RIGA TECHNICAL UNIVERSITY

Emīls Bolmanis

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

Summary of the Doctoral Thesis

RTU Press
Riga 2026

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

Emīls Bolmanis
Doctoral Student of the Study Programme "Chemistry, Materials Science and Engineering"

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

Summary of the Doctoral Thesis

Scientific supervisors
Senior Researcher Dr. biol.
ANDRIS KAZĀKS

Professor Dr. sc. ing.
JURIS VANAGS

Consultant
Professor Dr. sc. ing.
VYTAUTAS GALVANAUSKAS

RTU Press
Riga 2026

Bolmanis, E. Towards Industry 4.0: Real-Time Monitoring, Hybrid Modeling and Predictive Control in *Pichia Pastoris* Fermentations. Summary of the Doctoral Thesis. Riga: RTU Press, 2026. 49 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

https://doi.org/10.7250/9789934372810
ISBN 978-9934-37-281-0 (pdf)

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

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.

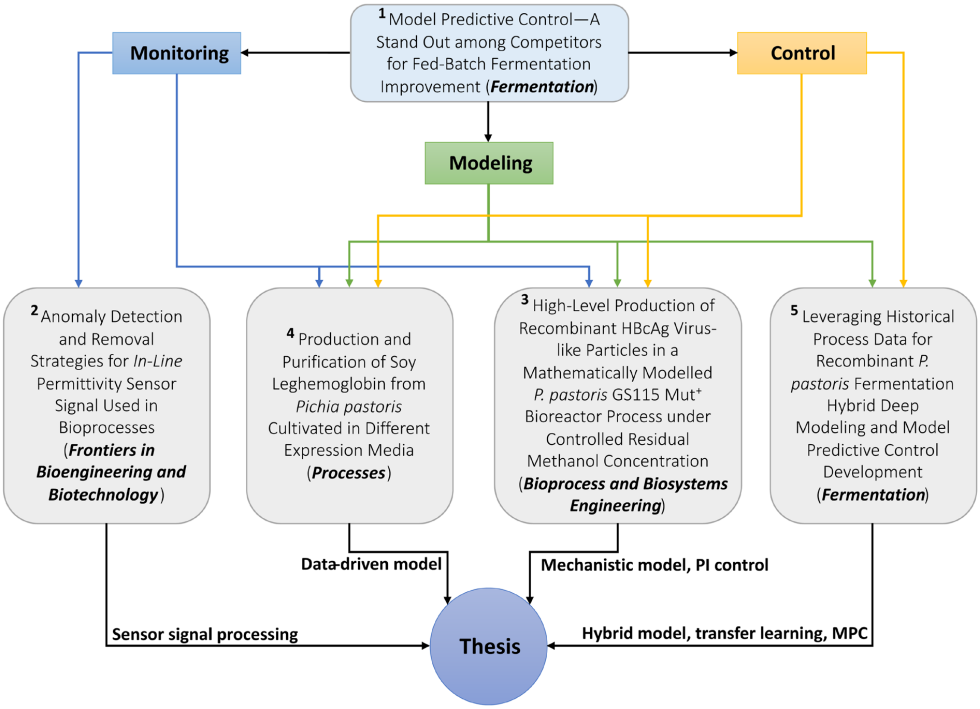
Practical Significance

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.

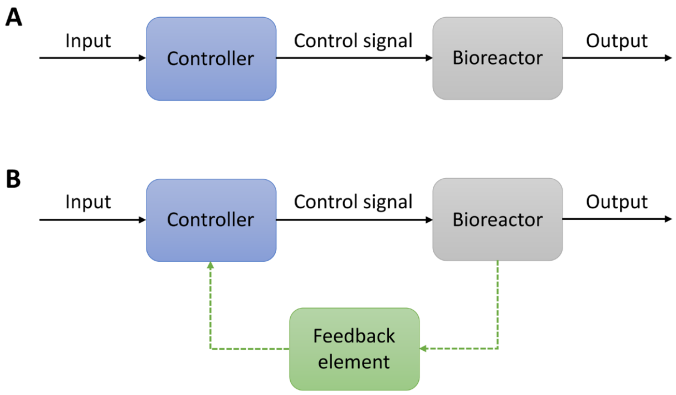


Fig. 1.1. Schematic of open-loop (A) and closed-loop (B) control architectures.

