84%
2	8.61	2.46%
3	24.44	6.69%
4 *	50.61	19.10%
5	10.66	2.85%
6	16.45	4.38%
7	9.33	2.45%
8	15.35	5.92%
9	7.28	2.16%
10 *	33.95	9.30%
Overall	13.36	3.72%

* Experiments omitted in overall accuracy calculation.

The overall precision of the developed ANN-based soft sensor for cell biomass estimation, is evaluated at $\pm$ 13.36 g/L WCW or 3.72%. Considering that the cultivation medium can be a significant factor in this case (e.g., Experiment 4), and that eight different media were employed in 10 performed cultivations, it speaks to the robustness of the developed sensor. Another factor that must be considered, is that the sensor does not use any additional sensor signals (e.g., CO_2 measurement), which, although, may reduce sensor accuracy, does not require the purchase of additional sensor systems. Overall, the sensor accuracy can be deemed as sufficient for application in recombinant *P. pastoris* cultivations.

4. Discussion

In this research, we investigated several reported cultivation media for recombinant LegH production with the yeast *P. pastoris*. For improved results interpretation, we estimated the elemental composition of each cultivation medium and compared the respective concentrations with the so-called Wegner ranges (Table A2, Appendix A).

The highest LegH yield was achieved in rich BMGY medium. Rich medium is known for improved cell growth, as the cells do not need to synthesize all of the required metabolic intermediates, as is the case in minimal media. Frequently, this impact can result in enhanced yields of recombinant proteins, as the Luedeking–Piret model suggests that, in numerous instances, the cell growth rate can be directly related to recombinant protein production. However, there are several drawbacks to using a complex cultivation medium, for example, increased costs, composition variability, and hindered purification. Also, it should be noted that the use of meat peptone in BMGY formulation for the production of a product mainly used as a vegan food supplement could be considered controversial. Substituting meat peptone with, for example, soy peptone could be a viable alternative; however, the changes in ingredient composition can not only result in reduced product yields, but potentially necessitate adjustments in the cultivation process itself.

The second best result was achieved, when cultivation was carried out in standard BSM medium, as the yields were comparable to those achieved in BMGY medium. As the elemental composition of BSM shows, it contains, per Wegner, all of the necessary elements for *P. pastoris* growth, most of them—even in excess of the preferred range. This, in part, accounts for the precipitation problems noted in BSM, as well as the increased osmotic

pressure that is considered as a stress factor on the cells [40]. However, as a reduced salt concentration (Experiment 5 (BSM/2)) did not result in an improved LegH yield, elevated osmotic pressure is probably not severely hindering LegH expression.

Considering the reported media formulations (FM22, D'Anjou, and RDM), an underwhelming LegH yield was achieved. The FM22 and RDM media are fairly similar to BSM; however, both have lower elemental concentrations in most cases. RDM is also supplemented with a mixture of vitamins suitable for yeast cultivation. Although, some salt precipitation was noted, when preparing these media, it was not on the same scale as with BSM. The D'Anjou medium has the least salts of any other media tested; hence, no precipitation was noted. The low LegH yield achieved in these media is somewhat perplexing. Considering the elemental compositions of the cultivation media, perhaps some of the excess elements in BSM amounted to an increased LegH expression.

The investigated addition of glycine or vitamin solution to BSM, did not yield any significant improvement to LegH yield. In both experiments, a lower specific productivity was achieved, even though a higher cell biomass concentration was recorded. We also noted that the glycine addition promoted excessive salt precipitation in this experiment. The addition of vitamins, did positively impact cell growth, as some authors have reported [24]; however, it did not have a positive effect on LegH production.

As testing different cultivation media formulations did not result in an improved LegH yield, we decided to test two of the more popular *P. pastoris* feed strategies—μ-stat and mixed feed induction with sorbitol. Unfortunately, neither of these strategies produced improved results. In both cases, LegH yield was lower than in a standard BSM cultivation, according to the Invitrogen protocol.

Methanol acts both as a growth substrate and product synthesis inducer in *P. pastoris* cultivations. Increased methanol feed (Experiment 9), however, did not amount to a higher LegH synthesis level, indicating that some bottleneck, probably in the heme biosynthesis pathway, may be present and consequently, limit LegH synthesis. Although *P. pastoris* has been defined as a GRAS (Generally Regarded As Safe) microorganism, some concerns, regarding the toxicity of residual methanol may arise for recombinant product use in food applications and promote the consideration of other promoters for biosynthesis induction, such as the galactose-induced *LAC4* promoter in *Kluyveromyces lactis* expression vector. However, these concerns are offset by the several studies that investigated the toxicity and allergenicity of LegH produced by Impossible foods Inc. and found no significant risks [4–6].

Taking into consideration the results from the previous experiments, we can conclude that process-specific optimization strategies did not have a positive impact on LegH yield, as the best result was achieved in the "unoptimized" cultivation, according to the Invitrogen guidelines. The results suggest that LegH expression in this particular case is most likely not hindered by expression conditions, but for strain-specific reasons. Strain engineering of *P. pastoris* is likely the key to improved LegH production, as clearly illustrated by the research of Shao et al. [13]. In a recent article, we also developed a metabolic model for *P. pastoris* LegH production, suggesting the reactions to up-/downregulate with the most potential for improved LegH production [38].

An efficient purification procedure was developed to ensure a purity level of at least 90% for the expressed LegH. Although three chromatography columns are involved in this method, the overall process takes less than one day. No detectable losses of target protein were observed during purification. The quality of the purified LegH was confirmed by PAAG. The inclusion of the heme group to LegH is proven by the characteristic reddish color of peak protein fractions.

For this method, a 7.0 g wet cell portion was chosen, considering the volume limitations of the utilized French press for cell disruption. Theoretically, this press currently acts as the bottleneck for the purification method. Scaling up cell lysis would enable the expansion of the purification process and the utilization of larger chromatography columns, thereby purifying a significantly greater quantity of LegH. However, given that the primary focus

of this research was to examine LegH production at the laboratory scale, we consider this purification method adequate for the stated objective.

Shao et al. used an Ni–NTA agarose column to purify and Amicon Ultra 3 K centrifugal filter units to desalt the LegH secreted in the culture medium [13]. Both Ni agarose and the Amicon filters are very expensive. The filtration procedure in this case is also volume limited and time intensive. Overall, this method is not suitable for large-scale production. Impossible foods Inc., on the other hand, did not employ chromatography at all in the purification of their product [7]. Insoluble material was removed by centrifugation and microfiltration. Then, ultrafiltration was used to concentrate the LegH, resulting in the end product purity of ~80%. Their approach was to identify all of the remaining contaminants in the product and to assess their toxicity and allergic properties. Although, this would be more time consuming and expensive at first, this approach is more suited to the large-scale commercial production of LegH.

Generated process data were used to establish and train a neural network model for cell biomass estimation. Several similar models have been previously reported [41,42]; however, the novelty of our approach is based on the absence of external sensor signals. Both reported examples utilize the CO_2 measurement signal, which requires an additional exhaust gas analyzer. However, our sensor is able to estimate cell biomass concentration, with sufficient precision, by only using standard real-time measurements by the bioreactor system itself.

The soft sensor can be used in cultivation processes in real time to estimate cell biomass concentration—one of the more important process parameters. Perhaps, it is most suited particularly for fed-batch fermentations, as feed rate profile calculation often requires precise and rapid biomass measurements. The inclusion of real-time cell biomass estimation can also benefit several advanced bioprocess control strategies, such as PID or model predictive (MPC) controllers [29].

5. Conclusions

In this research, we investigated recombinant *P. pastoris* LegH expression in several reported cultivation media (BSM, BMGY, FM22, D'Anjou, BSM/2, and RDM) and under different feeding strategies (μ-stat and mixed feed with sorbitol). Our results suggest that process-specific optimization techniques did not result in increased LegH yields; hence, strain-specific strategies should be investigated. Generated process data were used to establish and train a novel artificial neural network-based soft sensor for cell biomass estimation, utilizing only standard bioreactor measurements (stirrer speed, dissolved oxygen, O_2 enrichment, base feed, glycerol feed, methanol feed, and reactor volume). The developed soft sensor was robust and showed a good fit with experimental data (3.72% WCW).

Author Contributions: Conceptualization, E.B. and A.K.; methodology, E.B. and A.K.; formal analysis, E.B. and A.K.; investigation, E.B., J.B., I.A., T.K., and A.K.; resources, A.K.; writing—original draft preparation, E.B.; writing—review and editing, E.B. and A.K.; visualization, E.B.; supervision, A.K.; funding acquisition, A.K. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the European Regional Development Fund (ERDF) project No. 1.1.1.1/21/A/044 and by the European Social Fund within Project No. 8.2.2.0/20/I/008 «Strengthening of PhD students and academic personnel of Riga Technical University and BA School of Business and Finance in the strategic fields of specialization» of the Specific Objective 8.2.2 «To Strengthen Academic Staff of Higher Education Institutions in Strategic Specialization Areas» of the Operational Programme «Growth and Employment».

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. The compositions of explored cultivation media.

Component	Medium					
	BSM	**FM22**	**D'Anjou**	**BSM/2**	**RDM**	**BMGY**
Macro elements (g/L)						
Glycerol	40.00	40.00	50.00	40.00	40.00	40.00
H_3PO_4	45.66			22.83		
$CaSO_4$	0.93	1.00		0.47		
K_2SO_4	18.20	14.30		9.10		
$MgSO_4 \times 7H_2O$	14.90	11.70	4.70	7.45	4.70	
KOH	4.13	pH to 4.5	pH to 5.5	2.07	3.37	
$CaCl_2 \times 2H_2O$			0.36		0.36	
$(NH_4)_2SO_4$		5.00	20.00		1.65	
K_2HPO_4						2.30
KH_2PO_4		42.90	12.00		12.00	11.81
Glutamine					1.74	
Arginine					1.46	
Yeast Extract						10.00
Peptone						20.00
Yeast Nitrogen Base						13.40
Trace elements (mg/L)						
$CuSO_4 \times 5H_2O$	26.10	8.70		26.10	26.10	
NaI	0.35	0.35		0.35	0.35	
$MnSO_4 \times H_2O$	13.05	13.05	0.08	13.05	13.05	
$Na_2MoO_4 \times 2H_2O$	0.87	0.87	0.48	0.87	0.87	
$CoCl_2 \times 6H_2O$	4.00	4.00		4.00	4.00	
H_3BO_3	0.09	0.09	0.05	0.09	0.09	
$ZnCl_2 \times 2H_2O$	109.97	38.49		109.97	109.97	
$FeSO_4 \times 7H_2O$	282.75	94.25		282.75	282.75	
Biotin	0.87	0.87		0.87	0.87	
H_2SO_4	0.022 mL	0.004 mL		0.022 mL	0.022 mL	
$CaSO_4 \times 5H_2O$		4.00	0.05			
KI			0.21			
$ZnSO_4 \times 7H_2O$			5.03			
$FeCl_3 \times 6H_2O$			12.00			
Vitamins (mg/L)						
Thiamine HCl (B1)					1.00	
Nicotinic acid (B3)					1.00	
Ca pantothenate (B5)					1.00	
Pyridoxine HCl (B6)					1.00	
Biotin (B7)					0.05	0.40
Myo-inositol (B8)					25.00	
p-aminobenzoic acid (B10)					0.20	

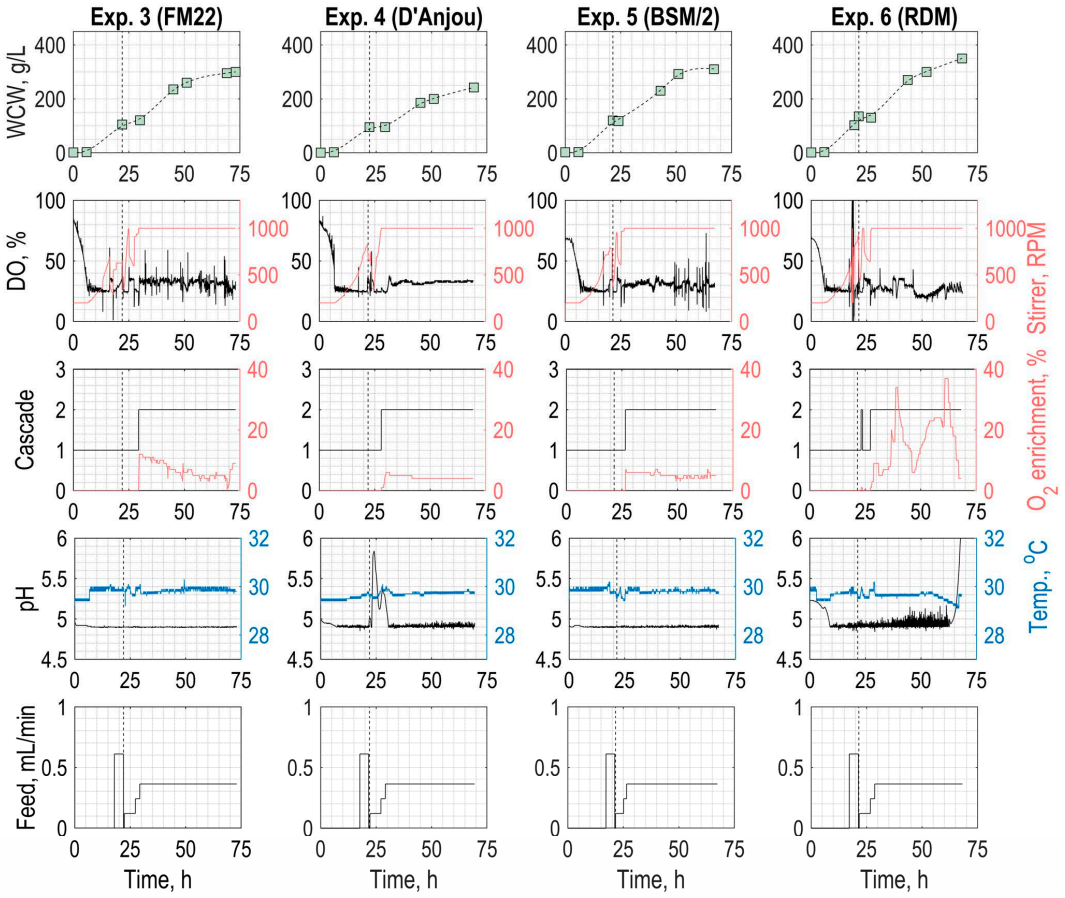


Figure A1. Cultivation parameters during *P. pastoris* LegH production processes in four reported cultivation media. Green squares indicate experimental WCW measurements. Black/colored lines correspond to their respective axes. Vertical dashed line indicates the onset of the methanol induction phase.

