

ADVANCING RESOURCE RECOVERY FROM WASTEWATER

**MECHANISTIC MODELING, HYBRID SYSTEM IDENTIFICATION,
AND ADAPTIVE PREDICTIVE CONTROL**

ADVANCING RESOURCE RECOVERY FROM WASTEWATER

**MECHANISTIC MODELING, HYBRID SYSTEM IDENTIFICATION,
AND ADAPTIVE PREDICTIVE CONTROL**

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Wednesday 22, January 2025 at 17:30 o'clock

by

Ali MORADVANDI

Master of Science in Chemical Engineering,
Sharif University of Technology, Iran,
born in Tehran, Iran.

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. B. De Schutter	Delft University of Technology, <i>promotor</i>
Dr. ir. E. Abraham,	Delft University of Technology, <i>promotor</i>
Dr. ir. R.E.F. Lindeboom,	Delft University of Technology, <i>copromotor</i>

Independent members:

Prof. dr. J.B. Jørgensen	Technical University of Denmark
Prof. dr. ir. T.A.E. Oomen	Eindhoven University of Technology
Prof. dr. S.E. Vlaeminck	University of Antwerp
Prof. dr. Z. Kapelan	Delft University of Technology
Prof. dr. ir. J.B. van Lier	Delft University of Technology, <i>reserved member</i>

The research presented in this thesis was performed in collaboration between the Sanitary Engineering Section, Department of Water Management, Faculty of Civil Engineering, and the Delft Center for Systems and Control, Faculty of Mechanical Engineering, at Delft University of Technology.

Keywords: Process control; Model predictive control; Adaptive control; Mechanistic modeling; System identification; Hybrid systems; Resource recovery; Anaerobic digestion; Purple bacteria.

Printed by: ipskampprinting

Cover: Variations in systems stabilized through control strategies.

Copyright © 2024 by A. Moradvandi

ISBN 978-94-6384-702-5

An electronic copy of this dissertation is available at
<https://repository.tudelft.nl/>.

"The first gulp from the glass of natural sciences will turn you into an atheist, but at the bottom of the glass, God is waiting for you."

Heisenberg

SUMMARY

This PhD thesis explores advancements in resource recovery from wastewater. Resource recovery from wastewater has gained significant attention in the last decades as besides water treatment. A variety of biological treatment technologies may, next to clean water, produce valuable by-products like biogas, microbial fertilizers, bio-plastic feedstock, and animal feed ingredients. More recently Purple Phototrophic Bacteria (PPB) raceway reactors have come to the limelight due to their waste-to-resource potential and low operational costs, especially as secondary treatment processes that enhance treatment efficiency, while recovering nutrients and organics, including nitrogen, phosphorous, and carbon. A key advantage of PPB cultivation is the high efficiency with which these microorganisms utilize volatile fatty acids (VFAs), a key intermediate in the anaerobic degradation of organic residues, as a carbon source. VFAs accumulation inside anaerobic digesters (ADs) results from the inhibition of the methane-producing methanogenesis. In anaerobic fermentation, the operational conditions can be deliberately steered towards inhibition of methanogenesis and thus VFA accumulation, and in malfunctioning ADs, VFA accumulates as a consequence of inadequate operational practices. The complexity of choosing the operational strategy lies in the fact that the operation of both AD and PPB raceway reactors is characterized by complicated responses and time-lags in the desired output objectives. We thus focus on optimizing PPB cultivation in raceway reactors, particularly when integrated with ADs that may otherwise produce incomplete byproducts, and on advancing resource recovery to assess whether AD can reliably serve its primary function in biogas production.

In this perspective, three parallel approaches are investigated: mechanistic modeling, hybrid system identification, and adaptive predictive control. These approaches address the challenges of bioprocess variability, operational fluctuations, limited operator availability, measuring and monitoring constraints, and the need for high-efficiency treatments. More specifically, mechanistic modeling provides detailed insights into microbial interactions and process dynamics, allowing for an enhanced understanding of complex bioprocesses. This increased insight supports improved design, operational strategies, scaling, and most importantly, control system design. However, these models are often too complex for real-time applications, such as data reconciliation and predictive modeling, due to excess details and measurement constraints. Hybrid system identification is, therefore, investigated to yield low-order models suited for the aforementioned applications, stepping beyond conventional methods. Simplified statistical models, derived from hybrid system identification, allow for practical data reconciliation and assimilation. Lastly, process-oriented control architectures are investigated based on model predictive control (MPC). While the MPC approach is a proven control method, this PhD research unlocks its full potential by developing

a hierarchical, process-oriented-designed adaptive MPC framework. This adaptive framework tailors control actions dynamically, addressing variability and maximizing process performance. Together, these approaches form a robust foundation for resource recovery systems that balance understanding, predictability, and operational control in three different research directions.

Based on the above motivations, this PhD thesis contributes to mechanistic modeling, hybrid system identification, and adaptive predictive control of PPB raceway reactors and ADs. First, we propose a mechanistic model for dynamical PPB selection in raceway reactors, introducing an empirical growth constant to capture distinct PPB growth pathways. For ADs, we extend the anaerobic digestion model no.1 (ADM1) with temperature-inhibition functions, based on the Cardinal model, to account for microbial activity under varying meteorological conditions. Both models enhance our understanding of microbial interactions and the impacts of potential disturbances. Second, to shed light on the state-of-the-art in hybrid system identification, a systematic survey is conducted, and we develop an online two-stage outer-bounded ellipsoid algorithm for identifying Switched Box-Jenkins systems. This approach has shown its capability of creating low-order, data-driven models for accurate approximation of the complex mechanistic models for data reconciliation and assimilation, and forecasting purposes. Lastly, the first-ever MPC control system integrated with a supervisory layer is developed for PPB growth in raceway reactors, which tackles model mismatches and disturbances, while optimizing performance. This hierarchical control architecture concept is also designed in a process-oriented manner for ADs under varying meteorological conditions to stabilize and enhance biogas production by adjusting feeding flow rate. Each of these parallel research directions is thoroughly investigated and discussed in the two case studies and shows the success in advancing resource recovery from wastewater in these three different intended aspects.

1

MODELING AND CONTROL: ADVANCING RESOURCE RECOVERY FROM WASTEWATER

1.1. INTRODUCTION

Wastewater-based resource recovery is a biological process to not only recycle wastewater but also to use it to produce raw materials (Puyol *et al.* 2017b). Wastewater treatment processes contribute to environmental sustainability by decreasing the amount of waste and by reducing the consumption of raw materials. They also play a role in renewable resources to address both energy demands to moderate the overuse of fossil fuels, and the production of high value-added products instead of removal and disposal. Despite wastewater treatment and resource recovery processes showing a high capacity for those aforementioned aspects in economically-rich societies, many under-developing societies face contamination of freshwater resources and shortage of nutrients, and energy resources due to the rapid growth of the human population (Ceron-Chafla and Lindeboom 2023). This has led to increased research into technological innovations in wastewater treatment and resource recovery plants. To address the aforementioned increases in research and development of biological Wastewater Resource Recovery Technologies (WRRTs), two broad fields of study can be taken into account: *bioprocess engineering*, and *bioprocess control* (Dochain 2013).

Bioprocess engineering, which can also be formulated as *advancing the best available technology (BAT)*, includes either developing or adapting the best identified available technology for specific environmental and biological conditions to enhance the process productivity. In other words, “BAT” refers to the best way of either treating wastewater or resource recovery, i.e. design and optimal operation in the context of *bioprocess engineering*. In the *SARASWATI 2.0* project, 10 different piloted technologies have been developed and tested across India from a *bioprocess engineering* point of view to identify and advance the “BAT”. This project was a follow-up to the 1st version, as many of the piloted plants in the 1st version of *SARASWATI* could not be considered “BAT” in terms of treatment efficiency and costs (Starkl *et al.* 2018). Moreover, low efficiency of treatment plants was also reported due to unskilled operations (Chatterjee *et al.* 2016). In other words, different levels of operator capacity and supervision can lead to operational failures and drastically disturb process performance. Therefore, besides the exploration of new technologies within the *bioprocess engineering* context, *bioprocess control* should also be taken into account.

Bioprocess control, which can also be formulated as *automation and control*, includes developing or adapting methods of monitoring and control to enhance the process productivity and to improve the process robustness. This also becomes particularly critical to consider, when transferring technologies from lab-scale to pilot- and full-scale. Seasonal variations, unexpected operational malfunctions, and environmental and biological fluctuations are examples of factors that can perturb process operations. If preventive or corrective actions are not taken by an operator, the process performance may deteriorate considerably or even fail. Therefore, the process can be integrated with a “smart” and “automatic” control system that detects operator faults and potential fluctuations and optimizes plant operation without operator interference. This is also considered in *SARASWATI 2.0*. The *bioprocess control* problem, as the main scope of the research in this thesis (as shown in Figure 1.1), has

two key aspects: modeling and control. Therefore, the aim of investigating *modeling and control for biological wastewater treatment technologies* is to demonstrate that *bioprocess engineering* and *bioprocess control* are complementary to one another, and by integrating an “automatic” control system, a biological treatment plant can be turned into a “BAT”. To validate this hypothesis, the technology of Pilot 5 of *SARASWATI 2.0* was selected, which will be explained in the next section. Thus, this thesis is dedicated to exploring this resource recovery technology of Pilot 5 from the aforementioned two aspects in the context of *bioprocess control*.

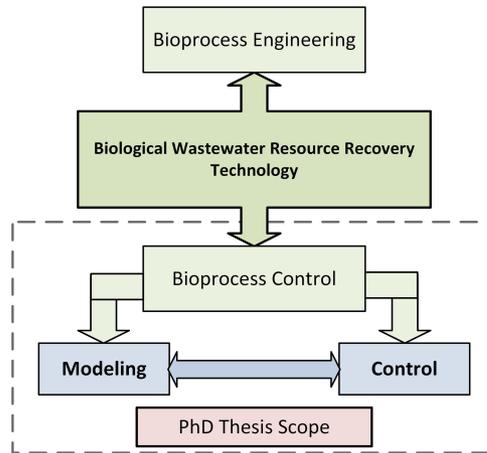


Figure 1.1: The two sides of a biological wastewater treatment process problem, i.e. bioprocess engineering and bioprocess control, and the scope of this thesis.

1.2. RESOURCE RECOVERY FROM WASTEWATER

Various types of anaerobic digestion plants were proposed and tested in the 1st version of *SARASWATI* to produce biogas as a valuable recovered resource from urban wastewater that can be used as biofuel. However, as mentioned earlier, the results showed that process efficiency was lowest in those plants not due to malfunction, but because of a lack of experts for process supervision (Chatterjee *et al.* 2016). Malfunction in anaerobic digestion typically results in inefficient conversion of influent organic matter into biogas, leaving some soluble organic acids in the effluent (Bornhöft *et al.* 2013). This malfunction is not only due to faulty design and unskilled operators but also fluctuating operational conditions such as variations in incoming influent composition and seasonal meteorological changes. To demonstrate how a “smart” control system can transform a technology into a “BAT”, we will discuss the management of biogas production in anaerobic digestion using smart feed control under varying meteorological conditions as a separate case study for wastewater resource recovery. The simplified schematization of an anaerobic digester is shown in Figure 1.2.

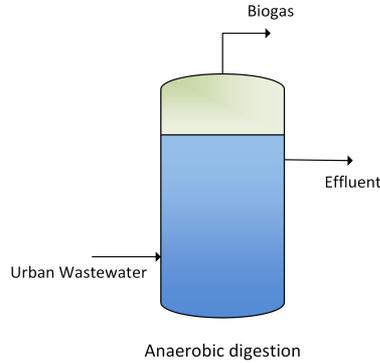


Figure 1.2: Schematization of an anaerobic digester.

Nevertheless, conditions under which the anaerobic digesters are not working properly could be turned into an advantage if their output can be used to produce high value products like VFAs (Kleerebezem *et al.* 2015). The VFAs produced can serve as a feed for a photobioreactor to cultivate PPB. One of the primary reasons to use PPB is their capability to achieve simultaneous nutrient and COD recovery from wastewater (Hülse *et al.* 2018). PPB are commonly applied as a secondary wastewater treatment to recover nutrients and to assimilate carbon for high-efficiency treatment. High-value products derived from PPB include microbial fertilizers, bio-stimulants, animal feed ingredients, and bio-plastics feedstock (Alloul *et al.* 2023). Microbial selectivity and growth are better controlled in a closed anaerobic photobioreactor, but considering scale-up capacity, a raceway-pond reactor offers a cost-friendly alternative in terms of construction and operation (Alloul *et al.* 2021). Therefore, integration of VFA-based anaerobic digestion with a raceway-pond photobioreactor for PPB production is a candidate of “BAT” in *SARASWATI 2.0*. The simplified schematization of a PPB raceway-pond reactor is depicted in Figure 1.3. This technology is relatively new and still under research and development, from both *bioprocess engineering* and *bioprocess control* standpoints. In this PhD thesis, we will focus on the *bioprocess control* aspects for these two aforementioned resource recovery from wastewater, namely anaerobic digestion for biogas production and a raceway reactor for PPB growth, and develop modeling and control approaches for them.

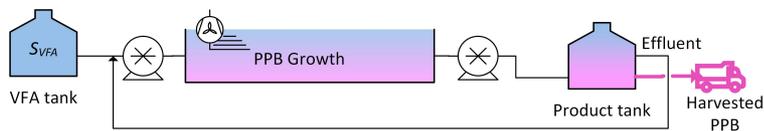


Figure 1.3: Schematization of a PPB raceway-pond reactor.

1.3. HOW TO MODEL AND CONTROL?

Modeling a biological wastewater resource recovery process is a primary means for process design, scale-up implementation, optimal operation determination, and process control (Solon *et al.* 2019). Dynamical models provide an insightful understanding of biological wastewater resource recovery processes as a virtual benchmark to interpret input-output relationships for decision-making in the context of process optimization, monitoring, and control. Depending on the main purpose of modeling, we can either use mechanistic (first-principles) modeling as “*detailed mathematical modeling for simulation and understanding*” or use statistical modeling as “*structural mathematical modeling for monitoring and control*” (Jeppsson 1996).

Mechanistic (first-principles) models, also known as white-box models, are dynamical models in the form of ordinary differential equations (and partial differential equations). These describe the time-wise evolution of a bioprocess based on biological and physicochemical interactions among components. Input-output relationships are usually modeled on the basis of a mass balance using different reaction metabolisms. Because of the various types of components in a bioprocess, the metabolic interactions among them, and the different phases of biological and physicochemical interactions, mechanistic models are usually complex, nonlinear, and stiff. Therefore, these models are suitable models for “*simulation, process understanding, and optimal decision-making operation*”, but, they are complex and computationally expensive for “*identification and model-based control*” (Donoso-Bravo *et al.* 2011; Yoshida *et al.* 2019).

On the other hand, with the increase of digital twins and model-based control, other alternatives for modeling biological wastewater resource recovery processes should also be considered. Structural models, such as statistical input-output models and state-space models, provide a meaningful framework for data assimilation and reconciliation, identification and prediction, and model-based control. These models are also known as gray-box models and have been used to represent bioprocesses (Capodaglio *et al.* 1991). As mentioned, bioprocesses are inherently complex, and representing these processes with overly simplistic models would not be realistic. Because of the capabilities of adaptive and hybrid models to capture highly nonlinear dynamics, they can be considered as appropriate candidates to model these bioprocesses for “*identification and model-based control*”. Therefore, these types of models along the mechanistic modeling approach are studied in this PhD thesis.

The other face of *bioprocess control* is designing and configuring an automatic control system for a biological wastewater resource recovery process. So far, by modeling it, we can provide a framework for understanding, analysis, identification, and control. Moving forward, a control system should be developed that enhances the productivity and performance of the process. Despite the increase in industrialization and the increasing importance of automatic control, managing biological processes is challenging. This complexity arises from the inherent variability in the behavior of living organisms and the lack of reliable real-time measurements and the lengthy procedure of off-line measurements (Dochain 2013). Even though a wide range of attempts have been made to either develop or advance

control strategies for biological wastewater resource recovery processes (Iratni and Chang 2019), the primary question still being explored is: “*How to configure a direct control strategy to improve process performance and productivity?*”. In other words, most control strategies have been dedicated to managing environmental conditions (e.g., pH, temperature, etc.) that regulate operational settings to favor optimal operation, while approaches that directly steer biological interactions, such as feed control, are more challenging, yet promising (Gaida *et al.* 2017).

Advancing the control strategies for biological wastewater resource recovery processes requires the adoption of optimization-based control strategies, where physical and performance objectives and constraints can be explicitly dealt with. Given the nature of biological wastewater resource recovery processes, and the long response time to changes (Anand *et al.* 2021), model predictive control (MPC) as a well-established advanced control method, has been tested for decades in the context of *bioprocess control*. (Chaib Draa *et al.* 2018; Craven *et al.* 2014; Muñoz *et al.* 2009). The key point is adapting the MPC control for a specific process and configuring a control system based on it to enhance process performance. In other words, this PhD thesis investigates how the design of a biological wastewater resource recovery process can be improved by integrating an adapted MPC control system for that process. The critical control system configuration is how to use *bioprocess engineering* knowledge and understanding, mostly through modeling, for *bioprocess control*, which will be discussed in this PhD thesis. In summary, advancing control strategies for biological wastewater resource recovery processes is more than utilizing an advanced control technique. The design of a “direct” control method and the configuration of the control system should also be taken into account.

1.4. RESEARCH THEMES AND STRUCTURE

As discussed above, the importance of modeling and control of a biological wastewater resource recovery process is taken into consideration for the aforementioned case studies in this PhD thesis. In other words, the central question in this thesis is:

What tools and approaches can be employed to advance resource recovery from wastewater, and how do these approaches contribute to achieving the desired outcome for PPB raceway reactors and anaerobic digesters?

According to the elaborated *bioprocess control* problem, the main research scheme has been divided into four subprojects as shown in Figure 1.4. As a primary answer to this research question, three main domains have been chosen for detailed explorations in this PhD thesis as listed below. A summary of the main contributions of each chapter to PPB raceway reactors and anaerobic digesters, and to the intended central research question along with highlighting the main gaps is also mentioned below. The detailed literature reviews are provided in each chapter individually.

- **Mechanistic modeling:** modeling mechanistic behaviors of the WRRTs of this study for the purpose of understanding and decision-making.

- **Hybrid system identification:** Identifying low-order statistical models of the WRRTs under consideration using system identification methods for the purpose of prediction, and data assimilation and reconciliation.
- **Adaptive predictive control:** Developing an automatic control system based on the adaptive MPC controller for the given WRRTs in order to enhance process efficiency.

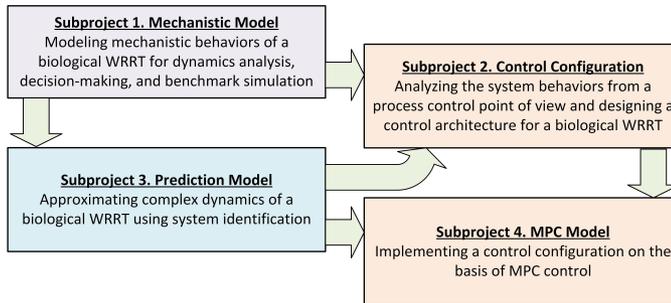


Figure 1.4: Step-by-step subprojects from modeling to control as a proposed overall research scheme for designing an advanced control system for a bioprocess.

Although modeling and control of anaerobic digestion have been studied extensively from various aspects (Batstone *et al.* 2002; Nguyen *et al.* 2015), the technology of the PPB raceway reactor is relatively new (Capson-Tojo *et al.* 2020) and requires further investigation. To provide a virtual benchmark for simulations and system dynamics analysis, mechanistic models are well-known candidates. The Anaerobic Digestion Model no. 1 (ADM1) (Batstone *et al.* 2002; Mo *et al.* 2023) is a well-established first-principles model describing the anaerobic digestion process in four stages. This model has been utilized in various contexts, such as model development (Flores-Alsina *et al.* 2016), model simplification (Weinrich and Nelles 2021), experimental design and interpretation (Donoso-Bravo *et al.* 2011), and control design (Ahmed and Rodríguez 2020). However, as mentioned, inoperative Anaerobic Digestion (AD) processes may result from the lack of an appropriate control system that addresses various environmental perturbations. One of the susceptible disturbances is meteorological fluctuation influencing operating temperature. To investigate this effect, a temporal temperature extension is introduced in the ADM1 model, and a feed MPC control framework is developed not only to handle varying operating temperatures but also to address biogas management by integrating a self-consumption heating system. The model extension and biogas management through a control system configuration are discussed in **Chapter 6**. Therefore, the main contributions of this chapter are the development of a temporal temperature-based ADM1 model and the design of an adaptive model predictive control system for feed rate adjustment to mitigate temperature fluctuations caused

by meteorological variations. These advancements demonstrate improvements in biogas production and treatment efficiency for anaerobic digestion.

Chapter 2 presents a novel mechanistic model for the growth of PPB in raceway reactors. While PPB behavior has been mechanistically modeled by Puyol *et al.* (2017a) and further extended by Capson-Tojo *et al.* (2023), the Purple Bacteria Model (PBM) has been exclusively developed for PPB cultivation in raceway-pond reactors. Microbial selection dynamics have also been mechanistically defined using an empirical parallel metabolic growth constant. This model serves further dynamics analysis and the required supervisory knowledge for the control system configuration. Thus, in **Chapter 5**, a control system based on MPC control is developed to manage microbial competition and environmental perturbations and to enhance process performance by proposing a supervisory layer based on quantity-driven and quality-driven operational scenarios. The developed control strategy has been evaluated through simulation studies using the PBM. Therefore, the main contributions of these two chapters are the development of a model for PPB dynamics that considers various parallel growth pathways and the design of an adaptive model predictive control system to mitigate operational disturbances and to enhance the performance of the raceway reactor for PPB cultivation.

Although conventional system identification approaches have been well-studied for various applications, including biological wastewater resource recovery processes, this PhD thesis explores the novel idea of hybrid system identification methods for them under modeling for prediction. In other words, hybrid systems represented as either piece-wise affine systems or switched linear systems can capture the high nonlinearity of process dynamics in a few linear subsystems with a simplified structure. These systems can be used not only for model-based control but also for prediction, data reconciliation, and digital twins. Therefore, **Chapter 3** provides an overview of proposed approaches in the literature through a systematic survey. Then, based on the state-of-the-art and properties of bioprocesses, a novel identification approach for switched Box-Jenkins models, considered to be the most complex form of input-output models that incorporate disturbances into modeling and that can be simplified for simpler forms of models, is developed in **Chapter 4**. The proposed method are also evaluated for approximating mechanistic models, namely the ADM1 and PBM models. According to the developed method, complex bioprocesses such as ADs and PPB raceway reactors can be accurately approximated using a low-order model. This simplification facilitates forecasting purposes, as certain detailed dynamics are relaxed, making the model more computationally efficient while still capturing the essential behavior of the system.

The overall aim of this thesis is *developing modeling methods and advancing control strategies for the given case studies*. Therefore, a general research plan is proposed, as shown in Figure 1.4, utilizing this framework and associated subprojects for the considered case studies. The main contribution of this thesis can be summarized as *advancing resource recovery from wastewater in PPB raceway reactors and anaerobic digestions* by exploring (i) *mechanistic modeling to enhance understanding of process behaviors*, (ii) *predictive modeling to approximate complex models for data assimilation and reconciliation*, and (iii) *a hierarchical adaptive*

predictive control system to improve efficiency under varying conditions. Given the aforementioned research directions and the contributions of each chapter, this PhD thesis has a paper-based organization. The structure of the chapters is shown in Figure 1.5. **Chapters 2-6** are based on peer-reviewed publications. Each chapter aims to discuss the knowledge gaps and corresponding contributions. Finally, **Chapter 7** summarizes the findings and discusses the reliability of the proposed research, along with potential future research directions.

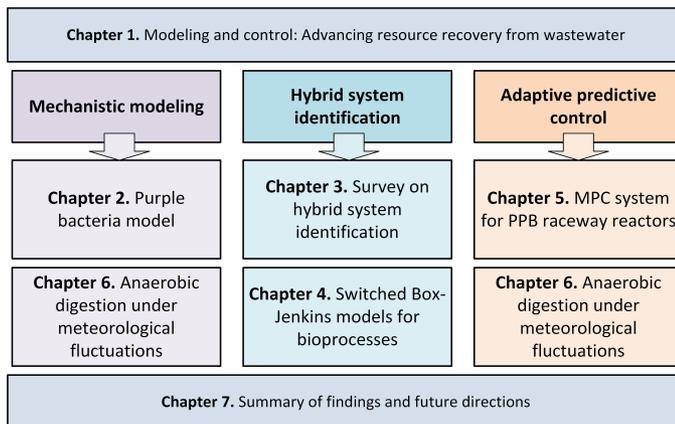


Figure 1.5: Chapter outline of the thesis and thematic organization of chapters from modeling to control of the PPB raceway reactor and of anaerobic digestion.

REFERENCES

- Ahmed, W. and J. Rodríguez (2020). "A model predictive optimal control system for the practical automatic start-up of anaerobic digesters". In: *Water Research* 174, p. 115599. ISSN: 0043-1354.
- Alloul, A., M. Cerruti, D. Adamczyk, D. G. Weissbrodt, and S. E. Vlaeminck (2021). "Operational Strategies to Selectively Produce Purple Bacteria for Microbial Protein in Raceway Reactors". In: *Environmental Science & Technology* 55.12, pp. 8278–8286. ISSN: 1520-5851.
- Alloul, A., A. Moradvandi, D. Puyol, R. Molina, G. Gardella, S. E. Vlaeminck, B. De Schutter, E. Abraham, R. E. Lindeboom, and D. G. Weissbrodt (2023). "A novel mechanistic modelling approach for microbial selection dynamics: Towards improved design and control of raceway reactors for purple bacteria". In: *Bioresource Technology* 390, p. 129844.
- Anand, K., A. P. Mittal, and B. Kumar (2021). "Modelling and simulation of dual heating of substrate with centralized temperature control for anaerobic digestion process". In: *Journal of Cleaner Production* 325, p. 129235. ISSN: 0959-6526.
- Batstone, D., J. Keller, I. Angelidaki, S. Kalyuzhnyi, S. Pavlostathis, A. Rozzi, W. Sanders, H. Siegrist, and V. Vavilin (2002). "The IWA Anaerobic Digestion Model No 1 (ADM1)". In: *Water Science and Technology* 45.10, pp. 65–73.
- Bornhöft, A., R. Hanke-Rauschenbach, and K. Sundmacher (2013). "Steady-state analysis of the Anaerobic Digestion Model No. 1 (ADM1)". In: *Nonlinear Dynamics* 73.1–2, pp. 535–549. ISSN: 1573-269X.
- Capodaglio, A. G., H. V. Jones, V. Novotny, and X. Feng (1991). "Sludge bulking analysis and forecasting: application of system identification and artificial neural computing technologies".

- In: *Water Research* 25.10, pp. 1217–1224. ISSN: 0043-1354.
- Capson-Tojo, G., D. J. Batstone, M. Grassino, S. E. Vlaeminck, D. Puyol, W. Verstraete, R. Kleerebezem, A. Oehmen, A. Ghimire, I. Pikaar, J. M. Lema, and T. Hülsen (2020). “Purple phototrophic bacteria for resource recovery: Challenges and opportunities”. In: *Biotechnology Advances* 43, p. 107567. ISSN: 0734-9750.
- Capson-Tojo, G., D. J. Batstone, and T. Hülsen (2023). “Expanding mechanistic models to represent purple phototrophic bacteria enriched cultures growing outdoors”. In: *Water Research* 229, p. 119401. ISSN: 0043-1354.
- Ceron-Chafla, P. and R. E. Lindeboom (2023). “Frugal engineering solutions for recovery of resources from wastewater”. In: *Handbook on Frugal Innovation*. Edward Elgar Publishing, pp. 197–219. ISBN: 9781788118873.
- Chaib Draa, K., A. Zemouche, M. Alma, H. Voos, and M. Darouach (2018). “A discrete-time nonlinear state observer for the anaerobic digestion process”. In: *International Journal of Robust and Nonlinear Control* 29.5, pp. 1279–1301. ISSN: 1099-1239.
- Chatterjee, P., M. M. Ghangrekar, and S. Rao (2016). “Low efficiency of sewage treatment plants due to unskilled operations in India”. In: *Environmental Chemistry Letters* 14.3, pp. 407–416. ISSN: 1610-3661.
- Craven, S., J. Whelan, and B. Glennon (2014). “Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller”. In: *Journal of Process Control* 24.4, pp. 344–357. ISSN: 0959-1524.
- Dochain, D. (2013). *Automatic control of bioprocesses*. John Wiley & Sons.
- Donoso-Bravo, A., J. Mailier, C. Martin, J. Rodríguez, C. A. Aceves-Lara, and A. V. Wouwer (2011). “Model selection, identification and validation in anaerobic digestion: A review”. In: *Water Research* 45.17, pp. 5347–5364. ISSN: 0043-1354.
- Flores-Alsina, X., K. Solon, C. Kazadi Mbamba, S. Tait, K. V. Gernaey, U. Jeppsson, and D. J. Batstone (2016). “Modelling phosphorus (P), sulfur (S) and iron (Fe) interactions for dynamic simulations of anaerobic digestion processes”. In: *Water Research* 95, pp. 370–382. ISSN: 0043-1354.
- Gaida, D., C. Wolf, and M. Bongards (2017). “Feed control of anaerobic digestion processes for renewable energy production: A review”. In: *Renewable and Sustainable Energy Reviews* 68, pp. 869–875. ISSN: 1364-0321.
- Hülsen, T., K. Hsieh, Y. Lu, S. Tait, and D. J. Batstone (2018). “Simultaneous treatment and single cell protein production from agri-industrial wastewaters using purple phototrophic bacteria or microalgae – A comparison”. In: *Bioresource Technology* 254, pp. 214–223. ISSN: 0960-8524.
- Iratni, A. and N.-B. Chang (2019). “Advances in control technologies for wastewater treatment processes: status, challenges, and perspectives”. In: *IEEE/CAA Journal of Automatica Sinica* 6.2, pp. 337–363. ISSN: 2329-9274.
- Jeppsson, U. (1996). “Modelling aspects of wastewater treatment processes”. In:
- Kleerebezem, R., B. Jooisse, R. Rozendal, and M. C. M. Van Loosdrecht (2015). “Anaerobic digestion without biogas?” In: *Reviews in Environmental Science and BioTechnology* 14.4, pp. 787–801. ISSN: 1572-9826.
- Mo, R., W. Guo, D. Batstone, J. Makinia, and Y. Li (2023). “Modifications to the anaerobic digestion model no. 1 (ADM1) for enhanced understanding and application of the anaerobic treatment processes – A comprehensive review”. In: *Water Research* 244, p. 120504. ISSN: 0043-1354.
- Muñoz, C., D. Rojas, O. Candía, L. Azocar, C. Bornhardt, and C. Antileo (Jan. 2009). “Supervisory control system to enhance partial nitrification in an activated sludge reactor”. In: *Chemical Engineering Journal* 145.3, pp. 453–460. ISSN: 1385-8947.
- Nguyen, D., V. Gadhamshetty, S. Nitayavardhana, and S. K. Khanal (2015). “Automatic process control in anaerobic digestion technology: A critical review”. In: *Bioresource Technology* 193, pp. 513–522. ISSN: 0960-8524.
- Puyol, D., E. Barry, T. Hülsen, and D. Batstone (2017a). “A mechanistic model for anaerobic phototrophs in domestic wastewater applications: Photo-anaerobic model (PanM)”. In: *Water Research* 116, pp. 241–253. ISSN: 0043-1354.
- Puyol, D., D. J. Batstone, T. Hülsen, S. Astals, M. Peces, and J. O. Kromer (2017b). “Resource Recovery from Wastewater by Biological Technologies: Opportunities, Challenges, and Prospects”. In: *Frontiers in Microbiology* 7. ISSN: 1664-302X.
- Solon, K., E. I. P. Volcke, M. Spérandio, and M. C. M. van Loosdrecht (2019). “Resource recovery and wastewater treatment modelling”. In: *Environmental Science: Water Research & Technology* 5.4, pp. 631–642. ISSN: 2053-1419.
- Starkl, M., J. Anthony, E. Aymerich, N. Brunner, C. Chubilleau, S. Das, M. M. Ghangrekar, A. A. Kazmi, L. Philip, and A. Singh (2018). “Interpreting best available technologies more flexibly: A policy perspective for municipal wastewater

- management in India and other developing countries". In: *Environmental Impact Assessment Review* 71, pp. 132–141. ISSN: 0195-9255.
- Weinrich, S. and M. Nelles (2021). "Systematic simplification of the Anaerobic Digestion Model No. 1 (ADM1) – Model development and stoichiometric analysis". In: *Bioresource Technology* 333, p. 125124. ISSN: 0960-8524.
- Yoshida, K., K. Kametani, and N. Shimizu (2019). "Adaptive identification of anaerobic digestion process for biogas production management systems". In: *Bioprocess and Biosystems Engineering* 43.1, pp. 45–54. ISSN: 1615-7605.