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.

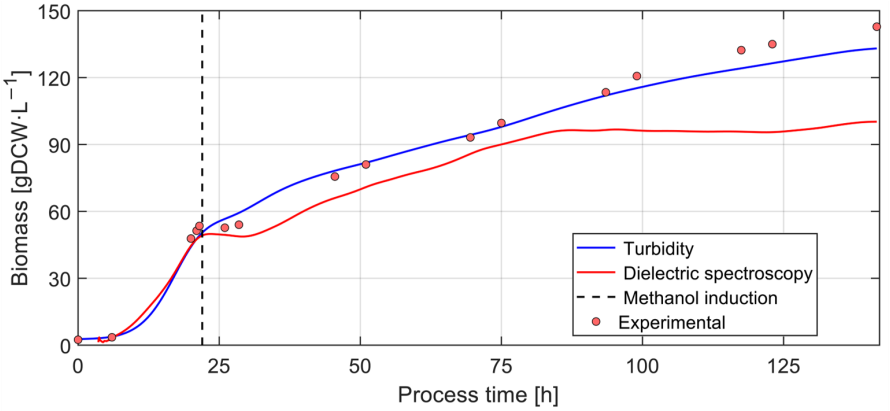


Fig. 2.1. Comparison of optical (turbidity) and dielectric spectroscopy sensor probe cell biomass measurements in a *P. pastoris* fermentation.

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.

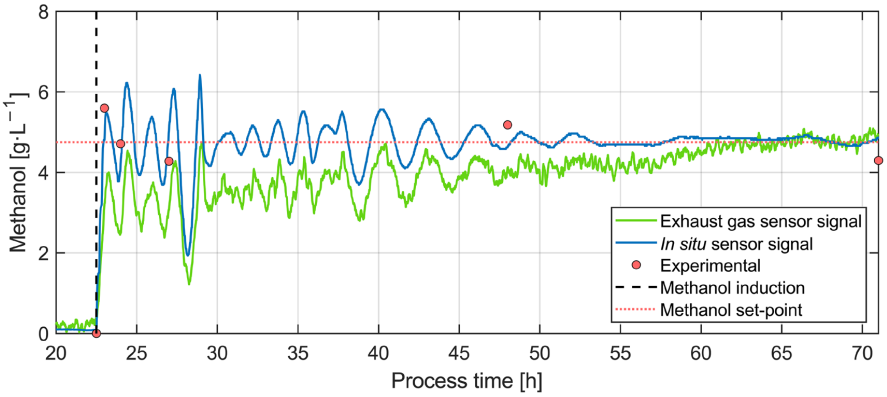


Fig. 2.2. Comparison of exhaust gas and *in situ* methanol sensor performance in a *P. pastoris* fermentation with residual methanol concentration control.

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.

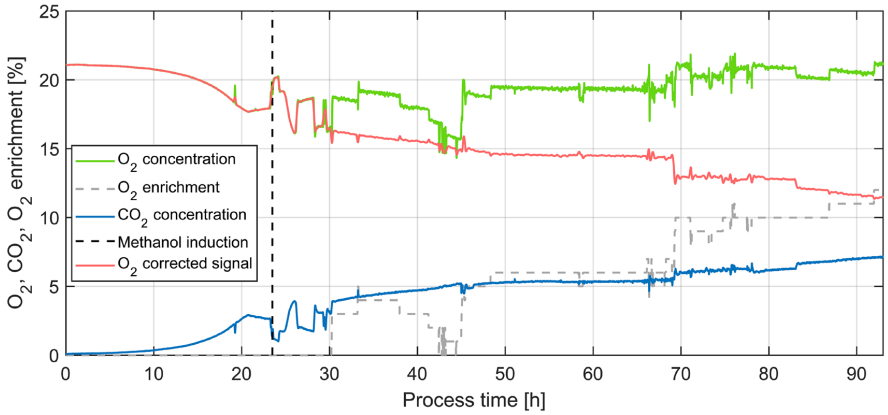


Fig. 2.3. O_2 and CO_2 concentrations in the bioreactor exhaust gas stream during fermentation and inlet air enrichment with pure O_2.