Table A2. Elemental compositions of reported *P. pastoris* cultivation media.

	Wegner			BSM	BMGY *	FM22	D'Anjou	BSM/2	RDM
	Min	Broad	Preferred						
Macro elements (g/L)									
N	-	-	-	NH_4OH	5.72 + NH_4OH	1.06 + NH_4OH	4.24 + NH_4OH	NH_4OH	1.15 + NH_4OH
P	1.90	2.9–20.0	2.2–10.0	12.27	4.44	9.76	2.73	6.13	2.73
K	1.00	1.0–20.0	1.5–10.0	11.05	5.75	18.74	3.45	5.52	5.80
Mg	0.15	0.15–3.0	0.3–1.2	1.47	0.22	1.15	0.46	0.73	0.46
Ca	0.006	0.006–1.60	0.08–0.8	0.27	0.08	0.29	0.10	0.14	0.10
S	0.10	0.1–8.0	0.2–5.0	5.56	2.80	5.61	5.47	2.80	1.06

Table A2. *Cont.*

	Wegner			BSM	BMGY *	FM22	D'Anjou	BSM/2	RDM
	Min	Broad	Preferred						
Trace elements (mg/L)									
Fe	6.0	6.0–140.0	9.0–80.0	56.80	0.90	8.84	2.48	56.80	56.80
Zn	2.0	2.0–100.0	3.0–40.0	41.72	2.16	5.31	1.14	41.72	41.72
Cu	0.6	0.6–16.0	1.0–10.0	6.64	0.04	1.02	-	6.64	6.64
Mn	0.6	0.6–20.0	0.9–8.0	4.24	0.31	1.95	0.02	4.24	4.24
Na	-	-	-	0.24	811.82	0.08	0.10	0.24	0.24
Co	-	-	-	0.99	-	0.46	-	0.99	0.99
B	-	-	-	0.02	0.17	0.01	0.01	0.02	0.02
I	-	-	-	0.29	0.15	0.14	0.16	0.29	0.29
Cl	-	-	-	46.44	249.42	6.31	178.35	46.44	218.88
Mo	-	-	-	0.38	0.19	0.18	0.21	0.38	0.38

* The composition varies for a complex medium and may not be entirely accurate. Estimation based on reported compositions for YNB [43], peptone, and yeast extract [44].

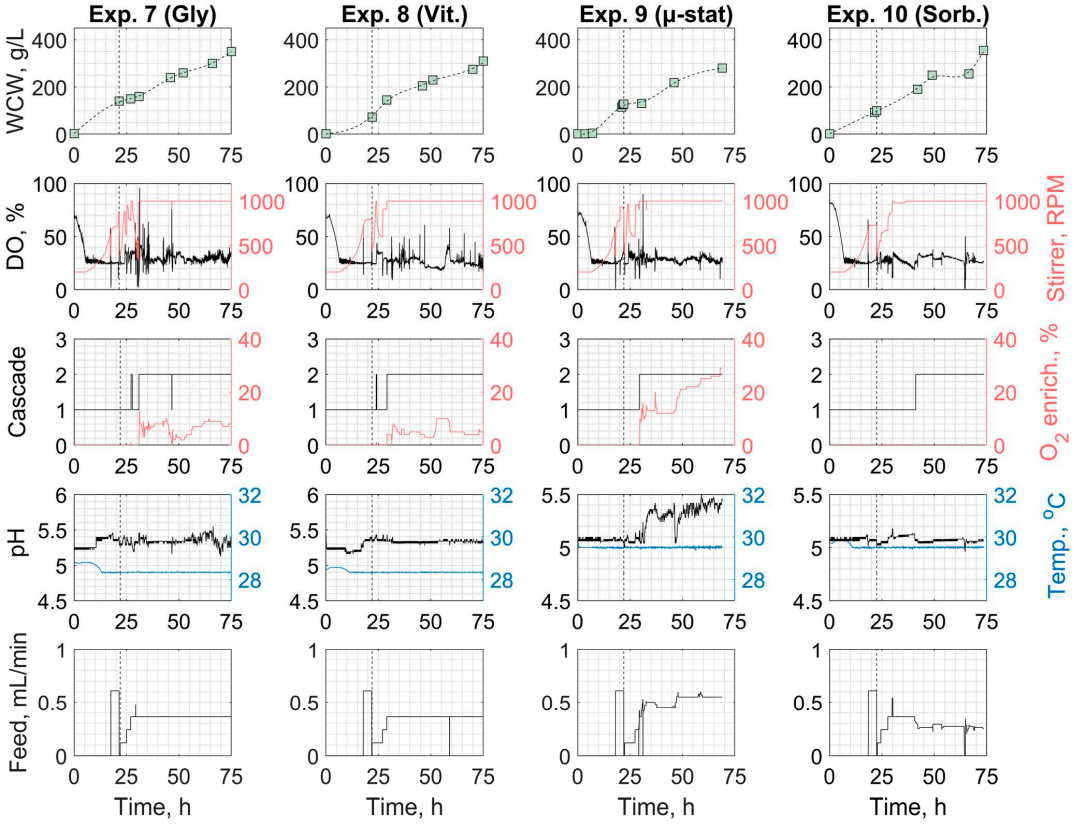


Figure A2. Cultivation parameters from *P. pastoris* LegH production processes. Green squares indicate experimental WCW measurements. Black/colored lines correspond to their respective axes. Vertical dashed line indicates the onset of the methanol induction phase.

References

1. Ahmad, M.; Qureshi, S.; Akbar, M.H.; Siddiqui, S.A.; Gani, A.; Mushtaq, M.; Hassan, I.; Dhull, S.B. Plant-Based Meat Alternatives: Compositional Analysis, Current Development and Challenges. *Appl. Food Res.* **2022**, *2*, 100154. [CrossRef]
2. Andreani, G.; Sogari, G.; Marti, A.; Froldi, F.; Dagevos, H.; Martini, D. Plant-Based Meat Alternatives: Technological, Nutritional, Environmental, Market, and Social Challenges and Opportunities. *Nutrients* **2023**, *15*, 452. [CrossRef] [PubMed]
3. Singh, A.; Sit, N. Meat Analogues: Types, Methods of Production and Their Effect on Attributes of Developed Meat Analogues. *Food Bioprocess Technol.* **2022**, *15*, 2664–2682. [CrossRef]
4. Reyes, T.F.; Chen, Y.; Fraser, R.Z.; Chan, T.; Li, X. Assessment of the Potential Allergenicity and Toxicity of *Pichia* Proteins in a Novel Leghemoglobin Preparation. *Regul. Toxicol. Pharmacol.* **2021**, *119*, 104817. [CrossRef]
5. Jin, Y.; He, X.; Andoh-Kumi, K.; Fraser, R.Z.; Lu, M.; Goodman, R.E. Evaluating Potential Risks of Food Allergy and Toxicity of Soy Leghemoglobin Expressed in *Pichia Pastoris*. *Mol. Nutr. Food Res.* **2018**, *62*, 1700297. [CrossRef]
6. Fraser, R.Z.; Shitut, M.; Agrawal, P.; Mendes, O.; Klapholz, S. Safety Evaluation of Soy Leghemoglobin Protein Preparation Derived From *Pichia Pastoris*, Intended for Use as a Flavor Catalyst in Plant-Based Meat. *Int. J. Toxicol.* **2018**, *37*, 241–262. [CrossRef] [PubMed]
7. *GRAS Notification No. 737-Soy Leghemoglobin Preparation from a Strain of Pichia Pastoris*; Impossible Foods Inc.: Redwood City, CA, USA, 2016; p. 1063.
8. Risk and Technical Assessment Report—Application A1186: Soy Leghemoglobin in Meat Analogue Products. 2020; p. 1. Available online: https://www.foodstandards.gov.au/code/applications/Documents/A1186%20Approval%20Report.pdf (accessed on 11 November 2023).
9. Vainshtein, B.K.; Harutyunyan, E.H.; Kuranova, I.P.; Borisov, V.V.; Sosfenov, N.I.; Pavlovsky, A.G.; Grebenko, A.I.; Konareva, N.V. Structure of Leghaemoglobin from Lupin Root Nodules at 5 Å Resolution. *Nature* **1975**, *254*, 163–164. [CrossRef] [PubMed]
10. Bogusz, D.; Appleby, C.A.; Landsmann, J.; Dennis, E.S.; Trinick, M.J.; Peacock, W.J. Functioning Haemoglobin Genes in Non-Nodulating Plants. *Nature* **1988**, *331*, 178–180. [CrossRef]
11. Hargrove, M.S.; Barry, J.K.; Brucker, E.A.; Berry, M.B.; Phillips, G.N.; Olson, J.S.; Arredondo-Peter, R.; Dean, J.M.; Klucas, R.V.; Sarath, G. Characterization of Recombinant Soybean Leghemoglobin a and Apolar Distal Histidine Mutants. *J. Mol. Biol.* **1997**, *266*, 1032–1042. [CrossRef]
12. Brucker, E.A.; Hargrove, M.S.; Phillips, G.N., Jr. Leghemoglobin A (Acetomet). 1997. Available online: https://www.rcsb.org/ (accessed on 10 November 2023).
13. Shao, Y.; Xue, C.; Liu, W.; Zuo, S.; Wei, P.; Huang, L.; Lian, J.; Xu, Z. High-Level Secretory Production of Leghemoglobin in *Pichia Pastoris* through Enhanced Globin Expression and Heme Biosynthesis. *Bioresour. Technol.* **2022**, *363*, 127884. [CrossRef]
14. Xue, J.; Zhou, J.; Li, J.; Du, G.; Chen, J.; Wang, M.; Zhao, X. Systematic Engineering of *Saccharomyces Cerevisiae* for Efficient Synthesis of Hemoglobins and Myoglobins. *Bioresour. Technol.* **2023**, *370*, 128556. [CrossRef]
15. Kosmachevskaya, O.V.; Nasybullina, E.I.; Shumaev, K.B.; Topunov, A.F. Expressed Soybean Leghemoglobin: Effect on *Escherichia coli* at Oxidative and Nitrosative Stress. *Molecules* **2021**, *26*, 7207. [CrossRef] [PubMed]
16. Jones, K.D.; Badii, R.; Rosell, I.F.; Lloyd, E. Bacterial Expression and Spectroscopic Characterization of Soybean Leghaemoglobin A. *Biochem. J.* **1998**, *330*, 983–988. [CrossRef] [PubMed]
17. Cregg, J.M. Recombinant Protein Expression in *Pichia Pastoris*. *Appl. Biochem. Biotechnol. Part B Mol. Biotechnol.* **2000**, *16*, 23–52. [CrossRef] [PubMed]
18. Macauley-Patrick, S.; Fazenda, M.L.; McNeil, B.; Harvey, L.M. Heterologous Protein Production Using the *Pichia Pastoris* Expression System. *Yeast* **2005**, *22*, 249–270. [CrossRef] [PubMed]
19. Invitrogen Corporation Pichia Fermentation Process Guidelines. Available online: https://tools.thermofisher.com/content/sfs/manuals/pichiaferm_prot.pdf (accessed on 26 October 2023).
20. Looser, V.; Bruhlmann, B.; Bumbak, F.; Stenger, C.; Costa, M.; Camattari, A.; Fotiadis, D.; Kovar, K. Cultivation Strategies to Enhance Productivity of *Pichia Pastoris*: A Review. *Biotechnol. Adv.* **2014**, *33*, 1177–1193. [CrossRef] [PubMed]
21. Cos, O.; Ramón, R.; Montesinos, J.L.; Valero, F. Operational Strategies, Monitoring and Control of Heterologous Protein Production in the Methylotrophic Yeast *Pichia Pastoris* under Different Promoters: A Review. *Microb. Cell Fact.* **2006**, *5*, 17. [CrossRef] [PubMed]
22. Stratton, J.; Chiruvolu, V.; Meagher, M. High Cell-Density Fermentation. *Methods Mol. Biol.* **1998**, *103*, 107–120. [CrossRef]
23. D'Anjou, M.C.; Daugulis, A.J. Mixed-Feed Exponential Feeding for Fed-Batch Culture of Recombinant Methylotrophic Yeast. *Biotechnol. Lett.* **2000**, *22*, 341–346. [CrossRef]
24. Matthews, C.B.; Kuo, A.; Love, K.R.; Love, J.C. Development of a General Defined Medium for *Pichia Pastoris*. *Biotechnol. Bioeng.* **2018**, *115*, 103–113. [CrossRef]
25. Pais-Chanfrau, J.M.; Trujillo-Toledo, L.E. Optimization of Culture Medium for Large-Scale Production of Heterologous Proteins in *Pichia Pastoris* to Be Used in Nanoscience and Other Biotechnological Fields. *Biol. Med.* **2016**, *8*, 279. [CrossRef]
26. Zhu, W.; Xu, R.; Gong, G.; Xu, L.; Hu, Y.; Xie, L. Medium Optimization for High Yield Production of Human Serum Albumin in *Pichia Pastoris* and Its Efficient Purification. *Protein Expr. Purif.* **2021**, *181*, 105831. [CrossRef] [PubMed]
27. Brady, C.P.; Shimp, R.L.; Miles, A.P.; Whitmore, M.; Stowers, A.W. High-Level Production and Purification of P30P2MSP119 an Important Vaccine Antigen for Malaria, Expressed in the Methylotropic Yeast *Pichia Pastoris*. *Protein Expr. Purif.* **2001**, *23*, 468–475. [CrossRef] [PubMed]