2

PURPLE BACTERIA MODEL

Purple phototrophic bacteria (PPB) show an under-explored potential for resource recovery from wastewater. Raceway reactors offer a more affordable full-scale solution on wastewater and enable useful additional aerobic processes. Current mathematical models of PPB systems provide useful mechanistic insights, but do not represent the full metabolic versatility of PPB and thus require further advancement to simulate the process for technology development and control. In this study, a new modeling approach for PPB that integrates the photoheterotrophic, and both anaerobic and aerobic chemoheterotrophic metabolic pathways through an empirical parallel metabolic growth constant was proposed. It aimed the modeling of microbial selection dynamics in competition with aerobic and anaerobic microbial community under different operational scenarios. A sensitivity analysis was carried out to identify the most influential parameters within the model and calibrate them based on experimental data. Process perturbation scenarios were simulated, which showed a good performance of the model.

This chapter is an adapted version of a novel mechanistic modeling approach for microbial selection dynamics: towards improved design and control of raceway reactors for purple bacteria, Abbas Alloul¹, Ali Moradvandi¹, et al. (2023), *Bioresource Technology*.

¹Equally contribution

2.1. INTRODUCTION

Purple phototrophic bacteria (PPB) offer great potential to recover resources from wastewater like organics and the nutrients nitrogen and phosphorus. The phototrophically produced biomass have circular economy applications as microbial fertilizer, biostimulant, animal feed ingredient, and bioplastics feedstock (Alloul *et al.* 2023; Capson-Tojo *et al.* 2020). PPB exhibit metabolic diversity and versatility, capable of thriving on various energy sources (photo- and chemotrophy), electron sources (organo- and lithotrophy), and carbon sources (auto- and heterotrophy). (Imhoff 2006). As phototrophs, they have the unique ability to grow on infrared light, are anoxygenic and proliferate swiftly in anaerobic mixed-culture systems fed with volatile fatty acids (VFAs). For wastewater treatment and resource recovery, this photoorganoheterotrophic mode is mostly studied (Capson-Tojo *et al.* 2020).

In research labs, membrane (photo)bioreactors, tubular photobioreactors, flat-plate photobioreactors, as well as stirred-tank photobioreactors have been mainly used to achieve high PPB selectivity (Alloul *et al.* 2021a; Capson-Tojo *et al.* 2020; Cerruti *et al.* 2020; Hülsen *et al.* 2022). Successful selection of PPB has been mainly obtained in closed configurations to maintain an anaerobic reactor environment. However, the high capital expenditure of these systems may prevent implementation at full scale (Acién *et al.* 2012; Alloul *et al.* 2021a). Open raceway reactors reduced total investment and operational cost 5 to 10-fold, relative to closed anaerobic photobioreactors (Alloul *et al.* 2021a). However, open raceway ponds are exposed to passive oxygenation ($217\text{--}226\text{ mgO}_2\text{L}^{-1}\text{h}^{-1}$) through the combination of the paddle wheel rotation used to circulate the wastewater and their high atmospheric surface-to-volume ratio used to maximize light accessibility ($5\text{ m}^2\text{m}^{-3}$). The uncontrolled supply of dissolved oxygen in this system not only enables aerobic conversions but also creates a competitive environment where PPB compete with aerobic bacteria (AEB) for available organic substrates that may be contaminated by external bacteria (Alloul *et al.* 2021a). Achieving high PPB selectivity is, therefore, more challenging than in closed anaerobic systems. Nonetheless, changing the operational conditions in terms of oxygen supply, light availability, sludge retention time (SRT), and chemical oxygen demand (COD) loading successfully boosted the PPB abundance from 14% to 78% in a 100-L raceway reactor operated on synthetic wastewater (Alloul *et al.* 2021a).

Mechanistic simulation models are needed to improve the design and operation of PPB processes, boost piloting activities, scale up implementations, and control the processes. On a more fundamental basis, genome-scale metabolic models have been developed to describe biohydrogen production in pure-culture systems (Golomysova *et al.* 2010; Imam *et al.* 2011); however, they cannot be directly used for environmental biotechnology applications. Puyol *et al.* (2017) have translated the activated sludge model formalism to predict nutrient conversions driven by PPB, which has been discussed by Henze *et al.* (2015). The Photo-anaerobic model (PANM) is limited to photo-anaerobic conditions and does not take microbial competition with non-PPB guilds like aerobic and fermentative chemoorganoheterotrophs and photolithoautotrophs into account. The extended version of PANM (ePANM) has been proposed by (Capson-Tojo *et al.* 2023), which integrates eight different types

of microorganisms, i.e. PPB, AEB, acidogenic and acetogenic fermentors, aerobic predators, heterotrophic and autotrophic sulphate reducing bacteria, and microalgae and takes diverse metabolic capabilities of PPB into account. Compared to PAnM, PPB photoheterotrophic, aerobic and anaerobic uptake rates and yields have been used to simulate PPB growth in batch and semi-continuous processes in the ePAnM. As far as observed, the difference in maximum growth rate in relation to PPBs metabolic versatility under various environmental conditions were not explicitly addressed Capson-Tojo *et al.* (2023) and Puyol *et al.* (2017). While the ePAnM is capable of adapting to various environmental conditions, it has been suggested recalibrating specifically for the operation of open pond raceway reactors (Capson-Tojo *et al.* 2023), which was investigated in this study. A detailed comparison among the PAnM, the ePAnM, and the PBM models can be found in Table 2.1. With the technological development of PPB raceway systems, a more comprehensive model is necessary to simulate wastewater treatment along with microbial selection to assess operational and control scenarios for raceway reactors and also taking parallel metabolic growth into account. Such a model will allow to better engineer, implement, and control raceway reactors by predicting process conditions, variations, and perturbations that affect the PPB abundance and the treatment performance.

In this study, a new mechanistic purple bacterial model (PBM) was constructed that considers the most relevant metabolic growth modes that PPB and competing microbes including aerobic and anaerobic heterotrophic bacteria within a raceway reactor for the purpose of process simulation. This study introduced a novel approach aimed at incorporating diverse growth pathways of PPB, considering their simultaneous occurrence by hypothesizing a phenomenon observed in previous research—namely, the coexistence of multiple growth pathways. Therefore, an empirical parallel metabolic growth constant was defined to account for the contribution of alternative pathways to PPB growth alongside the dominant pathway. After model construction, a sensitivity analysis was conducted to identify the most influential parameters and to assess the impact of parameter variations on the model outputs. Calibration of the important factors was carried out through iterative error minimization. Short- and long-term perturbations of incoming soluble organic matters, volatile fatty acid (VFA), suspended solids and light variations were simulated to their effects on PPB abundance and COD removal rate.

2.2. MATERIALS AND METHODS

2.2.1. MODEL DESCRIPTION

The PBM was constructed to simulate the wastewater treatment performance and PPB selectivity in open raceway reactors for design and control purposes. The PBM can be transferred to other reactor systems or used as add-on to existing models such as the Activated Sludge Model (ASM), Anaerobic Digestion Model no.1 (ADMn1) or algae-bacteria models e.g., the ALBA model (Batstone *et al.* 2002; Casagli *et al.* 2021; Henze *et al.* 2015), since the PBM is units compatible to the ASM and ADMn1 series of the International Water Association (IWA). The PBM consists of 15 state

Table 2.1: Different versions of mechanistic models simulating the utilization of PPB in WWT.

Name	PnM	PBM	ePnM
Conditions	Anaerobic	Semi-aerobic	Aerobic
Processes	6	16	30
<i>PPB-based</i>	5	6	11
<i>Other biomass</i>	-	5	19
<i>Hydrolysis</i>	1	1	1
<i>Physical processes</i>	-	4	2
State variables	10	15	21
<i>PPB-based</i>	1	3	1
Substrate	Acetate Other organics	VFAs Other organics	Acetate Other VFAs Other organics
Biomass	PPB	Photoheterotrophic PPB (Anaerobic chemoheterotrophic PPB Aerobic heterotrophic Acidogenic	PPB Aerobic heterotrophic Acidogenic Acetogenic Aerobic predators Heterotrophic sulphate reducing bacteria Autotrophic sulphate reducing bacteria Microalgae biomass
PPB metabolisms	Photoheterotrophic Photoautotrophic Anaerobic chemoheterotrophic	Photoheterotrophic (An)aerobic chemoheterotrophic	Photoheterotrophic Photoautotrophic (An)aerobic chemoheterotrophic Fermentative
Inhibition functions	Competitive inhibitions Inorganic nitrogen Inorganic phosphor Free ammonia Light	Competitive inhibitions Inorganic nitrogen Inorganic phosphor Photoheterotrophic light Chemoheterotrophic light Photoheterotrophic oxygen Chemoheterotrophic oxygen	Inorganic nitrogen Inorganic phosphor Free ammonia Light Oxygen pH Temperature
Physical processes	-	Carbon dioxide dissolution/stripping Oxygen dissolution/stripping Stripping of hydrogen Stripping of ammonium	Carbon dioxide losses Oxygen supply
Operational mode	Continuous	SBR/Constant	Continuous
pH	Dynamic	Constant	Dynamic
Temperature	Constant	Constant	Cardinal temperature model
Light	Constant	Lambert-Beer's law	Lambert-Beer's law

variables with 3 state variables associated with PPB community, and 16 processes including 12 biological and 4 physical processes.

STATE VARIABLES, BALANCES, AND PROCESSES

In this section, the structure of the proposed PBM model is explained by describing state variables, balances, and biological and physical processes. The PBM state variables are summarized in Table 5.1, and biological and physical processes are described in Table 2.3. Moreover, the PBM Peterson matrix is given in Table 2.4.

Table 2.2: State variables included in the PBM categorized in a particulate (X) and soluble (S) fraction. Microbial biomass in forms of purple phototrophic bacteria (indexed by PB), aerobic bacteria (AEB) and anaerobic bacteria (ANB). PPB can be grown photoheterotrophically (ph), aerobic chemoheterotrophically (aec), and anaerobic chemoheterotrophically (anc).

Symbol	Description	Unit
Microbial biomass		
$X_{PB,ph}$	Photoheterotrophic grown PPB	mgCODL ⁻¹
$X_{PB,aec}$	Aerobic chemoheterotrophic grown PPB	mgCODL ⁻¹
$X_{PB,anc}$	Anaerobic chemoheterotrophic grown PPB	mgCODL ⁻¹
X_{AEB}	Aerobic bacteria	mgCODL ⁻¹
X_{ANB}	Anaerobic bacteria	mgCODL ⁻¹
Substrates and products		
X_S	Slowly biodegradable organic matter	mgCODL ⁻¹
X_I	Inert particulate organic matter	mgCODL ⁻¹
S_S	Readily biodegradable organic matter	mgCODL ⁻¹
S_{VFA}	Volatile fatty acids	mgCODL ⁻¹
S_I	Inert soluble organic matter	mgCODL ⁻¹
S_{H2}	Soluble hydrogen	mgCODL ⁻¹
S_{IC}	Total inorganic carbon	mmolHCO ₃ ⁻ L ⁻¹
S_{IN}	Total inorganic nitrogen	mgNL ⁻¹
S_{IP}	Total inorganic phosphorus	mgPL ⁻¹
S_{O2}	Dissolved oxygen	mgO ₂ L ⁻¹

(i) *Microbial community*: The microbial biomass were divided in three main categories, namely purple phototrophic bacteria (PPB), aerobic bacteria (AEB) and anaerobic bacteria (ANB), reflecting the main microbial groups thriving in a raceway reactor covered with a selective infrared cover operated on COD-rich wastewater (Alloul *et al.* 2021a). The PPB-based bacteria were also divided into three subgroups consisted of photoheterotrophic grown PPB ($X_{PB,ph}$), aerobic chemoheterotrophic grown PPB ($X_{PB,aec}$), and anaerobic chemoheterotrophic grown PPB ($X_{PB,anc}$).

(ii) *Substrates and products*: The incoming COD consists of both particulate and soluble matters. Readily biodegradable COD was divided in a non-volatile fatty acid (S_S) and C2-C6 volatile fatty acid (S_{VFA}) fractions. The growth kinetics and biochemical pathways for both substrates are different (Batstone *et al.* 2002; Puyol

Table 2.3: Biological and physical process rates in the PBM model.

$p(i)$	Description	Rate equation
Purple bacteria (X_{PrB})		
1	Photoheterotrophic growth on soluble organics such as carbohydrates and alcohols (excluding VFA).	$\mu_{m,SS,PrB,ph} \frac{S_E}{S_E + K_{L,E}} \frac{K_{I,O2,PrB}}{S_O_2 + K_{I,O2,PrB}} \frac{S_{IN}}{S_{IN} + K_{S,IN}} \frac{S_{IP}}{S_{IP} + K_{S,IP}} \frac{S_S}{S_S + S_{VFA}} \frac{S_S}{S_S + K_{SS,PrB}} f_{ph}(X_{PrB}, M_S)$
2	Photoheterotrophic growth on VFA. The main products are biomass and carbon dioxide.	$\mu_{m,VFA,PrB,ph} \frac{S_E}{S_E + K_{L,E}} \frac{K_{I,O2,PrB}}{S_O_2 + K_{I,O2,PrB}} \frac{S_{IN}}{S_{IN} + K_{S,IN}} \frac{S_{IP}}{S_{IP} + K_{S,IP}} \frac{S_S}{S_S + S_{VFA}} \frac{S_S}{S_S + K_{SS,PrB}} f_{ph}(X_{PrB}, M_S)$
3	Aerobic chemoheterotrophic growth on soluble organics with oxygen as terminal electron acceptor.	$\mu_{m,SS,PrB,aer} \frac{K_{S,E}}{S_E + K_{S,E}} \frac{K_{O_2,PrB}}{S_{O_2} + K_{O_2,PrB}} \frac{S_{IN}}{S_{IN} + K_{S,IN}} \frac{S_{IP}}{S_{IP} + K_{S,IP}} \frac{S_S}{S_S + S_{VFA}} \frac{S_S}{S_S + K_{SS,aer}} f_{aer}(X_{PrB}, M_S)$
4	Aerobic chemoheterotrophic growth on VFA.	$\mu_{m,VFA,PrB,aer} \frac{K_{S,E}}{S_E + K_{S,E}} \frac{K_{O_2,PrB}}{S_{O_2} + K_{O_2,PrB}} \frac{S_{IN}}{S_{IN} + K_{S,IN}} \frac{S_{IP}}{S_{IP} + K_{S,IP}} \frac{S_S}{S_S + S_{VFA}} \frac{S_S}{S_S + K_{SS,aer}} f_{aer}(X_{PrB}, M_S)$
5	Anaerobic chemoheterotrophic growth and acidogenic fermentation reactions on soluble organics. The main products are biomass, hydrogen, carbon dioxide and VFA.	$\mu_{m,SS,PrB,anc} \frac{K_{S,E}}{S_E + K_{S,E}} \frac{K_{I,O_2,anc}}{S_{O_2} + K_{I,O_2,anc}} \frac{S_{IN}}{S_{IN} + K_{S,IN}} \frac{S_{IP}}{S_{IP} + K_{S,IP}} \frac{S_S}{S_S + S_{VFA}} f_{anc}(X_{PrB}, M_S)$
6	Decay of PFB biomass into biodegradable (X_S) and inert (X_I) organic matter and release of ammonium, phosphate, and bicarbonate.	$b_{m,PrB,dec}(X_{PrB,ph} + X_{PrB,aer} + X_{PrB,anc})$
Aerobic bacteria (X_{AEB})		
7	Aerobic chemoheterotrophic growth on soluble organics with oxygen as terminal electron acceptor.	$\mu_{m,SS,AEB} \frac{K_{S,O_2,anc}}{S_{O_2} + K_{S,O_2,anc}} \frac{K_{O_2,AEB}}{S_{O_2} + K_{O_2,AEB}} \frac{S_{IN}}{S_{IN} + K_{S,IN}} \frac{S_{IP}}{S_{IP} + K_{S,IP}} \frac{S_S}{S_S + S_{VFA}} \frac{S_S}{S_S + K_{SS,AEB}} X_{AEB}$
8	Aerobic chemoheterotrophic growth on VFA.	$\mu_{m,VFA,AEB} \frac{K_{S,O_2,anc}}{S_{O_2} + K_{S,O_2,anc}} \frac{K_{O_2,AEB}}{S_{O_2} + K_{O_2,AEB}} \frac{S_{IN}}{S_{IN} + K_{S,IN}} \frac{S_{IP}}{S_{IP} + K_{S,IP}} \frac{S_S}{S_S + S_{VFA}} \frac{S_S}{S_S + K_{SS,AEB}} X_{AEB}$
9	Decay of AEB biomass into biodegradable (X_S) and inert (X_I) organic matter and release of ammonium, phosphate, and bicarbonate.	$b_{m,AEB,dec} X_{AEB}$
Anaerobic bacteria (X_{ANB})		
10	Anaerobic chemoheterotrophic growth and acidogenic fermentation reactions on soluble organics. The main products are biomass, hydrogen, carbon dioxide and VFA.	$\mu_{m,VFA,ANB} \frac{K_{S,O_2,anc}}{S_{O_2} + K_{S,O_2,anc}} \frac{K_{O_2,ANB}}{S_{O_2} + K_{O_2,ANB}} \frac{S_{IN}}{S_{IN} + K_{S,IN}} \frac{S_{IP}}{S_{IP} + K_{S,IP}} \frac{S_S}{S_S + S_{VFA}} X_{ANB}$
11	Decay of ANB biomass into biodegradable (X_S) and inert (X_I) organic matter and release of ammonium, phosphate, and bicarbonate.	$b_{m,ANB,dec} X_{ANB}$
Hydrolysis		
12	Hydrolysis of biodegradable particulates (X_S) into soluble (S_S and S_{VFA}) and inert organic carbon (X_I), ammonium, hydrogen and bicarbonate.	$\mu_{Hyd} X_S$
Physical processes		
13	Stripping/dissolution of oxygen in raceway reactor.	$K_L a_{O_2} (O_2^{sat} - S_{O_2})$
14	Stripping/dissolution of carbon dioxide in raceway reactor.	$K_L a_{CO_2} (CO_2^{sat} - S_{CO_2})$
15	Stripping of hydrogen in raceway reactor.	$K_L a_{H_2} (H_2^{sat} - S_{H_2})$
16	Stripping of ammonium in raceway reactor.	$K_L a_{NH_3} (NH_3^{sat} - S_{NH_3})$

Table 2.4: Peterson matrix of the PBM model.

Process (1)	State Variable (-)	S_{O_2}	S_5	S_{VFA}	S_{IC}	S_{I2}	S_{IP}	S_{IN}	S_{I1}	$X_{Pp,ph}$	$X_{Pp,acc}$	$X_{Pb,acc}$	X_{Pb}	X_{ANB}	X_S	Process rate (1)
Purple bacteria (X_{Pp})																
Photoheterotrophic growth on S_5	0	0	$-1/Y_{Pp,ph}$	0	$f_{IC,Pb,SS}$	0	$-Y_{Pp,Pb}$	$-Y_{Pp,Pb}$	0	1	0	0	0	0	0	1
Photoheterotrophic growth on S_{VFA}	0	0	$-1/Y_{Pp,ph}$	$-1/Y_{Pp,ph}$	$f_{IC,Pb,VFA}$	0	$-Y_{Pp,Pb}$	$-Y_{Pp,Pb}$	0	0	0	0	0	0	0	2
Aerobic chemoheterotrophic growth S_5	0	0	0	$-1/Y_{Pb,acc}$	$f_{IC,acc,SS}$	0	$-Y_{Pb,Pb}$	$-Y_{Pb,Pb}$	0	0	0	0	0	0	0	3
Aerobic chemoheterotrophic growth S_{VFA}	0	0	0	$-1/Y_{Pb,acc}$	$f_{IC,acc,VFA}$	0	$-Y_{Pb,Pb}$	$-Y_{Pb,Pb}$	0	0	0	0	0	0	0	4
Anaerobic chemoheterotrophic growth S_5	0	0	0	$-1/Y_{Pb,acc}$	$f_{IC,acc,SS}$	0	$-Y_{Pb,Pb}$	$-Y_{Pb,Pb}$	0	0	0	0	0	0	0	5
Decay of $X_{Pp} = X_{Pp,ph} + X_{Pp,acc} + X_{Pb,acc}$	0	0	0	0	$f_{IC,dec}$	0	$f_{Pb,dec}$	$f_{Pb,dec}$	0	-1	-1	0	0	0	0	6
Aerobic bacteria (X_{Aer})																
Aerobic chemoheterotrophic growth on S_5	0	0	$-1/Y_{Aer}$	0	$f_{IC,Aer,SS}$	0	$Y_{Aer,Pb}$	$Y_{Aer,Pb}$	0	0	0	0	1	0	0	7
Aerobic chemoheterotrophic growth on S_{VFA}	0	0	$-1/Y_{Aer}$	0	$f_{IC,Aer,VFA}$	0	$Y_{Aer,Pb}$	$Y_{Aer,Pb}$	0	0	0	0	1	0	0	8
Decay of X_{Aer}	0	0	0	0	$f_{IC,dec}$	0	$f_{Pb,dec}$	$f_{Pb,dec}$	0	0	0	0	-1	0	0	9
Anaerobic bacteria (X_{ANB})																
Anaerobic chemoheterotrophic growth on S_5	0	0	$-1/Y_{ANB}$	0	$f_{IC,ANB,SS}$	$f_{VFA,ANB}/Y_{ANB}$	$-Y_{ANB}$	$-Y_{ANB}$	0	0	0	0	0	1	0	10
Anaerobic chemoheterotrophic growth on S_{VFA}	0	0	$-1/Y_{ANB}$	0	$f_{IC,ANB,SS}$	$f_{VFA,ANB}/Y_{ANB}$	$-Y_{ANB}$	$-Y_{ANB}$	0	0	0	0	0	-1	0	11
Decay of X_{ANB}	0	0	0	0	$f_{IC,dec}$	0	$f_{Pb,dec}$	$f_{Pb,dec}$	0	0	0	0	0	0	0	12
Hydrolysis																
Oxygen dissolution	1	0	0	0	f_{S,O_2}	$f_{S,I2,SS}$	$f_{S,I2,SS}$	$f_{S,I2,SS}$	$f_{S,I2,SS}$	0	0	0	0	0	-1	13
Carbon dioxide stripping/dissolution	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	14
Hydrogen stripping	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	15
Nitrogen stripping	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	16

et al. 2017). VFAs and other dissolved organics follow a different uptake pattern, and lead to different selections (Cerruti *et al.* 2023). Furthermore, subdividing soluble solids based on individual VFA might also be reasonable because VFAs induce different maximal specific growth rates, yet might increase complexity, thereby, making the model impractical for wastewater treatment applications. As nutrients, only inorganic nitrogen (S_{IN}) and phosphorus (S_{IP}) were included, since organic fractions can be hydrolyzed in the preceding fermentation as suggested by Alloul *et al.* (2021a) for operational functionality of raceway reactors in producing PPB.

Since raceway systems are open reactors, the concentration of dissolved oxygen (S_{O_2}) plays an important role, especially since AEB can outcompete PPB. The effect of the paddle wheel on the mass transfer of dissolved oxygen in the liquid phase was also taken into account for modeling. To close carbon, electron and mass balances, the production of carbon dioxide (S_{CO_2}) and hydrogen (S_{H_2}), which can come from non-VFA and VFAs, were included in the model.

The mentioned biochemical and physical processes take place in the raceway reactor. The lab-scale and even pilot-scale raceway reactors can be modelled as a single continuous-flow stirred tank (CSTR) due to their reduced dimensions and the high water circulation flow. Sequencing batch with cycles of filling the reactor with influent once a day and turning on the paddle wheel to promote gas transfer during the day, and stopping the paddle wheel to slow down the reactions and settle the biomass during nights prior to purging the excess sludge, is assumed as a modeling of the reactor regime to simulate the real situation in physical models in labs. The model can also be used either for other reactor measures by changing volume (V) and surface area (A) in the model implementation or for other reactor geometry systems such as tubular or flat plate photobioreactors with some modifications to consider biofilm formation. Raceway reactors typically have a liquid depth of 20 cm and a surface-to-volume ratio of $5 \text{ m}^2 \text{ m}^{-3}$ (Norsker *et al.* 2011).

(iii) *Balances and system dynamics equations:* The reactor volume may be varied during filling and extracting phases, especially if they are not done at the same time. Its variability, therefore, can be written as a following differential equation:

$$\frac{dV}{dt} = Q_{\text{inflow}}(t) - Q_{\text{outflow}}(t), \quad (2.1)$$

where, V , Q_{inflow} , Q_{outflow} denote the reactor volume and the input and output flow rates, respectively. Therefore, the mass balances of the soluble materials can be written as follows:

$$\frac{dVS_i}{dt} = S_i^{\text{input}}Q_{\text{inflow}}(t) - S_iQ_{\text{outflow}}(t) + \sum v_i\rho_i, \quad (2.2)$$

where v_i , ρ_i denote the reactor volume and the input and output flow rates, the initial volume, the stoichiometric coefficient, and the process rate of the corresponding conversion, respectively, in which the subscript i denotes component name.

The mass balances for the particulate materials can be written in a similar way as follows, while integrating a factor related to hydraulic (HRT) and sludge (SRT) retention times to model the effect of the paddle wheel activation, the effluent

extraction, and the water recirculation if needed. This factor was defined as the HRT/SRT ratio ($f_{H/S}$) to consider the fraction of removed particles. The HRT/SRT ratio provides insights into the performance and stability of the reactor, which should be optimized. The mass balance equation for particulate materials becomes the following, considering transport and conversion terms:

$$\frac{dV X_i}{dt} = X_i^{\text{input}} Q_{\text{inflow}}(t) - X_i(Q_{\text{outflow}}(t) + (f_{H/S} - 1) \frac{dV}{dt}) + \sum v_i \rho_i. \quad (2.3)$$

In addition, the mass balance equation of oxygen was modified to take the effect of oxygen dissolution into account when the paddle wheel is working. A switching function associated with the on/off conditions of the paddle wheel was, therefore, integrated to consider oxygen promotion.

(iv) *Gas-liquid transfers*: These systems are open to air and agitated with a paddle wheel. Gas-liquid mass transfer of four components, was, therefore, included in the model, namely: oxygen dissolution from the air through paddle wheel rotation; carbon dioxide dissolution from the air or produced through biological processes; hydrogen dissolution from the air or produced through biological processes; and ammonia origination from the incoming wastewater. The mass transfer kinetics for all gases was described through volumetric mass transfer rate and the gas saturation concentration through Henry's law.

(v) *Inhibitory factors*: As for the activated sludge model and anaerobic digestion model, kinetics expressions are multiplicative and based on Monod type functions to describe limitations of organics, ammonium, phosphate, light, and oxygen. In comparison with the PAnM (Puyol *et al.* 2017) and the ePAnM (Capson-Tojo *et al.* 2023), since, the PBB community was not considered as a single cell in the PBM, light inhibitory factor is differentiated between photoheterotrophy and chemoheterotrophy. Moreover, different oxygen inhibitions were taken into account for PBB photoheterotrophy and chemoheterotrophy as well as aerobic and anaerobic microbial communities. Competitive inhibition function between VFAs and other soluble organics was also included into the PBM as PBB growth competition has been reported by (Cerruti *et al.* 2020).

PPB AND ITS COMPETITION

(i) *Metabolic versatility of PPB*: Six biological processes were assigned to PPB, namely: two photoheterotrophic growths on soluble organics and VFAs, respectively; two aerobic chemoheterotrophic growths on soluble organics and VFAs, respectively; one anaerobic chemoheterotrophic growth on soluble organics; and biomass decay into biodegradable materials and inerts while releasing inorganic nitrogen, phosphorus and carbon. The PPB biomass therefore comprises grow metabolisms by photoheterotrophy ($X_{PB,ph}$), aerobic chemoheterotrophy ($X_{PB,aec}$), and anaerobic chemoheterotrophy ($X_{PB,anc}$). This subdivision reflects the different types of metabolisms that PPB conduct in a raceway reactor subjected to varying environmental conditions, e.g., availability of light, oxygen, fermentable organics, etc. (Alloul *et al.* 2021a). The model was written to account for the ability of PPB to grow on different substrates (electron donors and carbon sources) or

energy sources (light or chemical redox reactions) in parallel and to hypothesize a combined metabolic-mechanistic understanding to the most opportune metabolism. The parallel metabolic growth of PPB was implemented in the model by introducing a novel factor and inhibition-type functions with light and oxygen to control photoheterotrophy and chemoheterotrophies, respectively. Therefore, besides the light and oxygen inhibition functions, a parallel metabolic growth constant (M_S) between the three PPB biomass types was included, so that biomass is grown phototrophically and chemotrophically in parallel during shifting between days and nights. This factor plays a pivotal role in accounting for the contribution of alternative pathways to PPB growth alongside the dominant pathway. This novel factor is the main difference between the PBM proposed and the PAnM and ePAnM developed by (Capson-Tojo *et al.* 2023; Puyol *et al.* 2017) and can be written as follows:

$$f_{ph} = X_{PB,ph} + M_S(X_{PB,aec} + X_{PB,anc}) \quad (2.4a)$$

$$f_{aec} = X_{PB,aec} + M_S(X_{PB,ph} + X_{PB,anc}) \quad (2.4b)$$

$$f_{anc} = X_{PB,anc} + M_S(X_{PB,ph} + X_{PB,aec}) \quad (2.4c)$$

where f_{ph} , f_{aec} , and f_{anc} represent state variables with regard to photoheterotrophic, aerobic chemoheterotrophic, and anaerobic chemoheterotrophic growths, respectively. This constant means that biomass grown photoheterotrophically is able to use chemoheterotrophic conversion and vice versa, which have been experimentally observed by Alloul *et al.* (2021b) and Cerruti *et al.* (2023). PPB are known to divide their metabolic growth pathways over photo- and chemotrophy. An M_S of zero implies that PPB biomass grown photoheterotrophically cannot switch to chemoheterotrophy and is, therefore, completely independent of chemoheterotrophy. Without the inclusion and proper calibration of this constant, the model is not able to accurately predict the production of PPB.

(ii) *Microbial competitors*: Microbial processes that compete with PPB metabolisms in the raceway biomass are mainly aerobic chemoheterotrophy by AEB and anaerobic chemoheterotrophy by ANB. Nitrifiers, methanogens, denitrifiers and sulfate reducing bacteria were not implemented in the model since not prevalent in open raceway reactors (Alloul *et al.* 2021a). It should be highlighted that the model structure can be adapted with additional processes (such as described in ASM and ADM) by users, depending on the local conditions to be investigated.

2.2.2. MODEL ANALYSIS

In this section, the methodology employed to analyze the PBM is described. Sensitivity analysis was carried out to first identify the influential parameters of the model, then calibrate these impactful model parameters using data from controlled experiments. The robustness of the model was assessed by simulating the model under different operational scenarios.

SENSITIVITY ANALYSIS

After model construction, a local sensitivity analysis was conducted to assess the impact of each parameter on the PBM. The PBM contains various parameters, ranging from physical to kinetic parameters, in which physical parameters and yields have been broadly researched and validated. Total kinetic parameters including maximal specific growth rates (μ_m), substrate affinities (K_S), inhibition constants (K_I), and specific decay rate (b_m) associated with microbial community are covered in the sensitivity analysis, which a part of them is directly linked to PPB.

As baseline, kinetic parameters were obtained from previous researches. The maximal specific and decay rates were given from the activated sludge model (Henze *et al.* 2015), the ADM1 (Batstone *et al.* 2002), the PAnM (Puyol *et al.* 2017), and the substrate affinities from Capson-Tojo *et al.* (2021) and Katsuda *et al.* (2000). Newly-defined parameters, such as PPB oxygen affinities, were chosen based on the experiments (Alloul *et al.* 2021a), trial and error, as well as the expert knowledge. Therefore, to measure the impact of parameters on the model output, sensitivity functions were computed one-at-the-time (OAT). Each parameter was uniformly perturbed ten times (the sensitivity coefficient) higher and lower of the default values (Manhaeghe *et al.* 2020). The duration time of each simulation was set to 50 days to ensure reaching the steady-state condition. The relative abundance of PPB, COD removal rate ($\text{mgCODL}^{-1}\text{d}^{-1}$), biomass productivity ($\text{mgCODL}^{-1}\text{d}^{-1}$) and biomass yield ($\text{mgCODbiomass mg}^{-1}\text{CODremoved}$) were selected as outputs because of their feasibility to measure for the experiments. The mean of the model output for the last five days (when the process reaches steady-state) for each iteration was calculated, and to be able to compare the sensitivity functions, the absolute sensitivity should be converted into relative sensitivity by dividing the output results by the baseline (default values). Therefore, the outcome can be compared between parameters and outputs. The wastewater composition for the sensitivity analysis was considered based on the synthetic medium proposed by Alloul *et al.* (2021a).