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.

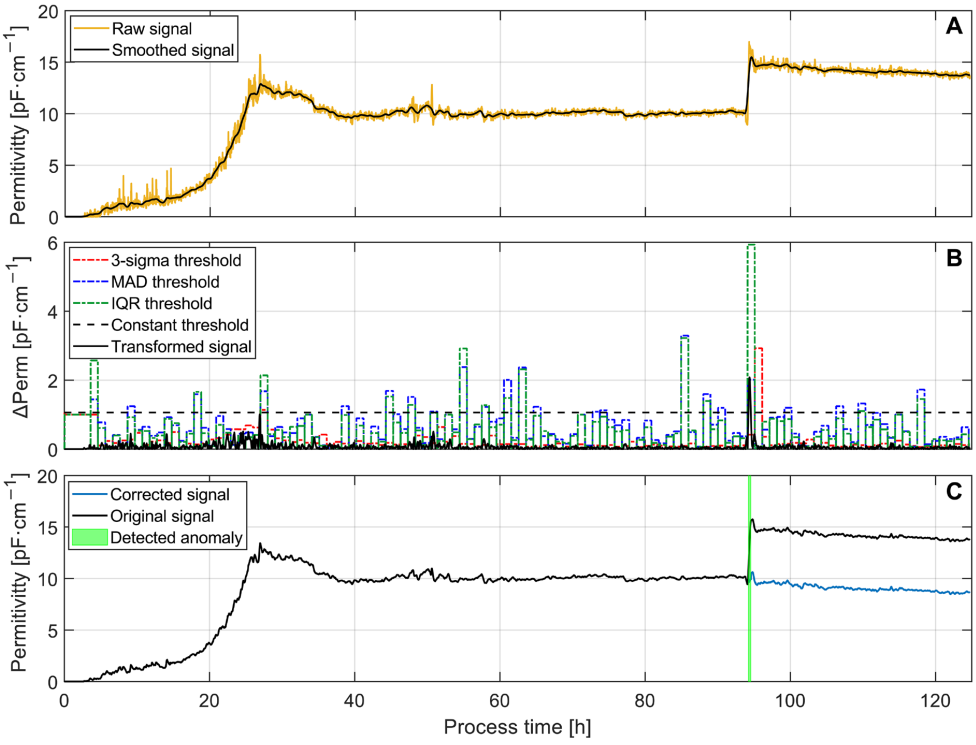


Fig. 2.4. Visual overview of the algorithm's performance: A – signal enhancement through preprocessing, B – comparison of anomaly detection method thresholds applied to the DRA-transformed permittivity signal, and C – performance in a real-time process simulation.

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.