28. Wegner, E.H. Biochemical Conversions by Yeast Fermentation at High Cell Densities. U.S. Patent No 4,414,329, 8 November 1983.
29. Bolmanis, E.; Dubencovs, K.; Suleiko, A.; Vanags, J. Model Predictive Control—A Stand Out among Competitors for Fed-Batch Fermentation Improvement. *Fermentation* **2023**, *9*, 206. [CrossRef]
30. Beiroti, A.; Hosseini, S.N.; Aghasadeghi, M.R.; Norouzian, D. Comparative Study of μ-stat Methanol Feeding Control in Fed-batch Fermentation of *Pichia Pastoris* Producing HBsAg: An Open-loop Control versus Recurrent Artificial Neural Network-based Feedback Control. *J. Chem. Technol. Biotechnol.* **2019**, *94*, 3924–3931. [CrossRef]
31. Chen, L.Z.; Nguang, S.K.; Chen, X.D.; Li, X.M. Modelling and Optimization of Fed-Batch Fermentation Processes Using Dynamic Neural Networks and Genetic Algorithms. *Biochem. Eng. J.* **2004**, *22*, 51–61. [CrossRef]
32. Jenzsch, M.; Simutis, R.; Eisbrenner, G.; Stückrath, I.; Lübbert, A. Estimation of Biomass Concentrations in Fermentation Processes for Recombinant Protein Production. *Bioprocess Biosyst. Eng.* **2006**, *29*, 19–27. [CrossRef]
33. Zhang, A.H.; Zhu, K.Y.; Zhuang, X.Y.; Liao, L.X.; Huang, S.Y.; Yao, C.Y.; Fang, B.S. A Robust Soft Sensor to Monitor 1,3-Propanediol Fermentation Process by *Clostridium Butyricum* Based on Artificial Neural Network. *Biotechnol. Bioeng.* **2020**, *117*, 3345–3355. [CrossRef]
34. Krishna, V.; Pappa, N.; Vasantharani, S.P.J. Realization of Deep Learning Based Embedded Soft Sensor for Bioprocess Application. *Intell. Autom. Soft Comput.* **2022**, *32*, 781–794. [CrossRef]
35. Ferreira, L.S.; De Souza, M.B.; Folly, R.O.M. Development of an Alcohol Fermentation Control System Based on Biosensor Measurements Interpreted by Neural Networks. *Sens. Actuators B Chem.* **2001**, *75*, 166–171. [CrossRef]
36. Grigs, O.; Bolmanis, E.; Galvanauskas, V. Application of *In-Situ* and Soft-Sensors for Estimation of Recombinant *P. Pastoris* GS115 Biomass Concentration: A Case Analysis of HBcAg (Mut^+) and HBsAg (Mut^S) Production Processes under Varying Conditions. *Sensors* **2021**, *21*, 1268. [CrossRef] [PubMed]
37. Niu, H.; Jost, L.; Pirlot, N.; Sassi, H.; Daukandt, M.; Rodriguez, C.; Fickers, P. A Quantitative Study of Methanol/Sorbitol Co-Feeding Process of a *Pichia Pastoris* Mut^+/P_{AOX1}-LacZ Strain. *Microb. Cell Fact.* **2013**, *12*, 33. [CrossRef] [PubMed]
38. Pentjuss, A.; Bolmanis, E.; Suleiko, A.; Didrihsone, E.; Suleiko, A.; Dubencovs, K.; Liepins, J.; Kazaks, A.; Vanags, J. *Pichia Pastoris* Growth—Coupled Heme Biosynthesis Analysis Using Metabolic Modelling. *Sci. Rep.* **2023**, *13*, 15816. [CrossRef] [PubMed]
39. Joseph, J.A.; Akkermans, S.; Cornillie, E.; Deberlanger, J.; Van Impe, J.F.M. Optimal Culture Medium Selection and Supplementation for Recombinant Thaumatin II Production by *Komagataella Phaffii*. *Food Bioprod. Process.* **2023**, *139*, 190–203. [CrossRef]
40. Damasceno, L.M.; Huang, C.J.; Batt, C.A. Protein Secretion in *Pichia Pastoris* and Advances in Protein Production. *Appl. Microbiol. Biotechnol.* **2012**, *93*, 31–39. [CrossRef]
41. Wang, B.; Nie, Y.; Zhang, L.; Song, Y.; Zhu, Q. An Soft-Sensor Method for the Biochemical Reaction Process Based on LSTM and Transfer Learning. *Alex. Eng. J.* **2023**, *81*, 170–177. [CrossRef]
42. Beiroti, A.; Hosseini, S.N.; Norouzian, D.; Aghasadeghi, M.R.; Jajan, L.H.G. Development of Soft Sensors for Online Biomass Prediction in Production of Hepatitis B Vaccine. *Biointerface Res. Appl. Chem.* **2023**, *13*, 195. [CrossRef]
43. Yeast Nitrogen Base without AA & with AS Composition. Available online: https://www.usbio.net/media/Y2037/Yeast%20Nitrogen%20Base%20w~2Fo%20AA%20&%20w~2FAS/data-sheet (accessed on 26 October 2023).
44. Abelovska, L.; Bujdos, M.; Kubova, J.; Petrezselyova, S.; Nosek, J.; Tomaska, L. Comparison of Element Levels in Minimal and Complex Yeast Media. *Can. J. Microbiol.* **2007**, *53*, 533–535. [CrossRef]

Appendix V
5. Pielikums

Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development

Emils Bolmanis, Vytautas Galvanauskas, Oskars Grigs, Juris Vanags, Andris Kazaks
Fermentation **2025**, *11*, DOI: 10.3390/fermentation11070411

E.B. input: Conceptualization, formal analysis, investigation, methodology, software, validation, data curation, writing – original draft preparation, writing – review and editing, visualization, project administration.

fermentation

MDPI

Article

Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development

Emils Bolmanis [1,2], Vytautas Galvanauskas [3], Oskars Grigs [4], Juris Vanags [2] and Andris Kazaks [1,*]

[1] Latvian Biomedical Research and Study Centre, LV-1067 Riga, Latvia; emils.bolmanis@biomed.lu.lv
[2] Institute of Biomaterials and Bioengineering, Faculty of Natural Sciences and Technology, Riga Technical University, LV-1658 Riga, Latvia; juris.vanags@bioreactors.net
[3] Department of Automation, Faculty of Electrical and Electronics Engineering, Kaunas University of Technology, LT-51367 Kaunas, Lithuania; vytautas.galvanauskas@ktu.lt
[4] Laboratory of Bioengineering, Latvian State Institute of Wood Chemistry, LV-1006 Riga, Latvia; oskars.grigs@kki.lv
* Correspondence: andris@biomed.lu.lv

Abstract

Hybrid modeling techniques are increasingly important for improving predictive accuracy and control in biomanufacturing, particularly in data-limited conditions. This study develops and experimentally validates a hybrid deep learning model predictive control (MPC) framework for recombinant *P. pastoris* fed-batch fermentations. Bayesian optimization and grid search techniques were employed to identify the best-performing hybrid model architecture: an LSTM layer with 2 hidden units followed by a fully connected layer with 8 nodes and ReLU activation. This design balanced accuracy (NRMSE 4.93%) and computational efficiency (AICc 998). This architecture was adapted to a new, smaller dataset of bacteriophage Qβ coat protein production using transfer learning, yielding strong predictive performance with low validation (3.53%) and test (5.61%) losses. Finally, the hybrid model was integrated into a novel MPC system and experimentally validated, demonstrating robust real-time substrate feed control in a way that allows it to maintain specific target growth rates. The system achieved predictive accuracies of 6.51% for biomass and 14.65% for product estimation, with an average tracking error of 10.64%. In summary, this work establishes a robust, adaptable, and efficient hybrid modeling framework for MPC in *P. pastoris* bioprocesses. By integrating automated architecture searching, transfer learning, and MPC, the approach offers a practical and generalizable solution for real-time control and supports scalable digital twin deployment in industrial biotechnology.

Keywords: *Pichia pastoris*; hybrid process model; deep learning; Bayesian optimization; hybrid model architecture screening; transfer learning; model predictive control; hybrid MPC

Academic Editor: Nicolai S. Panikov

Received: 25 June 2025
Revised: 11 July 2025
Accepted: 15 July 2025
Published: 17 July 2025

Citation: Bolmanis, E.; Galvanauskas, V.; Grigs, O.; Vanags, J.; Kazaks, A. Leveraging Historical Process Data for Recombinant *P. pastoris* Fermentation Hybrid Deep Modeling and Model Predictive Control Development. *Fermentation* **2025**, *11*, 411. https://doi.org/10.3390/fermentation11070411

1. Introduction

The inherent complexity of biological systems—characterized by interconnected subsystems and nonlinear dynamics—presents enduring challenges for effective bioprocess monitoring and control. Mathematical modeling offers a powerful means to capture and manage this complexity [1]. Over the past few decades, such models have become indispensable in understanding and optimizing bioprocesses, benefiting from advances in

computational capabilities and analytical tools. This progress has led to the adoption of advanced modeling approaches, such as genome-scale metabolic models and computational fluid dynamics simulations [2,3]. With the advent of Industry 4.0, modeling has taken on an even more prominent role in the digital transformation of biomanufacturing [4]. However, the limitations of mechanistic models—particularly their reliance on complete system knowledge—have prompted interest in alternative strategies. As a result, machine learning techniques are gaining traction, offering flexible, data-driven alternatives that can extract insights without relying on fully defined system knowledge—supporting next-generation bioprocess digitalization [5–8].

To bridge this gap, hybrid neural network (HNN) models have emerged as a compelling solution, integrating domain knowledge with data-driven flexibility [9–14]. These models combine the structure of mechanistic frameworks (e.g., mass or energy balances) with the flexibility of data-driven components such as ANNs, effectively leveraging both physical laws and empirical data [15,16]. A typical use case involves modeling unknown or complex kinetics using neural networks within a differential equation-based framework. Compared to purely nonparametric models, hybrid approaches often yield more accurate, generalizable, and interpretable results—leading to more robust bioprocess operation and control [17,18]. The recent surge in deep learning methodologies further enhanced hybrid modeling by enabling neural networks to approximate intricate biological functions, as deep neural networks (DNNs) with multiple hidden layers demonstrated superior capacity for learning hierarchical and compositional functions with fewer parameters and reduced sample complexity compared to shallow architectures [9,10,19]. As such, hybrid deep learning models are emerging as powerful tools in the development of digital twins and advanced bioprocess monitoring systems.

However, training deep models requires extensive, high-frequency datasets that are often unavailable in bioprocessing due to high costs, long cultivation times, and sensor limitations [20]. In early-stage bioprocess development or pilot-scale operations, datasets typically only contain a small number of replicates per condition, resulting in limited coverage of the process space and insufficient diversity to support generalizable deep models [21]. Bioprocess data is typically noisy, heterogeneous, and expensive to generate, posing a major bottleneck for training deep architectures. Such data limitations have accelerated the adoption of hybrid modeling frameworks, which leverage both prior process knowledge and empirical data. The promising approach of integrating historical process data through transfer learning frameworks can significantly improve prediction accuracy and reduce the need for extensive new experimental data [22,23].

Gathering experimental bioprocess data is often time-consuming and labor-intensive, limiting the volume of new datasets available for model development. Therefore, leveraging historical data from similar experiments can significantly benefit the development of hybrid process models for new or evolving bioprocesses. Transfer learning (TL) in deep neural networks involves repurposing models trained on a source task to enhance learning on a related target task, particularly when data are scarce. By leveraging pretrained models—often developed on large, generic datasets—TL facilitates improved convergence speed, generalization, and computational efficiency compared to training from scratch [24,25]. In bioprocess engineering, TL enables the adaptation of models trained on well-characterized systems to predict dynamics in novel or data-limited processes. For instance, TL has been successfully applied to model microalgal bioprocess dynamics using limited time-series data, achieving high accuracy in forecasting process behavior [26] and for the quantification and identification of cellular phenotypes from high-content microscopy images [27]. Effective transfer requires the careful selection of source models, layer-freezing strategies, and learning rate tuning to preserve useful features while adapting to the new task [28].

In bioprocess applications, these considerations are critical due to the heterogeneity of biological systems and the frequent lack of large datasets. Nevertheless, TL remains a promising strategy—especially when combined with hybrid modeling approaches that integrate mechanistic insights with data-driven learning [29].

Deep neural network architecture screening typically involves systematic strategies to identify optimal model configurations and hyperparameters. Common methods include grid search, which exhaustively evaluates combinations within a predefined parameter grid [30]. While easy to implement and parallelize, grid search becomes computationally expensive as the number of hyperparameters increases. Random search improves efficiency by sampling configurations at random, often outperforming grid search in high-dimensional spaces where only a few parameters significantly affect performance [30]. A more advanced and efficient approach is Bayesian optimization, which builds a probabilistic surrogate model (e.g., Gaussian Process) to predict performance and selects promising configurations using acquisition functions like Expected Improvement [31]. This strategy significantly reduces the number of required evaluations and is especially valuable when model training is computationally costly. Several studies confirm that Bayesian optimization typically outperforms traditional approaches in terms of sample efficiency and final model performance [30,31].

As the dominant production mode, fed-batch fermentation remains widely used due to its robustness and high product yield, with most biotherapeutics in clinical and commercial use produced using this mode [32,33]. However, maintaining optimal substrate feeding remains a major challenge, requiring precise control to ensure consistent performance. Model predictive control (MPC) has emerged as a powerful strategy to address this, leveraging predictive models to optimize feeding decisions in real time [34–36]. Yet, the nonlinear and dynamic nature of high-cell-density fermentations often limits the accuracy of purely mechanistic models, as parameter estimation and unforeseen biological interactions degrade model reliability [37,38]. Hybrid modeling strategies help mitigate these challenges by improving parameter adaptability, accounting for process nonlinearities, and enhancing real-time prediction accuracy [39]. Integrating hybrid bioreactor process models with MPC enhances the optimization and control of bioprocesses, particularly in complex systems like high-cell-density fermentations, resulting in improved modeling accuracy, increased adaptability to changing process conditions, and real-time feedback for increased process stability [40–42].