OVERVIEW OF THE EXPERIMENTS

Based on operational strategies to selectively produce PPB raceway reactors discussed by Alloul *et al.* (2021a), three scenarios were selected as (i) 24 h stirring at a surface-to-volume ratio of $5\text{ m}^2\text{ m}^{-3}$ with a 12 h light and 12 h dark regime (scenario 1), (ii) 12 h stirring during the light period at a surface-to-volume ratio of $5\text{ m}^2\text{ m}^{-3}$ with a 12 h light and 12 h dark regime (scenario 2), (iii) 24 h stirring at a surface-to-volume ratio of $10\text{ m}^2\text{ m}^{-3}$ with a 12 h light and 12 h dark regime (scenario 3).

These three controlled experiment were conducted in a 100-L pilot scale raceway reactor with a mixed PPB culture dominated by *Rhodobacter capsulatus* and a *Rhodospseudomonas* (Alloul *et al.* 2021a). The reactor has been operated for 40 days on synthetic wastewater at a temperature of 28°C and illuminated artificially with halogen lamps (50 W m^{-2}). The raceway had a depth of 10–20 cm and a surface-to-volume ratio of $5\text{--}10\text{ m}^2\text{ m}^{-3}$. A VFAs solution was used as substrate composed by acetate, propionate and butyrate in a ratio of 1/1/1 gCODL^{-1} . All experiments started with a volatile suspended solid (VSS) concentration of

0.02 gVSSL⁻¹. The initial and final total COD and soluble COD were measured in order to analyze the COD removal, final biomass and yield of the reactions. Moreover, optical density at 660 nm (OD660) was measured to extrapolate growth. The ratio between absorbance (A660) and TSS was also taken from (Cerruti *et al.* 2020). The focus was on measuring PPB abundance, which aligns with practical feasibility.

CALIBRATION AND VALIDATION

A relevant range for each of selected parameters was selected based on expert knowledge and sensitivity analysis to be calibrated. All possible combinations of the five parameters were then computed for the three mentioned operational strategies. The mean of the last five simulated days when the process reaches steady-state was, then, calculated for each iteration of every individual outputs, i.e. relative abundance of PPB, COD removal rate, biomass productivity and biomass yield. Therefore, the relative error (e_{rel}) was calculated as the absolute difference between the simulation and the experiment divided by the experimental value, given as follows:

$$e_{rel} = \frac{|y^{\text{simulation}} - y^{\text{experiment}}|}{y^{\text{experiment}}}, \quad (2.5)$$

where, $y^{\text{simulation}}$ and $y^{\text{experiment}}$ denote each of the mentioned outputs from the simulation and experiment, respectively. This was achieved by implementing a nested loop structure using the “*for*” programming construct to identify a parameters’ combination that exhibited the lowest relative error for each model output, while also demonstrating a consistent trend across the three strategies. This procedure led to find the calibrated parameters, while assessing the model performance with the relative error as summarized in Table 4.

PERTURBED SCENARIO ASSESSMENT

The calibrated model can be used to improve the design and operation of PPB-based reactors. In pilot and full-scale systems, for example, process and environmental perturbation can influence the stability of the PPB community and wastewater treatment performance. Four short-term perturbations, likely to occur in a full-scale PPB system, were simulated to assess the effect of the perturbations on the model, namely as: (i) incoming VFA concentration (250, 500, and 3000 mgCODL⁻¹d⁻¹), (ii) incoming soluble organic matters (0, 100, and 3000 mgCODL⁻¹d⁻¹), (iii) incoming suspended solids (0, 250, 1000 mgCODL⁻¹d⁻¹), and (iv) the light intensity (14, 54, 108 Wm⁻²). The simulation was run for 25 days at a 12h light and 12h dark condition and stirring regime of 12h on and 12h off to reach steady-state and the perturbations were, then, implemented.

2.3. RESULTS AND DISCUSSION

The lab-scale and even pilot-scale raceway reactors can be modelled as a single continuous-flow stirred tank (CSTR) due to their reduced dimensions and the high

water circulation flow. Sequencing batch with cycles of filling the reactor with influent once a day and turning on the paddle wheel to promote gas transfer during the day, and stopping the paddle wheel to slow down the reactions and settle the biomass during nights prior to purging the excess sludge, was assumed as a modeling of the reactor regime to simulate the real situation in physical models in labs.

The average light intensity in the bioreactor was determined based on Lambert-Beer's Law, which was formulated in the form of an inhibition function to differentiate light illumination between day and night. Light attenuation (Solimeno *et al.* 2017) was also taken into account in function of the biomass concentration to formulate a close-to-real-world condition for light intensity. The model is capable of simulating the process under real-world conditions by incorporating historical solar irradiation data as well as other operational scenarios with artificial lighting, as light intensity serves as an input parameter for the model's implementation. In the following, the results are presented in three distinct sections: sensitivity analysis (Table 2.5), calibration and validation (Table 2.6), and perturbed scenario assessment (Figure 2.1).

2.3.1. SENSITIVITY ANALYSIS

The sensitivity analysis revealed the influential parameters for further calibration. Maximal specific growth rates (μ_m), specific decay rates (b_m) and kinetic parameters related to oxygen as well as light constants had the greatest impact on the PPB abundance. Substrate half-saturation constants (K_S) such as the soluble organic for phototrophic and chemotrophic growth of PPB and VFA for chemotrophic and aerobic heterotrophic growth of PPB did not show any or only a very low impact (close to zero). Similar results have been also reported by Biase *et al.* (2021), for a high-rate moving bed biofilm reactor model, which showed that the aerobic decay rate and maximal specific growth rate of AEB had a 5–12 times stronger impact on the model compared to substrate half-saturation constants.

The effect of parameter variation on the model output was only impactful in a specific range, typically between 0.5–545 times the baseline values. Furthermore, the sensitivity analysis showed that the relative PPB abundance was more sensitive to parameter variations compared to the other model outputs. This is because of splitting PPB community and differentiating among light and oxygen inhibitions to build a model more accurate in terms of predicting PPB abundance. This also reflected the challenge to construct a model, which predicts relative PPB abundances.

The most impactful parameter is the maximal specific growth rate of aerobic chemoheterotrophic AEB ($\mu_{m,SS,AEB}$), which directly affects the microbial community distribution and COD removal rates. Managing passive oxygen entry into a raceway reactor, especially during nighttime through paddle wheel control, can boost PPB abundance significantly.

Additionally, the maximal specific aerobic chemoheterotrophic growth rate of PPB on VFA ($\mu_{m,VFA,PPB,aec}$) is essential for PPB to compete with AEB under dark aerobic conditions. Parameters related to oxygen, such as the oxygen half-saturation constants ($K_{S,O_2,PPB}$ and $K_{S,O_2,AEB}$) and oxygen inhibitory constant for phototrophic

growth of PPB ($K_{I,O_2,PPB}$), also have a substantial impact, particularly in low dissolved oxygen raceway systems.

Furthermore, light-related parameters, including the light half-saturation constant of PPB ($K_{S,E}$) and the light inhibitory constant for chemotrophic growth of PPB ($K_{I,E}$), play a critical role. These parameters determine the balance between aerobic chemoheterotrophy and phototrophy in response to varying light intensities. A summary of sensitivity analysis is provided in Table 2.5.

Overall, the sensitivity analysis uncovered the impactful parameters of the PBM. Specifically the parameters $K_{S,O_2,PPB}$, $K_{I,O_2,PPB}$, $K_{S,E}$, $K_{I,E}$ and M_S are thus far unreported or only limited-studied, yet have a strong impact on the model. Therefore, the calibration of these five parameters and comparison of the simulated output with experimental data of a raceway reactor were presented in the following.

2.3.2. CALIBRATION AND VALIDATION

According to the sensitivity analysis, five impactful PPB-related parameters were identified that have not been described in literature and not yet calibrated for a raceway reactor, namely the oxygen half-saturation constant for aerobic chemoheterotrophic growth of PPB ($K_{S,O_2,PPB}$), oxygen inhibitory constant for phototrophic growth of PPB ($K_{I,O_2,PPB}$), light half-saturation constant of PPB ($K_{S,E}$), light inhibitory constant for chemotrophic growth of PPB ($K_{I,E}$) and the parallel metabolic growth factor (M_S) that is newly introduced to the model. $K_{I,O_2,PPB}$, $K_{I,E}$, and M_S have not been reported in literature, yet are crucial for the proper functioning of the model based on their impact on the PPB abundance and COD removal.

Model calibration combined with data of the three controlled experiments explained in "overview of experiments" section, resulted in a value for $K_{S,O_2,PPB}$, $K_{I,O_2,PPB}$, $K_{S,E}$, $K_{I,E}$ and M_S of $0.05 \text{ mgO}_2\text{L}^{-1}$, $5 \text{ mgO}_2\text{L}^{-1}$, 4 Wm^{-2} , 135 Wm^{-2} , and 0.28, respectively. Overall, the simulated and experimental outputs are in good agreement and show a similar trend over the three operational strategies, based on the summarized results in Table 2.6. In terms of relative error, the model has predicted the third scenario of the experiments, i.e. the half-day light and half-day dark condition along with constantly stirring, more accurately than the other two. In terms of COD output accuracy, relative errors of COD removal rate, biomass productivity, and PPB abundance are in an acceptable range.

Two oxygen-PPB-related parameters were calibrated, namely the oxygen half-saturation constant for aerobic chemoheterotrophic growth of PPB ($K_{S,O_2,PPB}$) and oxygen inhibitory constant for phototrophic growth of PPB ($K_{I,O_2,PPB}$). The oxygen half-saturation constant was calibrated to $0.05 \text{ mgO}_2\text{L}^{-1}$, which is aligned with the oxygen half-saturation constant for AEB in the activated sludge model (Henze *et al.* 2015). The second oxygen-PPB-related parameter, i.e. $K_{I,O_2,PPB}$, is responsible for the direct oxygen suppression of the photoheterotrophic growth. A value of $5 \text{ mgO}_2\text{L}^{-1}$ is assigned to the parameter after calibration. Calibrating the oxygen inhibitory constant through a dedicated experimental setup at different oxygen concentrations is challenging. PPB are able to grow both photo- and chemoheterotrophically. Increasing the dissolved oxygen concentration even results in an enhancement of

Table 2.5: Sensitivity analysis of the most impactful kinetic parameters of the PBM. Sensitivity analysis is computed by changing the default value of each parameter one by one, ten times higher and lower, while keeping the other parameters constant. The mean for the PPB abundance, chemical oxygen demand (COD) removal rate (mg COD removed $L^{-1} d^{-1}$), biomass productivity (mg COD biomass d^{-1}) and the biomass yield (mg COD biomass mgCODremoved) of the last five days for each iteration is calculated and divided by the output results of the baseline (default values) to obtain the relative error.

Parameter (unit)	Symbol	Default	PPB abundance	COD removal	Impact on		Biomass yield
					Biomass productivity	Biomass yield	
Parallel metabolic growth (-)	M_S	0.300	13.98	0.33	1.72		-0.74
Maximal specific of AEB on soluble organics (d^{-1})	$\mu_{m,SS,AEB}$	0.076	-10.97	-10.97	-2.24		0.04
Maximal specific aerobic chemotrophic growth rate of PPB on volatile fatty acids (d^{-1})	$\mu_{m,VFA,AEB}$	0.053	10.94	-10.94	-0.60		0.07
Specific decay rate of PPB (d^{-1})	$D_{m,PPB,dec}$	0.011	-10.11	-10.11	-2.29		0.54
Oxygen half-saturation constant for chemotrophic growth of PPB (mgO ₂ L ⁻¹)	$K_{S,O_2,PP}$	0.050	-9.92	-9.92	0.53		-0.08
Maximal specific phototrophic growth rate of PPB on volatile fatty acids (d^{-1})	$\mu_{m,VFA,PP,ph}$	0.078	9.90	9.90	2.09		-0.93
Oxygen half-saturation constant for AEB (mgO ₂ L ⁻¹)	$K_{S,O_2,AEB}$	0.050	-9.26	-9.26	-0.50		-0.06
Specific decay rate of AEB (mgO ₂ L ⁻¹)	$D_{m,AEB,dec}$	0.016	8.99	8.99	0.49		-0.06
Light half-saturation constant of PPB (Wm ²)	$K_{S,E}$	4.500	-7.44	-7.44	0.39		2.29
Maximal specific aerobic chemotrophic growth rate of PPB on soluble organics (d^{-1})	$\mu_{m,SS,PP,acc}$	0.050	2.87	2.87	-0.16		-0.03
Maximal specific anaerobic chemotrophic growth rate of PPB on soluble organics (d^{-1})	$\mu_{m,SS,PP,anc}$	0.012	2.36	0.14	-0.13		0.07
Light inhibitory constant for chemotrophic growth of PPB (Wm ²)	$K_{I,E}$	200.000	1.40	1.40	-0.09		-0.02
Maximal specific phototrophic growth rate of PPB on soluble organics (d^{-1})	$\mu_{m,SS,PP,ph}$	0.053	1.31	1.31	-0.07		0.04
Oxygen inhibitory constant for phototrophic growth of PPB (mgO ₂ L ⁻¹)	$K_{I,O_2,PP}$	5.000	1.20	0.00	-0.08		-0.02

the growth rate due to additional chemoheterotrophic conversion, e.g. 10% increase in maximal specific growth rate subject to increase in oxygen transfer from 72 to 336 mgO₂L⁻¹ (Alloul *et al.* 2021a). It should be highlighted that next to dissolved oxygen concentration, the oxygen transfer rate is also a key factor steering the microbial community in a raceway reactor. Together with the oxygen uptake rate, it influences the actual dissolved Oxygen concentration in the system.

Two light-related parameters, i.e. $K_{S,E}$ and $K_{I,E}$ were also calibrated. The calibrated value of $K_{I,E}$ is 4 Wm⁻², in the range (i.e. 4.58±7.40 Wm⁻²) reported by Capson-Tojo *et al.* (2022). In comparison with other work and their experiments such as Capson-Tojo *et al.* (2022) and Katsuda *et al.* (2000), it can be concluded that different wavelengths and inoculum result in other type of pigment responsible for light capturing, which might also affect the effective the light half-saturation constant. The second calibrated light-related parameter of the model inhibits the (an)aerobic chemoheterotrophic growth of PPB. Along with $K_{I,O_2,PB}$, it enables the PPB community to allocate their metabolic growth pathway between photo- and chemoheterotrophy.

The parallel metabolic growth factor (M_S) as the final parameter considered for calibration, allows PPB to account different metabolisms in parallel. The final calibrated value was assigned to 0.28. This suggests that the PPB growth in a raceway reactor is probably not governed by independent subpopulations, however individual PPB cells may switch between metabolism. The experiments done by Alloul *et al.* (2021b) showed that the PPB species *Rhodobacter capsulatus*, *Rb. sphaeroides*, *Rhodopseudomonas palustris* and *Rhodospirillum rubrum* are able to switch from photoheterotrophy to aerobic chemoheterotrophy and from photoheterotrophy to photoautotrophy, which supports the idea of integrating the metabolic constant to the PBM.

2.3.3. PERTURBED SCENARIO ASSESSMENT

The effect of process and environmental perturbations, which are likely-to-occur in pilot- and full-scale raceway systems, was assessed with respect to the stability of the PPB community. Four different perturbation scenarios, as described, were simulated and the results were depicted in Figure 2.1. The fluctuations observed in the curves represent daily variations in PPB abundance. PPB are produced through both photoheterotrophic and aerobic and anaerobic chemoheterotrophic processes. In the context of an open reactor, the contribution of anaerobic chemoheterotrophic growth to PPB abundance is minimal. Instead, the primary metabolic pathway for PPB growth is through photoheterotrophy. This growth steadily increases when exposed to solar radiation and decreases in the absence of light, that causes daily fluctuation in PPB abundance.

The wastewater composition can change throughout the day. The effect of influent VFA variability (500–3000 mgCODL⁻¹) was, therefore, assessed on the model performance. This scenario is depicted in Figure 1 (A) showing that the PPB abundance is immediately affected when the incoming VFA concentration drops from 3000 to 500 mgCODL⁻¹. When organic carbon is limited in the reactor, dissolved oxygen can freely increase up to 3 mgO₂L⁻¹. This oxygen increase eventually

Table 2.6: Calibration results including calibrated model outputs for relative purple phototrophic bacteria abundance, chemical oxygen demand (COD) removal rate, biomass productivity, and biomass yield and associated experimental data for three defined operational scenarios. Mean relative errors for the scenarios as well as mean relative errors for the selected outputs are also provided.

Scenario	Experiment and simulation conditions		Purple bacteria (Relative abundance %)		COD removal rate (mgCODL ⁻¹)		Biomass productivity (mgCODL ⁻¹)		Biomass yield (mgCOD _{biomas} mg ⁻¹ COD _{removed})		Relative error (%)	
	Illumination (light/dark)	Stirring (on/off)	Surface-to-volume (m ² m ⁻³)	Experiment	Model	Experiment	Model	Experiment	Model	Experiment		Model
1	12h/12h	24h on	5	14	14	721.83	702.66	0.45	0.64	326	459	21.83
2	12h/12h	12h/12h	5	46	27	462.50	593.52	0.65	0.62	302	390	26.04
3	12h/12h	24h on	10	78	90	785.17	822.31	0.58	0.53	452	577	14.08
Relative error (→)				19.50		11.90		18.57		32.62		-

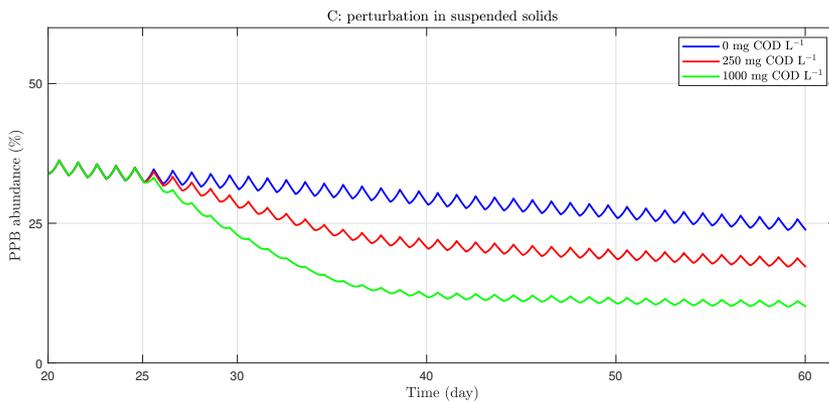
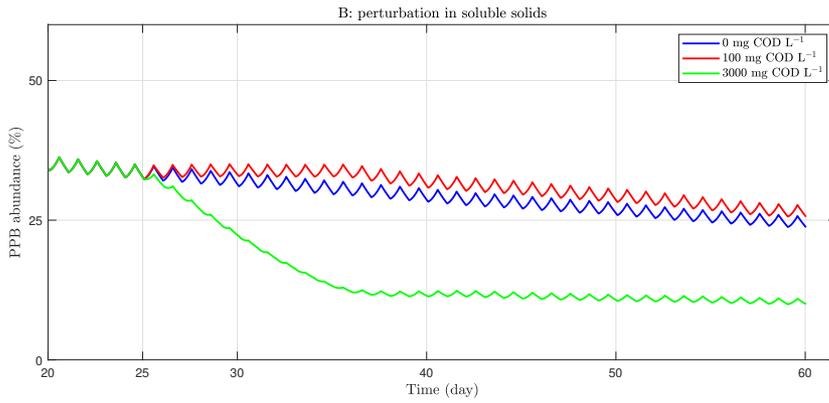
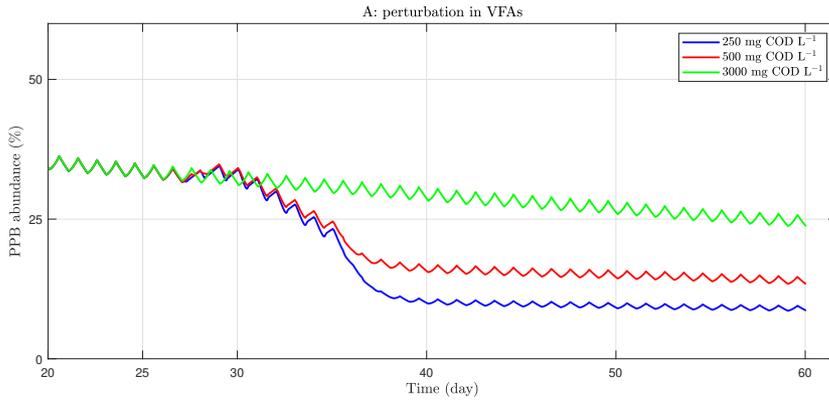
inhibits the phototrophic growth of PPB. Lowering the HRT and, thus, increasing the volumetric loading rate can be a contingency measure, yet the raceway reactor is typically designed for a certain range in flow rate. Increasing the SRT is a better solution as it increases the retention of PPB in the system, thereby, lowering the impact of oxygen inhibition. The increase in oxygen still occurs, yet PPB is less heavily affected. A faster contingency strategy would be decreasing paddle wheel rotation or even stopping it completely. An oxygen sensor with paddle wheel speed control can help to steer the oxygen transfer to the system to cope with changes in loading rates during operation.

The second perturbation considered the effect of soluble organic matters (non-VFA matters) in the incoming wastewater ($0\text{--}3000\text{mgCODL}^{-1}$) as depicted in Figure 1 (B). For PPB systems, several articles recommended to preferment the wastewater, thereby, separating acidogenic fermentation and PPB biomass production (Alloul *et al.* 2021a, 2018; Cerruti *et al.* 2023). This would result in lower competition with acidogenic fermentative microorganisms and provides more accessible organic carbon in the form of VFA for PPB to grow. PPB are also able to ferment, yet their maximal specific anaerobic chemoheterotrophic growth rate is lower compared to non-PPB, i.e. 0.3 vs 0.6d^{-1} (Alloul *et al.* 2021a). For instance, PPB cannot compete with acidogenic fermenters on sucrose-rich synthetic wastewater (Cerruti *et al.* 2023). The perturbations reconfirmed these experimental observations as shown in Figure 1(B), in which a drop in relative PPB abundance from 30 % to 11 % when the wastewater composition changes from 100 % VFA to 100 % non-VFA.

The third perturbation simulated variations in influent suspended solids. These organics can cause higher turbidity of the water and, thereby, lower the light availability for phototrophs and less PPB abundance as depicted in Figure 1 (C). Suspended solids concentrations as low as 250mgCODL^{-1} drops the PPB abundance from 26 % to 18 %. Suspended solids concentrations of 1000mgCODL^{-1} resulted in a complete collapse of the PPB community. In this case, stopping the feed and operate the reactor in semi-continuous mode until full recovery achieved again would be a solution. The results also showed that a good solid/liquid separation is crucial to prevent suspended material from entering the raceway reactor.

Light intensity was the last studied perturbation. The results again showed a serious disruption on the PPB community, as depicted in Figure 1 (D). As expected, variations of light intensity had an important impact on the stability of the phototrophic community; when the light intensity decreases from 108 to 14Wm^{-2} for 10 days, the final PPB abundance decreases from 27 % to 7 %.

The effect of SRT and HRT on the PPB abundance and the treatment performance is also crucial, particularly SRT that can cope with sudden perturbations that may destabilize the reactor performance as a contingency measure. The extension of the SRT would minimize the effect of TSS in influent to increase the PPB abundance, while increasing the COD removal efficiency. For an incoming concentration of 3000mgCODL^{-1} , an SRT between 3–4 d can be chosen to minimally reach a COD removal efficiency of 60 %, considering HRT between 3–4 d. Higher COD removal efficiencies up to 100 % are achievable for SRT and HRT, with less HRT compared to SRT. The related figure to the effect of SRT and HRT on the PPB abundance and the



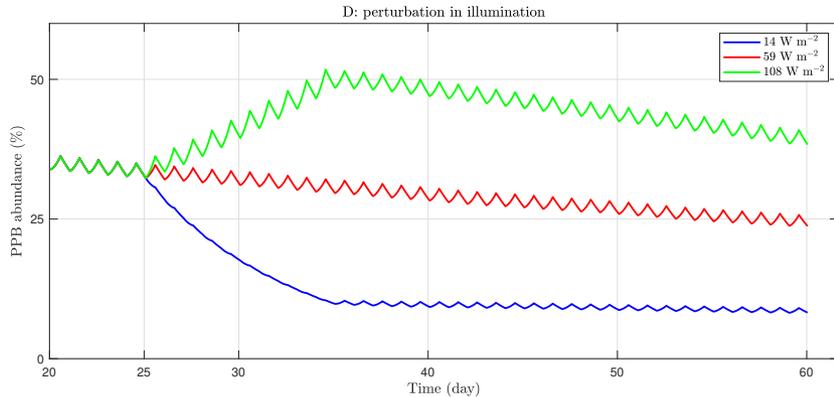


Figure 2.1: Effects of the perturbation scenarios on relative PPB abundance based on variations on (A) influent VFA (250, 500, and 3000 mg COD L⁻¹), (B) influent soluble organic (non-VFA) (0, 100, and 3000 mg COD L⁻¹), (C) influent suspended solid (0, 250, and 1000 mg COD L⁻¹), and (D) light (14, 54, and 108 Wm⁻²). The perturbation was applied on day 25.

treatment performance can be seen in Figure 2.2.

2.3.4. MODEL IMPLICATIONS AND FURTHER DEVELOPMENT

To reduce the multiscale complexity of the PBM, simplifications on metabolisms that PPB can adopt in nature were made. For instance, photolithoautotrophic growth by PPB on carbon dioxide as a carbon source and dihydrogen or iron II as electron donors, and N_2 -fixation by PPB were not included. These latter metabolisms are less likely to occur in the raceway reactor. However, depending on local conditions to be studied, the PBM structure can be adapted. Species or strain specific metabolic traits of PPB such as denitrification, and autotrophic uptake of hydrogen were also not adopted in the model for the sake of simplicity and the type of raceway reactor.

For further modeling improvement, other PPB metabolisms and microbial community can be included, considering losing simplicity for simulation application to increasing accuracy for detailed biological analysis. The integration of temperature dependency and dynamic pH of the biological and other physical processes can also be added depending on an operational condition under consideration, since the temperature and pH were considered to be fixed in the PBM.

This model is primarily focused on offering a mechanistic framework within the realm of modeling to elucidate a specific observation derived from previously reported experimental studies. Since PBM is a new mechanistic model for PPB community in terms of metabolic-mechanistic understanding, the transfer of biotechnology from a proof-of-concept to practical application requires a proof-of-feasibility. Calibrating and verifying the model under various conditions and configurations would enable expanding its applicability and accuracy by designing some dedicated experiments

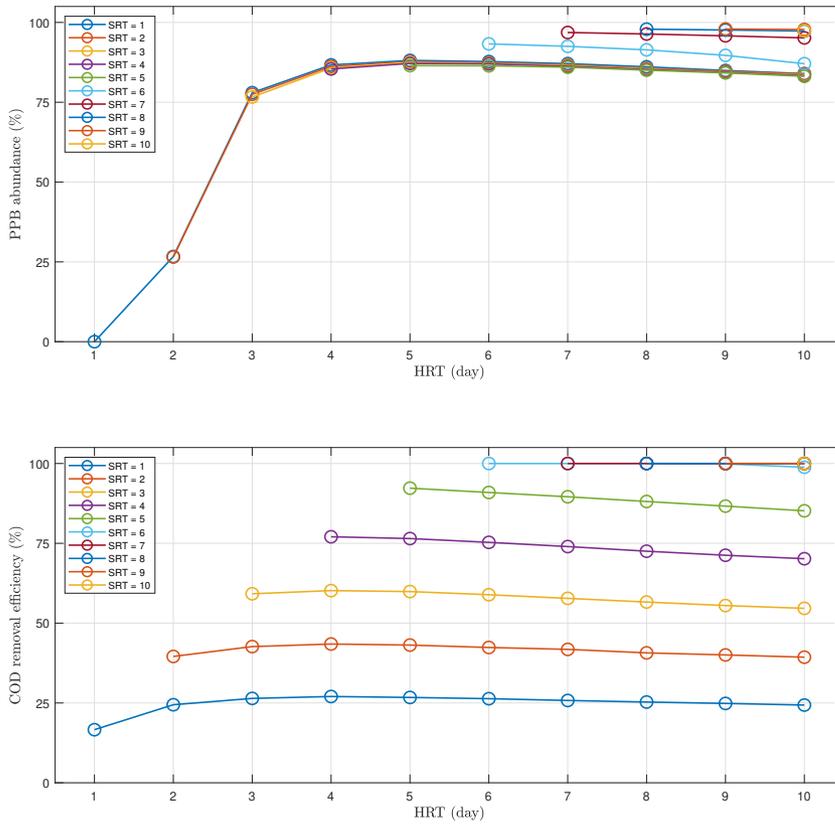


Figure 2.2: Effects of various HRT and SRT on PPB abundance and COD removal efficiency.

for different metabolisms. Additionally, kinetic parameters can also be experimentally estimated using different substrates and light ranges, considering a dominant metabolic growth pathway alongside different simultaneous potential growth ways, in which more experiments should be carried out to evaluate these options as a future research.

2.4. CONCLUSIONS

In this chapter, a new mechanistic model to describe the mechanisms of microbial selection, competition and conversions in an open raceway reactor was proposed. The model integrates three different metabolisms of PPB, and competing aerobic and anaerobic organisms. Integration of parallel metabolic growth was modeled. The model was shown to be able to accurately predict changes in the PPB abundance and COD removal rates with the relative errors 19.5% and 12%, respectively. The outputs

of the simulated perturbation scenarios were shown to capture both theoretical understanding and experimental observations.