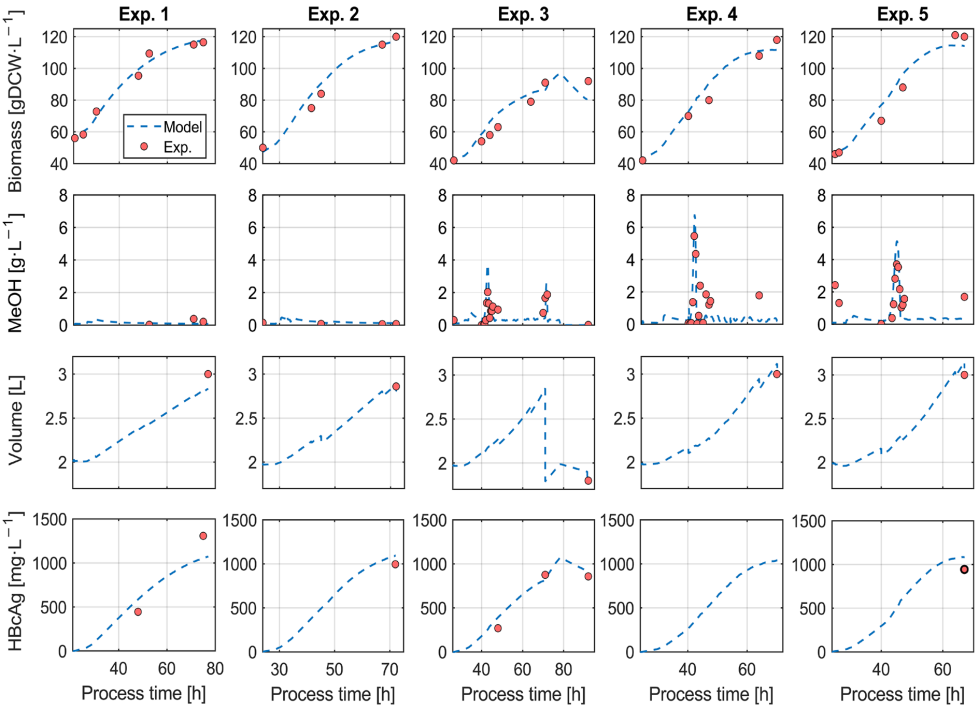


Fig. 3.1. Mechanistic modeling results, showing cell biomass, methanol concentration, culture volume and product concentration dynamics from five *P. pastoris* fermentations producing HBcAg.

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.

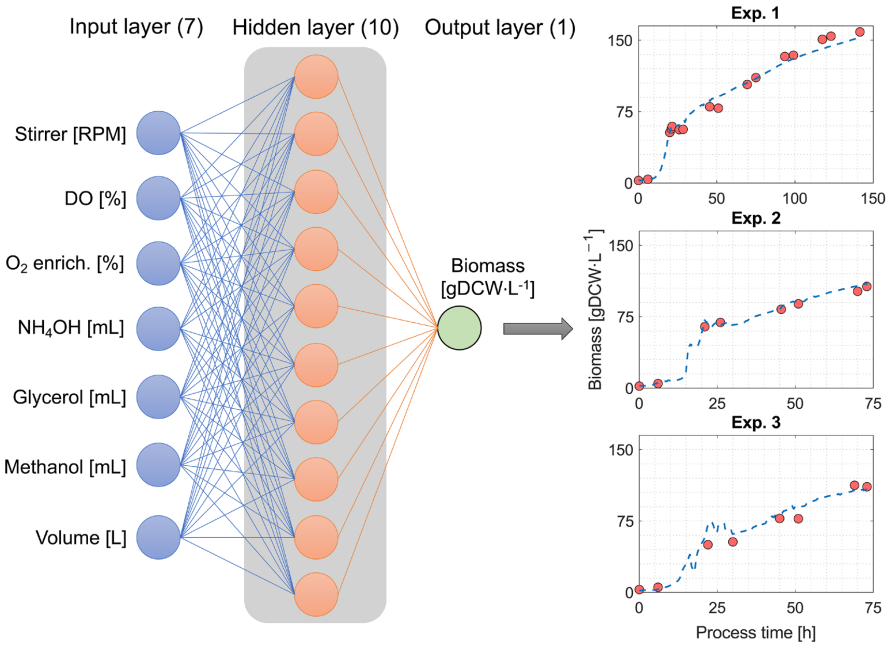


Fig. 3.2. Structure and the biomass modeling results of the developed ANN-based biomass soft sensor.

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.