Recent work has explored hybrid modeling in recombinant *P. pastoris* cultivations, demonstrating performance gains in process control, generalization, and scalability. Ferreira et al. used a serial HNN, consisting of a three-layer feedforward neural network (FFNN) combined with material balance equations, for the dynamic modeling of *P. pastoris* GS115 expressing scFv in a 50 L pilot bioreactor. This hybrid model was then applied for iterative batch-to-batch control, resulting in a 4-fold increase in the titer after four optimization cycles [12]. However, no network architecture screening was performed, and the model was trained using a fixed configuration without evaluating alternative structures. Pinto et al. revisited the general bioreactor hybrid model and introduced deep learning techniques. Multi-layer networks with varying depths were combined with First Principles equations to form deep hybrid models. Techniques like ADAM, stochastic regularization, and depth-dependent weight initialization were evaluated in this context. The methods were applied to a synthetic *E. coli* dataset and a 50 L Mut$^+$ *P. pastoris* process, expressing a single-chain antibody fragment. Results showed significant improvements in generalization, with an 18.4% increase in prediction accuracy and a 43.4% reduction in CPU time compared to shallow models [10]. In another study, Pinto et al. developed a hybrid deep modeling method with state-space reduction, applied to a *P. pastoris* GS115 Mut$^+$

strain expressing scFv. Deep FFNNs of varying depths were combined with bioreactor material balance equations and trained using ADAM, semidirect sensitivity equations, and stochastic regularization. A state-space reduction method, based on principal component analysis (PCA) of the cumulative reacted amount, reduced model complexity by 60%, and improved predictive accuracy by 18.5%. Validation with data from nine fed-batch *P. pastoris* 50 L cultivations highlighted optimization opportunities, with potential increases in an scFv titer of 30% and 80% achieved by optimizing methanol feed and inorganic elements, respectively [9]. Despite these advances, none of the studies addressed transfer learning, nor did they integrate the hybrid models into real-time control architectures. While two studies investigated optimal network architecture selection, none carried out exhaustive design space exploration—such as varying activation functions, layer placement, or layer types—leaving the potential of more performant architectures underexplored.

The present study aims to advance previous work in hybrid modeling of *P. pastoris* bioprocesses by addressing several key limitations. Specifically, it expands the search for optimal neural network architectures through a systematic screening approach, enabling the identification of robust and efficient model structures tailored to process dynamics. Transfer learning is introduced as a strategy to adapt pretrained hybrid models—originally developed on historical bioprocess data—to a new, much smaller fermentation dataset, thereby reducing the need for extensive retraining while preserving predictive accuracy. Beyond model development, the practical application of the hybrid modeling approach is demonstrated through its integration into an MPC framework. The resulting hybrid MPC system is experimentally validated, showcasing its effectiveness for the real-time optimization and control of recombinant *P. pastoris* fed-batch fermentations.

2. Materials and Methods

2.1. Experimental Conditions

Cultivations were performed using a recombinant *Pichia pastoris* X-33 wild-type strain, producing Qβ coat protein VLPs. The construction of the expression vector and the selection of clones for this specific producer are described in detail elsewhere [43].

The batch and feed media formulations were prepared following the "*Pichia* Fermentation Process Guidelines" provided by the Invitrogen Corporation [44]. Fermentations were carried out in a 5 L bench-top bioreactor (Bioreactors.net, EDF-5.4/BIO-4, Riga, Latvia) with a working volume of 2–4 L. The pH was continuously monitored using a calibrated pH probe (Hamilton, EasyFerm Bio, Bonaduz, Switzerland) and adjusted to 5.0 ± 0.1 with a 28% NH_4OH solution prior to inoculation; it was then maintained at this value throughout the process. Temperature was regulated at 30.0 ± 0.1 °C using a temperature sensor and jacketed vessel control. A thermostatic circulator (LKB Bromma, Multitemp II, Bromma, Sweden) maintained the cooling water at a preset temperature of 6 °C during experiments. Dissolved oxygen (DO) levels were measured using a DO sensor (Hamilton, Oxyferm Bio, Bonaduz, Switzerland) and maintained above 30 ± 5% via a dual cascade strategy, adjusting the stirrer speed between 200 and 1000 RPM (Cascade 1), and supplementing the inlet air with pure O_2 when necessary (Cascade 2). Throughout the fermentation, we sustained a constant airflow or an air/oxygen mixture at 3.0 slpm. A condenser was employed to capture moisture from exhaust gases, and Antifoam 204 (Sigma-Aldrich, Burlington, MA, USA) was added as needed to suppress excessive foam formation. Substrate feeding was controlled with a high-precision peristaltic pump (Longer-Pump, BT100–2J, Baoding, China).

Methanol feeding proceeded in three phases: initially at 0.12 $mL \cdot min^{-1}$ for 5 h, then at 0.24 $mL \cdot min^{-1}$ for 2 h, and finally at 0.36 $mL \cdot min^{-1}$. This continued either until the end of the experiment or for 2–3 h until the hybrid MPC control was activated so that the cells could effectively adapt to methanol utilization.

In experiments where real-time biomass concentration was monitored using an in-situ turbidity probe (Optek-Danulat, ASD19-EB-01, Essen, Germany), the turbidity signal was correlated with biomass concentration using an exponential calibration equation reported previously [45]. To enhance the in-line biomass measurement quality for hybrid model training, an algorithm was implemented to correct for abrupt, anomalous spikes caused by sudden process disturbances. The details of this algorithm are provided in a separate publication [46].

Although the cultivation conditions were generally similar across all historical experiments, some minor differences were present. For detailed information on these variations, as well as the construction of the expression vectors and clone selection, we refer the reader to the original publications [47,48].

2.2. Downstream Processing of Qβ VLPs

Overall, 4.0 g of wet cells was resuspended in 20 mL of lysis buffer (20 mM Tris 8.0, 100 mM NaCl) and disrupted using a French press (4 × 10,000 psi). The suspension was then centrifuged for 30 min at 18,500× *g* (4 °C). Ammonium sulfate was added to the supernatant to 40% saturation and proteins were precipitated at 4 °C for 60 min. The suspension was then centrifuged for 20 min at 18,500× *g* (4 °C) and the supernatant was discarded. The precipitate was dissolved in a 20 mM Tris 8.0 buffer and loaded onto the Sepharose 4 Fast Flow size-exclusion column (12 mL volume) in a lysis buffer at 0.3 mL/min. Peak fractions were pooled and loaded onto an anion-exchange Fractogel DEAE (M) column (5 mL volume) in a lysis buffer and eluted with a linear gradient of 20 mM Tris-HCl, 1 M NaCl pH 8.0 at 2 mL/min.

2.3. Analytical Measurements

Cell growth was observed by offline measurements of dry cell weight (DCW), determined gravimetrically. Biomass samples were placed in pre-weighted Eppendorf® tubes and centrifuged at 15,500× *g* for 5 min. Afterwards, the supernatant was discarded and the remaining wet cell biomass was weighted. DCW was calculated using a previously determined correlation coefficient.

$$DCW = WCW \times 0.27 \quad (1)$$

2.4. Dataset for Hybrid Model Training

The characteristics of the dataset of *P. pastoris* fermentation data used for hybrid model training are compiled in Table 1.

Table 1. Dataset of *P. pastoris* fermentations and parameter ranges used for hybrid model training.

Exp No.	Strain, Product	Induction Time, h	DCW, g·L^{-1}	Feed Rate, mL·min^{-1}	V_{end}, L	Reference
1 *	GS115, HBcAg	65	37.5–101.6	0.12–0.78	2.85	[45,47]
2 *	GS115, HBcAg	45	40.6–113.5	0.12–1.00	3.09	
3 *	GS115, HBcAg	43	41.2–120.1	0.12–0.98	3.13	

Table 1. *Cont.*

Exp No.	Strain, Product	Induction Time, h	DCW, g·L^{-1}	Feed Rate, mL·min^{-1}	V_{end}, L	Reference
4 *	GS115, HBcAg	50	59.2–120.1	0.12–0.36	2.54	
5 *	GS115, HBcAg	51	41.4–96.6	0.12–0.36	2.87	
6	GS115, HBcAg	48	49.1–120.0	0.12–0.50	2.88	
7	GS115, HBcAg	43	53.7–101.5	0.12–0.36	2.74	Unpublished data
8	GS115, CA IX	54	44.1–84.0	0.12–0.56	2.75	
9 *	X-33, LegH	65	55.4–123.2	0.12–0.36	2.57	
10 *	X-33, LegH	46	49.5–95.4	0.12–0.60	2.98	
11	X-33, LegH	65	48.9–111.2	0.12–0.36	2.85	
12	X-33, LegH	48	56.4–105.3	0.12–0.50	2.63	
13	X-33, LegH	50	45.3–101.3	0.12–0.36	2.61	[48]
14	X-33, LegH	45	52.9–103.1	0.12–0.36	2.55	
15	X-33, LegH	46	45.1–101.3	0.12–0.36	2.52	
16	X-33, LegH	65	51.0–101.7	0.12–0.36	2.66	
17	X-33, LegH	46	50.6–92.4	0.12–0.60	3.00	
18	X-33, Qβ	65	52.5–117.6	0.12–0.49	3.23	
19	X-33, Qβ	48	49.3–117.2	0.12–1.00	3.40	This research
20	X-33, Qβ	55	50.1–107.7	0.12–0.36	2.84	
21	X-33, Qβ	52	52.9–112.6	0.12–0.87	3.45	

* Experiments with real-time turbidity measurements.

2.5. Hybrid Process Model Structure and Training

The general structure of the hybrid process model (Figure 1) is similar to the structure previously explored by Pinto et al. [9,10]. The input layer comprises five variables: substrate feed rate (Fs, mL·min^{-1}), dry cell biomass concentration (X, g·L^{-1}), product concentration (P, mg·L^{-1}), culture medium volume (V, L), and an empirical shock factor (Sh). The shock factor equation (Sh(0) = 1) confers the cumulative toxic effect of methanol feeding on the cells and is an unmeasured internal state variable. A similar equation was proposed by Pinto et al. and Lee & Ramirez [9,10,49]. To enhance model training efficiency and convergence, the sequence input layer incorporates the normalization of the input features by scaling each sequence sample to the [0, 1] range using the minimum and maximum values computed over the entire dataset. The output layer provides three variables: the specific cell growth rate (μ, h^{-1}), production rate (q_P, h^{-1}), and the rate of change of the shock factor (k_1). The optimal composition of hidden layer structure was investigated in further steps.

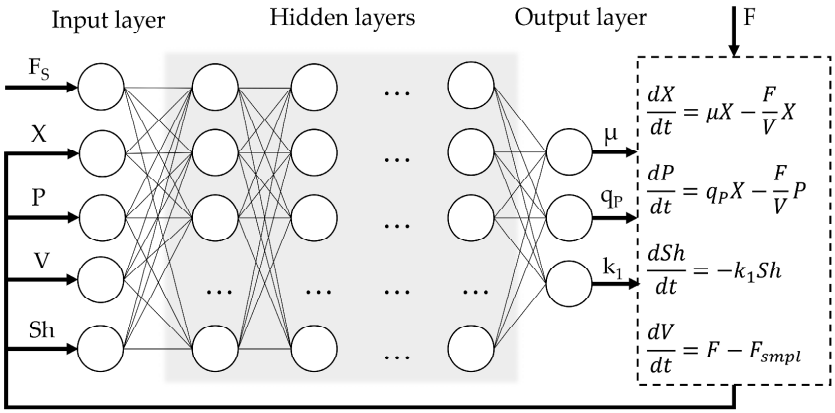


Figure 1. Overview of hybrid model structure.

The outputs of the nonparametric model are then passed to the parametric component, which captures the dynamics of the state variables using a system of ordinary differential equations (ODEs) derived from macroscopic and/or intracellular material balances, as well as other relevant physical assumptions. The only external inputs to the model are the aforementioned substrate feed rate (*Fs*) and the volumetric flow rate (*F*). This involves the following equation:

$$F = F_s + F_b + F_{AF} - F_{evp} \tag{2}$$

where F_b and F_{AF} are the added base and antifoam solution flow rates (mL·min^{-1}) and F_{evp} is the determined culture evaporation rate (0.11 mL·min^{-1}).

The loss function was defined as the normalized root mean square error (*NRMSE*):

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2}{n}}}{y_{max} - y_{min}} 100\% \tag{3}$$

where n is the number of training samples, y_i represents the measured values, y_i^* is the corresponding predicted variables, and y_{min} and y_{max} denote the minimum and maximum values for the dataset, respectively.

To fully utilize the real-time cell biomass sensor data, each process dataset was segmented into 60 equally spaced batches using an interleaved batching approach. Given a time-series dataset consisting of N total time points $\{t_1, t_2, \ldots, t_N\}$ and a chosen number of segments $k = 60$, each batch B_j (for $j = 1, 2, \ldots, 60$) was constructed by selecting every k-th time point starting from offset j. This can be expressed mathematically as follows:

$$B_j = \{t_i | i = j + nk,\ n \in \mathbb{N}_0,\ j + nk \le N\} \tag{4}$$

This approach ensures that each batch contains a temporally distributed subset of the full dataset, preserving temporal variability and aiding in model generalization. For instance, Batch 1 contains $\{t_1, t_{61}, t_{121}, \ldots\}$, Batch 2 includes $\{t_2, t_{62}, t_{122}, \ldots\}$ and so on, up to Batch 60. Because time-series data were recorded at 1-min intervals, this batching scheme effectively introduced a consistent 60 min gap between successive data points within each batch. As a result, it enabled the estimation of average dynamic rates (e.g., biomass growth or product formation) on an hourly basis—an appropriate timescale for bioprocess interpretation and modeling. This method is particularly suitable for sequence-

based machine learning models, as it ensures diverse temporal representation in each training segment.