REFERENCES

- Acién, F., J. Fernández, J. Magán, and E. Molina (2012). "Production cost of a real microalgae production plant and strategies to reduce it". In: *Biotechnology Advances* 30.6, pp. 1344–1353.
- Alloul, A., N. Blansaer, P. C. Segura, R. Wattiez, S. E. Vlaeminck, and B. Leroy (2023). "Dehazing redox homeostasis to foster purple bacteria biotechnology". In: *Trends in Biotechnology* 41.1, pp. 106–119.
- Alloul, A., M. Cerruti, D. Adamczyk, D. G. Weissbrodt, and S. E. Vlaeminck (2021a). "Operational Strategies to Selectively Produce Purple Bacteria for Microbial Protein in Raceway Reactors". In: *Environmental Science & Technology* 55.12, pp. 8278–8286.
- Alloul, A., R. Ganigué, M. Spiller, F. Meerburg, C. Cagnetta, K. Rabaey, and S. E. Vlaeminck (2018). "Capture-Ferment-Upgrade: A Three-Step Approach for the Valorization of Sewage Organics as Commodities". In: *Environmental Science & Technology* 52.12, pp. 6729–6742.
- Alloul, A., M. Muys, N. Hertoghs, F.-M. Kerckhof, and S. E. Vlaeminck (2021b). "Cocultivating aerobic heterotrophs and purple bacteria for microbial protein in sequential photo- and chemotrophic reactors". In: *Bioresource Technology* 319, p. 124192.
- Batstone, D., J. Keller, I. Angelidaki, S. Kalyuzhnyi, S. Pavlostathis, A. Rozzi, W. Sanders, H. Siegrist, and V. Vavilin (2002). "The IWA Anaerobic Digestion Model No 1 (ADM1)". In: *Water Science and Technology* 45.10, pp. 65–73.
- Biase, A., M. S. Kowalski, T. R. Devlin, and J. A. Oleszkiewicz (2021). "Modeling of the attached and suspended biomass fractions in a moving bed biofilm reactor". In: *Chemosphere* 275, p. 129937.
- Capson-Tojo, G., D. J. Batstone, M. Grassino, and T. Hülsen (2022). "Light attenuation in enriched purple phototrophic bacteria cultures: Implications for modelling and reactor design". In: *Water Research* 219, p. 118572.
- Capson-Tojo, G., D. J. Batstone, M. Grassino, S. E. Vlaeminck, D. Puyol, W. Verstraete, R. Kleerebezem, A. Oehmen, A. Ghimire, I. Pikaar, J. M. Lema, and T. Hülsen (2020). "Purple phototrophic bacteria for resource recovery: Challenges and opportunities". In: *Biotechnology Advances* 43, p. 107567.
- Capson-Tojo, G., D. J. Batstone, and T. Hülsen (2023). "Expanding mechanistic models to represent purple phototrophic bacteria enriched cultures growing outdoors". In: *Water Research* 229, p. 119401.
- Capson-Tojo, G., S. Lin, D. J. Batstone, and T. Hülsen (2021). "Purple phototrophic bacteria are outcompeted by aerobic heterotrophs in the presence of oxygen". In: *Water Research* 194, p. 116941.
- Casagli, F., G. Zuccaro, O. Bernard, J.-P. Steyer, and E. Ficara (2021). "ALBA: A comprehensive growth model to optimize algae-bacteria wastewater treatment in raceway ponds". In: *Water Research* 190, p. 116734.
- Cerruti, M., G. Crosset-Perrotin, M. Ananth, J. L. Rombouts, and D. G. Weissbrodt (2023). "Syntrophy between fermentative and purple phototrophic bacteria to treat and valorize carbohydrate-rich wastewaters". In: *Bioresource Technology Reports*, p. 101348.
- Cerruti, M., B. Stevens, S. Ebrahimi, A. Alloul, S. E. Vlaeminck, and D. G. Weissbrodt (2020). "Enrichment and Aggregation of Purple Non-sulfur Bacteria in a Mixed-Culture Sequencing-Batch Photobioreactor for Biological Nutrient Removal From Wastewater". In: *Frontiers in Bioengineering and Biotechnology* 8.
- Golomysova, A., M. Gomelsky, and P. S. Ivanov (2010). "Flux balance analysis of photo-heterotrophic growth of purple nonsulfur bacteria relevant to biohydrogen production". In: *International Journal of Hydrogen Energy* 35.23, pp. 12751–12760.
- Henze, M., W. Gujer, T. Mino, and M. van Loosdrecht (2015). "Activated Sludge Models ASM1, ASM2, ASM2d and ASM3". In: *Water Intelligence Online* 5.0.
- Hülsen, T., C. Züger, Z. M. Gan, D. J. Batstone, D. Solley, P. Ochre, B. Porter, and G. Capson-Tojo (2022). "Outdoor demonstration-scale flat plate photobioreactor for resource recovery with purple phototrophic bacteria". In: *Water Research* 216, p. 118327.
- Imam, S., S. Yilmaz, U. Sohmen, A. S. Gorzalski, J. L. Reed, D. R. Noguera, and T. J. Donohue (2011). "iRsp1095: A genome-scale reconstruction of the Rhodobacter sphaeroides metabolic network". In: *BMC Systems Biology* 5.1.
- Imhoff, J. (2006). "The phototrophic alpha-proteobacteria. Prokaryotes: a handbook on the biology of bacteria, Vol 5: Proteobacteria: alpha and beta subclasses". In: *Springer, New York*. doi 10, pp. 0–387.
- Katsuda, T., T. Arimoto, K. Igarashi, M. Azuma, J. Kato, S. Takakuwa, and H. Ooshima (2000). "Light intensity distribution in the exter-

- nally illuminated cylindrical photo-bioreactor and its application to hydrogen production by *Rhodobacter capsulatus*". In: *Biochemical Engineering Journal* 5.2, pp. 157–164.
- Manhaeghe, D., T. Blomme, S. V. Hulle, and D. Rousseau (2020). "Experimental assessment and mathematical modelling of the growth of *Chlorella vulgaris* under photoautotrophic, heterotrophic and mixotrophic conditions". In: *Water Research* 184, p. 116152.
- Norsker, N.-H., M. J. Barbosa, M. H. Vermuë, and R. H. Wijffels (2011). "Microalgal production — A close look at the economics". In: *Biotechnology Advances* 29.1, pp. 24–27.
- Puyol, D., E. Barry, T. Hülsen, and D. Batstone (2017). "A mechanistic model for anaerobic phototrophs in domestic wastewater applications: Photo-anaerobic model (PANM)". In: *Water Research* 116, pp. 241–253.
- Solimeno, A., F. G. Acien, and J. García (2017). "Mechanistic model for design, analysis, operation and control of microalgae cultures: Calibration and application to tubular photo-bioreactors". In: *Algal Research* 21, pp. 236–246.

3

SURVEY ON HYBRID SYSTEM IDENTIFICATION

Dynamical systems and processes that either exhibit non-smooth behaviours (e.g. through logic control or natural phenomena) or work in different modes of operation are usually represented using hybrid systems models, i.e. mathematical models that combine continuous dynamics with discrete-event dynamics. Identification of a hybrid system includes finding switching patterns and identification of model parameters to obtain a data-driven model. This survey provides a systematic review of models (how to parameterize the system) and methods (how to identify unknown parameters) proposed for hybrid system identification with an exposition of recent advances and developments, and further research directions.

This chapter is an adapted version of *models and methods for hybrid system identification: a systematic survey*, Ali Moradvandi, et al. (2023), *IFAC-PapersOnLine*.

3.1. INTRODUCTION

Intertwined continuous and discrete behaviors can be represented as hybrid dynamical, which can either model systems with non-smooth behaviors or approximate systems with nonlinearities. The discrete behavior can be represented as a switching pattern (mode) using a finite number of values as countable state variables that orchestrates over corresponding continuous subsystems (submodels). Hybrid system identification is a twofold problem: 1) estimating the parameters of submodels and 2) determining the switching patterns. While estimation of parameters of the model is the objective in classical system identification, hybrid system identification also requires the estimation of switching patterns. In other words, the hybrid system identification problem, where the switching mechanisms are known, reduces to the conventional system identification.

The first step is the determination of the modeling structure, i.e. model parameterization. The structure of submodels and switching mechanisms can be parameterized based on: 1) identifying what model structure is well-fitted to capture the dynamical behavior of the system, and 2) whether the main purpose of system identification is prediction or model-based control. These decisions delineate the arena of hybrid system identification and once a parameterization has been chosen, a methodology can be chosen to solve the identification problem.

Parameterized submodels can be classified into three main groups: input-output (Piga *et al.* 2020b), state space (Du *et al.* 2021b), and probabilistic models (Breschi *et al.* 2019). The input-output model complexity ranges from Auto-Regressive exogenous (ARX) models (Du *et al.* 2018) to the more complex Box-Jenkins (BJ) models (Piga *et al.* 2020b) and nonlinear models (Bianchi *et al.* 2020b). Complexity in the structure of a parameterized model to include dynamical noises and disturbances, and delay is a trend in recent years. State-space models, which are a more control-oriented model structure, have been widely discussed for linear (Sefidmazgi *et al.* 2016), affine (Du *et al.* 2021b) or nonlinear (Du *et al.* 2021a) representations. Furthermore, hybrid model representations in a probabilistic setting have recently drawn attention, since they can describe parametric uncertainties and external disturbances, and the number of parameters can be adapted as more data is collected (Piga *et al.* 2020a).

Similarly, the switching mechanism can be represented differently, depending on the system dynamics, switching behaviors, and purpose of modeling. Switching logics can be generally classified in three groups: polyhedral partitions (linearly or affinely partitioned) (Breschi *et al.* 2016b), random switching (Liu *et al.* 2021), and event-driven (deterministic (Basiri *et al.* 2018)). For hybrid system identification, the switching logic is mostly modelled either randomly (state-independent) or piecewise affinely (state-dependent). In the first part of this survey, a systematic discussion on parameterization of hybrid systems will be made, and various classes of hybrid systems for identification will be reviewed.

As a second aspect, besides system description, new recently-developed identification methods are reviewed and categorized into four groups. Taking into account the groups discussed in the previous survey by Garulli *et al.* (2012), most identification approaches fit in the so-called optimization-based framework.

Clustering-based (Du *et al.* 2020) and probabilistic (Chen *et al.* 2020a,b) methods are the other identification methods that are widely discussed in recent publications. The other methods that cannot be covered by the aforementioned classes are algebraic, bounded-error-based, continuous-time, and neural network approaches. An innovative trend is either combination or generalization of the approaches (Liu *et al.* 2022; Piga *et al.* 2020b; Tang and Dong 2020) for either dealing with new forms of parameterization or improving efficiency of computational burden and accuracy of the methods. The theory and mathematics of various well-established hybrid system identification methods have been discussed by Lauer and Bloch (2018), which was an inspiration for systematically modifying and expanding the categorization in this survey in comparison with Garulli *et al.* (2012) and Lauer and Bloch (2018). Moreover, discussing the probabilistic parametrization and the associated likelihood-based methods are two additions to the aforementioned reviews. Furthermore, the detailed specification of switching patterns, specifically piecewise affine is reviewed in both the sections on parameterization and solution method of this survey.

The chapter is organized as follows. Models of hybrid systems are presented in Section 3.2. We discuss input-output models (Section 3.2.1), state-space models (Section 3.2.2, switching mechanisms (Section 3.2.3), and probabilistic models (Section 3.2.4), and a comparison in Section 3.2.5. A review of methods is given in Section 3.3. This section consists of optimization-based methods (Section 3.3.1, clustering-based methods (Section 3.3.2), likelihood-based methods (Section 3.3.3), other methods (Section 3.3.4), and a comparison of important papers (Section 3.3.5). In the last section, conclusions are drawn to show current and future research directions.

3.2. MODELS OF HYBRID SYSTEMS: PARAMETERIZATION

3.2.1. INPUT-OUTPUT MODELS

A quite wide spectrum of hybrid systems can be represented in input-output (I/O) form as

$$y_k = f_{q_k}(x_k) + \varepsilon_k, \quad (3.1)$$

where $x_k \in \mathbb{R}^{n_x}$ is the regressor or input, $y_k \in \mathbb{R}^{n_y}$ is the output, and $\varepsilon_k \in \mathbb{R}^{n_y}$ is a noise vector, in which k denotes the index (e.g. time step) of the sequence, and $q_k \in \{1, \dots, N\}$ is the switching signal that determines which vector field, f_{q_k} , is active at time step k . The vector field f_{q_k} can be either linear or nonlinear. In case of linear functionality of hybrid systems, the vector field can be represented by $f_i(x_k) = x_k^T \theta_i$, in which $q_k = i$ (i -th mode).

The Switched Auto-Regressive eXogenous (SARX) model is the simplest and widely-used parameterization of hybrid systems for the identification problem. SARX models are a combination of several different ARX models defined as submodels. Discrete-time Single-Input Single-Output (SISO) SARX dynamical system can be expressed as

$$y_k = x_k^T \theta_{q_k} + \varepsilon_k, \quad (3.2)$$

where $x_k = [y_{k-1}, \dots, y_{k-n_a}, u_{k-1}, \dots, u_{k-n_b}]^T \in \mathbb{R}^{n_a+n_b}$ is the regressor, in which n_a and

n_b are the orders of the system, and y_k and u_k are the output and input at time step k , respectively. $q_k \in \mathcal{S} = \{1, \dots, N\}$ denotes the active mode at time step k and $\theta_i = [a_{i1}, \dots, a_{in_a}, b_{i1}, \dots, b_{in_b}, c_i]^T \in \mathbb{R}^{n_a+n_b+1}$ represents the parameter vector for the i -th mode, in which $i \in \mathcal{S}$.

Jump Box-Jenkins (JB) models have a more general and flexible structure for representing hybrid systems in comparison with the mentioned auto-regressive (AR) models, since they include both the moving average and auto-regressive terms to model dynamics of disturbances and noises. For this class of systems, the output is noise-corrupted, while the noise term is a dynamical noise. The noise-corrupted output, y_k , can be written with respect to the noise-free output, \hat{y}_k , and the noise term, v_k as follows:

$$y_k = \hat{y}_k + \varepsilon_k, \quad (3.3)$$

where the noise-free output, \hat{y}_k , is modelled based on the input u_k as follows:

$$\hat{y}_k = G(q^{-1}, \theta_i) u_k, \quad (3.4)$$

and noise term, v_k , is written as follows:

$$\varepsilon_k = H(q^{-1}, \theta_i) v_k, \quad (3.5)$$

where $G(q^{-1}, \theta_i)$ and $H(q^{-1}, \theta_i)$ are the linear filters in which θ_i denotes the i -th mode at time step k . These linear filters are rational functions of the time shift operator q^{-1} (i.e. $q^{-d}x_k = x_{k-d}$ for $d \in \mathbb{Z}$) as written below

$$G(q^{-1}, \theta_i) = \frac{B(q^{-1}, \theta_i)}{A(q^{-1}, \theta_i)} = \frac{b_{i1}q^{-1} + \dots + b_{in_b}q^{-n_b}}{1 + a_{i1}q^{-1} + \dots + a_{in_a}q^{-n_a}}, \quad (3.6a)$$

$$H(q^{-1}, \theta_i) = \frac{C(q^{-1}, \theta_i)}{D(q^{-1}, \theta_i)} = \frac{1 + c_{i1}q^{-1} + \dots + c_{in_c}q^{-n_c}}{1 + d_{i1}q^{-1} + \dots + d_{in_d}q^{-n_d}}, \quad (3.6b)$$

where n_a , n_b , n_c , and n_d are the orders of the system. The parameters vector that should be identified for the i -th submodel in a compact form is as follows:

$$\theta_i = [a_{i1}, \dots, a_{in_a}, b_{i1}, \dots, b_{in_b}, c_{i1}, \dots, c_{in_c}, d_{i1}, \dots, d_{in_d}]^T. \quad (3.7)$$

Substituting (3.6a) and (3.6b) into (3.4) and (3.5) results in

$$\hat{y}_k = (1 - A(q^{-1}, \theta_i))\hat{y}_k + B(q^{-1}, \theta_i)u_k = x_{k1}^T \theta_{i1}, \quad (3.8a)$$

$$\varepsilon_k = (1 - C(q^{-1}, \theta_i))\varepsilon_k + D(q^{-1}, \theta_i)v_k = x_{k2}^T \theta_{i2} + v_k, \quad (3.8b)$$

where x_{k1} and x_{k2} are the regressor vector defined as follows:

$$x_{k1} = [-\hat{y}_{k-1}, \dots, -\hat{y}_{k-n_a}, u_{k-1}, \dots, u_{k-n_b}]^T \in \mathbb{R}^{n_a+n_b}, \quad (3.9a)$$

$$x_{k2} = [-\varepsilon_{k-1}, \dots, -\varepsilon_{k-n_c}, v_{k-1}, \dots, v_{k-n_d}]^T \in \mathbb{R}^{n_c+n_d}, \quad (3.9b)$$

$$x_k = [x_{k1}^T, x_{k2}^T]^T \in \mathbb{R}^{n_a+n_b+n_c+n_d}, \quad (3.9c)$$

and the parameters vector can be expressed as

$$\theta_{i1} = [a_{i1}, \dots, a_{in_a}, b_{i1}, \dots, b_{in_b}]^T \in \mathbb{R}^{n_a+n_b}, \quad (3.10a)$$

$$\theta_{i2} = [c_{i1}, \dots, c_{in_c}, d_{i1}, \dots, d_{in_d}]^T \in \mathbb{R}^{n_c+n_d}, \quad (3.10b)$$

$$\theta_i = [\theta_{i1}^T, \theta_{i2}^T]^T \in \mathbb{R}^{n_a+n_b+n_c+n_d}, \quad (3.10c)$$

Therefore, according to (3.3), the model can be parameterized for the identification problem considering (3.2) as follows:

$$\begin{aligned} y_k &= x_{k1}^T \theta_{i1} + x_{k2}^T \theta_{i2} + v_k \\ &= x_k^T \theta_i + v_k. \end{aligned} \quad (3.11)$$

According to the definitions of $G(q^{-1}, \theta_i)$, $H(q^{-1}, \theta_i)$, $A(q^{-1}, \theta_i)$, $B(q^{-1}, \theta_i)$, $C(q^{-1}, \theta_i)$, and $D(q^{-1}, \theta_i)$ in (3.6a) and (3.6b), other classes of hybrid systems can also be defined. The switched Finite Impulse Response (SFIR) model is the simpler than SARX. The switched Auto-Regressive Moving-Average with eXogenous input (SARMAX) model is a well-structured model representing hybrid systems subject to disturbances. The switched Output-Error (SOE) model is a suitable hybrid model structure for systems subject to output measurement noise. The Error-in-Variable SARX (EIV-SARX) model is a class of hybrid systems, where input measurements are also corrupted with noise. Moreover, time-delayed models have drawn attention, since several real-world applications are subject to delays. The parameterization of these submodels are presented in Table 3.1.

Table 3.1: Switching input-output linear models.

Name	$G(q^{-1}, \theta_i)$	$H(q^{-1}, \theta_i)$	Reference
SFIR	$B(q^{-1}, \theta_i)$	1	Liu <i>et al.</i> (2021)
SOE	$\frac{B(q^{-1}, \theta_i)}{A(q^{-1}, \theta_i)}$	1	Goudjil <i>et al.</i> (2017b)
SARX	$\frac{B(q^{-1}, \theta_i)}{A(q^{-1}, \theta_i)}$	$\frac{1}{A(q^{-1}, \theta_i)}$	Du <i>et al.</i> (2018)
Delay-SARX	$\frac{B(q^{-1-i}, \theta_i)}{A(q^{-1}, \theta_i)}$	$\frac{1}{A(q^{-1}, \theta_i)}$	Chen <i>et al.</i> (2017)
EIV-SARX	$\frac{B(q^{-1}, \theta_i)}{A(q^{-1}, \theta_i)}$	$\frac{1}{A(q^{-1}, \theta_i)}$	Ozbay <i>et al.</i> (2019)
SARMAX	$\frac{B(q^{-1}, \theta_i)}{A(q^{-1}, \theta_i)}$	$\frac{C(q^{-1}, \theta_i)}{A(q^{-1}, \theta_i)}$	Hojjatinia <i>et al.</i> (2020)
SBJ	$\frac{B(q^{-1}, \theta_i)}{A(q^{-1}, \theta_i)}$	$\frac{C(q^{-1}, \theta_i)}{D(q^{-1}, \theta_i)}$	Piga <i>et al.</i> (2020b)

Besides the switching linear I/O systems, identification of switched nonlinear I/O systems has also been addressed in the literature. The switched nonlinear ARX (SNARX) model is a type of switched I/O nonlinear system that is represented by a finite set of nonlinear maps of ARX model. Considering (3.1), the nonlinear map, i.e. f_{q_k} , can be expressed as either polynomial expansion of all monomials of x_k up to a given order or any other (nonlinear) basis function as follows:

$$f_{q_k}(x_k) = \sum_{j=1}^n \vartheta_{ij} \varphi_j(x_k), \quad (3.12)$$

where $\varphi_j(x_k)$, $j = 1, \dots, n$, is a nonlinear regressor, and $\vartheta_i \in \mathbb{R}^n$ is the parameter vector of the i -th submodel. The vector field, f_{q_k} , can also be expressed by a Takagi-Sugeno (TS) model or a Neural Network (NN). TS models and NNs, however, are not nonlinear in principle and the nonlinearities come from membership and activation functions in TS models and NNs, respectively. Considering weighted Gaussian membership function and element-wise sigmoid with hyperbolic tangent activation functions, TS-SARX and NN-SARX are nonlinear models. The block-oriented model is represented by Wiener and Hammerstein structure, that consists of a Hammerstein block and a Wiener block in series with a linear block in between. The switched nonlinear system in this form is formulated based on SARX as a middle linear block. These input-output nonlinear switching models are summarized in Table 3.2.

Table 3.2: Switching input-output nonlinear models.

Name	Basis	Nonlinearity	Reference
SNARX	SARX	basis function	Bianchi <i>et al.</i> (2020b)
TS-SARX	SARX	membership function	Wagner and Kroll (2014)
NN-SARX	SARX	activation function	Brusafferri <i>et al.</i> (2020)
WH-SARX	SARX	Wiener, Hammerstein	Wang <i>et al.</i> (2019) Zhang <i>et al.</i> (2016)

3.2.2. STATE-SPACE MODELS

Switched State-Space (SS) models provide a more meaningful representation for physical applications in comparison with switched I/O models. Moreover, most control approaches and dynamical analysis rely on SS models as simple and compact form for theoretical developments. A general form of hybrid model in SS structure is described by

$$\begin{cases} x_{k+1} &= F_{q_k}(x_k, u_k) + w_k \\ y_k &= G_{q_k}(x_k, u_k) + v_k \end{cases} \quad (3.13)$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^p$, and $y_k \in \mathbb{R}^l$ are the continuous state, input and output of the system, respectively, $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^l$ are noise terms, $q_k \in \{1, \dots, N\}$ is the switching signal that determines which vector fields, F_{q_k} and G_{q_k} , are active at time step k . The vector fields, $F_{q_k} : \mathbb{R}^{n+p} \rightarrow \mathbb{R}^n$ and $G_{q_k} : \mathbb{R}^{n+p} \rightarrow \mathbb{R}^l$ can be either linear (and affine) or nonlinear.

Linearization of nonlinear form of (3.13) around operating points yields affine models. "Affine" systems can be represented in "linear" forms (i.e. without affine or bias constant), if the equilibrium points are known. The more accurate the approximation of a complex system, the higher the number of submodels required, which hinders the identification problem due to the increasing number of modes. Switched nonlinear systems in SS form can also be rewritten as a linear combination of the basis functions. SS forms represent a more informative model for a system in comparison with input-output models. Table 3.3 provides an overview of state-space models.

Table 3.3: Switching state-space models.

Name	type of F_i	type of G_i	Reference
Switched affine input-to-state	affine	-	Du <i>et al.</i> (2021b)
Switched linear state-space	linear	linear	Sefidmazgi <i>et al.</i> (2016)
Switched affine state-space	affine	linear	Rui <i>et al.</i> (2016)
Switched nonlinear input-to-state	basis function	-	Du <i>et al.</i> (2021a)

3.2.3. SWITCHING MECHANISMS

It is also worth to discuss switching mechanisms to review one of the important model group, i.e. piecewise affine models. Switching behavior is determined by the switching mechanism. The switching signal is defined as discrete state, i.e. $q_k \in \mathcal{S} = \{1, \dots, N\}$ that determines which submodel is active at time step k . There are a variety of switching mechanisms: exogenous, deterministic, state-driven, event-driven, time-driven, and completely random. In terms of hybrid system identification, according to what has been discussed in the literature, the switching mechanism can be considered as either continuous-state-dependent or -independent. The behaviour of an application (i.e. the source of switching) determines the mechanism of the switching signal.

The PieceWise Affine (PWA) model is a class of hybrid systems where in the discrete state depends on the continuous state. The simple and flexible structure as well as the universal properties of PWA model for approximation of any nonlinear functions with any accuracy has drawn researchers' attention to develop identification methods based on this model. In principle, the switching space in PWA model is partitioned into some regions based on the continuous states. According to (3.2), the mode of q_k for the i -th active mode is written based on the regressor vector, x_k , belonging to i -th set, χ_i , of a polyhedral partition $\{\chi_i\}_{i \in \mathcal{S}}$ of the regressor space χ as

$$q_k = i \iff x_k \in \chi_i. \quad (3.14)$$

In other words, a PWA function, $f: \chi \rightarrow \mathbb{R}^{n_b}$, approximates a nonlinear function with sufficient number of modes defined by a set of polyhedra, χ_i . The partition region χ_i , can be defined based on linear classifiers as follows, which is commonly-used formulation for the partition domain. The bounded polyhedron χ_i , can be represented by a hyperplane matrix $\mathcal{H}_i \in \mathbb{R}^{n_{p_i} \times (n_x+1)}$, in which n_{p_i} is the number of hyperplanes of the corresponding partition:

$$\chi_i = \{x \in \mathbb{R}^{n_x} : \mathcal{H}_i \begin{bmatrix} x \\ 1 \end{bmatrix} \leq 0\} \quad (3.15)$$

A few types of model have been used for hybrid systems with polyhedral partition as switching mechanism, which is given in Table 3.4. Voronoi-type partition (seeds generators), scheduling-variable space (Linear Parameter Varying (LPV)), time-partitioned region, , and input-to-state form of partition are the models of the switching mechanism. Therefore, different classes of hybrid systems such as PWFIR, PWARX (Breschi and Mejari 2020), PWAOE (Mejari *et al.* 2020b)), PWA LPV-ARX

(Mejari *et al.* 2018)), PWNL (Mazzoleni *et al.* 2021)), and state-space PWA (Rui *et al.* 2016)) models can be derived according to the switching pattern parameterization.

Besides the deterministic representation of PWA partitions, the random probability distribution can be represented over the discrete state model as a switching signal using the Dirichlet process. Moreover, for the state-independent domain, Markov switching is the other class of switching mechanism widely used. The orderly-switching pattern, in the form of a Markovian jump model, determines the switching between the submodels independent of the past (except the immediate one) information. The modes of a hybrid system can also be expressed as an event-based model, then the model is so-called event-driven. In this way, hybrid systems can mathematically model the physical “environment”. The summary of mode-switching models is given in Table 3.4.

Table 3.4: Mode-switching models.

Switching mechanism	Model	Reference
Polyhedral partition	Linear classifier	Regressor-vector dependent
	Center generator	Voroini-type
	Scheduling-variable space	Soft function
	Batch time-based	Time-partitioned
	State-space	State-input-vector dependent
Random	Dirichlet	Probabilistic
	Arbitrary	Non-modelled
	Markov	Probabilistic
Event-driven	Inferred mathematical	

3.2.4. PROBABILISTIC MODELS

Probabilistic models (also so-called non-parametric model) are models that take a probability distribution into account to describe the process. This kind of the model representation, that can be expressed in various settings, is suited for the systems with available priori physical knowledge. Generally, the distribution of the output in the discrete-time form can be expressed as follows:

$$y_k | (x_k, \theta_{I_k}, \sigma_{I_k}^2, v_{I_k} : I_k = i) \sim \mathcal{D}(\theta_i^T x_k, \sigma_i^2, v_i) \quad (3.16)$$

where v_i and σ_i^2 denote degree of freedom and the variance of the distribution, respectively. Moreover, I_k is the Markov chain and \mathcal{D} can be any form of distribution like Gaussian distribution and t -distribution to represent the distribution of the parameters (θ_{I_k}) based on given regressor vector (x_k). Other parameterization are derived to include some features such as delays and missing data. Various classes of hybrid systems that can be parameterized in this way range from SFIR and SARX to PWARX.

Moreover, the stable spline kernel is a way of stochastic modeling of hybrid systems via a Bayesian network. The stable spline kernel modeling can be used for PWFIR and PWARX models. The distribution of the model in the form of linear

stable spline kernel can be expressed as follows:

$$y_k | (\{\theta_i\}_{i=1}^M, \lambda, \alpha, \sigma^2, \{\omega_k\}_{k=1}^N) \sim \mathcal{N}(\Phi_k \theta_k, \frac{1}{\sigma^{-2}} I) \quad (3.17)$$

where λ and α are the scalar factor and the stability parameter, respectively as the stable spline hyperparameters, ω_k denotes other hyperparameters associated with the classification, Φ_k is the matrix with rows selected by the input regressor vector with respect to the type of the input regressor chosen based on the type of the model, and σ^{-2} and I are the noise precision and the identity matrix with the appropriate dimension, respectively.

Furthermore, the submodels also can be formulated in the stochastic setting with the deterministic derivations. In this way, the a priori prediction error, $\epsilon_{k|k-1}$ and the a posteriori prediction error, $\epsilon_{k|k}$ can be expressed as

$$\epsilon_{k|k-1} = Y_k - \hat{Y}_{k|k-1} \quad (3.18a)$$

$$\epsilon_{k|k} = Y_k - \hat{Y}_{k|k} \quad (3.18b)$$

where $Y_k \in \mathbb{R}^{N \times 1}$ is the collection of outputs for all submodels at time step k , $\hat{Y}_{k|k-1}$ and $\hat{Y}_{k|k}$ are the a priori and the a posteriori estimations, respectively. The partitions can also be represented deterministically. The representations of different model structure in the literature based on the above discussion have been summarized in Table 3.5.

Table 3.5: Representation of hybrid systems in probabilistic settings.

Model structure	Representation feature	Reference(s)
SFIR	Missing measurement included	Liu <i>et al.</i> (2021)
	Hammerstein nonlinearity included	Ma <i>et al.</i> (2019)
	Delay included	Chen <i>et al.</i> (2020b)
SARX	Based on t -distribution	Fan <i>et al.</i> (2017)
	Missing measurement included	Chen <i>et al.</i> (2020a)
	Error-based posteriori prediction	Goudjil <i>et al.</i> (2016)
SOE	Error-based posteriori prediction	Goudjil <i>et al.</i> (2017a)
PWFIR	Based on kernel hyperparameters	Pillonetto (2016)
	Hierarchical Bayesian	Wågberg <i>et al.</i> (2015)
PWARX	Based on kernel hyperparameters	Scampicchio and Pillonetto (2018)
	Error-based posteriori prediction	Yahya <i>et al.</i> (2020)
PWA	Bayesian inference	Piga <i>et al.</i> (2020a)
SBJ	Posterior distribution	Breschi <i>et al.</i> (2019)

3.2.5. APPLICATIONS OF MODELS

The advantage of the more complex models in terms of structure complexity like SBJ and EIV-SARX models is the accuracy in prediction due to consideration of either dynamical disturbances and noises or errors in the parameters.

Moreover, an advantage of switching patterns represented by polyhedron is regressor-dependent switching that can project physical behaviors of the process

such as different metabolic stages (Wang *et al.* 2020)), different operating modes (Song *et al.* 2020)), and different phases of a batch process (Xu *et al.* 2018)). Various forms of polyhedral partition may not have a specific advantage unless a specific problem for classification is defined, e.g. Bako and Yahya (2019) in which the centers of the partition map are intuitively interpreted as operating points. Arbitrary patterns can also be used for a system that has no information on switching instants (Zhang *et al.* 2018)).

In addition, for control-orientated purposes, state-space representation of the submodels and the partition region (Du *et al.* 2021b)) has the advantage over others because of its structure for observer and controller design.

For applications with no prior knowledge on their structures, non-parametric models have an advantage because of probability distribution over parameters and noises to probabilistic model uncertainties (Fan *et al.* 2017)). The range of complexity in probabilistic structures is the same as input-output models (Breschi *et al.* 2019; Liu *et al.* 2021)). Different advantages of model representation in probabilistic settings are based on including system restrictions in practice, such as missing measurement (Chen *et al.* 2020a)) and time delay (Chen *et al.* 2020b)). Probabilistic models in stable spline kernel also have advantage as they mitigate the difficulty of the model order selection based on defining hyperparameters (Scampicchio and Pillonetto 2018)).

In case of existing time-varying relation between input and output measurements, scheduling variable can be introduced to extend linear time-invariant models to linear time-varying models (Mejari *et al.* 2020c)).

Nonlinear models have the ability of representing a nonlinear system with fewer number of parameters in a wider range. Polynomial nonlinear models are discussed by Bianchi *et al.* (2021) and non-parametric piecewise nonlinear models are addressed by Mazzoleni *et al.* (2021). These kinds of hybrid systems are a new trend in the field of system identification.

3.3. METHODS OF HYBRID SYSTEM IDENTIFICATION

The method used to solve the identification problem depends on the parameterization used to model the hybrid system. Subsystem identification and switching rule detection can be performed either separately, or jointly. Identification of hybrid systems, therefore, needs classification of dataset into some clusters, and estimation of the parameters of submodels for each cluster. The systematic classification of the methods has been reviewed below.

3.3.1. OPTIMIZATION-BASED METHODS

The generic identification problem can be formulated in an optimization framework as follows:

$$\underset{\mathcal{X}_{i,k}, \theta_i}{\text{minimize}} \sum_k \sum_i \mathcal{X}_{i,k} (y_k - x_k^T \theta_i)^2, \quad (3.19a)$$

$$\text{s.t.} \sum_i \mathcal{X}_{i,k} = 1, \forall k \quad (3.19b)$$

where y_k is the actual system output, x_k and θ_i are the regressor and parameter vectors, respectively, which can be constructed based on the model parameterization discussed in Section 3.2, and $\mathcal{X}_{i,k} \in \{0,1\}$ denotes the submodel activation for all time steps. The goal of the optimization problem is to minimize the error of the identified output and actual output based on the available measurements i.e. the measured outputs $\{y_k\}_{k=1}^n$, and the measured inputs, $\{u_k\}_{k=1}^n$, where n is the number of measurement. It is, however, an NP-hard optimization problem, a reformulation can be made for different classes of model parameterization to make the problem computationally feasible.

Wang *et al.* (2019) have solved the optimization problem based on the least squares criterion for switched Hammerstein ARX models with a long-horizon and a different horizon iteration to detect the process rapid changes. Hu *et al.* (2015) have proposed a reformulation of the cost function in order to identify the subsystem parameters of a SARX model based on the least geometric mean squares algorithm, which is followed by a neural network classifier to label the training data.