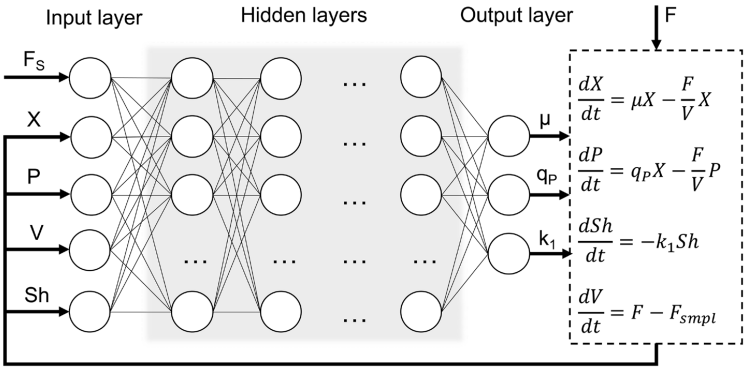


Fig. 3.3. General structure of the developed hybrid process model.

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.

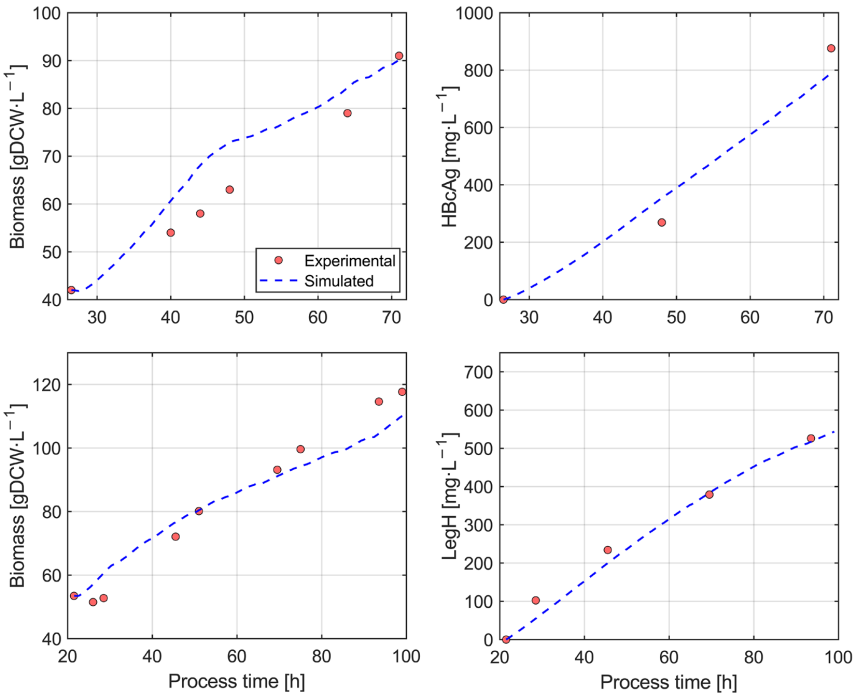


Fig. 3.4. Cell biomass and product concentrations estimated using the selected hybrid process model for HBcAg (upper row) and LegH (lower row).

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.