In experiments where only sparse experimental measurements were available, interpolation was necessary to ensure that each segmented batch contained the corresponding measurement values at the correct time points. To achieve this, piecewise cubic Hermite interpolating polynomial (PCHIP) interpolation was applied to the time-series data. This approach generated estimated values at all necessary time points, thereby ensuring that each batch contained a continuous and temporally consistent signal aligned with the original, sparsely sampled experimental measurements. Importantly, to preserve the integrity of model evaluation, NRMSE was only computed at the original measurement time points, ensuring that model performance was assessed strictly against experimentally observed data rather than interpolated estimates.

Prior to model training, a validation dataset was created by randomly selecting 10% of the original time-series data. Specifically, six batches were sampled from each fermentation experiment to form the validation partition. This subset was held out during training and used exclusively to monitor model performance and assess generalization to unseen data, and as an early-stopping criterion for training.

The hybrid process model was trained using the Adaptive Moment Estimation (ADAM) optimization algorithm, which combines the advantages of both momentum-based and adaptive learning rate methods [50]. The optimizer was configured with standard recommended parameters: the initial learning rate $\alpha_0 = 0.001$; the first moment decay rate $\beta_1 = 0.9$; the second moment decay rate $\beta_2 = 0.999$; the small numerical stability constant $\varepsilon = 10^{-8}$. To facilitate stable convergence and mitigate overfitting, an exponential learning rate decay strategy was employed, where the learning rate was gradually decreased every 100 epochs from 0.001 to 0.0001 by a calculated decay factor over the course of training, as per the following formula:

$$\alpha(t) = \alpha_0 * \gamma^{\left[\frac{t}{100}\right]} \quad (5)$$

where $\alpha(t)$ is the learning rate at epoch t and the learning rate decay factor $\gamma = 0.9007$.

This gradual decay enabled the model to make larger updates early in training and finer adjustments in later stages, facilitating both rapid convergence and precise parameter tuning. After each epoch, the training dataset was randomly shuffled and divided into six minibatches to support optimization using the ADAM algorithm. This randomization reduced the risk of learning spurious temporal or sequential dependencies and enhanced generalization. The use of minibatches, combined with ADAM's adaptive learning rate mechanism, further improved training efficiency and convergence reliability.

Hybrid model training was performed using a custom training script developed in the MATLAB environment (MathWorks, R2024b, Natick, MA, USA), leveraging the Deep Learning Toolbox (Scheme S1). Training was conducted on a personal computer equipped with an Intel i5-6600 CPU (3.30–3.90 GHz) and 16 GB of RAM. The training script was parallelized to enable the simultaneous training of multiple networks. For parallel training tasks, the High-Performance Computing (HPC) cluster of Riga Technical University (RTU) was utilized in conjunction with the personal computer. Training for one epoch took approximately 2.5 s.

2.6. Hybrid Model Architecture Screening

To efficiently identify the optimal hidden layer architecture for the hybrid model, a multi-step strategy was implemented. First, a Bayesian optimization approach was employed. This method systematically explores the hyperparameter space by building a probabilistic model of the objective function, enabling informed and efficient searches for

the best-performing network configurations with fewer training iterations compared to traditional grid or random search methods.

To accelerate the screening process, networks were trained in parallel for a limited duration of 10 epochs (corresponding to 90 iterations) using an elevated initial learning rate of $\alpha_0 = 0.01$. This higher learning rate was chosen to promote faster convergence during early training, enabling the quicker identification of promising model architectures without the need for extensive training. Only historical experimental data (Exps. 1–17) was used for hybrid model architecture screening, ensuring the model adapted its parameters based on well-established process dynamics before being fine-tuned with the new Qβ dataset, thereby improving generalization and robustness during transfer learning. Validation loss was used as the primary performance criterion during the grid search, as it provided a more reliable measure of the model's ability to generalize to unseen data and helped to prevent the selection of overfitted architectures. The corrected Akaike information criterion (AICc) was used alongside validation loss as a performance criterion to account for model complexity, ensuring that selected architectures not only fit the data well but also avoid overparameterization, which can hinder generalization:

$$AICc = n\ln(L) + 2k + \frac{2k(k+1)}{n-k-1} \quad (6)$$

where k is the number of model parameters, L is the loss function value (NRMSE, %), and n is the number of observations (sample size).

For the initial optimization phase, several key hyperparameters were selected to systematically investigate their impact on model performance. These included the choice of the first hidden layer type—either LSTM or fully connected (FC). The number of subsequent fully connected layers was varied between one and two to assess the effect of network depth. Additionally, the activation functions within the fully connected layers were optimized, considering options such as ReLU, Leaky ReLU (0.01), Tanh, or no activation, to evaluate how different nonlinearities affect learning. Finally, the number of hidden units or nodes in each layer was explored within a range of 1 to 5, enabling fine-grained control over model capacity and complexity. This amounts to 8000 possible parameter combinations. The Bayesian optimization algorithm was executed 10 times, each run consisting of 200 iterations. The best-performing model architecture with the lowest validation loss from each run was saved for further evaluation.

In the second step, the scope of the parameter search was narrowed to focus on the activation function (ReLU, Leaky ReLU, Tanh, or none), the number of LSTM hidden units (ranging from 1 to 5), and the number of fully connected layer nodes (ranging from 1 to 10). The upper limit for LSTM hidden units was intentionally kept low to prevent overparameterization, as each additional LSTM unit substantially increased the total number of trainable parameters. In contrast, the number of FC layer nodes was expanded up to 10 based on favorable results observed in the previous optimization step, where larger FC layers contributed to improved model performance without incurring excessive computational costs. With only 200 possible combinations, a full grid search was conducted during the second screening step. Each network architecture was evaluated across 10 training runs, and only the best-performing candidate was retained for each parameter combination. To balance validation loss with model complexity and mitigate overparameterization, a loss vs. AICc plot was generated to identify the Pareto front (Figure A1). From this front, five network architectures that best balanced low validation loss and favorable AICc values were selected for further evaluation.

In the third step, the selected models were trained in full for 20,000 iterations, and their predictive performance was assessed using validation loss. The best-performing

model was chosen as the optimal network configuration for the specific hybrid process model. Finally, the impact of incorporating a dropout layer within the finalized hybrid model architecture was thoroughly investigated to assess its effect on model robustness and generalization performance.

2.7. Hybrid Model Transfer Learning

To leverage the selected trained hybrid model for a new experimental dataset involving Qβ production, a transfer learning approach was applied. The previously optimized model architecture, trained on the historical dataset, served as the starting point. All model weights were initialized from this pretrained network to retain learned temporal and process dynamics. To adapt the model to the Qβ dataset, only the final fully connected layer was unfrozen and fine-tuned using the new data, while the LSTM unit—capturing general process behavior—was kept fixed. This strategy allows the model to efficiently adapt to the specific characteristics of the Qβ process while preserving useful generalizations from the original training, thereby reducing training time and improving predictive accuracy with limited new data. Experiment 20 was used for testing loss calculation.

2.8. Model Predictive Control (MPC) Architecture

A comprehensive model predictive control framework was developed and implemented, leveraging the advanced hybrid process model. The main task of the MPC controller was to track a pre-selected growth trajectory close to the maximum specific growth rate of the cells.

The MPC algorithm was implemented to dynamically estimate the optimal substrate feed rate, $F_s(t)$, required to maintain a desired specific growth rate, $\mu_{set}(t)$. In the developed hybrid modeling framework, the substrate feed rate served as an input, while the specific growth rate was one of the predicted outputs. Due to the non-invertible structure of the hybrid model, analytical inversion to compute $F_s(t)$ from a given $\mu_{set}(t)$ was not feasible. Consequently, a numerical optimization approach was employed at each control step using MATLAB's bounded nonlinear optimization function *fminbnd*, which searched for the optimal feed rate within specified operational constraints. These constraints, set between 0.36 and 1.00 mL min^{-1}, ensured that the substrate feed rate remained within the safe operating range of the bioreactor. The optimization problem at each time step was formulated as follows:

$$\underset{Fs\in[0.36,\ 1.00]mLmin^{-1}}{min} \sum_{k=1}^{Np} [\mu(k) - \mu_{set}(k)]^2 \tag{7}$$

This was subject to the hybrid model dynamics:

$$x(k+1) = f_{hybrid}(x(k), F_S(k)) \tag{8}$$

where $x(k)$ is the vector of the input state variables, and $\mu(k)$ is the predicted specific growth rate at step k.

The control horizon was set to N_c = 1 h to match the hybrid model's prediction timestep, while the prediction horizon was set to N_p = 12 h—twice the estimated dominant time constant of 6 h required for a typical *P. pastoris* fermentation. The hybrid model itself was simulated with a finer sampling time of 1 min to ensure accurate forward predictions, while the MPC made decisions on an hourly basis.

To improve model accuracy and adaptability, the hybrid model was retrained after each sampling (approx. three times per day). Each time, the newly measured biomass concentration, $X_{meas}(t)$, was appended to the training dataset, and the model parameters

were updated to reflect the latest process dynamics. The pretrained network was fine-tuned for 100 epochs per retraining cycle, requiring on average 5 min per update.

The MPC framework was experimentally validated in a *P. pastoris* fed-batch fermentation under the methanol-inducible AOX1 promoter. Real-time process data, including substrate feed rate, base and antifoam addition, were integrated into MATLAB via an OPC server, enabling seamless bidirectional communication with the SCADA system. A more technical description is available elsewhere [32]. This real-time data integration allowed the MPC to adjust the substrate feed rate based on actual process conditions.

MPC was initiated after methanol adaptation, typically 8–10 h after induction. The growth rate setpoint $\mu_{set}(t)$ was scheduled in a step-wise fashion to reflect the physiological limits of the cells: an initial value of 0.04 h^{-1} was maintained for the first 12 h, followed by reductions to 0.02 h^{-1} and 0.01 h^{-1} at 12 h intervals.

3. Results

3.1. Optimal Hybrid Model Architecture Screening

The Bayesian optimization approach was effectively employed to identify optimal hybrid model architectures. Although not all configurations within the design space were explored, the method yielded valuable insights while substantially narrowing the range of candidate architectures. This efficiency stems from the Bayesian framework's ability to prioritize promising regions of the hyperparameter space, reducing the number of models requiring evaluation. The ten best-performing network architectures, ranked by their validation loss, are summarized in Table 2.

Table 2. Summary of ten best-performing architectures from Bayesian optimization screening.

First Layer Type	No. of FC Layers	Hidden Units	Nodes	Activation	Validation Loss (%)	No. of Parameters	AICc
LSTM	1	5	5	Tanh	9.73	268	2417
		5	5	LeakyReLu	10.07	268	2431
		5	4	Tanh	10.17	259	2312
		1	5	ReLu	10.26	56	1126
		5	5	Tanh	10.37	268	2444
		3	5	Tanh	10.42	146	1448
		4	5	None	10.51	203	1783
		5	5	Tanh	10.68	268	2457
		4	5	ReLu	10.75	203	1792
		5	4	Tanh	11.30	259	2358

The results of the Bayesian optimization screening indicate that while individual performance metrics varied across architectures, consistent structural trends emerged among the top-performing models. Notably, all ten selected architectures include an LSTM layer as the first hidden layer, emphasizing the importance of temporal feature extraction in capturing the dynamics of the process. This is followed in each case by at least one fully connected layer, which likely contributes to nonlinear transformation and the mapping of the LSTM outputs to the target outputs. This consistent architectural pattern suggests that the hybrid model benefits from a sequence-aware representation (via LSTM), followed by flexible function approximation (via FC layers). Although the activation functions and number of units varied, these two structural components formed the core of the most effective configurations, reinforcing their critical role in model performance.

In the subsequent step, a comprehensive grid search was conducted across the full domain of feasible network architectures to investigate the optimal number of hidden units, node counts, and activation functions for the hidden layers. This systematic approach enabled an exhaustive evaluation of all combinations of relevant hyperparameters, ensuring that no potentially optimal configuration within the predefined design space was overlooked. To mitigate the risk of network overparameterization, AICc values were evaluated in parallel with validation loss. Candidate architectures from the Pareto front of the loss vs. AICc plot (Figure A1) were identified as the most promising, and their configurations are detailed in Table 3.

Table 3. Summary of the best-performing architectures from grid search screening.

Hidden Units	Nodes	Activation	Validation Loss (%)	No. of Parameters	AICc
5	9	LeakyReLu	8.85	304	3050
4	10	LeakyReLu	9.27	243	2092
3	6	LeakyReLu	9.29	153	1427
3 *	**5**	**LeakyReLu**	**9.37**	**146**	**1403**
2 *	**10**	**LeakyReLu**	**9.68**	**127**	**1334**
2 *	**9**	**Tanh**	**9.99**	**121**	**1316**
2 *	**8**	**ReLu**	**10.03**	**115**	**1302**
1 *	**9**	**Tanh**	**10.26**	**76**	**1189**
1	6	LeakyReLu	10.21	61	1139
1	4	ReLu	10.25	51	1112
1	1	Tanh	10.36	36	1079

* Network architectures selected for the next optimization step.

To ensure that there was a balanced trade-off between predictive accuracy and model complexity, the middle five architectures from the Pareto front were selected for further optimization. These architectures, indicated with an asterisk in Table 3, offer a pragmatic compromise: they demonstrate competitive validation losses (ranging from 9.37% to 10.26%) while maintaining a relatively low number of trainable parameters (between 76 and 146). This subset effectively spans the central region of the Pareto front, avoiding both the highly complex models at the top—which, despite slightly better accuracy, incur significantly higher parameter counts and AICc values—and the simplest models at the bottom, which show diminishing returns in terms of validation loss. By focusing on this middle range, the selected architectures are expected to generalize well while remaining computationally efficient and less prone to overfitting.

In the final step, the five selected network architectures underwent an extensive evaluation to rigorously assess their stability and predictive performance. Each architecture was trained independently across 10 separate runs, with each training session lasting for 20,000 iterations. This repeated training procedure helped account for variability due to random initialization and stochastic optimization effects, ensuring the robust comparison of model reliability and convergence behavior. The results, including key performance metrics such as validation loss and generalization capability, from every network's best session are summarized in Table 4, with the training curves shown in Figure 2.