A sum-of-norm regularized convex optimization problem for SARX models has been discussed by Hartmann *et al.* (2015), which is combined with Expectation-Maximization (EM) approach to cluster preliminary estimates and formulate a quadratic program to complete switching detection and parameter identification procedure. The application of EM has been also used by Tang and Dong (2020) as a first step of the two-step approach to solve a convex optimization problem for simultaneous clustering and identification. The proposed approach has been compared with a non-convex optimization algorithm proposed by Lauer (2013) and a recently-developed general optimization-based approach addressed by Yuan *et al.* (2019). Moreover, Xiujun *et al.* (2020) have developed a weighted multi-innovation the least squares algorithm for Hammerstein SARX models, which is based on EM approach for clustering.

An optimization problem has been formulated for a class of nonlinear switched ARX (SNARX) models and solved in an iterative way by Bianchi *et al.* (2020b). While the structure of the nonlinear model is characterized in a probability setting, a randomized method is employed to address the formulated combinatorial optimization. Since prior sample-mode assignment is required, Bianchi *et al.* (2021) have addressed this problem based on a two-stage randomized approach in a general framework with no a priori limit on the number of switching time instants by using a cost function that alternates between parameter and mode identification. The other heuristic approach for solving a typical heterogeneous optimization problem efficiently is similar to the discussed two-stage iterative approaches for both SARX and PWARX models proposed by Bianchi *et al.* (2020a).

Combining the prediction error method with a coordinate descent approach has been taken into account to address the identification problem of SARMAX by Breschi *et al.* (2018) in both batch and recursive ways. A close-to-optimal four-step solution has been also proposed by Amaldi *et al.* (2016) for the mixed-integer linear programming to fit a k -piecewise affine model with a piecewise linear separability problem. Domain partitioning based on multi-category linear classification and submodel fitting have been addressed simultaneously to guarantee solutions of the

k -hyperplane clustering problem.

Paoletti *et al.* (2019) have formulated the identification problem of PWA models in the framework of bi-level programming, in which data classification and partition estimation are addressed in the upper level and subsystem parameters are identified in the lower level based on a prediction error criterion. An optimization-based method has been formulated by Breschi and Mejari (2020) for structure selection and identification of PWARX models, using regularization-based shrinking strategies within a coordinate-descent identification method to determine the parameters of the submodels along with their structures.

If the partition of the PWA model is considered as a Voronoi type, the least harmonic mean approach can be employed that has been discussed by Bako and Yahya (2019). Moreover, identification of time-partitioned PWA-OE models has been tackled by Xu *et al.* (2018) for batch processes in the framework of optimization-based algorithms. Similarly, identification of PWA-OE models has been also discussed by Mejari *et al.* (2020a), that a recursive bias-correction scheme to correct the bias in the ordinary least square method has been presented. While simultaneous clustering and parameter estimation are achieved within the first stage by applying bias-corrected the least squares, partitioning the regressor space is obtained via a convex optimization problem known as multi-category discrimination.

Recursive multiple least squares for simultaneous clustering and parameter estimation has been proposed by Breschi *et al.* (2016b) for PWA models. A linear multi-category discrimination algorithm has been considered via a Newton-like approach and an averaged stochastic gradient descent for solving the unconstrained optimization problem for batch and recursive ways. The proposed method has been extended for LPV-ARX models with linear partitioning by Breschi *et al.* (2016a). The convex optimization problem has been solved using a sparse estimation approach as a likelihood-based methodology in stochastic framework by Mattsson *et al.* (2016).

Another two-stage optimization-based method has been discussed with an iterative regularized moving-horizon approach by Naik *et al.* (2017) for PWARX models, and Mejari *et al.* (2020c) for switched LPV-ARX models. Active modes and the parameters of the submodels are optimally and recursively found by solving small-size mixed-integer quadratic-programming problems and polyhedral partitions are identified using linear multi-category discrimination.

A quite general jump model, which is PWA models with hidden Markov jumps, has been formulated in an optimization-based framework by Bemporad *et al.* (2018) and solved by alternating between minimizing a loss function of fitting submodel parameters with a generalized k -means algorithm and minimizing a discrete objective function for determining active modes.

3.3.2. CLUSTERING-BASED METHODS

Clustering aims to divide a dataset into different subsets based on how similar they are to one another. This idea is very close to the hybrid system identification problem. Papers of this section have contributions to clustering, even they are combined with other methods.

A hierarchical clustering method based on the gap metric has been proposed by

Table 3.6: Optimization-based approaches.

Method	Switching mechanism	Hybrid model	Reference(s)
Different horizons least squares	Arbitrary (slow switching)	Hamm-SARX	Wang <i>et al.</i> (2019)
Combinatorial optimization	Arbitrary	SNARX	Bianchi <i>et al.</i> (2020b) Bianchi <i>et al.</i> (2021)
Constrained optimization	Arbitrary	SARX/PWARX	Bianchi <i>et al.</i> (2020a)
Prediction error method	Arbitrary/Markov	SARMAX	Breschi <i>et al.</i> (2018)
Discrete optimization	Linear classifier	PWA	Amaldi <i>et al.</i> (2016)
Nested optimization	Linear classifier	PWA	Paoletti <i>et al.</i> (2019)
Regularization-based optimization	Linear classifier	PWARX	Breschi and Mejari (2020)
Least harmonic mean approach	Center generator (Voronoi)	PWA	Bako and Yahya (2019)
Separable nonlinear least-squares	Time-based	PWA-OE	Xu <i>et al.</i> (2018)
Bias-correction approach	Linear classifier	PWA-OE	Mejari <i>et al.</i> (2020a)
Multiple least squares	Linear classifier Scheduling-variable space	PWA LPV-ARX	Breschi <i>et al.</i> (2016b) Breschi <i>et al.</i> (2016a)
Sparse estimation approach	Linear classifier	PWARX	Mattsson <i>et al.</i> (2016)
Regularized moving-horizon approach	Linear classifier Scheduling-variable space	PWARX LPV-ARX	Naik <i>et al.</i> (2017) Mejari <i>et al.</i> (2020c)
Fitting algorithm	Regression models (PWA) Statistical models (Markov jump)		Bemporad <i>et al.</i> (2018)
Sum-of-norm regularized optimization	Arbitrary	SARX	Hartmann <i>et al.</i> (2015)
EM-based sparse method	Linear classifier	PWARX	Tang and Dong (2020)
Weighted multi-innovation least squares	Arbitrary	Hamm-SARX	Xiujun <i>et al.</i> (2020)

Wang *et al.* (2020). Similarly, simultaneous submodel and optimal operation region partition estimation have been addressed based on output-error minimization in order to improve accuracy by Song *et al.* (2020). In this approach, local models are initially found with the least squares and clustering of local models and parameters identification are done based on the initialization using the feature vectors and weighted least squares, respectively.

Bounded-switching clustering has been discussed by Sefidmazgi *et al.* (2015) for SARX systems and Sefidmazgi *et al.* (2016) for switched state-space systems to convert the non-convex optimization problem into a binary integer programming problem using an innovative clustering method. However, the problem still includes optimization, by using bounded-switching technique, it has been easily formulated and solved by least squares and subspace identification by Sefidmazgi *et al.* (2015) and Sefidmazgi *et al.* (2016), respectively.

For PWA models, a semi-supervised clustering approach has been proposed by Du *et al.* (2020) to obtain the number of submodels, the initial clustered dataset, and the corresponding parameters of each submodels. The output of the clustering stage is used for a modified self-training Support Vector Machine (SVM) algorithm to identify the polyhedral partitions and the parameters of submodels. Moreover, a self-adaptive clustering algorithm has been addressed by Sellami *et al.* (2016). The sequential estimation procedure of the switching signal is based on an unsupervised self-adaptive classification algorithm. The core of the proposed approach is clustering based on three steps consisting of cluster creation, online cluster adaptation, and cluster fusion. Hure and Vasak (2017) developed a clustering-based identification algorithm for PWA models based on feature vectors and clustering them, in which the k-mean++ algorithm is adapted for initialization. Feature vector transformation is introduced to reduce and in some cases omit partitioning in some dimensions.

Li *et al.* (2016) have proposed a subspace clustering approach that removes

the requirement on convex regions in the conventional k-mean clustering. A block-diagonal matrix permutation algorithm is the proposed subspace algorithm to reduce the computational complexity in handling arbitrarily shaped regions. Another subspace clustering approach has been proposed by Li and Liu (2017), who employ a spectral clustering algorithm with a relaxed-permutation structure. The spectral clustering method has also been addressed by Zhang *et al.* (2018) for EIV-SARX models. Based on the proposed method, data points are partitioned into subsets and a manifold distance between the dynamics of each segment is computed via a Riemannian distance-like function to assign segments to clusters, and finally a common identification method is used to identify parameters of each cluster. Another subspace clustering algorithm for identification of EIV-SARX models has been presented by Ozbay *et al.* (2019), using sum-of-squares polynomial with Christoffel's functions to perform singular value decomposition independently of the number of data points. A subspace clustering algorithm for state-space switched systems has been proposed by Lopes *et al.* (2016). The hybrid Kalman filter as an interacting multiple model algorithm is used for reclassification to assign the original dataset to a specific mode and to refine model estimation at the end of the procedure.

A prototype-shaped clustering-based algorithm has been proposed by Wagner and Kroll (2014) for partitioning nonlinear Takagi-Sugeno systems. The identification process includes fuzzy c-means and Gustafson-Kessel algorithms for clustering and identification. Another application of fuzzy c-means clustering as an efficient unsupervised partitioned technique has been discussed by Shah and Adhyaru (2014) for PWARX models. The number of submodels is estimated by the proposed fuzzy clustering approach and submodel parameters are identified by weighted least squares approach based on a fuzzy distance weight matrix. Likewise, an incremental c-regression approach has been addressed by Blazic and Skrjanc (2020) as an online identification procedure. Furthermore, while PCA-guided k-Means clustering approach is a conventional clustering approach in which clusters are derived in a PCA-guided process, a fuzzy PCA-guided clustering technique has been proposed by Khanmirza *et al.* (2016) as a modified robust clustering method.

Classification and clustering with evolutionary algorithms for estimating switching patterns modeled by a Gaussian mixture before parameter identification with weighted and extended least squares algorithms is one of the innovative methods to address the identification problem of PWARX and PWARMAX models proposed by Barbosa *et al.* (2019). Classical first-order algorithms such as mirror descent algorithm or Nesterov's optimal scheme can be employed to solve the reformulated determination problem of the regions as a multi-class classification, which is discussed by Jianwang and Ramirez-Mendoza (2020). In this work, parameter estimation has been addressed via zonotope parameter identification.

A constrained clustering approach for time-partitioned PWARX model has been developed by Liu *et al.* (2022). The clustering optimization problem has been formulated by imposing the complete and non-overlapping partition constraints and it has been efficiently solved by employing a greedy iterative approach.

For PWNL models, a semi-supervised clustering setting has been proposed by

Mazzoleni *et al.* (2021), which is based on a data augmentation strategy to deal with a situation when unsupervised data are not basically provided. This work is a pioneer of piecewise nonlinear regression in the domain of hybrid system identification.

3.3.3. LIKELIHOOD-BASED METHODS

Likelihood-based methods are formulated based on the models represented in the probabilistic form. Expectation-maximization is one of the well-and-widely-studied algorithms, which not only can be used for clustering (as reviewed before), but also for the maximum-likelihood estimation. The EM algorithm consists of two steps: E-step and M-step. Considering the unknown parameter vector, Θ , defined based on the model structure, and observed and unobserved dataset, C_{obs} and C_{uno} , the E-step calculates the conditional expectation of the log-likelihood function (known as Q -function) formulated as follows:

$$Q(\Theta | \Theta_{\text{old}}) = E_{C_{\text{uno}} | (C_{\text{obs}}, \Theta_{\text{old}})} (\log P(C_{\text{uno}}, C_{\text{obs}} | \Theta_{\text{old}})) \quad (3.20)$$

where Θ_{old} is the parameter set calculated in the previous iteration. Besides the E-step, the M-step maximizes the Q -function with respect to parameter set written as follows:

$$\Theta = \underset{\Theta}{\text{argmax}} Q(\Theta | \Theta_{\text{old}}) \quad (3.21)$$

Rui *et al.* (2016) have proposed a framework based on the EM algorithm to identify the parameters of the model represented in PWA state-space form. A cumulative distribution function is employed to compute the probability of each submodel based on the measured samples at that time step, and then the latent discrete state is estimated using a Kalman smoother for the computed submodel, and finally parameters are identified based on the maximization of a surrogate function for the likelihood. Fan *et al.* (2017) have addressed a robust identification problem of the model parameterized by a hidden Markov ARX model using EM algorithm in which student's t -distribution is imposed to the noise model for more accurate estimation. The extension from a batch to a recursive EM algorithm has been proposed by Chen *et al.* (2020b) for delayed SARX models to identify the parameters of the submodels, the Markov chain transition, and the time delays simultaneously.

Variational Bayesian (VB) method is another Bayesian optimization-based strategy that is used to approximate high dimensional posterior distributions instead of point-wise estimations of the parameters. The VB scheme is a more general approach in comparison with the EM approach because of approximation of parameter densities.

The identification problem of Markov-switching Hammerstein ARX models has been addressed via the VB approach by Ma *et al.* (2019). Estimating the unknown number of submodels and switching signals as well as approximating the distributions of the unknown submodels parameters have been tackled for SFIR models by Liu *et al.* (2021). Estimation of parameters distributions and point-estimation of the transition probabilities of switched Markov ARX models and construction of missing measurements have been discussed by Chen *et al.* (2020a) under the VB framework. Similarly, a robust VB approach for SARX models with a

Table 3.7: Clustering-based approaches.

Method	Switching mechanism	Hybrid model	Reference(s)
Hierarchical clustering	Scheduling-variable space	PWARX	Song <i>et al.</i> (2020) Wang <i>et al.</i> (2020)
Bounded-switching clustering	Arbitrary	SARX Switched SS	Sefidmazgi <i>et al.</i> (2015) Sefidmazgi <i>et al.</i> (2016)
Semi-supervised clustering	Linear classifier	PWARX	Du <i>et al.</i> (2020)
Unsupervised clustering	Arbitrary	SARX	Sellami <i>et al.</i> (2016)
Enhanced k-means++	Center generator	PWARX	Hure and Vasak (2017)
Subspace clustering	Arbitrary	bi-model PWL	Li <i>et al.</i> (2016)
		Switched affine	Li and Liu (2017)
		EIV-SARX	Zhang <i>et al.</i> (2018) Ozbay <i>et al.</i> (2019)
		Switched SS	Lopes <i>et al.</i> (2016)
Fuzzy clustering	Linear classifier	PWA Takagi-Sugeno	Wagner and Kroll (2014)
	Linear classifier	PWARX	Shah and Adhyaru (2014)
Fuzzy PCA-guided clustering	Linear classifier	PWARX	Khanmirza <i>et al.</i> (2016)
Genetic-based clustering	Linear classifier	PWARX/PWARMAX	Barbosa <i>et al.</i> (2019)
Multi-class classification	Arbitrary	PWARX	Jianwang and Ramirez-Mendoza (2020)
Constrained K-means clustering	Batch time-based	PWARX	Liu <i>et al.</i> (2022)
Greedy semi-supervised	Linear classifier	PWNL	Mazzoleni <i>et al.</i> (2021)

combination of an adjusted-tail t -distribution to deal with contaminated data with outliers has been proposed by Lu *et al.* (2016). Furthermore, the application of a VB approach for Markov SARX models with slowly varying time delay has been extended to identify the parameter distributions of submodels besides the transition probability matrix and unknown delays by Chen and Liu (2019).

Kernel-based stable spline is another algorithm for non-parametric models in Bayesian framework. The hybrid systems in the form of stable spline kernel can be identified by optimizing marginal likelihood via a stochastic simulation scheme. Pillonetto (2016) has proposed this approach for identification of hybrid systems. The two-step kernel-based stable spline procedure consists of data classification and distribution in the marginal likelihood optimization by exploiting the Bayesian interpretation of regularization, and reconstruction of subsystems. While the performance of the proposed method has been assessed through a Markov chain Monte Carlo approach by Pillonetto (2016), the Gibbs sampling scheme has been employed by Scampicchio and Pillonetto (2018). Scampicchio *et al.* (2018) have also extended it for nonlinear hybrid systems, which is capable of automatic discrimination among linear and nonlinear submodels.

In addition, non-parametric representation of hierarchical PWARX models in the Bayesian framework with respect to Dirichlet clustering properties provides probabilistic predictions with confidence intervals, which has been addressed by Wågberg *et al.* (2015) within a Gibbs sampling process. Furthermore, two Rao-Blackwillized sampling algorithms in batch and recursive manners for PWA models represented in a Bayesian setting have been addressed by Piga *et al.* (2020a). The parameters of regressor-space partition formulated based on marginal posterior are approximated via Markov chain Monte Carlo sampling for offline learning and particle filters for online learning in batch and recursive ways. The identification problem of SBJ models has been tackled by Breschi *et al.* (2019) and Piga *et al.* (2020b) using a maximum-a-posterior estimation approach. Embedding the prediction error algorithm in the likelihood framework tailored by stochastic Markov chains within

a coordinate ascent method enables the identification procedure to iteratively and computationally effective (due to a suboptimal moving-horizon approach) alternate between local parameter identification and mode sequence reconstruction.

Table 3.8: Likelihood-based approaches.

Method	Switching mechanism	Hybrid model	Reference(s)
Expectation-Maximization	Linear classifier	PWA-SS	Rui <i>et al.</i> (2016)
	Markov	SARX	Fan <i>et al.</i> (2017)
	Markov	Delay-SARX	Chen <i>et al.</i> (2020b)
Variational Bayesian	Markov	Hamm-SARX	Ma <i>et al.</i> (2019)
	Arbitrary	SFIR	Liu <i>et al.</i> (2021)
	Markov	SARX	Chen <i>et al.</i> (2020a)
	Markov	Delay-SARX	Chen and Liu (2019)
Kernel-based stable spline	Linear classifier	PWFIR/PWARX	Pillonetto (2016)
	Linear classifier	PWFIR/PWARX	Scampicchio and Pillonetto (2018)
	Arbitrary	Nonlinear dynamics	Scampicchio <i>et al.</i> (2018)
Bayesian	Dirichlet	PWARX	Wågberg <i>et al.</i> (2015)
Maximum-a-posteriori	Linear classifier	PWA	Piga <i>et al.</i> (2020a)
	Markov	SBJ	Piga <i>et al.</i> (2020b) Breschi <i>et al.</i> (2019)

3.3.4. OTHER METHODS

Algebraic methods: Recent trends show a combination of algebraic methods with clustering (e.g. subspace clustering) and optimization-based algorithms. The set membership identification problem has been addressed for SARX models with prior information on the number of submodels by Ozay *et al.* (2015), using an algebraic procedure and combining it with a polynomial function of the unknown noise to recast the problem into constrained rank minimization form, which is a convex optimization problem. Likewise, matrix rank minimization along with an iterative partial matrix shrinkage algorithm has been presented by Konishi (2015) for identification of SRAX models. Hojjatinia *et al.* (2020) have proposed a similar approach but for cases with a very large number of samples affected by large levels of noise for both SARX and SARMAX models. Similarly, a non-convex optimization problem has been computationally efficiently solved using an algebraic procedure and a polynomial optimization approach with sparse reformulation of the problem to jointly identify a kernel-based mapping and the corresponding continuous-state evolution of Wiener SARX models by Zhang *et al.* (2016).

Furthermore, an iterative algebraic geometric approach has been proposed by Nazari *et al.* (2016), which is built upon stochastic hybrid decoupling polynomial construction and it is shown that the problem of the linear regression can be transferred into stochastic hybrid decoupling polynomial. An algebraic procedure by constructing Hankel-like matrices and performing singular value decomposition of the Hankel matrices results in parameter estimates, which has been discussed by Sarkar *et al.* (2019) for switched state-space models.

Outer Bounding Ellipsoid (OBE) methods: OBE type algorithms are set membership real-time identification algorithms under assumption of unknown-but-bounded noises or disturbances. The OBE algorithm has been presented for SARX models by Goudjil *et al.* (2016) PWARX models by Yahya *et al.* (2020), and SOE models

by Goudjil *et al.* (2017b). A two-step algorithm that proposed by Du *et al.* (2018) for SRAX models is another method similar to OBE. . Data assignment is carried out based on incorporating both the residual error and an upper bound of the subsystem estimation error, which is followed by a randomized algorithm to update simultaneously the parameters of the submodels.

Continuous-time identification methods: The literature reviewed above is for hybrid systems represented in discretized form, while there can be found some literature on continuous-time hybrid system identification methods. The concurrent learning technique, which has been developed in a recursive manner for PWA state space models by Kersting and Buss (2017), has been proposed to identify online continuous-time system dynamics, using the recorded and current data concurrently for adaption.

Due to necessity of state derivatives for the concurrent learning technique, extended integral concurrent learning identifier has been presented by Du *et al.* (2021a). The two-stage online identification of switched state space models consists of recognition of the active modes based on the projection matrix inspired by a recursive projection subspace method followed by an integral concurrent learning technique for identification of the system dynamics. Likewise, using an integral concurrent learning method for continuous-time PWA state space models has been addressed by Du *et al.* (2021b). Polyhedral regions are estimated by solving an optimization problem based on the parameter identification and mode recognition steps. A continuous-time identification method has been proposed by Goudjil *et al.* (2020) in which consistent submodel outputs are constructed based on a sum of sinusoids as an appropriate input signal and then parameter vector estimation is carried out by conventional continuous-time identification methods under assumption of a given number of submodels. Furthermore, Keshvari-Khor *et al.* (2018) have proposed an identification method for continuous-time switched state space models. The advantage of this approach is detection of switching time between two sampling instants even with low-rate sampled data.

Neural network (NN) methods: Due to the capability of NNs to represent nonlinear systems in a simpler structure with precision in approximation, Yang *et al.* (2017) have proposed a way to use and train NNs in modeling of nonlinear ARMA models in the form of hybrid systems, which is called multiple NARMA-L2 model. In addition, Brusafferri *et al.* (2020) have proposed Mixture of Expert (MoE) NN architecture that can represent the feature of hybrid systems in a NN structure for one-step-ahead identification of SARX models. MoE layer along with a gated recurrent units with softmax output plays a role as neural switching machine, while feed-forward NNs have been chosen as a structure to represent continuous dynamics. Moreover, a method to decompose an NN into a PWA model by using weight pruning to reduce the number of linear classifier has been discussed by Robinson (2021).

3.3.5. RESEARCH TRENDS IN METHODS

Recently, heuristic approaches combine different methods. For instance, Mejari *et al.* (2020b) have discussed a recursive manner of the least squares method with bias correction to deal with unknown noises based on the clustering method. However,

this online method estimates the submodel parameters and the switching signal as well as unknown noise variance with high accuracy, but clustering leads to misclassification that should be dealt with. Wang *et al.* (2020) have also addressed weighted least squares method, but with improvement on accuracy of clustering by introducing gap metric to find similar measurement and minimize the number of the submodels. Moreover, a combination of cluster-based algorithm and self-training SVM algorithm has been proposed by Du *et al.* (2020). In this method, the clustering outputs initialize the modified SVM algorithm that reduces the computational complexity and increases the precision of partitioning. Furthermore, using algebraic procedure as a starting point and combing it with a polynomial optimization is another example of a heuristic combined method to address high dimensional dataset affected by large level of noise by Hojjatinia *et al.* (2020).

Generalization is the other important trend. Using a maximum-a-posteriori algorithm by Piga *et al.* (2020b) in a general way in terms of fitting SBJ models with time varying coefficients and equivalent state space models is a pioneer, and relaxation of user-dependent regularization of hyperparameters and the derivation of confidence intervals are for further extension and generalization. Similarly, the proposed method by Piga *et al.* (2020a) has been derived for PWA models, and it can be extended for polynomial nonlinear models in this trend. The priority of these papers is using a maximum-a-posteriori method that also obtains the distribution of the model parameters and the predicted output.

Moreover, VB methods have drawn attention since they also provide the distribution of the parameters and solve the optimization problem effectively. With less generality but taking missing measurements and delays into account for two simple model classes, i.e. SFIR and SARX, Liu *et al.* (2021) and Chen and Liu (2019) have discussed the application of VB methods, which should be extended for other types of hybrid systems to complete this chain of trends.

Table 3.9: Other approaches.

Method category	Hybrid model	Algorithm	Reference(s)
Algebraic	SARX	Set membership Matrix rank minimization with partial matrix shrinkage	Ozay <i>et al.</i> (2015) Konishi (2015)
	SARX/SARMAX	Geometric approach with stochastic hybrid decoupling polynomial	Nazari <i>et al.</i> (2016)
	Wiener SARX	Veronese-map-embedded GPCA with polynomial optimization	Hojjatinia <i>et al.</i> (2020)
	Switched SS	Kernel-based mapping with polynomial optimization	Zhang <i>et al.</i> (2016)
		SVD based on Hankel-like matrices	Sarkar <i>et al.</i> (2019)
OBE	SARX	Real-time set membership algorithm	Goudjil <i>et al.</i> (2016)
	PWARX	based on unknown-but-bounded noises	Yahya <i>et al.</i> (2020)
	SOE		Goudjil <i>et al.</i> (2017b)
Continuous-time	PWA-SS	Concurrent learning	Kersting and Buss (2017)
	Switched SS	Integral concurrent learning	Du <i>et al.</i> (2021b)
	Continuous input-output	Integral concurrent learning	Du <i>et al.</i> (2021a)
	Switched SS	Conventional continuous-time identification Conversion of discrete-time to continuous-time	Goudjil <i>et al.</i> (2020) Keshvari-Khor <i>et al.</i> (2018)
NN	Multiple NARMA-L2	One-step-ahead	Yang <i>et al.</i> (2017)
	SARX	Mixture of expert with softmax output	Brusaferrri <i>et al.</i> (2020)
	PWA	Neural network decomposition	Robinson (2021)

3.4. CONCLUSIONS

As discussed, hybrid system identification is an active research field, and it can be used for a wide range of real-world applications for modeling and control. In this survey a systematic review on models and methods has been proposed to show the state-of-the-art in hybrid system identification. Current and future research directions include the following objectives to:

- increase and generalize model complexity in terms of parameterization and find a method to solve its identification problem in a computationally efficient way,
- extend the probabilistic model parameterization for other classes of hybrid systems and find a solving algorithm for it,
- parameterize the switching domain in a nonlinear form,
- take practical problems such as delay and missing measurement into account for other types of hybrid systems,
- solve the mis-classification problem with innovative clustering approaches, and
- explore nonlinear hybrid system identification to increase accuracy, while decrease the number of modes for nonlinear processes with a wide range of operation.

In summary, this survey has highlighted that the methodological focus has slightly changed from the use of innovative optimization and novel clustering approaches towards problem generalization, multi-method combination, and the use of novel probabilistic methods. Moreover, other important issues related to hybrid system identification such as experiment design and identifiability can be reviewed in future research.

REFERENCES

- Amaldi, E., S. Coniglio, and L. Taccari (2016). “Discrete optimization methods to fit piecewise affine models to data points”. In: *Computers and Operations Research* 75, pp. 214–230.
- Bako, L. and O. Yahya (2019). “Piecewise Affine System identification: A least harmonic mean approach”. In: *2019 IEEE 58th Conference on Decision and Control*.
- Barbosa, B. H., L. A. Aguirre, and A. P. Braga (2019). “Piecewise affine identification of a hydraulic pumping system using evolutionary computation”. In: *IET Control Theory and Applications* 13.9, pp. 1394–1403.
- Basiri, M. H., J. G. Thistle, and S. Fischmeister (2018). “A framework for inference and identification of hybrid-system models: Mixed event-/time-driven systems (METS)”. en. In: *IFAC-PapersOnLine* 51.15, pp. 287–292.
- Bemporad, A., V. Breschi, D. Piga, and S. P. Boyd (2018). “Fitting jump models”. In: *Automatica* 96, pp. 11–21.
- Bianchi, E., V. Breschi, D. Piga, and L. Piroddi (2021). “Model structure selection for switched NARX system identification: A randomized approach”. In: *Automatica* 125, p. 109415.
- Bianchi, E., A. Falsone, L. Piroddi, and M. Prandini (2020a). “An alternating optimization method for switched linear systems identification”. In: *IFAC-PapersOnLine* 53.2, pp. 1071–1076.
- Bianchi, E., M. Prandini, and L. Piroddi (2020b). “A randomized two-stage iterative method for switched nonlinear systems identification”. In: *Nonlinear Analysis: Hybrid Systems* 35, p. 100818.
- Blazic, S. and I. Skrjanc (2020). “Hybrid System Identification by Incremental Fuzzy C-regression Clustering”. In: *2020 IEEE International Conference on Fuzzy Systems*.
- Breschi, V., A. Bemporad, and D. Piga (2016a). “Identification of hybrid and linear parameter varying models via recursive piecewise affine regression and discrimination”. In: *2016 European Control Conference*.
- Breschi, V. and M. Mejeri (2020). “Shrinkage Strategies for Structure Selection and Identification of Piecewise Affine Models”. In: *2020 59th IEEE Conference on Decision and Control*.
- Breschi, V., D. Piga, and A. Bemporad (2016b). “Piecewise affine regression via recursive multiple least squares and multicategory discrimination”. In: *Automatica* 73, pp. 155–162.
- (2019). “Maximum-a-posteriori estimation of jump Box-Jenkins models”. In: *2019 IEEE 58th Conference on Decision and Control*.
- Breschi, V., A. Bemporad, D. Piga, and S. Boyd (2018). “Prediction error methods in learning jump ARMAX models”. In: *2018 IEEE Conference on Decision and Control*.
- Brusaferrri, A., M. Matteucci, P. Portolani, S. Spinelli, and A. Vitali (2020). “Hybrid system identification using a mixture of NARX experts with LASSO-based feature selection”. In: *2020 7th International Conference on Control, Decision and Information Technologies*.
- Chen, X. and F. Liu (2019). “A Variational Bayesian Approach for Identification of Time-Delay Markov Jump Autoregressive Exogenous Systems”. In: *Circuits, Systems, and Signal Processing* 39.3, pp. 1265–1289.
- Chen, X., S. Zhao, and F. Liu (2017). “Identification of time-delay Markov jump autoregressive exogenous systems with expectation-maximization algorithm”. In: *International Journal of Adaptive Control and Signal Processing* 31.12, pp. 1920–1933.
- (2020a). “Identification of jump Markov autoregressive exogenous systems with missing measurements”. In: *Journal of the Franklin Institute* 357.6, pp. 3498–3523.
- (2020b). “Online identification of time-delay jump Markov autoregressive exogenous systems with recursive expectation-maximization algorithm”. In: *International Journal of Adaptive Control and Signal Processing* 34.3, pp. 407–426.
- Du, Y., F. Liu, J. Qiu, and M. Buss (2020). “A Semi-Supervised Learning Approach for Identification of Piecewise Affine Systems”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 67.10, pp. 3521–3532.
- (2021a). “A novel recursive approach for online identification of continuous-time switched nonlinear systems”. In: *International Journal of Robust and Nonlinear Control* 31.15, pp. 7546–7565.
- (2021b). “Online Identification of Piecewise Affine Systems Using Integral Concurrent Learning”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 68.10, pp. 4324–4336.
- Du, Z., L. Balzano, and N. Ozay (2018). “A Robust Algorithm for Online Switched System Identification”. In: *IFAC-PapersOnLine* 51.15, pp. 293–298.
- Fan, L., H. Kodamana, and B. Huang (2017). “Robust Identification of Switching Markov ARX Models Using EM Algorithm.” In: *IFAC-PapersOnLine* 50.1, pp. 9772–9777.
- Garulli, A., S. Paoletti, and A. Vicino (2012). “A survey on switched and piecewise affine system