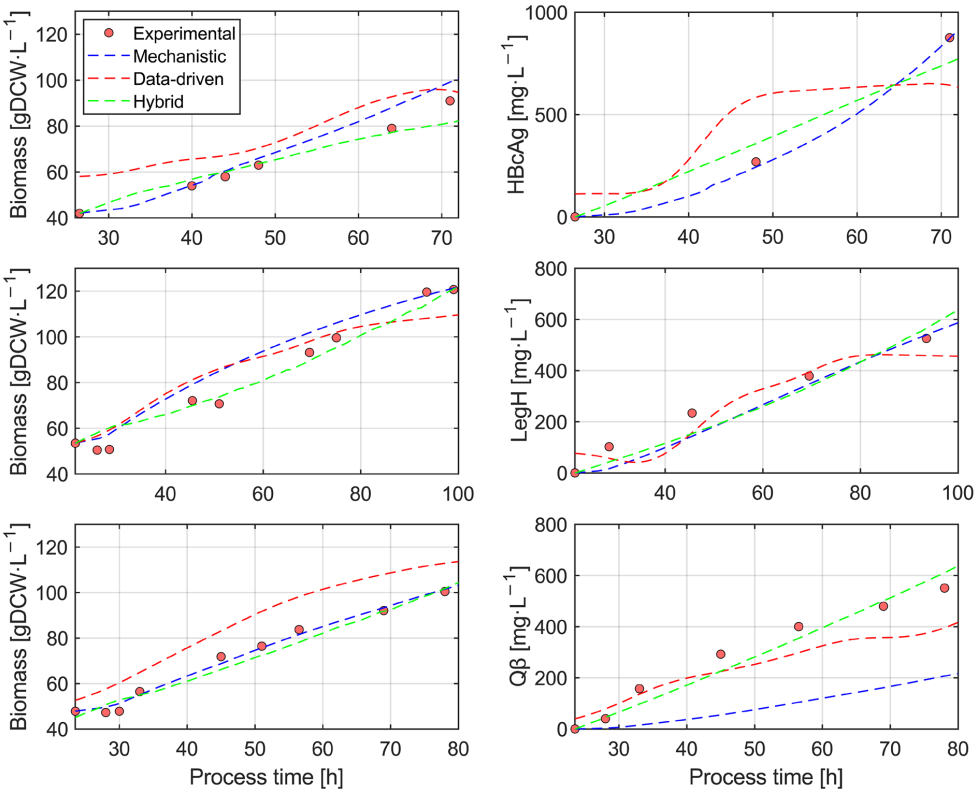


Fig. 3.5. Comparison of model performance in estimating biomass and product concentrations for one experiment from each recombinant protein producer.

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.

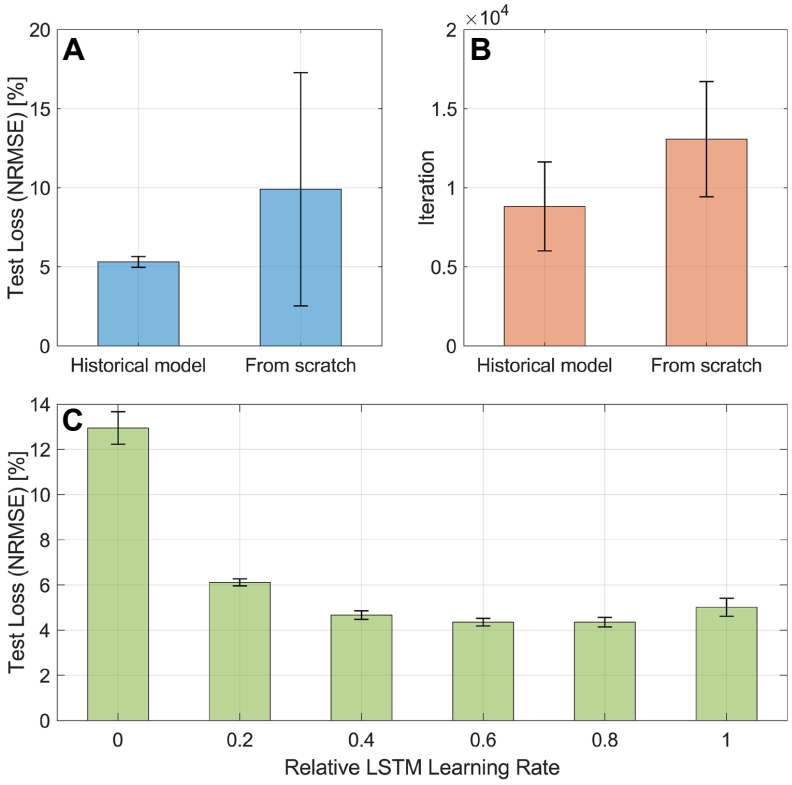


Fig. 3.6. Mean test losses (A) and number of training iterations (B) for the hybrid Qβ process model when adapted from a historical model or trained from scratch, and mean test loss as a function of the LSTM layer relative learning rate (C).

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).