Table 4. Summary of best-performing network architectures.

Hidden Units	Nodes	Activation	Validation Loss (%)	No. of Parameters	AICc
3	5	LeakyReLu	7.28	146	1294
2	10	LeakyReLu	6.37	127	1155
2	9	Tanh	8.14	121	1236
2	8	ReLu	4.93	115	998
1	9	Tanh	8.27	76	1090

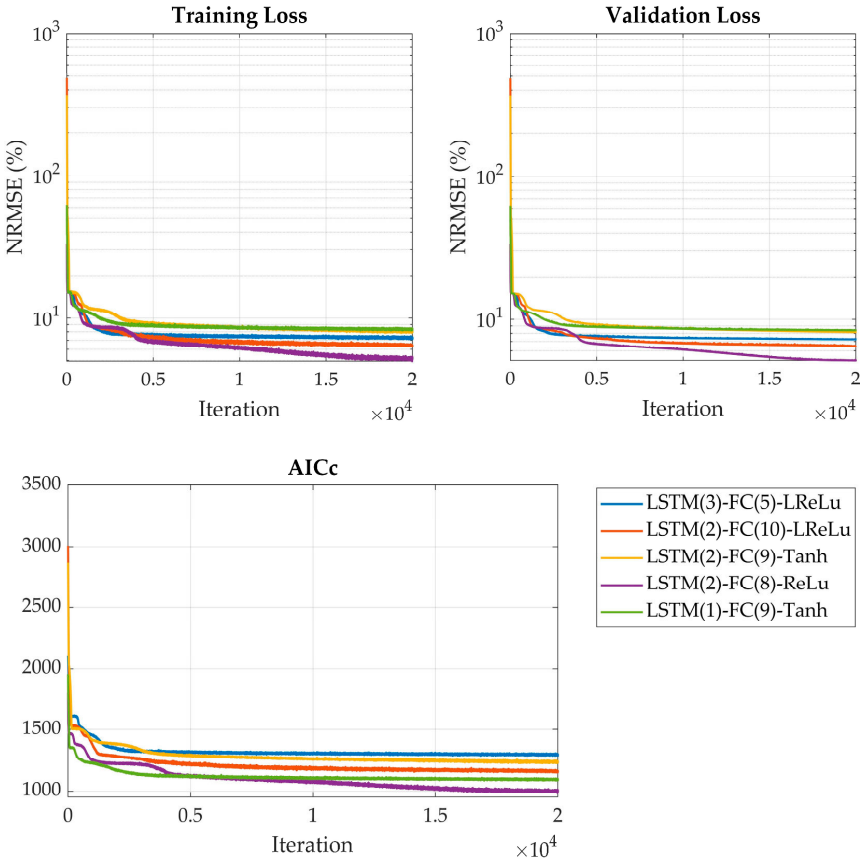


Figure 2. Training and validation loss curves, along with AICc evolution over 20,000 iterations, for the five best-performing network architectures.

Training results indicate that the optimal network architecture comprises 2 hidden units in the LSTM layer, 8 fully connected layer nodes, and utilizes an ReLU activation function. This configuration achieved the best performance, with a validation loss of 4.93% and the lowest AICc value of 998 among all evaluated models, demonstrating excellent balance between predictive accuracy and model complexity. Among activation functions, ReLU showed significantly better performance than both LeakyReLU and Tanh, aligning with the findings of Pinto et al. [9].

The effects of including a dropout layer with varying probabilities (ranging from 0.1 to 0.5) were systematically evaluated. The results showed that increasing the dropout rate consistently led to higher validation losses. Even a modest dropout probability of 0.1 negatively impacted model performance, suggesting that the selected network architecture is already sufficiently regularized and relatively simple. In such cases, the additional noise introduced by dropout appears to hinder learning rather than mitigate overfitting. This outcome implies that the model generalizes well without further regularization, and applying dropout may introduce unnecessary instability or lead to underfitting. Moreover, the observation may reflect the nature of the dataset, which is likely clean and of limited size—conditions under which every training instance is valuable and even minimal information loss (through activation masking) can reduce training efficiency.

3.2. Adapting the Hybrid Model to Qβ Dataset with Transfer Learning

Finally, the already trained optimal hybrid model was adapted to the new dataset (Qβ fermentations) using transfer learning. The LSTM layer weights were frozen, allowing the network to retain useful generalizations from the original training.

The resulting network demonstrated strong predictive performance, achieving a training loss of 3.18%, a validation loss of 3.53%, and a testing loss of 5.61%. In comparison, the purely historical data-based hybrid model from the previous step, resulted in a testing loss of 11.9% for Exp. 20, showcasing a more than 2-fold increase in predictive power. Together, these metrics indicate that the hybrid model effectively captured the underlying dynamics of the experimental dataset while maintaining generalization to unseen data (Figure 3).

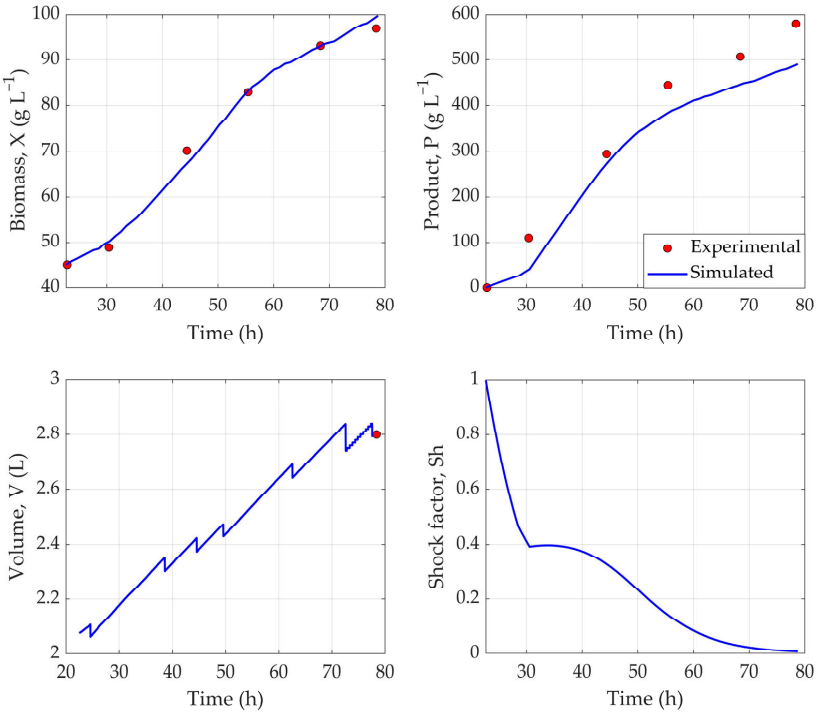


Figure 3. Predicted dynamic profiles of biomass (X), product (P), volume (V), and shock factor (Sh) using hybrid network trained via transfer learning. Experiment 20 was used as test dataset.

As illustrated in Figure 3, the hybrid model demonstrated strong predictive capability by closely capturing the dynamic behavior of the modeled variables throughout the fermentation process. Quantitatively, the model achieved an NRMSE of 1.68% for biomass concentration (X), indicating high accuracy in cell growth prediction. The product concentration (P), however, was slightly underestimated, resulting in a 9.54% error. Lower product prediction accuracy is to be expected when using such a small dataset, as the whole design space cannot be effectively explored in just 2–3 experiments. Using a larger, more comprehensive experimental dataset would help to improve the recombinant protein concentration prediction accuracy.

3.3. Hybrid MPC Experimental Validation

To evaluate the practical applicability and effectiveness of the hybrid MPC framework, experimental validation was carried out, controlling the feed rate in an actual fermentation run. This allowed for the assessment of the system's ability to predict and regulate key bioprocess variables in real-time. The results can be seen in Figure 4.

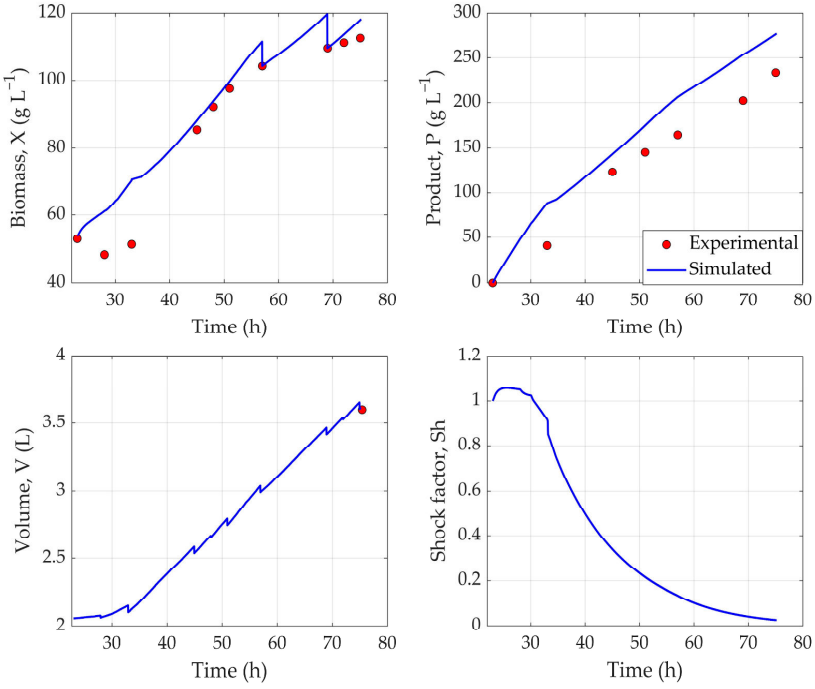


Figure 4. Predicted dynamic profiles of biomass (X), product (P), volume (V), and shock factor (Sh) generated by hybrid model during experimental validation (Exp. 21).

As can be seen in Figure 4, the tracking accuracy of cell biomass (X) by the hybrid model can be characterized as moderate. Notably, the model failed to adequately capture the physiological adaptation phase of *P. pastoris* to methanol during the first 8–12 h following induction—a period in which cellular growth is minimal or absent. As a result, predictions during this early phase were consistently higher than observed values. However, at the next sampling times (48–51 h), the model accurately predicts the cell biomass concentration, indicating that it does somewhat take into account the reduced growth during methanol adaptation. In the later stages of fermentation, the model exhibited a tendency to overestimate biomass concentration. To compensate for this discrepancy, manual

adjustments to the biomass profile were introduced during model retraining, following the availability of offline sampling data. Despite these limitations, the model achieved an overall NRMSE of 6.51% for biomass prediction, indicating acceptable predictive performance.

With respect to product concentration (P), the model similarly demonstrated a tendency to overestimate values across the fermentation timeline. This resulted in an average NRMSE of 14.65%, suggesting moderate accuracy in modeling recombinant protein production dynamics during an actual experimental validation run.

Although the predictive accuracy of the model was not exemplary—particularly in tracking certain state variables with high precision—it nonetheless demonstrated strong performance in its control functionality. Specifically, the model was able to generate robust and reliable feed rate profiles, effectively regulating substrate addition to maintain the desired specific growth rate throughout the fermentation process. This highlights the strength of the hybrid MPC framework in achieving process objectives, even in the presence of moderate prediction errors. The μ tracking results are presented in Figure 5.

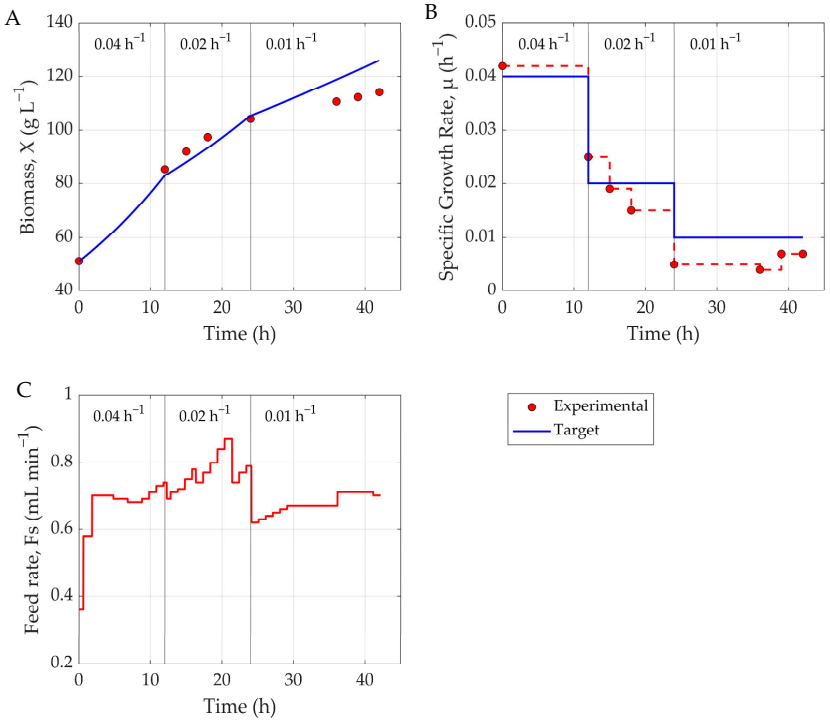


Figure 5. Comparison of target and experimental profiles during MPC-controlled fermentation. (**A**) Sampled biomass concentration versus target growth curve, (**B**) experimental versus target-specific growth rate, and (**C**) substrate feed rate dynamics. Vertical lines indicate setpoint changes in specific growth rate (μ).

As Figure 5 illustrates, the hybrid MPC system demonstrated good performance in tracking the specific growth rate setpoint over the course of the fermentation with a tracking error of 10.64% NRMSE. However, similar to the behavior observed in the prediction model, a slight deviation from the target biomass trajectory was noted during the final 12 h of the process—particularly after the specific growth rate was reduced to 0.001 h^{-1}. This deviation is likely attributable to the cytotoxic effects of methanol accumulation, which can impair cellular metabolism and inhibit biomass formation in the later stages of fermentation. This

phenomenon was further reflected in the growth rate tracking plot, where the measured specific growth rate was consistently lower than the target value during this terminal phase. As the purpose of the shock factor is to account for the cytotoxic effects of methanol during the early (adaptation) and late stages of cultivation, revisiting the equation may be warranted.