- identification". In: *IFAC Proceedings Volumes* 45.16, pp. 344–355.
- Goudjil, A., M. Pouliquen, E. Pigeon, and O. Gehan (2016). "A real-time identification algorithm for switched linear systems with bounded noise". In: *2016 European Control Conference*.
- (2017a). "Identification algorithm for MIMO switched output error model in presence of bounded noise". In: *2017 IEEE 56th Annual Conference on Decision and Control*.
- Goudjil, A., M. Pouliquen, E. Pigeon, O. Gehan, and T. Bonargent (2020). "Continuous-time Identification for a class of Switched Linear Systems". In: *2020 European Control Conference*.
- Goudjil, A., M. Pouliquen, E. Pigeon, O. Gehan, and B. Targui (2017b). "Recursive Output Error Identification Algorithm for Switched Linear systems with Bounded Noise". In: *IFAC-PapersOnLine* 50.1, pp. 14112–14117.
- Hartmann, A., J. M. Lemos, R. S. Costa, J. Xavier, and S. Vinga (2015). "Identification of switched ARX models via convex optimization and expectation maximization". In: *Journal of Process Control* 28, pp. 9–16.
- Hojjatnia, S., C. M. Lagoa, and F. Dabbene (2020). "Identification of switched autoregressive exogenous systems from large noisy datasets". In: *International Journal of Robust and Nonlinear Control* 30.15, pp. 5777–5801.
- Hu, Q., Q. Fei, H. Ma, Q. Wu, and Q. Geng (2015). "Identification for Switched Systems." In: *IFAC-PapersOnLine* 48.28, pp. 514–519.
- Hure, N. and M. Vasak (2017). "Clustering-based identification of MIMO piecewise affine systems". In: *2017 21st International Conference on Process Control*.
- Jianwang, H. and R. A. Ramirez-Mendoza (2020). "Zonotope parameter identification for piecewise affine system". In: *Systems Science and Control Engineering* 8.1, pp. 232–240.
- Kersting, S. and M. Buss (2017). "Recursive estimation in piecewise affine systems using parameter identifiers and concurrent learning". In: *International Journal of Control* 92.6, pp. 1264–1281.
- Keshvari-Khor, H., A. Karimpour, and N. Pariz (2018). "Identification of continuous-time switched linear systems from low-rate sampled data". In: *IET Control Theory and Applications* 12.14, pp. 1964–1973.
- Khanmirza, E., M. Nazarahari, and A. Mousavi (2016). "Identification of piecewise affine systems based on fuzzy PCA-guided robust clustering technique". In: *EURASIP Journal on Advances in Signal Processing* 2016.1.
- Konishi, K. (2015). "Multiple low rank matrix approach to switched autoregressive exogenous system identification". In: *2015 10th Asian Control Conference*.
- Lauer, F. and G. Bloch (2018). "Hybrid system identification: Theory and algorithms for learning switching models, vol. 478". In: *Cham, Switzerland: Springer*.
- Lauer, F. (2013). "Estimating the probability of success of a simple algorithm for switched linear regression". In: *Nonlinear Analysis: Hybrid Systems* 8, pp. 31–47.
- Li, L., W. Done, and Y. Ji (2016). "A subspace approach to the identification of MIMO piecewise linear systems". In: *2016 35th Chinese Control Conference*.
- Li, L. and J. Liu (2017). "Subspace clustering on parameter estimation of switched affine models". In: *2017 36th Chinese Control Conference*.
- Liu, J., Z. Xu, J. Zhao, and Z. Shao (2022). "Identification of piecewise affine model for batch processes based on constrained clustering technique". In: *Chemical Engineering Research and Design* 181, pp. 278–286.
- Liu, X., X. Yang, and M. Yu (2021). "Identification of switched FIR systems with random missing outputs: A variational Bayesian approach". In: *Journal of the Franklin Institute* 358.1, pp. 1136–1151.
- Lopes, R. V., J. Y. Ishihara, and G. A. Borges (2016). "Identification of state-space switched linear systems using clustering and hybrid filtering". In: *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 39.2, pp. 565–573.
- Lu, Y., B. Huang, and S. Khatibisepehr (2016). "A Variational Bayesian Approach to Robust Identification of Switched ARX Models". In: *IEEE Transactions on Cybernetics* 46.12, pp. 3195–3208.
- Ma, J., B. Huang, and F. Ding (2019). "Parameter estimation of Markov-switching Hammerstein systems using the variational Bayesian approach". In: *IET Control Theory and Applications* 13.11, pp. 1646–1655.
- Mattsson, P., D. Zachariah, and P. Stoica (2016). "Recursive Identification Method for Piecewise ARX Models: A Sparse Estimation Approach". In: *IEEE Transactions on Signal Processing* 64.19, pp. 5082–5093.
- Mazzoleni, M., V. Breschi, and S. Formentin (2021). "Piecewise nonlinear regression with data augmentation". In: *IFAC-PapersOnLine* 54.7, pp. 421–426.
- Mejari, M., V. Breschi, V. V. Naik, and D. Piga (2020a). "A Bias-Correction Approach for the Identification of Piecewise Affine Output-Error Models". In: *IFAC-PapersOnLine* 53.2, pp. 1096–1101.
- Mejari, M., V. Breschi, and D. Piga (2020b). "Recursive Bias-Correction Method for Identification of

- Piecewise Affine Output-Error Models". In: *IEEE Control Systems Letters* 4.4, pp. 970–975.
- Mejari, M., V. V. Naik, D. Piga, and A. Bemporad (2018). "Regularized Moving-Horizon PWA Regression for LPV System Identification". In: *IFAC-PapersOnLine* 51.15, pp. 1092–1097.
- (2020c). "Identification of hybrid and linear parameter-varying models via piecewise affine regression using mixed integer programming". In: *International Journal of Robust and Nonlinear Control* 30.15, pp. 5802–5819.
- Naik, V. V., M. Mejari, D. Piga, and A. Bemporad (2017). "Regularized moving-horizon piecewise affine regression using mixed-integer quadratic programming". In: *2017 25th Mediterranean Conference on Control and Automation*.
- Nazari, S., B. Rashidi, Q. Zhao, and B. Huang (2016). "An Iterative Algebraic Geometric Approach for Identification of Switched ARX Models with Noise". In: *Asian Journal of Control* 18.5, pp. 1655–1667.
- Ozay, N., C. Lagoa, and M. Sznaier (2015). "Set membership identification of switched linear systems with known number of subsystems". In: *Automatica* 51, pp. 180–191.
- Ozbay, B., O. Camps, and M. Sznaier (2019). "Efficient Identification of Error-in-Variables Switched Systems via a Sum-of-Squares Polynomial Based Subspace Clustering Method". In: *2019 IEEE 58th Conference on Decision and Control*.
- Paoletti, S., I. Savelli, A. Garulli, and A. Vicino (2019). "A bilevel programming framework for piecewise affine system identification". In: *2019 IEEE 58th Conference on Decision and Control*.
- Piga, D., A. Bemporad, and A. Benavoli (2020a). "Rao-Blackwellized sampling for batch and recursive Bayesian inference of Piecewise Affine models". In: *Automatica* 117, p. 109002.
- Piga, D., V. Breschi, and A. Bemporad (2020b). "Estimation of jump Box-Jenkins models". In: *Automatica* 120, p. 109126.
- Pillonetto, G. (2016). "A new kernel-based approach to hybrid system identification". In: *Automatica* 70, pp. 21–31.
- Robinson, H. (2021). "Approximate Piecewise Affine Decomposition of Neural Networks". In: *IFAC-PapersOnLine* 54.7, pp. 541–546.
- Rui, R., T. Ardeshiri, and A. Bazanella (2016). "Identification of piecewise affine state-space models via expectation maximization". In: *2016 IEEE Conference on Computer Aided Control System Design*.
- Sarkar, T., A. Rakhlin, and M. Dahleh (2019). "Nonparametric System identification of Stochastic Switched Linear Systems". In: *2019 IEEE 58th Conference on Decision and Control*.
- Scampicchio, A., A. Giaretta, and G. Pillonetto (2018). "Nonlinear Hybrid Systems Identification using Kernel-Based Techniques". In: *IFAC-PapersOnLine* 51.15, pp. 269–274.
- Scampicchio, A. and G. Pillonetto (2018). "A New Model Selection Approach to Hybrid Kernel-Based Estimation". In: *2018 IEEE Conference on Decision and Control*.
- Sefidmazgi, M. G., M. M. Kordmahalleh, A. Homaifar, and A. Karimoddini (2015). "Switched linear system identification based on bounded-switching clustering". In: *2015 American Control Conference*.
- Sefidmazgi, M. G., M. M. Kordmahalleh, A. Homaifar, A. Karimoddini, and E. Tunstel (2016). "A bounded switching approach for identification of switched MIMO systems". In: *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE.
- Sellami, L., S. Zidi, and K. Abderrahim (2016). "Identification of switched linear systems using self-adaptive SVR algorithm". In: *2016 24th Mediterranean Conference on Control and Automation*.
- Shah, A. K. and D. M. Adhyaru (2014). "Parameter identification of PWARX models using fuzzy distance weighted least squares method". In: *Applied Soft Computing* 25, pp. 174–183.
- Song, C., J. Wang, X. Ma, and J. Zhao (2020). "A PWA model identification method based on optimal operating region partition with the output-error minimization for nonlinear systems". In: *Journal of Process Control* 88, pp. 1–9.
- Tang, X. and Y. Dong (2020). "Expectation maximization based sparse identification of cyberphysical system". In: *International Journal of Robust and Nonlinear Control* 31.6, pp. 2044–2060.
- Wågberg, J., F. Lindsten, and T. B. Schön (2015). "Bayesian nonparametric identification of piecewise affine ARX systems." In: *IFAC-PapersOnLine* 48.28, pp. 709–714.
- Wagner, M. and A. Kroll (2014). "A method to identify hybrid systems with mixed piecewise affine or nonlinear models of Takagi-Sugeno type". In: *2014 European Control Conference*.
- Wang, J., C. Song, J. Zhao, and Z. Xu (2020). "A PWA model identification method for nonlinear systems using hierarchical clustering based on the gap metric". In: *Computers and Chemical Engineering* 138, p. 106838.
- Wang, Z., H. An, and X. Luo (2019). "Switch detection and robust parameter estimation for slowly switched Hammerstein systems". In: *Nonlinear Analysis: Hybrid Systems* 32, pp. 202–213.

- Xiujun, C., W. Hongwei, W. Lin, and X. Zhengqing (2020). "Identification of switched nonlinear systems based on EM algorithm". In: *2020 39th Chinese Control Conference*.
- Xu, Z., Y. Huang, J. Zhao, C. Song, and Z. Shao (2018). "Time-Partitioned Piecewise Affine Output Error Model for Batch Processes". In: *Industrial and Engineering Chemistry Research* 57.5, pp. 1560–1568.
- Yahya, O., Z. Lassoued, and K. Abderrahim (2020). "Identification of PWARX Model Based on Outer Bounding Ellipsoid Algorithm". In: *2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering*.
- Yang, Y., C. Xiang, S. Gao, and T. H. Lee (2017). "Data-driven identification and control of nonlinear systems using multiple NARMA-L2 models". In: *International Journal of Robust and Nonlinear Control* 28.12, pp. 3806–3833.
- Yuan, Y., X. Tang, W. Zhou, W. Pan, X. Li, H.-T. Zhang, H. Ding, and J. Goncalves (2019). "Data driven discovery of cyber physical systems". In: *Nature Communications* 10.1.
- Zhang, X., Y. Cheng, Y. Wang, M. Sznaier, and O. Camps (2016). "Identification of switched Wiener systems based on local embedding". In: *2016 IEEE 55th Conference on Decision and Control*.
- Zhang, X., M. Sznaier, and O. Camps (2018). "Efficient Identification of Error-in Variables Switched Systems Based on Riemannian Distance-Like Functions". In: *2018 IEEE Conference on Decision and Control*.

4

SWITCHED BOX-JENKINS MODELS FOR BIOPROCESSES

This chapter focuses on the development of linear Switched Box–Jenkins (SBJ) models for approximating complex dynamical models of biological wastewater treatment processes. We discuss the adaptation of these processes to the SBJ framework, showing the model’s generality and flexibility as a class of switched systems that can offer accurate predictions for complex and nonlinear dynamics. This approach of modeling enables real-time data reconciliation from experiments and allows the design of model-based control strategies previously inaccessible with conventional complex wastewater treatment models. Through the extension of the Outer Bounding Ellipsoids (OBEs) algorithm, the chapter introduces an online two-stage parameter identification algorithm that effectively handles bounded disturbances for SBJ models. Using the OBE method relaxes the stochastic assumptions on disturbances, which may not be satisfied in practice, particularly for biological and environmental fluctuations. The proposed decomposed OBE algorithm separately identifies the switching patterns and parameters of linear submodels, conducting parameter identification in two distinct phases for input/output and disturbance/output submodels. The efficacy of this approach is shown via simulation results validated against both ADM1 and PBM, demonstrating the proposed algorithm’s capability to accurately predict outputs from different bio-process models.

This chapter is an adapted version of *an identification algorithm of switched Box-Jenkins systems in the presence of bounded disturbances: An approach for approximating complex biological wastewater treatment models*, Ali Moradvandi, et al. (2024), *Journal of Water Process Engineering*.

4.1. INTRODUCTION

Hybrid (switched) dynamical systems capture interconnected continuous and discrete behaviors, serving to model processes with non-smooth behaviors or to approximate systems with high-order nonlinearities. Biological treatment processes are described by interconnected and competing bio- and physico-chemical reactions for substrate consumption and growth of different trophic groups within a microbial community, resulting in nonlinear model behaviors. This type of complex nonlinear behavior can be simplified in terms of modeling using hybrid systems. Switched systems, as a well-known class of hybrid systems, consist of a switching pattern (or mode) and a finite number of values (countable state variables) that coordinates with corresponding continuous and linear subsystems (or submodels) (Lauer and Bloch 2018).

4

Hybrid system identification methods, as a tool to find a switched system to approximate a highly nonlinear biological treatment model, involve two steps: 1) estimating the parameters of the submodels, and 2) determining the switching patterns. Furthermore, hybrid system identification methods as a data-driven modeling approach avoid the complexity inherent in mechanistic modeling of input-output relations. Moreover, using a set of linear models to approximate a nonlinear dynamic of a biological treatment process not only is straightforward to implement in comparison with Neural Networks but also holds significant accuracy in comparison with non-switched systems.

The input-output model complexity ranges from relatively simple Auto-Regressive exogenous (ARX) models to more complex general Box-Jenkins (BJ) models. Input-output models consist of two parts, i.e. auto-regressive (depending on the previous forecasts) and moving-average (depending on the error of previous forecasts). Box-Jenkins (BJ) models have the advantage of describing stochastic systems in a more general and flexible way, since they include the output error model (Ding *et al.* 2010), the output error moving average model (Wang 2011), and the output error autoregressive model (Wang *et al.* 2010) as special cases. Moreover, switched finite impulse response, SFIR, (Liu *et al.* 2021) switched autoregressive exogenous, SARX (Du *et al.* 2018), switched autoregressive and moving-average, SARMAX (Hojjatnia *et al.* 2020), switched output error, SOE (Goudjil *et al.* 2017), and error-in-variable SARX, EIV-SARX (Ozbay *et al.* 2019) models can be mathematically considered as subclasses of a switched Box-Jenkins (SBJ) model. In other words, the mentioned simple model structures can be driven with simplification of a BJ model.

The BJ structure, also, has been widely and effectively used for time series prediction due to its generality and efficiency in prediction (Box *et al.* 2015). As summarized in Table 4.1, some biological processes have been modelled by switched systems in the literature. The foundation of the submodels in these papers is ARX. The identification problem has been addressed using different approaches in these articles, including optimization-based methods by Hartmann *et al.* (2015) and Song *et al.* (2020), likelihood-based methods by Chen *et al.* (2020a,b), clustering-based methods by Wang *et al.* (2020), and Outer Bounding Ellipsoid (OBE) methods by Yahya *et al.* (2020). Since all these papers deal with ARX models, the identification approaches cannot be directly extended for the more general SBJ models.

Table 4.1: Applications of biological processes modelled by different hybrid systems.

Application	Hybrid model	Method	Reference
pH neutralization process	PWARX	Clustering-based	Wang <i>et al.</i> (2020)
Diauxic bacterial growth	SARX	Optimization-based	Hartmann <i>et al.</i> (2015)
CSTR with exothermic reaction	PWARX	Clustering-based	Song <i>et al.</i> (2020)
Continuous fermentation reactor	SARX	Likelihood-based	Chen <i>et al.</i> (2020a)
	Delay-SARX	Likelihood-based	Chen <i>et al.</i> (2020b)
Transesterification reactor	PWARX	OBE	Yahya <i>et al.</i> (2020)

In addition to the base model (parametrization), selecting a suitable algorithm for solving the identification problem is an integral part of hybrid system identification that should be developed based on the selected base model (Moradvandi *et al.* 2023). The approaches are classified into optimization-based techniques (Bianchi *et al.* 2021), clustering-based methods (Mazzoleni *et al.* 2021), likelihood-based methods (Chen *et al.* 2020a), algebraic methods (Hojjatnia *et al.* 2020), and Outer Bounding Ellipsoid (OBE) methods (Goudjil *et al.* 2023; Yahya *et al.* 2020). Comprehensive reviews of these techniques have been done by (Garulli *et al.* 2012; Moradvandi *et al.* 2023). The selection of an appropriate method depends on factors such as parametrization, available knowledge of the system, and the computational burden associated with the model. Optimization-based algorithms are the most commonly used, and they have recently been combined with other approaches such as clustering and classical algebraic methods (Du *et al.* 2020; Wang *et al.* 2020).

To select an approach, practical aspects of a biological treatment process should also be taken into account. The behavior of a biological process can be affected by random and unpredictable factors. Common examples are meteorological fluctuations and influent concentration perturbations. Under these situations, Piga *et al.* (2020a) showed that stochastic modeling can be an option. However, the assumption of a statistical consideration for disturbances or noises may not always be justified due to an unknown probability distribution or modeling mismatch (Goudjil *et al.* 2023). On the other hands, the stochastic assumption requires precise distribution information and employs a sequence of representative scenarios, which is hard to be satisfied in real-world applications. Alternatively, the assumption of bounded disturbances is considered less stringent and therefore a pragmatic solution.

Amongst the mentioned hybrid system identification methods, the OBE method is one of the methods that has the advantage of not requiring any stochastic noise assumption. Furthermore, since the basis of the OBE algorithm is matrix manipulation, the OBE algorithm is not only computationally efficient, but also well-suited for analyzing large datasets (Goudjil *et al.* 2023). This method has been developed for hybrid systems parametrized by SARX (Goudjil *et al.* 2016), SOE (Goudjil *et al.* 2017), and piecewise affine ARX (PWARX) (Yahya *et al.* 2020) models, not yet for the general models such as SBJ. The OBE algorithm encompasses two stages: (1) the procedure of assigning data by considering both the residual error and an upper bound for the estimation error of all the submodels, and (2) utilizing Recursive Least Squares (RLS) simultaneously to update the parameters of the active

submodel in each time step (Goudjil *et al.* 2023).

Motivated by the importance of BJ models, particularly for biological treatment processes as well as the current trend of extending other methods for SBJ models (Chai *et al.* 2020; Piga *et al.* 2020b), this chapter addresses the extension of OBE algorithms to SBJ systems. For this purpose, auxiliary model identification and decomposition techniques, which have been discussed for non-switched systems by Ding and Duan (2013), are adapted to the considered switched structure and the OBE framework. This adaptation deals with lack of availability of internal signals within the BJ structure. Inspired by the work done by Chai *et al.* (2020), the underlying principle involves the decomposition of a BJ system into two parts (the autoregressive part and the moving-average part), followed by the auxiliary model identification approach to determine the parameters of each part and the internal signals simultaneously. Therefore, a reformulation of the two-stage OBE algorithm based on adaptation of the decomposed technique is addressed in this study, and the active submodel detection and the parameter identification procedures are developed based on a decomposed OBE objective function for SBJ models.

The primary aim of the present work is, therefore, to develop the OBE algorithm for SBJ models. To achieve this objective, we present a mathematical exposition by adapting the decomposition technique to switched systems in order to formulate the identification problem posed by SBJ systems within the OBE framework. Furthermore, the approximation of biological treatment processes represented by complex mathematical models, is explored within the framework SBJ models by validating the proposed algorithm for Anaerobic Digestion Model 1 (ADM1) and Purple Bacteria Model (PBM). Through a comprehensive numerical assessment and interpretation, this research sheds light on the potential applications of the SBJ modeling approach, contributing valuable insights into real-time data reconciliation and control strategies of biological treatment processes.

The chapter is organized as follows. Materials and methods (Section 4.2) include the formulation of the identification problem in Section 4.2.1 and the OBE identification procedure in Section 4.2.2. Section 4.3 presents results and discussions. Formulating of biological models in the form of SBJ is discussed in this section, and the aforementioned case studies of biological wastewater treatment models are also analyzed. Limitations of the proposed method are discussed in Section 4.4, and in the last section, conclusions are drawn.

4.2. MATERIALS AND METHODS

4.2.1. PROBLEM FORMULATION

A switched discrete-time linear system parameterized by a BJ model is represented as follows:

$$y_k = \frac{B(q^{-1}, \theta_{z_k})}{A(q^{-1}, \theta_{z_k})} u_k + \frac{C(q^{-1}, \theta_{z_k})}{D(q^{-1}, \theta_{z_k})} v_k \quad (4.1)$$

where $y_k \in \mathbb{R}$, $u_k \in \mathbb{R}$, and $v_k \in \mathbb{R}$ denote the system output, the system input, and the disturbance (noise). Moreover, $A(q^{-1}, \theta_{z_k})$, $B(q^{-1}, \theta_{z_k})$, $C(q^{-1}, \theta_{z_k})$, and $D(q^{-1}, \theta_{z_k})$ are the linear filters. The discrete state, $z_k \in \{1, \dots, m\}$ indicates the active mode of m

number of parameterized submodels or modes at time step k . If we assume at time step k , the i -th mode is active, i.e. $z_k = i$, the linear filters that are rational functions of the time shift operator q^{-1} (i.e. $q^{-d}x_k = x_{k-d}$ for $d \in \mathbb{Z}$), can be written as follows:

$$\frac{B(q^{-1}, \theta_i)}{A(q^{-1}, \theta_i)} = \frac{b_{i1}q^{-1} + \dots + b_{in_b}q^{-n_b}}{1 + a_{i1}q^{-1} + \dots + a_{in_a}q^{-n_a}}, \quad (4.2a)$$

$$\frac{C(q^{-1}, \theta_i)}{D(q^{-1}, \theta_i)} = \frac{1 + c_{i1}q^{-1} + \dots + c_{in_c}q^{-n_c}}{1 + d_{i1}q^{-1} + \dots + d_{in_d}q^{-n_d}}, \quad (4.2b)$$

where n_a , n_b , n_c , and n_d are the orders of the filters ($A(\cdot)$, $B(\cdot)$, $C(\cdot)$, $D(\cdot)$) respectively, and the vectors of parameters can be expressed as

$$\theta_{1i} = [a_{i1}, \dots, a_{in_a}, b_{i1}, \dots, b_{in_b}]^T \in \mathbb{R}^{n_a+n_b}, \quad (4.3a)$$

$$\theta_{2i} = [c_{i1}, \dots, c_{in_c}, d_{i1}, \dots, d_{in_d}]^T \in \mathbb{R}^{n_c+n_d}, \quad (4.3b)$$

$$\theta_i = [\theta_{1i}^T, \theta_{2i}^T]^T \in \mathbb{R}^{n_a+n_b+n_c+n_d}, \quad (4.3c)$$

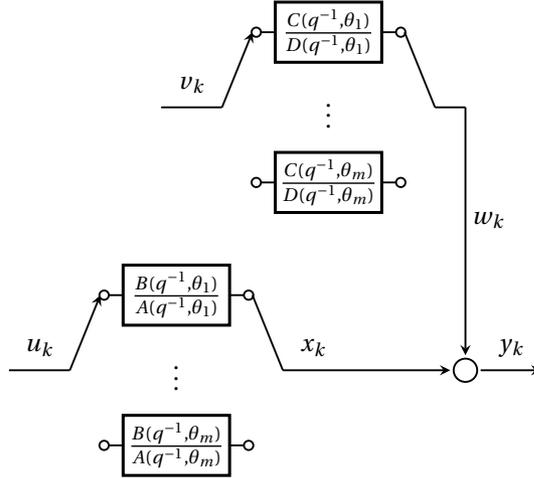


Figure 4.1: Schematization of the switched BJ system for m number of mode.

The block diagram of the switched BJ system is depicted in Figure 4.1. According to the block diagram, the two auxiliary variables x_k and w_k can be written as follows:

$$x_k = (1 - A(q^{-1}, \theta_{z_k}))x_k + B(q^{-1}, \theta_{z_k})u_k = \phi_k^T \theta_{1, z_k}, \quad (4.4a)$$

$$w_k = (1 - C(q^{-1}, \theta_{z_k}))w_k + D(q^{-1}, \theta_{z_k})v_k = \psi_k^T \theta_{2, z_k} + v_k, \quad (4.4b)$$

where ϕ_k and ψ_k are the regressor vectors:

$$\phi_k = [-x_{k-1}, \dots, -x_{k-n_a}, u_{k-1}, \dots, u_{k-n_b}]^T \in \mathbb{R}^{n_a+n_b}, \quad (4.5a)$$

$$\psi_k = [-w_{k-1}, \dots, -w_{k-n_c}, v_{k-1}, \dots, v_{k-n_d}]^T \in \mathbb{R}^{n_c+n_d}. \quad (4.5b)$$

Therefore, the model (4.1) can be rewritten as

$$\begin{aligned} y_k &= \phi_k^T \theta_{1,z_k} + \psi_k^T \theta_{2,z_k} + v_k \\ &= \Phi_k^T \theta_{z_k} + v_k. \end{aligned} \quad (4.6)$$

where $\Phi_k = [\phi_k^T, \psi_k^T]^T \in \mathbb{R}^{n_a+n_b+n_c+n_s d}$,

The decomposition technique is a tool that is used to deal with two-stage identification procedure (Ding and Duan 2013). In this study, we want to formulate it for switched systems. An intermediate variable is defined as

$$\omega_k = y_k - \psi_k^T \theta_{2,z_k} \quad (4.7)$$

and the main system in (4.6) can be decomposed into two subsystems as follows:

$$\omega_k = \phi_k^T \theta_{1,z_k} + v_k \quad (4.8a)$$

$$w_k = \psi_k^T \theta_{2,z_k} + v_k, \quad (4.8b)$$

$$\omega_k - \phi_k^T \theta_{1,z_k} = w_k - \psi_k^T \theta_{2,z_k} = v_k, \quad (4.8c)$$

and they can be rewritten as

$$\omega_k = y_k - \phi_k^T \theta_{1,z_k} \quad (4.9a)$$

$$w_k = y_k - \psi_k^T \theta_{2,z_k}, \quad (4.9b)$$

These decomposed functions will be utilized in the parameter identification stage later on. The identification objective should be defined in order to estimate the discrete state, z_k , and the parameter vectors, θ_{z_k} , $z_k = 1, \dots, m$, given a collection of input and output observations. If the estimations of the discrete state and the parameter vectors are defined as \hat{z}_k , $\hat{\theta}_{1,\hat{z}_k}$, and $\hat{\theta}_{2,\hat{z}_k}$, they should satisfy

$$|y_k - \Phi_k^T \hat{\theta}_{\hat{z}_k}| \leq \delta, \forall k \quad (4.10a)$$

$$|y_k - \phi_k^T \hat{\theta}_{1,\hat{z}_k} - \psi_k^T \hat{\theta}_{2,\hat{z}_k}| \leq \delta, \forall k \quad (4.10b)$$

where δ is an upper bound of v_k , i.e. $|v_k| \leq \delta, \forall k$. The objective can also be expressed according to (4.8c). The representation of the objective for the decomposed form of the switched system will be used to derive the parameter identification procedure in the next section.

To apply the OBE algorithm for the defined objective and to derive the estimation procedure of the discrete state, the system represented by (4.6), should be extended in the following format. If we assume that at time step k the submodel i is active, then it can be written as

$$\begin{cases} y_k = \phi_k^T \theta_{11} + \psi_k^T \theta_{21} + v_k + \phi_k^T (\theta_{1i} - \theta_{11}) + \psi_k^T (\theta_{2i} - \theta_{21}) \\ y_k = \phi_k^T \theta_{12} + \psi_k^T \theta_{22} + v_k + \phi_k^T (\theta_{1i} - \theta_{12}) + \psi_k^T (\theta_{2i} - \theta_{22}) \\ \vdots \\ y_k = \phi_k^T \theta_{1i} + \psi_k^T \theta_{2i} + v_k \\ \vdots \\ y_k = \phi_k^T \theta_{1m} + \psi_k^T \theta_{2m} + v_k + \phi_k^T (\theta_{1i} - \theta_{1m}) + \psi_k^T (\theta_{2i} - \theta_{2m}) \end{cases} \quad (4.11)$$

By defining the extended parameter vectors, $\Theta_1 \in \mathbb{R}^{(n_a+n_b)m \times 1}$ and $\Theta_2 \in \mathbb{R}^{(n_c+n_d)m \times 1}$, the extended noise vector, $V_k \in \mathbb{R}^{m \times 1}$, and the extended output vector, $Y_k \in \mathbb{R}^{m \times 1}$, as follows, the system (4.11) can be rewritten as follows:

$$\Theta_1 = [\theta_{11}, \dots, \theta_{1m}]^T \quad (4.12a)$$

$$\Theta_2 = [\theta_{21}, \dots, \theta_{2m}]^T \quad (4.12b)$$

$$Y_k = [y_k, \dots, y_k]^T \quad (4.12c)$$

$$V_{k,z_k=i} = \begin{bmatrix} v_k + \phi_k^T(\theta_{1i} - \theta_{11}) + \psi_k^T(\theta_{2i} - \theta_{21}) \\ \vdots \\ v_k \\ \vdots \\ v_k + \phi_k^T(\theta_{1i} - \theta_{1m}) + \psi_k^T(\theta_{2i} - \theta_{2m}) \end{bmatrix} \quad (4.12d)$$

$$Y_k = \bar{\phi}_k^T \Theta_1 + \bar{\psi}_k^T \Theta_2 + V_{k,z_k} \quad (4.12e)$$

where $\bar{\phi} = I_m \otimes \phi$ and $\bar{\psi} = I_m \otimes \psi$, in which \otimes and I_N denote the Kronecker product and the identity matrix of order m , respectively. If the estimations of z_k , and the parameter vectors, Θ_1 and Θ_2 are denoted by \hat{z}_k , $\hat{\Theta}_1$, and $\hat{\Theta}_2$, respectively, (4.12e) can be rewritten as

$$V_{k,\hat{z}_k} = Y_k - \bar{\phi}_k^T \hat{\Theta}_1 - \bar{\psi}_k^T \hat{\Theta}_2 \quad (4.13)$$

Therefore, if we define $v_k(j)$ as the j -th element of V_{k,z_k} , tanking (4.13) into account, the problem objective (4.10b) can be redefined as follows:

$$|v_k(\hat{z}_k)| \leq \delta, \forall k \quad (4.14)$$

where \hat{z}_k can be any integer values between 1 and m at time step k .

4.2.2. IDENTIFICATION ALGORITHM

The OBE method is a technique used in conventional identification algorithms to estimate the parameters of a model within a given set of constraints, where the feasible region (the set of possible solutions) is bounded. Using this technique for switched systems allows computing the ellipsoid bounds for all the submodels and finding the active one that fits inside the assigned ellipsoid bound. The proposed identification algorithm is based on two stages, i.e. we first estimate the discrete state (the switching pattern), then the parameter vectors, in a repetitive manner for each time step. The parameter vector estimation is also derived based on the decomposition technique in two stages, i.e. the parameter vector is primarily updated, then we estimate internal signals for next steps. To derive the algorithm, the estimates of the parameter vectors at time step k are denoted by $\hat{\Theta}_{1,k}$ and $\hat{\Theta}_{2,k}$. The a priori and the a posteriori predictors of Y_k can be written w.r.t. (4.12e), respectively, as

$$\begin{cases} Y_{k/k-1} = \bar{\phi}_k^T \hat{\Theta}_{1,k-1} + \bar{\psi}_k^T \hat{\Theta}_{2,k-1} \\ Y_{k/k} = \bar{\phi}_k^T \hat{\Theta}_{1,k} + \bar{\psi}_k^T \hat{\Theta}_{2,k} \end{cases} \quad (4.15)$$

Then a priori prediction error can be defined as follows:

$$V_{k/k-1} = Y_k - Y_{k/k-1} = Y_k - \bar{\phi}_k^T \hat{\Theta}_{1,k-1} - \bar{\psi}_k^T \hat{\Theta}_{2,k-1} \quad (4.16)$$

Therefore, the two-stage OBE algorithm can be described as follows:

Step 1 (estimation of \hat{z}_k): The first step estimates the discrete state, i.e. \hat{z}_k based on the smallest element of the vector $V_{k/k-1}$ that can be expressed by $\varrho_k = |v_{k/k-1}(\hat{z}_k)|$, in which $\hat{z}_k \in \{1, \dots, m\}$ is the detected active mode at time step k .