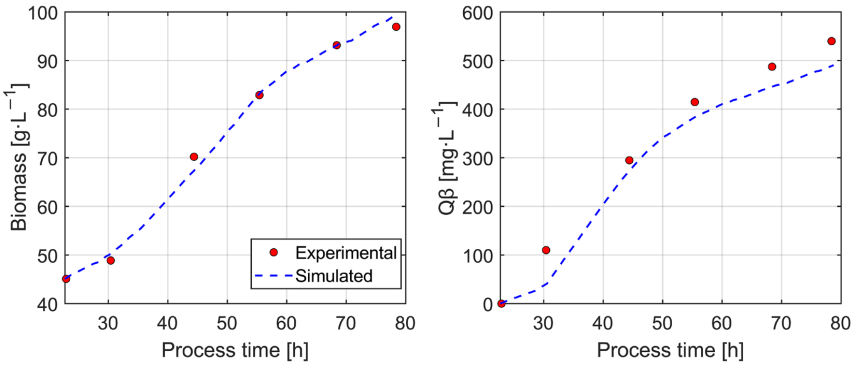


Fig. 3.7. Prediction of the dynamic profiles of cell biomass and product modeled with the trained hybrid process model.

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.

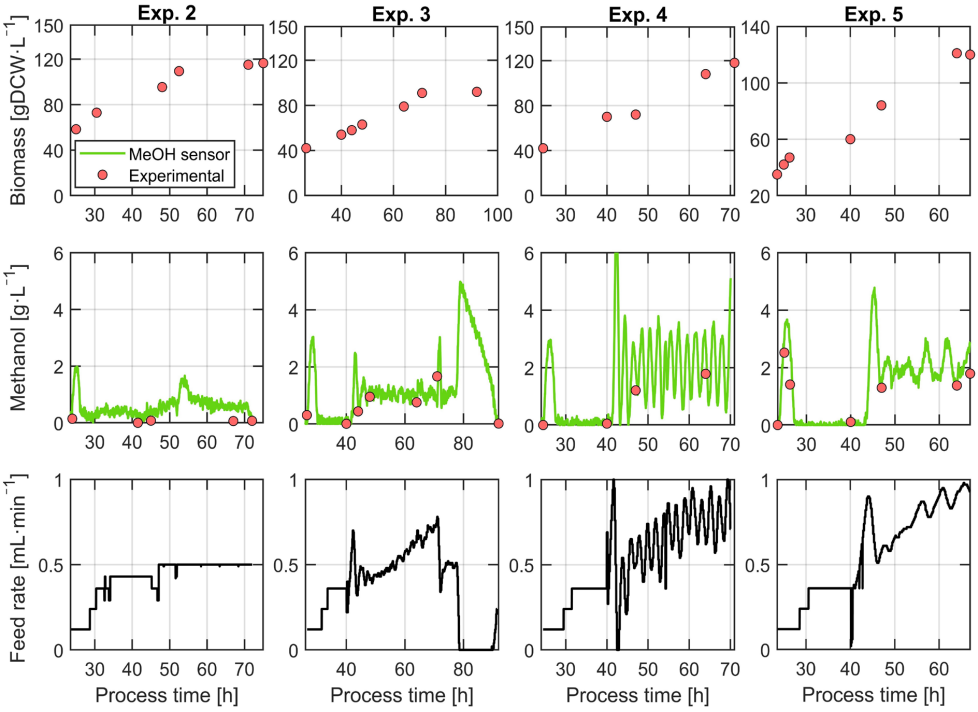


Fig. 4.1. Cell biomass, residual methanol concentration and methanol feed rate dynamics during *P. pastoris* fermentation experiments with residual methanol PI-control.

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}:

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

subject to the hybrid model dynamics:

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

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).