Overall, while minor discrepancies emerged toward the end of the process, the control system maintained a high degree of accuracy in regulating growth for the majority of the fermentation, underscoring its effectiveness and reliability under realistic bioprocessing conditions.

4. Discussion

This study demonstrates that Bayesian optimization efficiently identifies high-performing hybrid network architectures by focusing on promising hyperparameter regions, reducing the need for exhaustive searches [51]. A consistent architectural pattern emerged among the top models: an LSTM layer followed by one FC layer. This aligns with findings showing that temporal feature extraction via LSTM is critical in sequence modeling, especially in bioprocess contexts [52]. The FC layers then effectively map these temporal representations into nonlinear predictive outputs.

The grid search extended this exploration and employed AICc alongside validation loss to avoid overfitting. The Pareto front strategy is known to balance model complexity and accuracy [30]. Selecting architectures from the midsection of the Pareto front ensured efficient, generalizable models without an unnecessary computational burden. Robustness was confirmed through the repeated training of the selected architectures, mitigating variability from random initialization. The best-performing model—a minimal architecture with 2 LSTM units and 8 FC nodes—achieved the lowest validation loss (4.93%) and AICc (998), illustrating that compact models can deliver strong performance. This has practical implications for real-time applications, where computational efficiency and stability are critical. Interestingly, dropout regularization consistently worsened performance, in contrast to its typical effects in neural networks [53]. This suggests that the additional noise led to underfitting rather than preventing overfitting, corroborating findings in Bayesian LSTM studies [54]. In summary, the combined use of Bayesian optimization and grid search enabled the discovery of robust, efficient hybrid models—providing a practical strategy for developing bioprocess models.

The application of TL showcased its potential for bioprocess applications, as a robust hybrid process model, developed from a historical dataset, was successfully adapted to Qβ fermentations using only three experimental runs. By freezing the pretrained LSTM layer, the model effectively retained temporal feature representations learned from the original fermentation dataset—an approach aligned with the known benefits of sequential inductive transfer learning, where pretrained layers serve as robust feature extractors. Performance metrics indicate a robust predictive capability: a training loss of 3.18%, a validation loss of 3.53%, and a testing loss of 5.61%. Figure 3 illustrates that the hybrid model accurately tracks the dynamic behavior of key process variables throughout fermentation. Notably, the NRMSE of 1.68% for biomass reflects excellent predictive accuracy, while a 9.54% error for product concentration indicates the reliable estimation of recombinant protein output. The performance, especially for product estimation, could further be improved by using a larger experimental dataset that comprehensively presents the fermentation conditions. Overall, the strong performance of the Qβ model—trained on just three fermentations—demonstrates the hybrid model's generalizability and highlights transfer learning, particularly with LSTM architectures, as a powerful tool for rapid deployment in data-limited fermentation settings [26].

The effective tuning of MPC parameters is essential for achieving stable and responsive control in bioprocesses. In this study, MPC parameters were chosen based on the dynamics of *P. pastoris* fermentation, informed by operator experience, general rules of thumb, and hybrid model simulations. The typical prediction horizon for *P. pastoris* fermentations 2–3 times the dominant time constant (τ) is estimated by analyzing the process's dynamic response—typically from a step change in input. This involves measuring the time it takes the output to reach approximately 63% of its total change. The control horizon is typically 10–20% of the prediction horizon, and in this case is set to 1 h to match the hybrid model's prediction timestep [55]. In MPC, constraint handling ensures that control actions stay within operational limits, while signal smoothing prevents abrupt changes in control inputs. In this study, the substrate feed rate was constrained between 0.36 and 1.00 mL min^{-1} to ensure safe operation within the bioreactor's physical limits, particularly cooling capacity, based on operator experience. These constraints should be tailored for different bioreactor systems and scales. Process simulations with the hybrid model (digital twin) indicated that control signal smoothing was unnecessary, as no abrupt changes in the feed rate occurred. Nevertheless, smoothing should be considered when applying the model to new processes or when operating near the boundaries of the training design space.

To assess the practical utility of the hybrid MPC framework, a real-time control experiment was conducted, involving feed rate adjustment in an actual *P. pastoris* fermentation (Figure 4). The hybrid model achieved a biomass NRMSE of 6.51%, indicating moderate prediction accuracy. However, the model struggled during the 8–12 h adaptation phase following methanol induction—likely due to the challenge of capturing initial physiological delays during cell adaptation to methanol uptake. During later stages, persistent biomass overestimation, likely caused by the cumulative cytotoxic effect of methanol, required offline-informed manual corrections. This indicates that the shock factor (Sh) equation should be revisited to better capture the effects of methanol adaptation and cytotoxicity on cell growth. The model also tended to overestimate product concentration, with an average NRMSE of 14.65%, revealing moderate accuracy in tracking recombinant protein dynamics. As previously discussed, this could be addressed with a more comprehensive training dataset. Also, since product concentration cannot be measured at the line, a robust model—such as one based on inputs like biomass (X) or specific growth rate (μ), e.g., the Luedeking–Piret equation—could be used to correct the product inputs during MPC operation to further enhance predictive power.

Despite prediction inaccuracies, the hybrid MPC effectively maintained the target-specific growth rate via robust feed profiles throughout most of the fermentation (Figure 5), with an average tracking error of 10.64%. Only in the final 12 h—particularly after μ dropped to 0.001 h^{-1}—did deviations emerge. These were likely driven by methanol-induced cytotoxicity, which reduced cellular metabolism and hindered biomass formation. The hybrid MPC system showed strong real-time control capabilities, effectively regulating substrate feed to meet specific growth objectives throughout most of the fermentation. Prediction errors were manageable and did not significantly impair overall control performance. This framework can also support tasks beyond growth trajectory tracking—such as production maximization—but would require a more comprehensive dataset to achieve a reliable performance.

This validation confirms that hybrid MPC, built on data-driven hybrid models, is a viable strategy for real-world bioprocess control. Future work should aim to improve model representations of early-stage methanol adaptation, as well as cytotoxicity-induced dynamics towards the end of fermentation. This would further enhance both predictive fidelity and control reliability, supporting broader application in fermentation-based production workflows. Also, a reliable estimator for recombinant protein concentration should

be included in the MPC framework to provide accurate and reliable estimations to use for hybrid model retraining during operation.

5. Conclusions

This study presents a comprehensive framework for the hybrid modeling and control of *P. pastoris* fed-batch fermentations, integrating deep learning with model predictive control. The use of Bayesian optimization proved effective in identifying efficient and accurate hybrid neural network architectures, with consistent structural trends—namely, the inclusion of an LSTM layer followed by fully connected layers—emerging among top-performing models. A grid search guided by AICc and validation loss identified an optimal architecture, balancing accuracy and simplicity. This comprised an LSTM layer with 2 hidden units, followed by a fully connected layer with 8 nodes and ReLU activation. This configuration achieved the best performance, with a validation loss of 4.93% and the lowest AICc of 998.

Transfer learning was successfully used to adapt the hybrid model, originally trained on historical data, to a new Qβ fermentation dataset comprising just three experimental runs. The adapted model maintained strong predictive accuracy (5.61%) while preserving generalizable temporal features. This highlights the model's flexibility and potential for rapid adaptation to new, yet related, bioprocesses—an important capability in multiproduct biomanufacturing environments.

The experimental validation of the hybrid MPC framework demonstrated reliable real-time control of the fermentation process, despite moderate prediction errors in biomass (6.51%) and product (14.65%) concentration. Notably, the system maintained accurate regulation of the specific growth rate, with an average tracking error of just 10.64% throughout most of the process, deviating only in the final 10–12 h—underscoring its practical robustness.

In summary, this work establishes a robust, adaptable, and computationally efficient hybrid modeling approach for model predictive bioprocess control. The combination of automated architecture search, transfer learning, and MPC provides a scalable methodology for accelerating digital twin deployment in industrial biotechnology.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/fermentation11070411/s1, Scheme S1: Hybrid DNN Training Pseudocode.

Author Contributions: Conceptualization, E.B., V.G. and A.K.; methodology, E.B., V.G. and A.K.; software, E.B.; validation, E.B.; formal analysis, E.B.; investigation, E.B. and O.G.; resources, A.K.; data curation, E.B.; writing—original draft preparation, E.B.; writing—review and editing, E.B., V.G., A.K., O.G. and J.V.; visualization, E.B.; supervision, V.G. and A.K.; project administration, E.B. and J.V.; funding acquisition, J.V. and A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the EU Recovery and Resilience Facility within Project No 5.2.1.1.i.0/2/24/I/CFLA/003 "Implementation of consolidation and management changes at Riga Technical University, Liepaja University, Rezekne Academy of Technology, Latvian Maritime Academy and Liepaja Maritime College for the progress towards excellence in higher education, science and innovation" academic career doctoral grant No. 1094.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in Zenodo at 10.5281/zenodo.15855565.

Acknowledgments: We acknowledge Riga Technical University's HPC Center for providing access to their computing infrastructure. The authors would also like to acknowledge the contribution of Inara Akopjana for preparing seed inoculation cultures and performing SDS-PAGE analysis, and Janis Bogans for performing chromatography runs.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

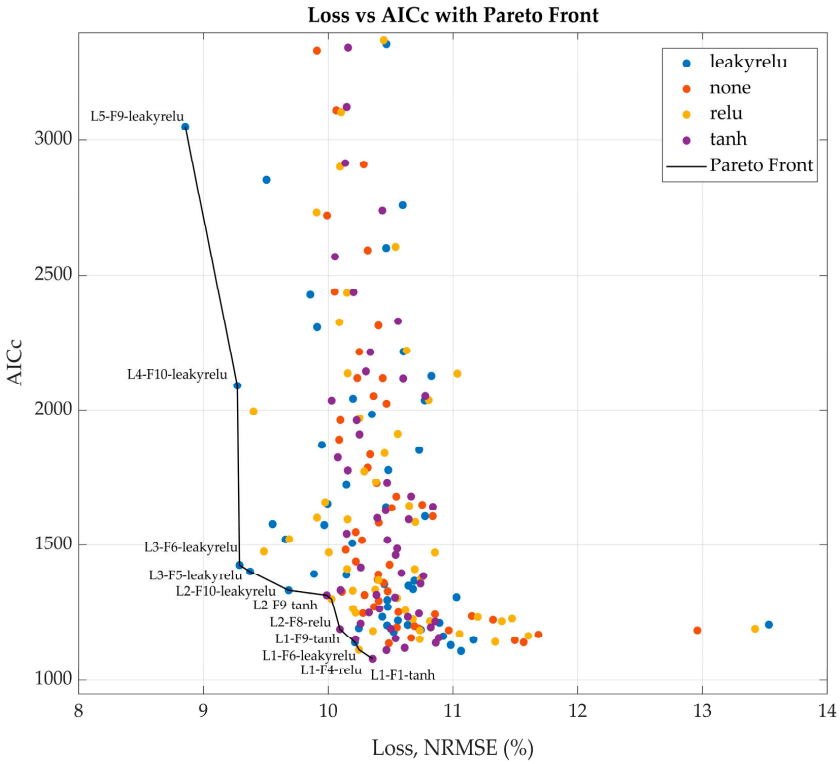


Figure A1. Pareto front, demonstrating the best model architectures, considering Loss and AICc. L stands for LSTM layer hidden units, F—for FC hidden nodes, followed by the activation function. For example, L4-F10-leakyrelu is an LSTM layer with 4 hidden units, followed by a FC layer with 10 nodes and Leaky ReLu activation.

References

1. Bailey, J.E. Mathematical Modeling and Analysis in Biochemical Engineering: Past Accomplishments and Future Opportunities. *Biotechnol. Prog.* **1998**, *14*, 8–20. [CrossRef] [PubMed]
2. Noll, P.; Henkel, M. History and Evolution of Modeling in Biotechnology: Modeling & Simulation, Application and Hardware Performance. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 3309–3323. [CrossRef] [PubMed]
3. Heinemann, M.; Panke, S. Synthetic Biology—Putting Engineering into Biology. *Bioinformatics* **2006**, *22*, 2790–2799. [CrossRef] [PubMed]
4. Lee, J.; Kao, H.-A.; Bagheri, B. Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics. In Proceedings of the International Conference on Industrial Informatics (INDIN), Porto Alegre, Brazil, 27–30 July 2014. [CrossRef]
5. Agharafeie, R.; Ramos, J.R.C.; Mendes, J.M.; Oliveira, R. From Shallow to Deep Bioprocess Hybrid Modeling: Advances and Future Perspectives. *Fermentation* **2023**, *9*, 922. [CrossRef]