Step 2 (estimations of $\hat{\Theta}_1$ and $\hat{\Theta}_2$): The second step is to identify the defined parameter vectors, i.e. $\hat{\Theta}_1$ and $\hat{\Theta}_2$. This step is derived based on the decomposition technique. According to the decomposed model written by (4.8c), the objective functions to derive a Recursive Least Square (RLS) minimization for the decomposed model can be defined as follows:

$$J_1(\theta_{1,z_k}) := \sum_{j=1}^k (\omega_k - \phi_k^T \theta_{1,z_k})^2 \quad (4.17a)$$

$$J_2(\theta_{2,z_k}) := \sum_{j=1}^k (w_k - \psi_k^T \theta_{2,z_k})^2 \quad (4.17b)$$

where $J_1 = J_2$ according to (4.8c). Assuming the i -th mode is active at time step k ($\hat{z}_k = i$), the update laws for the estimates of the parameters, i.e. $\hat{\theta}_{1i}$ and $\hat{\theta}_{2i}$ can be written as a result of the RLS minimization as follows:

$$\hat{\theta}_{1i,k} = \hat{\theta}_{1i,k-1} + L_{1,k} [y_k - \psi_k^T \hat{\theta}_{2i,k-1} - \phi_k^T \hat{\theta}_{1i,k-1}], \quad (4.18a)$$

$$\hat{\theta}_{2i,k} = \hat{\theta}_{2i,k-1} + L_{2,k} [y_k - \phi_k^T \hat{\theta}_{1i,k-1} - \psi_k^T \hat{\theta}_{2i,k-1}], \quad (4.18b)$$

where

$$L_{1,k} = P_{1,k-1} \phi_k [1 + \phi_k^T P_{1,k-1} \phi_k]^{-1}, \quad (4.19a)$$

$$L_{2,k} = P_{2,k-1} \psi_k [1 + \psi_k^T P_{2,k-1} \psi_k]^{-1}, \quad (4.19b)$$

and

$$P_{1,k} = [I_{n_a+n_b} - L_{1,k} \phi_k^T] P_{1,k-1}, \quad (4.20a)$$

$$P_{2,k} = [I_{n_c+n_d} - L_{2,k} \psi_k^T] P_{2,k-1}, \quad (4.20b)$$

Now, the solution of the decomposed RLS formulated above for the i -th mode can be extended for all m number of submodels to be able to apply the OBE algorithm. This is done considering the definitions of Θ_1 and Θ_2 expressed by (4.12a) and (4.12b). The extended matrices, $\bar{\phi}$ and $\bar{\psi}$, should also be used as defined by the Kronecker product of an identity matrix of the order m to ϕ and ψ stated in (4.5a) and (4.5b). To be able to update only the parameters of the active submodel, a symmetric matrix is defined such that the values of all the elements are zero except for the one element corresponding to the identified active submodel (Goudjil *et al.* 2023). Because we are using the decomposition technique in this chapter, we define

two matrices - one for the autoregressive part, $\Upsilon_{1,k} \in \mathbb{R}^{m \times m}$, and the other one for the moving average part, $\Upsilon_{2,k} \in \mathbb{R}^{m \times m}$:

$$\Upsilon_{1,k} = \begin{cases} \left(\bar{\phi}_k^T P_{1,k-1} \bar{\phi}_k \right)^{-1} (\Lambda_k - I_m); \\ \quad \text{if } \bar{\phi}_k^T P_{1,k-1} \bar{\phi}_k > 0 \text{ and } \varrho_k > \delta \\ \mathbf{0}_{m \times m}; \quad \text{else} \end{cases} \quad (4.21a)$$

$$\Upsilon_{2,k} = \begin{cases} \left(\bar{\psi}_k^T P_{2,k-1} \bar{\psi}_k \right)^{-1} (\Lambda_k - I_m); \\ \quad \text{if } \bar{\psi}_k^T P_{2,k-1} \bar{\psi}_k > 0 \text{ and } \varrho_k > \delta \\ \mathbf{0}_{m \times m}; \quad \text{else} \end{cases} \quad (4.21b)$$

in which $\Lambda_k \in \mathbb{R}^{m \times m}$ denotes the identity matrix at time step k , where the \hat{z}_k -th element on the diagonal is $\frac{\varrho_k}{\delta}$. Therefore, the parameters of the active submodel are updated, when the error of the output, ϱ_k , is not within the assigned ellipsoid bound, δ . The update gain is $\frac{\varrho_k}{\delta}$ in matrix Λ_k . On the other side, the adaptation is frozen when $\varrho_k \leq \delta$.

Considering the discussions above, the equations (4.18a)-(4.20b) can be reformulated for the extended version as follows:

$$\hat{\Theta}_{1,k} = \hat{\Theta}_{1,k-1} + L_{1,k} \left[Y_k - \bar{\psi}_k^T \hat{\Theta}_{2,k-1} - \bar{\phi}_k^T \hat{\Theta}_{1,k-1} \right] \quad (4.22a)$$

$$\hat{\Theta}_{2,k} = \hat{\Theta}_{2,k-1} + L_{2,k} \left[Y_k - \bar{\phi}_k^T \hat{\Theta}_{1,k-1} - \bar{\psi}_k^T \hat{\Theta}_{2,k-1} \right] \quad (4.22b)$$

$$L_{1,k} = \frac{1}{2} P_{1,k-1} \bar{\phi}_k \Upsilon_{1,k} \left[I_m + \bar{\phi}_k^T P_{1,k-1} \bar{\phi}_k \Upsilon_{1,k} \right]^{-1} \quad (4.22c)$$

$$L_{2,k} = \frac{1}{2} P_{2,k-1} \bar{\psi}_k \Upsilon_{2,k} \left[I_m + \bar{\psi}_k^T P_{2,k-1} \bar{\psi}_k \Upsilon_{2,k} \right]^{-1} \quad (4.22d)$$

$$P_{1,k} = \left[I_{m \times (n_a + n_b)} - L_{1,k} \bar{\phi}_k^T \right] P_{1,k-1} \quad (4.22e)$$

$$P_{2,k} = \left[I_{m \times (n_c + n_d)} - L_{2,k} \bar{\psi}_k^T \right] P_{2,k-1} \quad (4.22f)$$

The introduction of the factor $\frac{1}{2}$ in (4.22c) and (4.22d) allows us to prove the objective we defined in (4.14), which comes later. *Remark:* It should be noted that individual update equations (4.18a) and (6.9) are written based on this assumption that the system stays in one mode in two consecutive time instants $k-1$ and k . After the extension and defining $\Upsilon_{1,k}$ and $\Upsilon_{2,k}$, it is not the case for the extended update equations (4.22a) and (4.22b), since if the mode is changed from $k-1$ to k , the corresponding elements on diagonal of matrices $\Upsilon_{1,k}$ and $\Upsilon_{2,k}$ are also changed to the associated active mode to be updated at time step k and other

submodels remain frozen for the update process until they are detected active and the procedure continues.

The inner variables x and w and the variable v within the definition of ϕ_k , (4.5a), and ψ_k , (4.5b) and their extended corresponding matrices, i.e. $\bar{\phi}_k$ and $\bar{\psi}_k$ are unknown, which the estimates of these variables, i.e. \hat{x} , \hat{w} , and \hat{v} can be replaced (Ding and Duan 2013) as follows:

$$\hat{x}_k = X_k(\hat{z}_k) \quad (4.23a)$$

$$\hat{w}_k = y_k - \hat{x} \quad (4.23b)$$

$$\hat{v}_k = \hat{w}_k - W_k(\hat{z}_k) \quad (4.23c)$$

where $X_k = \bar{\phi}_k^T \hat{\Theta}_{1,k}$ and $W_k = \bar{\psi}_k^T \hat{\Theta}_{2,k}$ are the estimates of the unknown signals for all the submodels. If we assume the detected active submodel at time step k is i , i.e. $\hat{z}_k = i$, the i -th element of the vectors X_k and W_k should be used for the calculation of \hat{x}_k and \hat{v}_k , respectively, as stated in (4.23a) and (4.23c). Considering the explained procedure, the two-stage decomposed OBE algorithm can be summarized in Algorithm .

Algorithm Two-stage decomposed OBE algorithm

- 1: **Initialize:** $P_{1,0} = p_0 I_{m \times (n_a + n_b)}$, $P_{2,0} = p_0 I_{m \times (n_c + n_d)}$,
 - 2: $\hat{\Theta}_{1,0}$ and $\hat{\Theta}_{2,0}$ randomly initialized,
 - 3: $\hat{x}_k = \hat{w}_k = \hat{v}_k = 0 \quad \forall k \leq 0$
 - 4: **for** $k = 1$ **do**
 - 5: **step 1:** detect the active submodel \hat{z}_k
 - 6: **Receive** u_k and y_k
 - 7: **Form** $\bar{\phi}_k = \phi_k \otimes I_m$ and $\bar{\psi}_k = \psi_k \otimes I_m$
 - 8: based on (4.5a) and (4.5b)
 - 9: **Compute** $v_{k/k-1}$ as (4.16)
 - 10: **Compute** $\hat{z}_k = \underset{j=1, \dots, m}{\operatorname{argmin}} |v_{k/k-1}(j)|$
 - 11: **Compute** $\rho_k = |v_{k/k-1}(\hat{z}_k)|$
 - 12: **step 2:** estimate the parameters vectors $\hat{\Theta}_{1,k}$ and $\hat{\Theta}_{2,k}$
 - 13: **Compute** $\Upsilon_{1,k}$ and $\Upsilon_{2,k}$ as (4.21a) and (4.21b)
 - 14: **Compute** $L_{1,k}$, $L_{2,k}$, $P_{1,k}$, and $P_{2,k}$
 - 15: as (4.22c)-(4.22f)
 - 16: **Update** $\hat{\Theta}_{1,k}$ and $\hat{\Theta}_{2,k}$ as (4.22a) and (4.22b)
 - 17: **Compute** \hat{x}_k , \hat{w}_k , and \hat{v}_k as (4.23a)-(4.23c)
 - 18: $k = k + 1$
 - 19: **end for**
-

Remark: It can be shown that the objective defined in (4.14) is satisfied at each time step by implementing the proposed two-stage decomposed OBE algorithm. The a posteriori prediction error, i.e. $V_{k/k}$, can be written according to (4.16) as follows:

$$V_{k/k} = V_{k/k-1} - (\bar{\phi}_k^T L_{1,k} + \bar{\psi}_k^T L_{2,k}) V_{k/k-1} \quad (4.24)$$

Using the definitions of $L_{1,k}$ and $L_{2,k}$ as stated in (4.22c) and (4.22d) in (4.24) yields

$$V_{k/k} = V_{k/k-1} \left(I_m - \frac{1}{2} \bar{\phi}_k^T P_{1,k-1} \bar{\phi}_k Y_{1,k} [I_m + \bar{\phi}_k^T P_{1,k-1} \bar{\phi}_k Y_{1,k}]^{-1} - \frac{1}{2} \bar{\psi}_k^T P_{2,k-1} \bar{\psi}_k Y_{2,k} [I_N + \bar{\psi}_k^T P_{2,k-1} \bar{\psi}_k Y_{2,k}]^{-1} \right) \quad (4.25)$$

If the persistent excitation conditions (Ljung 1999) are satisfied, i.e. $\bar{\phi}_k^T P_{1,k-1} \bar{\phi}_k > 0$, and $\bar{\psi}_k^T P_{2,k-1} \bar{\psi}_k > 0$, according to the expressions of $Y_{1,k}$ and $Y_{2,k}$ stated by (4.21a) and (4.21b), we have

- either $\rho_k \leq \delta$: $Y_{1,k}$ and $Y_{2,k}$ become zero and (4.25) can be rewritten element-wise as follows for the detected active submodel:

$$|v_{k/k}(\hat{z}_k)| = |v_{k/k-1}(\hat{z}_k)| \quad (4.26)$$

which yields

$$|v_{k/k}(\hat{z}_k)| \leq \delta \quad (4.27)$$

- or $\rho_k > \delta$: by substituting $Y_{1,k}$ and $Y_{2,k}$ in (4.25) yields

$$V_{k/k} = V_{k/k-1} \left(I_m - \frac{1}{2} (I_m - \Lambda_k^{-1}) - \frac{1}{2} (I_m - \Lambda_k^{-1}) \right) \quad (4.28)$$

which can be rewritten element-wise for the detected active submodel as follows:

$$v_{k/k}(\hat{z}_k) = \Lambda_k^{-1}(\hat{z}_k) v_{k/k-1}(\hat{z}_k) \quad (4.29)$$

where $\Lambda_k^{-1}(\hat{z}_k)$ denotes \hat{z}_k -th element of matrix Λ_k^{-1} and since $|\Lambda_k^{-1}(\hat{z}_k)| = \frac{\delta}{\rho_k}$, it gives

$$|v_{k/k}(\hat{z}_k)| = \delta \quad (4.30)$$

Therefore, considering the two cases that can happen at each time step and according to (4.27) and (4.30), (4.14) is proved.

4.3. RESULTS AND DISCUSSIONS

4.3.1. NUMERICAL EXAMPLE

A numerical example is considered to assess the accuracy of the prediction using the proposed identification algorithm. The dynamics of this example as a two-mode SBJ system are provided in Table 4.2.

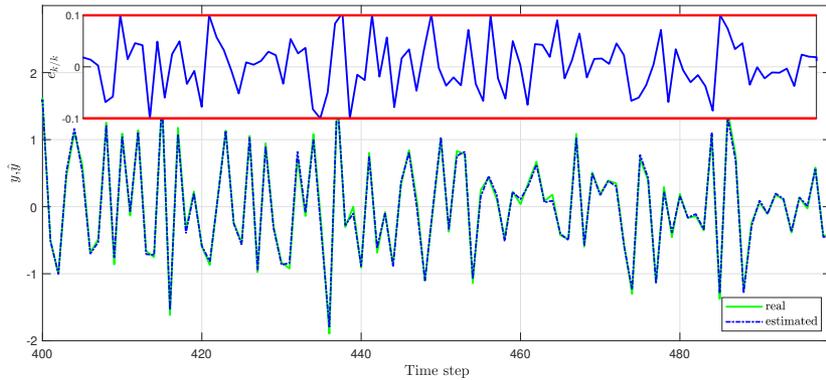
To satisfy the persistent excitation, the input sequence is generated randomly within the range of $[-1, 1]$. The lower and upper bounds of the noise sequence are considered -0.08 and 0.08 , respectively. Therefore, δ as the upper bound of the noise can be taken any value as larger as 0.08 , which it is set to 0.1 in this example. To reach and stay within the assigned bounds, 500 samples of the system are produced and given to the proposed algorithm for the purpose of prediction. The results are plotted for the last 100 samples. As depicted in Figure 4.2 (a), the

Table 4.2: Dynamics of the numerical example; a two-mode SBJ system.

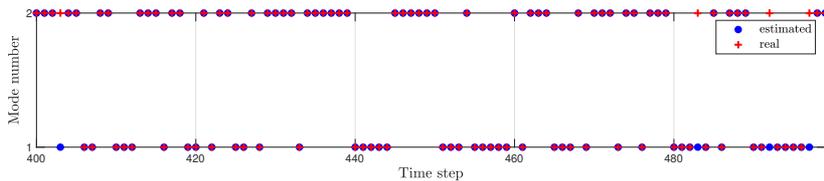
Subsystem dynamics	Subsystem 1	Subsystem 2
$A(q^{-1}, \theta_1 \text{ or } 2)$	$1 + 0.45q^{-1} - 0.2q^{-2}$	$1 - 0.15q^{-1} + 0.35q^{-2}$
$B(q^{-1}, \theta_1 \text{ or } 2)$	$-0.4 + 0.95q^{-1}$	$-0.5 + 1.15q^{-1}$
$C(q^{-1}, \theta_1 \text{ or } 2)$	$1 + 0.64q^{-1}$	$1 - 0.36q^{-1}$
$D(q^{-1}, \theta_1 \text{ or } 2)$	$1 - 0.32q^{-1}$	$1 - 0.50q^{-1}$

estimated output is capable to track the real output within the specified range. Figure 4.2 shows the prediction output and errors and the detection of the switching time instants. Switching instants have been also detected accurately, except at a few steps. To assess the performance of the algorithm, the *FIT* index is considered, which is the percentage fitting error between the true output, y , and the estimated output, \hat{y} , which is 95.2 for the last 100 samples and 88.4 for all the samples.

4



(a) The real system output, y , and the predicted system output, \hat{y} based on the estimated SBJ system. The inner figure shows the posteriori prediction error.



(b) Detection of the switching sequences of the SBJ.

Figure 4.2: Numerical example simulation.

Remark: A few factors can impact the performance and the accuracy of the proposed algorithm. The value of δ that comes from the main constraint of the objective, is one of the major parameters. If it is chosen close to the bound of the system noise, it can numerically destabilize the prediction, while by selecting it

too big, the accuracy is deteriorated. The other important factor is the forgetting procedure. The forgetting procedure is used to reduce the weight of past data and to avoid the matrices $P_{1,k}$ and $P_{2,k}$ from approaching zero, as this can affect the accuracy. Therefore, resetting the parameters $P_{1,k}$ and $P_{2,k}$ in a periodic time interval can affect the accuracy of the prediction, which should be taken into account.

4.3.2. BIOLOGICAL WASTEWATER TREATMENT PROCESSES

A key question in modeling of biological wastewater treatment processes is which modeling approach to choose. Using first principal knowledge to mechanistically drive a model is one of the common and well-known approach. Mechanistic models rely on chemical and biochemical insights and experimental studies, yet they can suffer model mismatch due to potential inaccuracies, occasional perturbations, and varying operational scenarios. Input-output modeling enables an alternative, since it is a data-driven approach. These models can be utilized as prediction models of model-based control systems like model-predictive control, even with the lack of poor interpretability in some cases.

Within input-output modeling approaches, switched system identification is worth exploring, particularly for approximating (highly-)nonlinear complex biological processes. As discussed in the introduction, a few limited real-world applications have been modelled by using simple switched system structures like SARX. Therefore, in this study, we open up a new window for further exploration of input-output switched system identification for the purpose of predictive modeling of biological treatment processes.

For approximating a complex process in the form of input-output models, a critical question arises: "how do we select influential inputs and their corresponding influenced outputs?" Upon this selection, inputs can be categorized as main inputs and disturbances. Taking (1) into account, main inputs are denoted as u , and disturbances as v . By identifying parameters related to their dynamics, represented by $A(\cdot)$, $B(\cdot)$, $C(\cdot)$, and $D(\cdot)$, the relationship between outputs and inputs/disturbances is modelled in a data-driven framework. This chapter sheds light on applications to be modelled using general SBJ models by illustrating this via two examples. Depending on the application, some simple structures would suffice for modeling of the process (Chen *et al.* 2020a; Hartmann *et al.* 2015; Wang *et al.* 2020; Yahya *et al.* 2020). For other cases, more complex structures may be needed.

In this section, we explore the implementation of the proposed prediction method through two wastewater treatment processes; anaerobic fermentation in a continuous stirred-tank reactor (CSTR) and microbial growth of purple phototrophic bacteria (PPB) in a raceway-pond reactor acting as sequencing batch reactor (SBR). Anaerobic fermentation in CSTR is chosen to discuss the importance of using a SBJ model for such a complex bioprocess widely-used in various operational scenarios. PPB biomass cultivation in an SBR is also selected not only because of dynamic complexity, but also for assessment of a potential application of the proposed algorithm in sequencing batch conditions. Moreover, the coupled anaerobic fermentation and purple bacteria raceway-pond reactors for the growth of PPB biomass is a resource recovery process, which has been designed as a pilot

plant in *SARASWATI2.0* project.

Anaerobic fermentation in CSTR: Anaerobic digestion is a multistage complex biological process for converting biodegradable organic matter into biogas through volatile fatty acid (VFA) intermediates in the absence of oxygen (Anukam *et al.* 2019; Batstone *et al.* 2002b). This process can be represented by comprehensive mechanistic models such as ADM1, with a high-degree nonlinearity and stiffness (Batstone *et al.* 2002a). The model, however, is bio- and physio-chemical-structured for the purposes of *process design and understanding*, but it is computationally expensive to use for the purposes of *predictive models* (Ghanavati *et al.* 2021; Kil *et al.* 2017). Its differential-algebraic equation sets consist of time-varying parameters, multiple variables with intricate interconnections, monod-type kinetics, inhibition functions, and competitive uptakes, which are the reasons for the nonlinear behavior. Furthermore, significant fluctuations in both inflow and the composition of incoming wastewater, that do reflect real-world behaviors, perturb both liquid and gas phases characteristics. Input-output system identification for such a typical nonlinear biological model in the framework of switched systems and BJ structure is worth investigating, and as far as authors are aware is reported in literature for the first time in this study.

It is challenging to select input and output variables of the process. As mentioned, output variables can be a function of different variables. As an example, the output to be predicted is chosen *acetate* as the process is fermentation and *acetate* is expected to be the main product of the anaerobic fermentation process. Moreover, prediction of *acetate* is worth considering due to its critical role, especially when the anaerobic digestion is designed for operation in a wider range (Wainaina *et al.* 2019). From a practical point of view, the most influential while easily being manipulating input on production of VFAs is the input flowrate. The flowrate affects the hydraulic retention time, and is one of the most feasible manipulators in terms of process control in practice. However, as mentioned earlier, producing *acetate* does not depend only on inflow. Considering the mechanistic equation describing the dynamic of *acetate* in the ADM1 model (Batstone *et al.* 2002b), its function can be expressed as follows:

$$S_{\text{acetate}} = f(q, X_{\text{lipid}}, X_{\text{protein}}, X_{\text{carbohydrate}}, S_{\text{sugar}}, S_{\text{amino acid}}, S_{\text{fatty acid}}, \dots) \quad (4.31)$$

where S_i and X_i stand for soluble and particulate concentrations of material i , respectively, q denotes and inflow rate. The composition of the influent is considered as disturbance to the process. In practice, the process is usually designed around a specific operating point by monitoring various bioreactor operating parameters. However, perturbations like sudden influent concentration changes may happen any time during operation, playing a role as a disturbance. Therefore, the input-output relations can be represented by a BJ model. It means that disturbances can be integrated in modeling with independent dynamics, which is biologically explainable due to different mechanistic effects between the input and the disturbance to the output. The dynamic between the input flowrate and *acetate* is completely different from the dynamic between other variables and *acetate* as described in the ADM1 model (Batstone *et al.* 2002b). Therefore, considering the schematic of a BJ structure

as depicted in Figure 4.1, dynamics of the disturbance is not the same as dynamics for input.

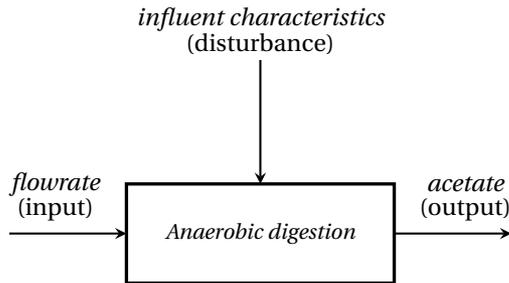
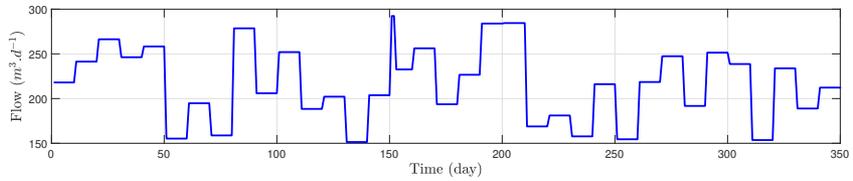


Figure 4.3: Simplified schematization of the anaerobic fermentation process for the purpose of estimation with a SBJ system.

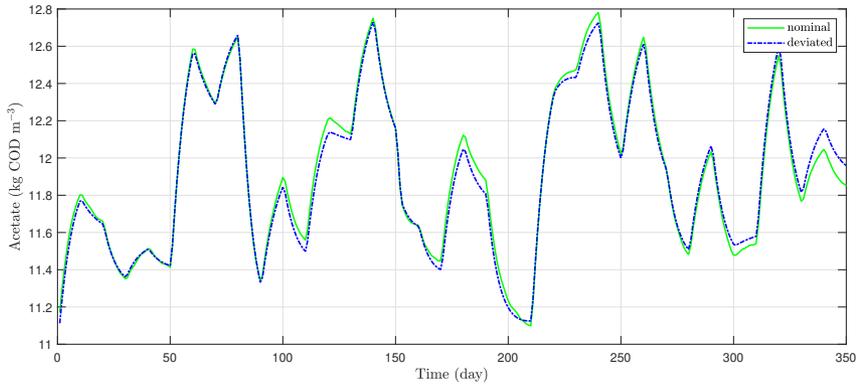
The three main components i.e. carbohydrate, protein, and lipid represents the influent characteristics, which can be considered as the disturbance. They highly impact the process output and are the potential perturbations due to lack of online measurement. Now, the schematic of the process can be drawn in Figure 4.3. The nominal operating condition as given in (Batstone *et al.* 2002a) are considered to generate the dataset, while the reactor environment (the initial conditions) is considered to be acidified at the start-up phase. To explore a wide domain of operation, the process is excited by the input flowrate produced by a pseudo random input signal, depicted in Figure 4.4 (a). The nominal values for carbohydrate, protein, and lipid are 5, 20, and 5 kgCODm^{-3} , respectively, while for fluctuation purposes, a random deviation from the nominal values in a range of $[-0.5, 0.5]$ is assumed. Therefore, the process output deviates from its designated nominal value, as shown in Figure 4.4 (b).

Considering the modeling structure explained above, the proposed algorithm is implemented to identify a parametric SBJ model, given the dataset generated from complex ADM1 model. A few design parameters, therefore, should be assigned. It should be noted that the process is not hybrid by its intrinsic nature and the algorithm is used to capture the dynamics within the designed operating space by a set of linear systems for simplicity for the purpose of prediction, not interpretation. The orders of the SBJ system, therefore, are assigned as *one* for all n_a , n_b , n_c , and n_d . While the higher order may result in higher accuracy, but no amelioration is observed when the complexity is increased. The bound of the disturbance, δ , should be set equal to or bigger than 0.05 due to the assigned range for the disturbance. The process dynamics can be captured accurately ($FIT \approx 95$) by adjusting the two major design parameters for different number of submodels. It is highlighted in *Remark 1* that the value of δ and the forgetting factor play important role for the numerical stability as well as the output accuracy. The effects of these aforementioned parameters on prediction accuracy are investigated in Table 6.1.

A comparison with the conventional two-stage BJ system identification (Ding and Duan 2013) is also made to explore the priority of using a SBJ system instead



(a) The input flowrate produced by a pseudo random input signal used for identification process.



(b) The process output (acetate) in nominal operating condition (green line) and deviated (blue line) by random perturbation within the main components used for identification process.

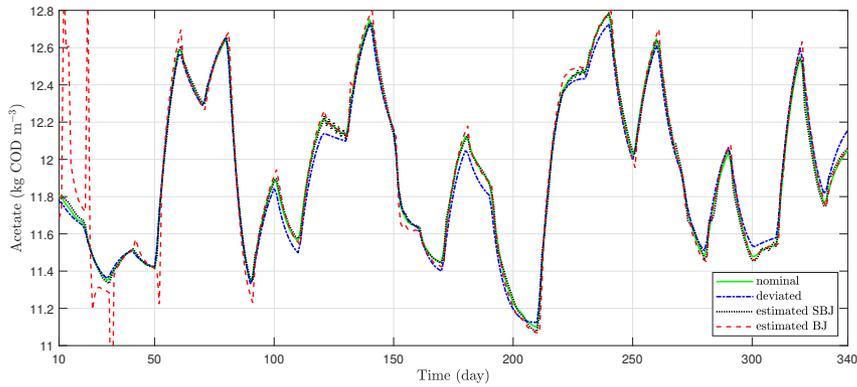
Figure 4.4: Input (flowrate) and output (acetate) of the anaerobic fermentation process.

of a non-switched system. The system orders are chosen the same for the both conventional BJ and SBJ systems. The number of modes and the ellipsoid bound for the SBJ system are assigned to 4 and 0.05, respectively. The initial values and other required parameters are set similarly. For the forgetting factor, a period of 60 days is chosen for this particular application. This setting suffices the need for accurate prediction with the desire for a reasonable rate of convergence. Generally, the proposed SBJ system identification algorithm outperforms the conventional BJ system identification method. The accuracy of the identified SBJ model is better during the whole of the operation and particularly the start-up as shown in Figure 4.5 (a). The OBE algorithm forces the system to stay within the assigned bound by jumping to other mode, while the conventional BJ system cannot keep the output error in the range accurately. As can be seen in Figs 4.5 (a) and (b), the spikes occur, when the direction of the response output is changed, which can be compensated by going to the other submodels in the SBJ system to keep the accuracy within the assigned bound.

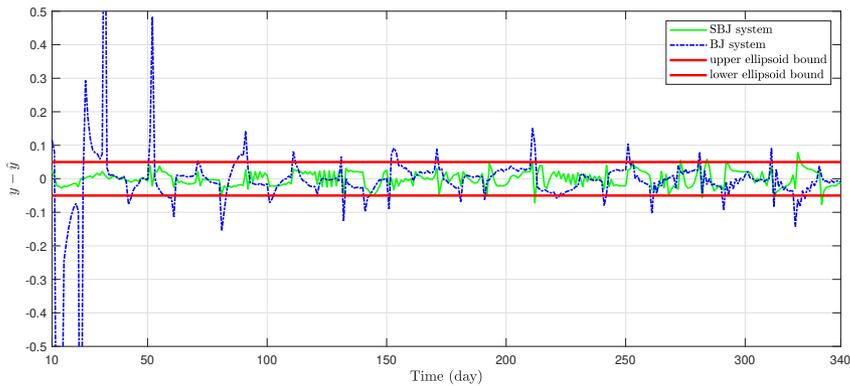
Remark: The anaerobic digestion process is not hybrid by its nature, but a highly nonlinear system. Approximation of the dynamics by using a SBJ model with the OBE algorithm has an advantage of capturing input-output relations with a limited

Table 4.3: Prediction accuracy of anaerobic fermentation process under different scenarios based on the proposed output prediction algorithm.

Number of modes (N)	Ellipsoid bound (δ)	Period of forgetting factor (day)	Accuracy (FIT)
2	0.2	40	94.9157
3	0.1	50	95.7483
4	0.05	60	96.7509
5	0.05	50	97.1826



(a) The process output (acetate) in nominal operating condition (green line) and deviated (blue line) and its output prediction by a SBJ system (black line) and a BJ system (red line).



(b) The error comparison between the output of the estimated SBJ system (green line) and the output of the estimated BJ system (blue line).

Figure 4.5: Prediction performance of the proposed identification algorithm on the anaerobic fermentation process.

number of linear submodels jumping among each other with a desired bound of accuracy in terms of prediction error. Moreover, the other advantage of using BJ structure is identifying different parameters for the moving average part, which is explainable because of different dynamical function of disturbance to output from mechanistic modeling point of view. Furthermore, the type of disturbance as it comes from a nonlinear dynamics in the real system cannot be fitted easily to the conventional stochastic assumption that is relaxed by proposing the developed OBE algorithm.

Growth of PPB biomass in an SBR: Purple phototrophic bacteria (PPB) as a group of microbes for resource recovery from wastewater can be cultivated by cost-effective raceway-pond bioreactors (Alloul *et al.* 2023a). A mechanistic model for PPB in raceway bioreactors has been proposed by Alloul *et al.* (2023b), known as the Purple Bacteria Model (PBM). This type of bioprocesses, i.e. sequencing batch, is selected to assess modeling in the SBJ framework with the proposed OBE algorithm. The cyclic nature of sequencing batch bioreactor operation is regularly applied in conventional wastewater treatment, like for example in aerobic granular sludge technology.

Besides hydraulic and sludge retention times, light also plays a critical role in growth of PPB. In a raceway-pond bioreactor, control over light, more specifically solar radiation, is not practically feasible, due to various hour-by-hour, day-by-day, and seasonal fluctuations. It should be, therefore, considered as a potential disturbance, especially for modeling of an open reactor. Furthermore, distribution of solar radiation is barely representable by the common distribution functions. For instance, illumination durations and radiation angles at a single day are not independent of subsequent days, which may violate the independence assumption required for probability distributions. It is, therefore, another motivation to employ the OBE algorithm for approximation of the process dynamics, since it is not subject to any assumptions for disturbances.

The dynamics of PPB in raceway reactors are also highly nonlinear (Alloul *et al.* 2023b). If the production of PPB is selected as an output to be predicted, flowrate that determines feeding of each sequence is considered as input, while solar irradiation fluctuation that deviates the process from the nominal operating is considered as disturbance. The schematic of an SBJ structure is depicted in Figure 4.6. Considering the mechanistic model proposed by Alloul *et al.* (2023b), PPB production is the function of a wide range of variables with different dynamics. Therefore, defining the problem of approximating this bioreactor in the frame of BJ model is reasonable, due to different dynamics for the input and the disturbance.