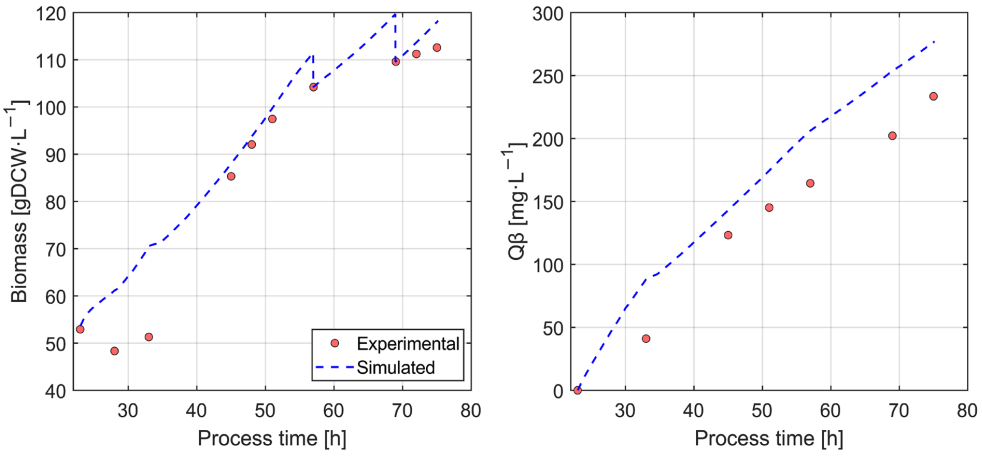


Fig. 4.2. Hybrid MPC predicted vs. experimentally determined biomass and product concentrations during experimental validation.

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.

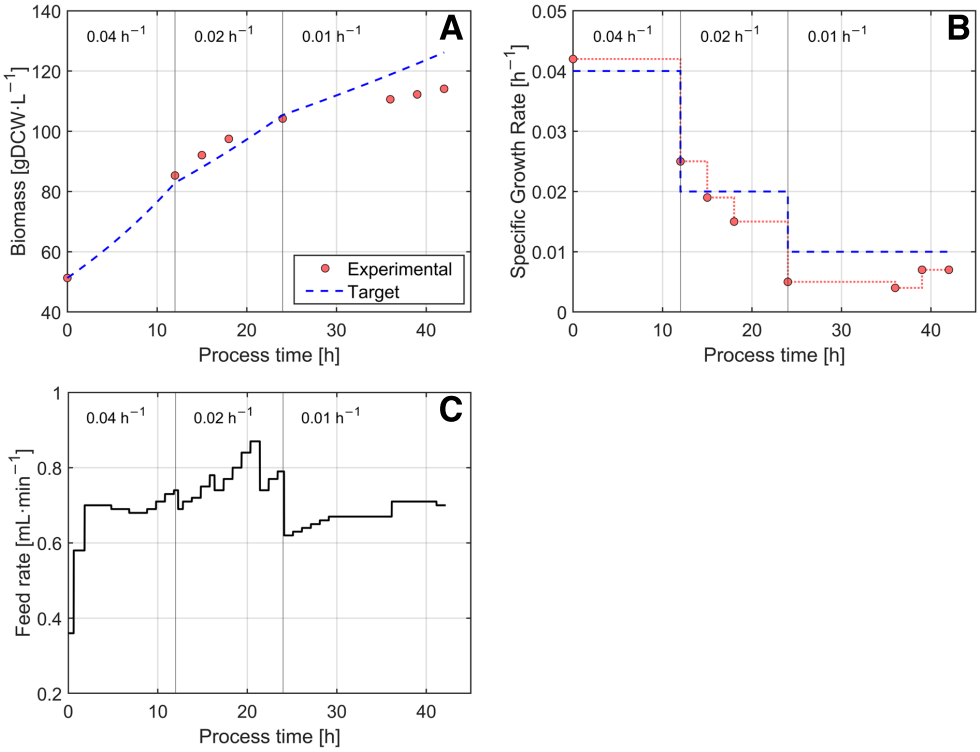


Fig. 4.3. Target vs experimental biomass growth (A), specific growth rate (B) and feed rate plots (C). Vertical lines indicate μ shifts.

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.

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).