6. Carbonell, P.; Radivojevic, T.; García Martín, H. Opportunities at the Intersection of Synthetic Biology, Machine Learning, and Automation. *ACS Synth. Biol.* **2019**, *8*, 1474–1477. [CrossRef] [PubMed]
7. Glassey, J.; Gernaey, K.V.; Clemens, C.; Schulz, T.W.; Oliveira, R.; Striedner, G.; Mandenius, C.F. Process Analytical Technology (PAT) for Biopharmaceuticals. *Biotechnol. J.* **2011**, *6*, 369–377. [CrossRef] [PubMed]
8. Sun, Y.; Nathan-Roberts, W.; Pham, T.D.; Otte, E.; Aickelin, U. Multi-Fidelity Gaussian Process for Biomanufacturing Process Modeling with Small Data. *arXiv* **2022**, arXiv:2211.14493. [CrossRef]
9. Pinto, J.; Ramos, J.R.C.; Costa, R.S.; Oliveira, R. Hybrid Deep Modeling of a GS115 (Mut+) *Pichia Pastoris* Culture with State–Space Reduction. *Fermentation* **2023**, *9*, 643. [CrossRef]
10. Pinto, J.; Mestre, M.; Ramos, J.; Costa, R.S.; Striedner, G.; Oliveira, R. A General Deep Hybrid Model for Bioreactor Systems: Combining First Principles with Deep Neural Networks. *Comput. Chem. Eng.* **2022**, *165*, 107952. [CrossRef]
11. Ramos, J.R.C.; Pinto, J.; Poiares-Oliveira, G.; Peeters, L.; Dumas, P.; Oliveira, R. Deep Hybrid Modeling of a HEK293 Process: Combining Long Short-Term Memory Networks with First Principles Equations. *Biotechnol. Bioeng.* **2024**, *121*, 1554–1568. [CrossRef] [PubMed]
12. Ferreira, A.R.; Dias, J.M.L.; Von Stosch, M.; Clemente, J.; Cunha, A.E.; Oliveira, R. Fast Development of *Pichia Pastoris* GS115 Mut$^+$ Cultures Employing Batch-to-Batch Control and Hybrid Semi-Parametric Modeling. *Bioprocess. Biosyst. Eng.* **2014**, *37*, 629–639. [CrossRef] [PubMed]
13. Narayanan, H.; von Stosch, M.; Feidl, F.; Sokolov, M.; Morbidelli, M.; Butté, A. Hybrid Modeling for Biopharmaceutical Processes: Advantages, Opportunities, and Implementation. *Front. Chem. Eng.* **2023**, *5*, 1157889. [CrossRef]
14. Schweidtmann, A.M.; Zhang, D.; von Stosch, M. A Review and Perspective on Hybrid Modeling Methodologies. *Digit. Chem. Eng.* **2024**, *10*, 100136. [CrossRef]
15. von Stosch, M.; Oliveira, R.; Peres, J.; Feyo de Azevedo, S. Hybrid Semi-Parametric Modeling in Process Systems Engineering: Past, Present and Future. *Comput. Chem. Eng.* **2014**, *60*, 86–101. [CrossRef]
16. Psichogios, D.C.; Ungar, L.H. A Hybrid Neural Network-First Principles Approach to Process Modeling. *AIChE J.* **1992**, *38*, 1499–1511. [CrossRef]
17. von Stosch, M.; Davy, S.; Francois, K.; Galvanauskas, V.; Hamelink, J.M.; Luebbert, A.; Mayer, M.; Oliveira, R.; O'Kennedy, R.; Rice, P.; et al. Hybrid Modeling for Quality by Design and PAT-Benefits and Challenges of Applications in Biopharmaceutical Industry. *Biotechnol. J.* **2014**, *9*, 719–726. [CrossRef] [PubMed]
18. Bangi, M.S.F.; Kao, K.; Kwon, J.S.-I. Physics-Informed Neural Networks for Hybrid Modeling of Lab-Scale Batch Fermentation for β-Carotene Production Using *Saccharomyces cerevisiae*. *Chem. Eng. Res. Des.* **2022**, *179*, 415–423. [CrossRef]
19. Mhaskar, H.N.; Poggio, T. Deep vs. Shallow Networks: An Approximation Theory Perspective. *Anal. Appl.* **2016**, *14*, 829–848. [CrossRef]
20. Helleckes, L.M.; Hemmerich, J.; Wiechert, W.; von Lieres, E.; Grünberger, A. Machine Learning in Bioprocess Development: From Promise to Practice. *Trends Biotechnol.* **2023**, *41*, 817–835. [CrossRef] [PubMed]
21. Alzubaidi, L.; Bai, J.; Al-Sabaawi, A.; Santamaría, J.; Albahri, A.S.; Al-dabbagh, B.S.N.; Fadhel, M.A.; Manoufali, M.; Zhang, J.; Al-Timemy, A.H.; et al. A Survey on Deep Learning Tools Dealing with Data Scarcity: Definitions, Challenges, Solutions, Tips, and Applications. *J. Big Data* **2023**, *10*, 46. [CrossRef]
22. How Invert Uses Meta-Learning to Leverage Old Bioprocess Data. Available online: https://invertbio.com/blogs/how-invert-uses-meta-learning-to-leverage-old-data-and-make-better-biomanufacturing-predictions (accessed on 25 June 2025).
23. Duong-Trung, N.; Born, S.; Kim, J.W.; Schermeyer, M.-T.; Paulick, K.; Borisyak, M.; Cruz-Bournazou, M.N.; Werner, T.; Scholz, R.; Schmidt-Thieme, L.; et al. When Bioprocess Engineering Meets Machine Learning: A Survey from the Perspective of Automated Bioprocess Development. *Biochem. Eng. J.* **2022**, *190*, 108764. [CrossRef]
24. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
25. Weiss, K.; Khoshgoftaar, T.M.; Wang, D.D. A Survey of Transfer Learning. *J. Big Data* **2016**, *3*, 9. [CrossRef]
26. Rogers, A.W.; Vega-Ramon, F.; Yan, J.; del Río-Chanona, E.A.; Jing, K.; Zhang, D. A Transfer Learning Approach for Predictive Modeling of Bioprocesses Using Small Data. *Biotechnol. Bioeng.* **2022**, *119*, 411–422. [CrossRef] [PubMed]
27. Kensert, A.; Harrison, P.J.; Spjuth, O. Transfer Learning with Deep Convolutional Neural Networks for Classifying Cellular Morphological Changes. *SLAS Discov.* **2019**, *24*, 466–475. [CrossRef] [PubMed]
28. Iman, M.; Rasheed, K.; Arabnia, H.R. A Review of Deep Transfer Learning and Recent Advancements. *Technologies* **2022**, *11*, 40. [CrossRef]
29. Narayanan, H.; Luna, M.; Sokolov, M.; Butté, A.; Morbidelli, M. Hybrid Models Based on Machine Learning and an Increasing Degree of Process Knowledge: Application to Cell Culture Processes. *Ind. Eng. Chem. Res.* **2022**, *61*, 8658–8672. [CrossRef]
30. Liashchynskyi, P.; Liashchynskyi, P. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. *arXiv* **2019**, arXiv:1912.06059v1. [CrossRef]
31. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. *arXiv* **2012**, arXiv:1206.2944. [CrossRef]

32. Bolmanis, E.; Dubencovs, K.; Suleiko, A.; Vanags, J. Model Predictive Control—A Stand Out among Competitors for Fed-Batch Fermentation Improvement. *Fermentation* **2023**, *9*, 206. [CrossRef]
33. Mears, L.; Stocks, S.M.; Sin, G.; Gernaey, K.V. A Review of Control Strategies for Manipulating the Feed Rate in Fed-Batch Fermentation Processes. *J. Biotechnol.* **2017**, *245*, 34–46. [CrossRef] [PubMed]
34. Rashedi, M.; Khodabandehlou, H.; Demers, M.; Wang, T.; Garvin, C. Model Predictive Controller Design for Bioprocesses Based on Machine Learning Algorithms. In Proceedings of the 13th IFAC Symposium on Dynamics and Control of Process Systems, Including Biosystems DYCOPS 2022, Busan, Republic of Korea, 14–17 June 2022; pp. 45–50. [CrossRef]
35. Eslami, T.; Jungbauer, A. Control Strategy for Biopharmaceutical Production by Model Predictive Control. *Biotechnol. Prog.* **2024**, *40*, e3426. [CrossRef] [PubMed]
36. Rathore, A.S.; Mishra, S.; Nikita, S.; Priyanka, P. Bioprocess Control: Current Progress and Future Perspectives. *Life* **2021**, *11*, 557. [CrossRef] [PubMed]
37. Lyubenova, V.; Ignatova, M.; Zoteva, D.; Roeva, O. Model-Based Adaptive Control of Bioreactors—A Brief Review. *Mathematics* **2024**, *12*, 2205. [CrossRef]
38. Albino, M.; Gargalo, C.L.; Nadal-Rey, G.; Albæk, M.O.; Krühne, U.; Gernaey, K. V Hybrid Modeling for On-Line Fermentation Optimization and Scale-Up: A Review. *Processes* **2024**, *12*, 1635. [CrossRef]
39. Mahanty, B. Hybrid Modeling in Bioprocess Dynamics: Structural Variabilities, Implementation Strategies, and Practical Challenges. *Biotechnol. Bioeng.* **2023**, *120*, 2072–2091. [CrossRef] [PubMed]
40. De Jong, H.; Casagranda, S.; Giordano, N.; Cinquemani, E.; Ropers, D.; Geiselmann, J.; Gouzé, J.L. Mathematical Modelling of Microbes: Metabolism, Gene Expression and Growth. *J. R. Soc. Interface* **2017**, *14*, 20170502. [CrossRef] [PubMed]
41. Tsopanoglou, A.; Jiménez del Val, I. Moving towards an Era of Hybrid Modelling: Advantages and Challenges of Coupling Mechanistic and Data-Driven Models for Upstream Pharmaceutical Bioprocesses. *Curr. Opin. Chem. Eng.* **2021**, *32*, 100691. [CrossRef]
42. Galvanauskas, V.; Simutis, R.; Lübbert, A. Hybrid Modeling of Biochemical Processes. In *Hybrid Modeling in Process Industries*, 1st ed.; Glassey, J., von Stosch, M., Eds.; CRC Press: Boca Raton, FL, USA, 2018; pp. 89–127.
43. Freivalds, J.; Dislers, A.; Ose, V.; Skrastina, D.; Cielens, I.; Pumpens, P.; Sasnauskas, K.; Kazaks, A. Assembly of Bacteriophage Qβ Virus-like Particles in Yeast *Saccharomyces cerevisiae* and *Pichia pastoris*. *J. Biotechnol.* **2006**, *123*, 297–303. [CrossRef] [PubMed]
44. Invitrogen Corporation Pichia Fermentation Process Guidelines. Available online: https://tools.thermofisher.com/content/sfs/manuals/pichiaferm_prot.pdf (accessed on 25 June 2025).
45. Grigs, O.; Bolmanis, E.; Galvanauskas, V. Application of In-Situ and Soft-Sensors for Estimation of Recombinant *P. Pastoris* GS115 Biomass Concentration: A Case Analysis of HBcAg (Mut^+) and HBsAg (Mut^S) Production Processes under Varying Conditions. *Sensors* **2021**, *21*, 1268. [CrossRef] [PubMed]
46. Bolmanis, E.; Uhlendorff, S.; Pein-Hackelbusch, M.; Galvanauskas, V.; Grigs, O. Anomaly Detection and Removal Strategies for In-Line Permittivity Sensor Signal Used in Bioprocesses. *Front. Bioeng. Biotechnol.* **2025**, *13*, 1609369. [CrossRef]
47. Bolmanis, E.; Grigs, O.; Kazaks, A.; Galvanauskas, V. High-Level Production of Recombinant HBcAg Virus-like Particles in a Mathematically Modelled *P. pastoris* GS115 Mut^+ Bioreactor Process under Controlled Residual Methanol Concentration. *Bioprocess. Biosyst. Eng.* **2022**, *45*, 1447–1463. [CrossRef] [PubMed]
48. Bolmanis, E.; Bogans, J.; Akopjana, I.; Suleiko, A.; Kazaka, T.; Kazaks, A. Production and Purification of Soy Leghemoglobin from *Pichia Pastoris* Cultivated in Different Expression Media. *Processes* **2023**, *11*, 3215. [CrossRef]
49. Lee, J.; Ramirez, W.F. Optimal Fed-Batch Control of Induced Foreign Protein Production by Recombinant Bacteria. *AIChE J.* **1994**, *40*, 899–907. [CrossRef]
50. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
51. White, C.; Ai, A.; Ai, C.; Neiswanger, W.; Savani, Y.; Ai, Y. BANANAS: Bayesian Optimization with Neural Architectures for Neural Architecture Search. *arXiv* **2020**, arXiv:1910.11858. [CrossRef]
52. Blume, S.; Benedens, T.; Schramm, D. Hyperparameter Optimization Techniques for Designing Software Sensors Based on Artificial Neural Networks. *Sensors* **2021**, *21*, 8435. [CrossRef] [PubMed]
53. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *arXiv* **2016**, arXiv:1506.02142. [CrossRef]
54. Vien, B.S.; Kuen, T.; Rose, L.R.F.; Chiu, W.K. Optimisation and Calibration of Bayesian Neural Network for Probabilistic Prediction of Biogas Performance in an Anaerobic Lagoon. *Sensors* **2024**, *24*, 2537. [CrossRef] [PubMed]
55. Keshav, S. Model Predictive Control. Available online: https://svr-sk818-web.cl.cam.ac.uk/keshav/wiki/images/e/e4/Model_Predictive_Control.pdf (accessed on 10 July 2025).

Emīls Bolmanis dzimis 1993. gadā Dobelē. Rīgas Tehniskajā universitātē ieguvis bakalaura (2019) un maģistra (2021, ar izcilību) grādu ķīmijas tehnoloģijā. Strādājis Latvijas Valsts koksnes ķīmijas institūta Bioinženierijas laboratorijā. Kopš 2018. gada strādā Latvijas Biomedicīnas pētījumu un studiju centrā (BMC), ieņemot laboranta un zinātniskā asistenta amatus. Patlaban ir pētnieks profesora K. Tāra grupā un BMC Rekombinanto biotehnoloģiju servisa centrā. Zinātniskās intereses saistītas ar biotehnoloģisko procesu inženieriju un fermentāciju optimizāciju ar uzsvaru uz reāllaika procesu monitoringu, datos balstītu modelēšanu (hibrīdā modelēšana, mašīnmācīšanās algoritmi) un prognozējošu vadību (MPC).

Emīls Bolmanis was born in 1993 in Dobele. He received his Bachelor's (2019) and Master's (2021, with distinction) degrees in Chemical Technology from Riga Technical University. Previously, he worked at the Laboratory of Bioengineering at the Latvian State Institute of Wood Chemistry. Since 2018, he has been a laboratory and research assistant at the Latvian Biomedical Research and Study Centre (BMC). Currently, he is a researcher in Professor K. Tars' group and in the BMC Core Facility of Recombinant Protein Biotechnology. His scientific interests include biotechnological process engineering and fermentation optimisation, with an emphasis on real-time process monitoring, data-driven modelling (hybrid modelling, machine learning algorithms), and predictive control (MPC).