To run the PBM model, the following conditions are considered; the sequential batch is designed to feed the reactor once a day at the midnight; influent filling and the effluent extraction are set at midnight, while feeding rate is set to one fourth of the volume per hydraulic retention time; the paddlewheel is considered working only during the light condition. Other operational parameters are set to the default values of the PBM (Alloul *et al.* 2023b).

The solar radiation is subject to fluctuation. Light intensity is depicted in Figure 4.7 (a) from day 21 to 42, when the process reaches steady state. It can be observed, finding a probability distribution is subject to some simplifications that

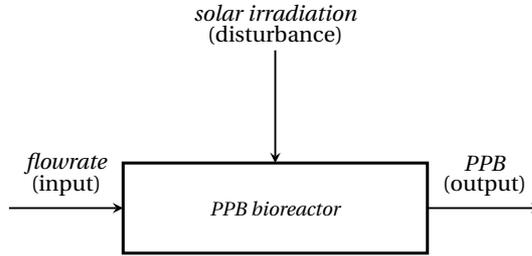


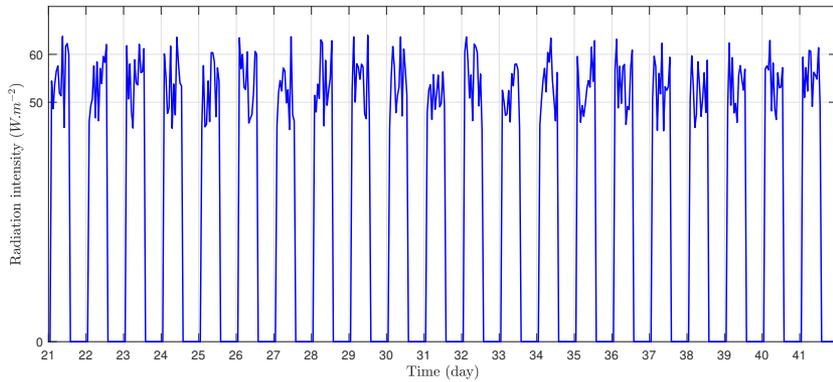
Figure 4.6: Simplified schematization of the purple bacteria raceway-pond photo-bioreactor process for the purpose of estimation with a switched BJ system.

may not be reliable. Therefore, the OBE algorithm that is not subject to probability of disturbance is practically and theoretically more reasonable.

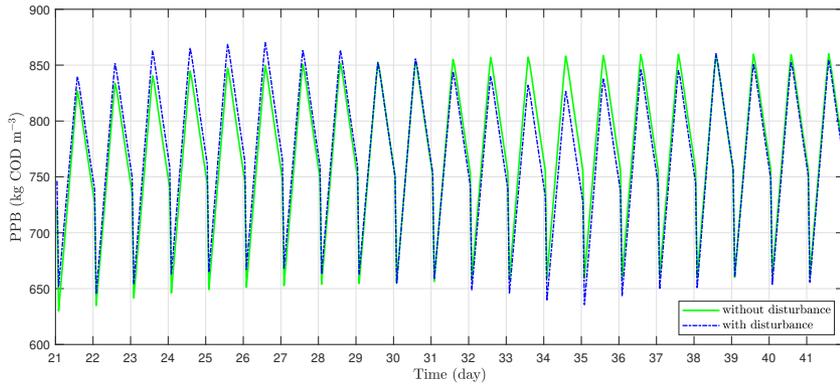
The deviation from nominal process operation with light variation as a disturbance to the operation is shown in Figure 4.7 (b) between day 21 to 42. The output to be predicted is considered purple bacteria produced from the three photoheterotrophic, anaerobic and aerobic chemoheterotrophic pathways. The proposed algorithm is implemented, given the dataset produced. Since the effect of ellipsoid bound and number of modes were investigated in the previous case study, and the same results were observed, the detected switching patterns and its interpretations are explored in this case study.

The orders of the estimated SBJ system are assigned as *one* for all n_a , n_b , n_c , and n_d . The bound of the ellipsoid, δ , the number of modes, and the forgetting period are set to 0.25, 2, and 60 h, respectively and the process behavior is acceptably approximated as depicted in Figure 4.8. Moreover, the switching patterns are shown in Figure 4.8 (b). As can be seen, the time of being in mode one is much longer than mode two. If only the subsystem one is active for prediction, the ellipsoid bound constraint is violated, as shown in Figure 4.8; sub-figures (a) and (b). In other words, using second mode assists the prediction process to stay within the bound.

Remark: Instants of jumping can be explained based on process operating conditions that they occurred around time of extraction, when the light goes off. As described above, biomass removal happens every 24h, and it is replaced by new influent. PPB are produced photoheterotrophically, aerobic and anaerobic chemoheterotrophically. As the reactor is an open system, the amount of PPB grown anaerobic chemoheterotrophically is negligible, while photoheterotrophic growth is the major metabolic growth pathway of PPB, which steadily increases when exposed to solar radiation and decreases when no light is available. A sudden decrease happens on the time extraction, and it is also affected negatively because of the absence of light availability. Therefore, the algorithm needs to switch to keep the accuracy within the assigned bound. In other words, this biomass withdrawal is behaving like a hybrid feature in this example that the algorithm is capable of capturing it.



(a) The solar radiation fluctuation over a 24-hour period, with zero radiation occurring for 12 hours followed by non-zero radiation for the next 12 hours each day.



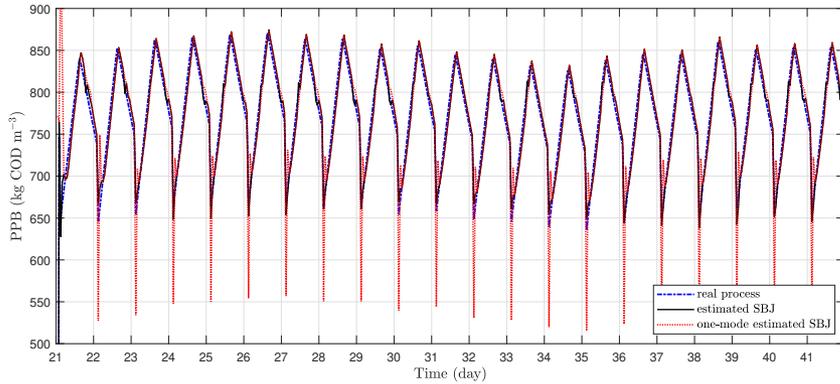
(b) the PPB production for the nominal-designed process (green line) and deviated PPB by disturbance caused by light intensity fluctuation (blue line).

Figure 4.7: Implemented disturbance (solar radiation) and observed output (PPB) of the raceway-pond photobioreactor.

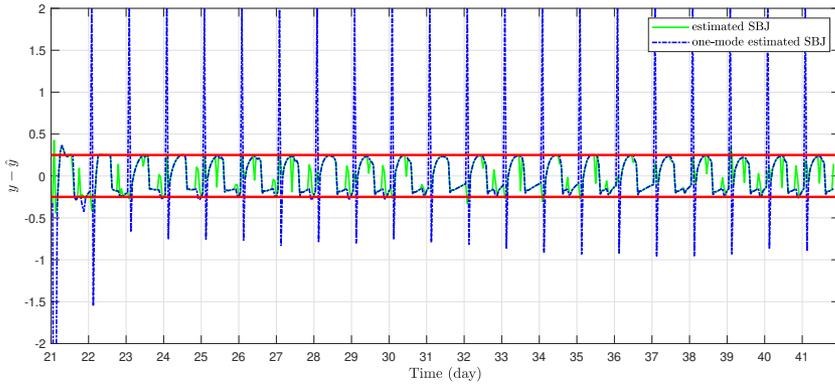
4.4. LIMITATIONS OF THE PROPOSED APPROACH AND FURTHER WORK

This chapter illustrates how SBJ models can be formulated for biological wastewater treatment process models by analyzing two ADM1 and PBM models. Depending on the application, some simple structures would suffice for process modeling (Chen *et al.* 2020a; Hartmann *et al.* 2015; Wang *et al.* 2020; Yahya *et al.* 2020). For other cases, more complex structures like SBJ may be more meaningful, as different dynamics could be fitted to represent the relation between disturbances and outputs.

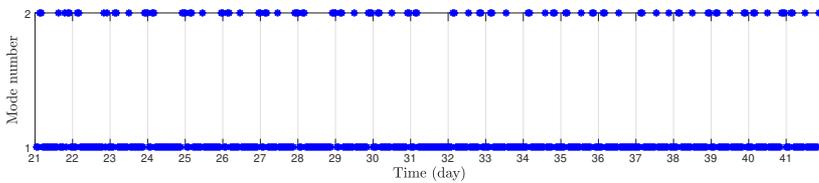
The identification algorithm used does not require an assumption on statistical



(a) The process output (blue line) and its output prediction by a SBJ system (black line) and the same estimated system with one mode, where more frequently mode is considered than the other one (red line).



(b) The error comparison between the output of the estimated SBJ system (green line) and the output of the estimated SBJ system with one mode, where that mode occurred more frequently is considered than the other (blue line).



(c) The switching patterns (mode occurrence) of two-mode estimated SBJ system.

Figure 4.8: Prediction performance of the proposed identification algorithm on the PPB photobioreactor.

distribution for disturbances, and only has the less strict assumption that they are bounded. Nonetheless, the proposed method is built upon an approach that needs a few design parameters influencing the accuracy of prediction. These parameters discussed in Remark 3 can be determined through trial and error simulations. Moreover, preprocessing of a dataset for some cases may be required to avoid numerical issues.

As a future research, the algorithm can be extended for processes that require a multiple inputs and multiple outputs system representation. Parametrizing the switching domain in the form of polyhedral partitions for better interpretation of switching behavior may also be considered as another extension, specially for biological wastewater treatment processes

4.5. CONCLUSIONS

In this chapter, the application of switched Box-Jenkins systems is investigated in the context of modeling biological treatment processes, using two widely-utilized complex models for benchmarking model performance, i.e. ADM1 and PBM. An identification method is introduced by extending the OBE identification algorithm for switched Box-Jenkins models. The algorithm builds upon the standard OBE approach as its foundation, eliminating the need for the assumption that a probability distribution of disturbances exists and relying solely on the assumption of bounded disturbances. This feature is particularly valuable in practical scenarios of treatment processes, where such distributions might not even be available due to unpredictable fluctuations. To tackle the mathematical challenges arising from the SBJ structure and its inner signals, we employ a decomposition technique. The resulting algorithm is recursive, enabling real-time data processing. This attribute makes it well-suited for systems dealing with extensive data volumes. The results underscore the algorithm's capacity to yield accurate predictions, thereby highlighting its potential for real-world implementation for biological treatment processes.

REFERENCES

- Alloul, A., N. Blansaer, P. C. Segura, R. Wattiez, S. E. Vlaeminck, and B. Leroy (2023a). “Dehazing redox homeostasis to foster purple bacteria biotechnology”. In: *Trends in Biotechnology* 41.1, pp. 106–119.
- Alloul, A., A. Moradvandi, R. M. Daniel Puyol, G. Gardella, S. E. Vlaeminck, B. D. Schutter, E. Abraham, R. E. F. Lindeboom, and D. G. Weissbrodt (2023b). “A novel mechanistic modelling approach for microbial selection dynamics: towards improved design and control of raceway reactors for purple bacteria”. In: *Submitted to Bioresource Technology*. DOI: 10.5281/zenodo.8277031.
- Anukam, A., A. Mohammadi, M. Naqvi, and K. Granström (2019). “A Review of the Chemistry of Anaerobic Digestion: Methods of Accelerating and Optimizing Process Efficiency”. In: *Processes* 7.8, p. 504.
- Batstone, D., J. Keller, I. Angelidaki, S. Kalyuzhnyi, S. Pavlostathis, A. Rozzi, W. Sanders, H. Siegrist, and V. Vavilin (2002a). “The IWA Anaerobic Digestion Model No 1 (ADM1)”. In: *Water Science and Technology* 45.10, pp. 65–73.
- Batstone, D. J., J. Keller, I. Angelidaki, S. Kalyuzhnyi, S. Pavlostathis, A. Rozzi, W. Sanders, H. a. Siegrist, and V. Vavilin (2002b). “The IWA anaerobic digestion model no 1 (ADM1)”. In: *Water Science and technology* 45.10, pp. 65–73.
- Bianchi, F., V. Breschi, D. Piga, and L. Piroddi (2021). “Model structure selection for switched NARX system identification: A randomized approach”. In: *Automatica* 125, p. 109415.
- Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Chai, X., H. Wang, X. Ji, and L. Wang (2020). “Identification of switched linear systems based on expectation-maximization and Bayesian algorithms”. In: *Transactions of the Institute of Measurement and Control* 43.2, pp. 412–420.
- Chen, X., S. Zhao, and F. Liu (2020a). “Identification of jump Markov autoregressive exogenous systems with missing measurements”. In: *Journal of the Franklin Institute* 357.6, pp. 3498–3523.
- (2020b). “Online identification of time-delay jump Markov autoregressive exogenous systems with recursive expectation-maximization algorithm”. In: *International Journal of Adaptive Control and Signal Processing* 34.3, pp. 407–426.
- Ding, F and H. Duan (2013). “Two-stage parameter estimation algorithms for Box–Jenkins systems”. In: *IET Signal Processing* 7.8, pp. 646–654.
- Ding, F, P. X. Liu, and G. Liu (2010). “Gradient based and least-squares based iterative identification methods for OE and OEMA systems”. In: *Digital Signal Processing* 20.3, pp. 664–677.
- Du, Y., F. Liu, J. Qiu, and M. Buss (2020). “A Semi-Supervised Learning Approach for Identification of Piecewise Affine Systems”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 67.10, pp. 3521–3532.
- Du, Z., L. Balzano, and N. Ozay (2018). “A Robust Algorithm for Online Switched System Identification”. In: *IFAC-PapersOnLine* 51.15, pp. 293–298.
- Garulli, A., S. Paoletti, and A. Vicino (2012). “A survey on switched and piecewise affine system identification”. In: *IFAC Proceedings Volumes* 45.16, pp. 344–355.
- Ghanavati, M. A., E. Vafa, and M. Shahrokh (July 2021). “Control of an anaerobic bioreactor using a fuzzy supervisory controller”. In: *Journal of Process Control* 103, pp. 87–99.
- Goudjil, A., M. Poulliquen, E. Pigeon, and O. Gehan (2016). “A real-time identification algorithm for switched linear systems with bounded noise”. In: *2016 European Control Conference*.
- (Jan. 2023). “On-line outer bounding ellipsoid algorithm for clustering of hyperplanes in the presence of bounded noise”. In: *Cluster Computing*.
- Goudjil, A., M. Poulliquen, E. Pigeon, O. Gehan, and B. Targui (2017). “Recursive Output Error Identification Algorithm for Switched Linear systems with Bounded Noise”. In: *IFAC-PapersOnLine* 50.1, pp. 14112–14117.
- Hartmann, A., J. M. Lemos, R. S. Costa, J. Xavier, and S. Vinga (2015). “Identification of switched ARX models via convex optimization and expectation maximization”. In: *Journal of Process Control* 28, pp. 9–16.
- Hojjatnia, S., C. M. Lagoa, and F. Dabbene (2020). “Identification of switched autoregressive exogenous systems from large noisy datasets”. In: *International Journal of Robust and Nonlinear Control* 30.15, pp. 5777–5801.
- Kil, H., D. Li, Y. Xi, and J. Li (2017). “Model predictive control with on-line model identification for anaerobic digestion processes”. In: *Biochemical Engineering Journal* 128, pp. 63–75.
- Lauer, F and G. Bloch (2018). “Hybrid system identification: Theory and algorithms for learning switching models, vol. 478”. In: *Cham, Switzerland: Springer*.
- Liu, X., X. Yang, and M. Yu (2021). “Identification of switched FIR systems with random missing

- outputs: A variational Bayesian approach". In: *Journal of the Franklin Institute* 358.1, pp. 1136–1151.
- Ljung, L. (1999). *System Identification - Theory for the User*. Prentice-Hall.
- Mazzoleni, M., V. Breschi, and S. Formentin (2021). "Piecewise nonlinear regression with data augmentation". In: *IFAC-PapersOnLine* 54.7, pp. 421–426.
- Moradvandi, A., R. E. F. Lindeboom, E. Abraham, and B. D. Schutter (2023). "Models and methods for hybrid system identification: a systematic survey". In: *Accepted by IFAC-PapersOnLine*.
- Ozbay, B., O. Camps, and M. Sznaier (2019). "Efficient Identification of Error-in-Variables Switched Systems via a Sum-of-Squares Polynomial Based Subspace Clustering Method". In: *2019 IEEE 58th Conference on Decision and Control*.
- Piga, D., A. Bemporad, and A. Benavoli (2020a). "Rao-Blackwellized sampling for batch and recursive Bayesian inference of Piecewise Affine models". In: *Automatica* 117, p. 109002.
- Piga, D., V. Breschi, and A. Bemporad (2020b). "Estimation of jump Box-Jenkins models". In: *Automatica* 120, p. 109126.
- Song, C., J. Wang, X. Ma, and J. Zhao (2020). "A PWA model identification method based on optimal operating region partition with the output-error minimization for nonlinear systems". In: *Journal of Process Control* 88, pp. 1–9.
- Wainaina, S., Lukitawesa, M. K. Awasthi, and M. J. Taherzadeh (2019). "Bioengineering of anaerobic digestion for volatile fatty acids, hydrogen or methane production: A critical review". In: *Bioengineered* 10.1, pp. 437–458.
- Wang, D. (2011). "Least squares-based recursive and iterative estimation for output error moving average systems using data filtering". In: *IET Control Theory & Applications* 5.14, pp. 1648–1657.
- Wang, D., Y. Chu, G. Yang, and F. Ding (2010). "Auxiliary model based recursive generalized least squares parameter estimation for Hammerstein OEAR systems". In: *Mathematical and Computer Modelling* 52.1-2, pp. 309–317.
- Wang, J., C. Song, J. Zhao, and Z. Xu (2020). "A PWA model identification method for nonlinear systems using hierarchical clustering based on the gap metric". In: *Computers and Chemical Engineering* 138, p. 106838.
- Yahya, O., Z. Lassoued, and K. Abderrahim (2020). "Identification of PWARX Model Based on Outer Bounding Ellipsoid Algorithm". In: *2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering*.

5

MPC SYSTEM FOR PPB RACEWAY REACTORS

Purple Phototrophic Bacteria (PPB) are increasingly being applied in resource recovery from wastewater. Open raceway-pond reactors offer a more cost-effective option, but subject to biological and environmental perturbations. This study proposes a hierarchical control system based on Adaptive Generalized Model Predictive Control (AGMPC) for PPB raceway reactors. The AGMPC uses simple linear models updated adaptively to project the complex process dynamics and capture changes. The hierarchical approach uses the AGMPC controller to optimize PPB growth as the core of the system. The developed supervisory layer adjusts set-points for the core controller based on two operational scenarios: maximizing PPB concentration for quality, or increasing yield for quantity through effluent recycling. Lastly, due to competing PPB and non-PPB bacteria during start-up phase, an override strategy for this transition is investigated through simulation studies. The Purple Bacteria Model (PBM) simulates this process, and simulation results demonstrate the control system's effective and robustness.

This chapter is an adapted version of *model predictive control of purple bacteria in raceway reactors: handling microbial competition, disturbances, and performance*, Ali Moradvandi, et al. (under review), *Computers and Chemical Engineering*.

5.1. INTRODUCTION

Cultivation of Purple Phototrophic Bacteria (PPB) has been gaining importance as a promising alternative for microalgae for nutrient and resource recovery in general. The beneficial aspect of nutrient recovery, coupled with the assimilation of Chemical Oxygen Demand (COD) in wastewater treatment, positions PPB as a viable solution for industrial wastewater resources, consequently contributing to the fertilizer-feed-food-fork chain (Capson-Tojo *et al.* 2020). It reflects the increasing interests into the application of PPB under various growth conditions over the last decade (Capson-Tojo *et al.* 2023a; Cerruti *et al.* 2020; Hulsen *et al.* 2016).

Owing to PPB's metabolic versatility, they can use a broad range of organic compounds for growth, both in the presence and absence of light (photoheterotrophic and chemoheterotrophic grown) and oxygen (aerobic and anaerobic conditions) (Capson-Tojo *et al.* 2020). PPB cultivation has proven effective in anaerobic closed photobioreactors. Puyol *et al.* (2017) have discussed the mechanistic growth metabolisms of PPB in this type of reactor environments and have proposed the Photo-Anaerobic Model (PanM). While closed controlled systems like membrane and tubular photobioreactors and illuminated stirred reactors offer ideal conditions for maximizing PPB microbial selectivity, open raceway-pond reactors require lower capital and operational expenses (Alloul *et al.* 2021). The PanM accurately represents PPB performances for controlled reactors in research labs, however, the extended PanM (ePanM) (Capson-Tojo *et al.* 2023b) and Purple Bacteria Model (PBM) (Alloul *et al.* 2023) are not limited to photo-anaerobic conditions by taking diverse metabolic capabilities of PPB across various varying environmental conditions into account. More specifically, the PBM mechanistically represents PPB growth in a sequencing batch configuration of raceway reactors, and has been calibrated for alternating aerobic and anaerobic conditions as well as various metabolic growth pathways of PPB (Alloul *et al.* 2023).

Although raceway reactors are potentially cost-effective industrial options for scale-up, biomass growth productivity can be easily perturbed due to limited control over operating conditions (de Andrade *et al.* 2016). Therefore, implementing automatic control systems can be a solution for ensuring bioreactor robustness against inevitable operational variations. In recent years, control of microalgae biomass production in tubular and raceway reactors have been studied by proposing various advanced control strategies, like linear active disturbance rejection control (Carreño-Zagarra *et al.* 2019), hierarchical optimization-based control (Fernández *et al.* 2016), and learning-based model predictive control (Pataro *et al.* 2023). These methods have been tailored for microalgae and their dynamics and metabolisms, while PPB responds differently to environmental conditions owing to their different and highly versatile metabolisms, like ability for high yield on organic carbon sources and utilization of the near infrared light spectrum. Alloul *et al.* (2019) have shown that efficient PPB production can be achieved utilizing fermented wastewater that is enriched in volatile fatty acids (VFAs). They have also experimentally investigated various operational strategies impacting the PPB growth in raceway reactors (Alloul *et al.* 2021).

The mechanistical understanding acquired through the aforementioned investiga-

tions has been incorporated in the PBM. Basically, the system configuration is made on a fixed daily sequencing batch through a natural 12 h dark and 12 h light regime, fed by VFAs, while taking the stirring effect of the paddle wheel into account. The growth of PPB encompasses three main pathways: photo-, aerobic, and anaerobic chemotrophic metabolisms. Observations have shown that due to the constant transition between dark and light conditions, all these pathways can contribute to PPB growth, even when a specific operational condition seems predominant (Alloul *et al.* 2021). This metabolic-mechanistic switch (Alloul *et al.* 2023) has been incorporated into the PBM by introducing an empirical constant for parallel metabolic growth (Alloul *et al.* 2023). Although this parameter can be fine-tuned through dedicated experiments, it remains a significant source of model mismatch. Furthermore, the open reactor environment of a raceway pond is conducive for microbial competition between PPB and non-PPB, when PPB are not the dominant trophic group. From the work by Alloul *et al.* (2021), as a consequence of being exposed to air with fluctuating diffusion causing by paddle wheel operation, the low but alternating dissolved oxygen concentration, seem to be particularly important. Light intensity and wavelength are other factors affecting PPB growth in raceway reactors that would affect the growth if not controlled (Cerruti *et al.* 2022).

Thus, design of a control system to optimize reactor performance under these challenging conditions is desired. As discussed above, such a control system would need to enhance the stability and efficiency of PPB cultivation under complex biological dynamics and meteorological fluctuations. Advanced control strategies like model predictive control (MPC) has shown its applicability and credibility for various biological wastewater treatment systems (Gupta *et al.* 2022; Han *et al.* 2021). The adaptive version of Generalized Model Predictive Control (GMPC) (Clarke *et al.* 1987) presents a control strategy for processes with intricate dynamics, such as the sophisticated microbial dynamics of PPB as described by the PBM. This approach involves simplifying the complex system into input-output dynamics and continuously updating parameters to mimic the evolving behavior of the actual process under varying operational conditions, uncertainties, and perturbations. Furthermore, the continuous changes in operational conditions pose the challenge of adapting the set-point to optimize process performance throughout the operation, which can be effectively tackled through a hierarchical control strategy assigning an appropriate set-point (Ghanavati *et al.* 2021; Sadeghassadi *et al.* 2018).

According to the authors' best knowledge, advanced control of PPB-based raceway reactor has not reported in the existing literature. Therefore, this chapter introduces a control configuration for a PPB-based raceway reactor. The primary controller is based on Adaptive GMPC (AGMPC), and a supervisory layer is responsible for determining an appropriate set-point given an operational decision strategy and current process status. An operational decision is made based on either a water quality-driven scenario, which reduces effluent VFA as much as possible to increase PPB concentration as well as treatment efficiency, or a quantity-driven scenario, which increases the production rate, and thereafter the yield, by recycling unconverted effluent VFA. Additionally, an override control strategy is integrated into the system to facilitate the transition from the start-up phase to the PPB-dominant

phase. The proposed control strategy is operationally advantageous as it makes it possible to do the following:

- Assigning appropriate time varying set-points for PPB concentration, employing a supervisory layer to determine based on two operational scenarios, i.e. quality-driven and quantity-driven under varying operational scenarios.
- Maintaining PPB concentration at the desired set-point under different illumination scenarios and light perturbations.
- Maintaining PPB concentration at the desired set-point even when parallel metabolic growth constant is unknown and the contribution of the different metabolic PPB pathways (photoheterotrophic, anaerobic, and aerobic chemoheterotrophic) to the overall PPB growth cannot be quantified.
- Suppressing the growth of other competing bacterial species, enabling moving towards the desired set-point for PPB concentration under non-steady-state conditions more swiftly (i.e. start up phase) using override phase-based control that regulates the paddlewheel.
- The proposed applied controller, based on an adaptively updated linear input-output model, ensures a low computational burden, and utilizes available measurements to effectively capture process variations and disturbances at each time step.

The chapter is organized as follows. Section 5.2 includes the PPB process description and the corresponding control challenges to be addressed. Section 6.3 presents the PPB control system by discussing the control configuration for PPB-based raceway reactor integrating adaptive generalized model predictive control, override phased-based control, and decision-making supervisory layer. Finally, the proposed control strategy is assessed via comprehensive simulation studies in Section 5.4, and in the last section, conclusions are drawn.

5.2. PPB PROCESS DESCRIPTION

A first-principle model, describing a biological wastewater treatment process, is a valuable tool to design, optimize, and control a process. Purple phototrophic bacteria (PPB) dynamics can be mechanistically represented by the Purple Bacteria Model (PBM) (Alloul *et al.* 2023). The PBM is the extended model based on the Photo-anaerobic model (PAnM) (Puyol *et al.* 2017) and the extended Photo-Anaerobic Model (ePAnM) (Capson-Tojo *et al.* 2023b) for growth of PPB in open raceway-pond reactors. The PBM thus serves as a reliable benchmark to analyze the PPB dynamics, considering the complexity of microbial versatility of PPB as well as competition between PPB and non-PPB. Therefore, in this work, it will be used as a benchmark to simulate the growth of PPB in a raceway-pond reactor and assess the performance of the proposed control system. In this section, a few notable behaviors of the process are described with respect to the PBM, which should be taken into account for designing a control system and assessing its performance.

5.2.1. METABOLIC VERSATILITY OF PPB

The PBM describes the PPB's microbial versatility among the photoheterotrophic ($X_{PB,ph}$), and both anaerobic ($X_{PB,anc}$) and aerobic chemoheterotrophic ($X_{PB,aec}$) growth of PPB. This mechanistic-metabolic microbial dynamical selection is modelled through an empirical constant called the parallel metabolic growth constant (M_S). This factor is responsible for the contribution of alternative pathways to PPB growth alongside the dominant pathway, resulting in model mismatches. These mismatches stem from its variations during operation, transitions between light and dark conditions, and the challenge of precisely determining the constant empirically through timely experiments (Alloul *et al.* 2021). Variations of M_S result in different values of PPB concentration during operation, as all these pathways can contribute to PPB growth.

5.2.2. PPB COMPETITORS

In addition to PPB, non-PPB are considered within the microbial biomass of the PBM. Non-PPB are divided into aerobic bacteria (X_{AEB}) and anaerobic bacteria (X_{ANB}). Since the raceway-pond reactor is an open system, aerobic bacteria are the main competitor of PPB. This competition can impact control performance, especially during the start-up phase when PPB concentration is not dominant. Although, the oxygen concentration in raceway reactors is nearly zero, using the paddlewheel to pass oxygen through the bulk, it affects the competition between PPB and non-PPB, particularly when PPB are not the dominant species.

5.2.3. LIGHT IRRADIANCE, ATTENUATION, AND DISTRIBUTION

Light is a crucial input factor to support the phototrophic growth of PPB. Light intensity is considered constant during daylight times; this assumption can be reliable if an artificial illumination system is used (Cerruti *et al.* 2022). Otherwise, the controller should be able to deal with a Gaussian-like illumination intensity that represents the real-world scenario in which the circadian rhythm is perturbed with cloud formation. Therefore, meteorological fluctuations and incoming suspended solids may disturb light distribution and attenuation.

5.2.4. MAXIMUM YIELD OF PPB

PPB is metabolically capable of using energy and carbon sources to grow (Imhoff 2006). In other words, light as an energy source and chemical oxygen demand (COD) in wastewater as a carbon source are required to efficiently cultivate PPB. In this sense, fermented wastewater including mostly volatile fatty acids (VFAs) has been considered by Alloul *et al.* (2019) and Capson-Tojo *et al.* (2020) as favorable carbon sources for PPB microbial selectivity. Depending on the availability of varying amounts of these two sources, the maximum yield achievable during operation can vary. Therefore, the control system should be designed in such way that it makes best use of available sources, subject to fluctuations, to enhance the process performance.

5.3. PPB CONTROL SYSTEM

In this section, a step-by-step design of an advanced control system aimed at tackling the mentioned control challenges for PPB utilization in raceway reactors is discussed. The primary control objective is to regulate the concentration of PPB during operation, subject to biological and meteorological fluctuations. Among the advanced control strategies, model predictive control (MPC), which has demonstrated its efficiency and applicability in various biological wastewater treatment processes, is selected as the core of the control system (Ghanavati *et al.* 2021; Han *et al.* 2021). To improve efficiency, a supervisory layer is also developed, accounting for an appropriate set-point to be assigned for the MPC controller based on two operational scenarios. An override control strategy is also proposed for the condition when the PPB are not dominant. The developed control system for the raceway reactor is illustrated in Figure (5.1). It includes three main components: (i) phased-based controller: it serves as an override control mechanism, facilitating the transition to the MPC controller for PPB concentration during the start-up operation; (ii) main controller: this component is dedicated to regulating PPB concentration and to manage process uncertainty and potential disturbances effectively; and (iii) decision-making supervisory layer: it acts as a supervisory layer to assign an appropriate PPB set-point concentration based on the preferred operational strategy and the process condition. In the following, the adaptation of the control architecture based on MPC for PPB cultivation in a raceway-pond reactor, and developing the supervisory layer and the override control strategy will be discussed.

5.3.1. CONTROL OF PPB RACEWAY REACTORS

PPB raceway reactors are modelled as a sequential batch process with daily cycles of filling and extracting the reactor with influent and effluent, respectively. As discussed by Alloul *et al.* (2021), a favorable operational strategy is 12h light/12h dark condition with 24h stirring, where the reactor is fed by the VFA-based medium before the start of the light condition. To maintain a constant reactor volume, the feeding and extraction rates are kept equal. From an automatic control point of view, practical manipulated variables include the concentration and the flowrate of influent. If the concentration of the incoming medium is assumed to be constant, the feeding flowrate is the feasible control action to regulate PPB concentration.

Given the operational conditions of the raceway reactor, the significance of employing MPC becomes evident. With the reactor being fed once a day and the complex behavior of microorganisms characterized by long response times, making predictions over a horizon and controlling the process accordingly becomes crucial. Therefore, while the simulation (process) time step is an hour, the controller time step is a day (24h). As depicted in Figure 5.2, at time step $24k$, where k is an integer value, measured PPB concentrations (measured X_{PB}) and implemented feeding rates (past u) at past times like $24(k-1)$, $24(k-2)$, $24(k-3)$, etc. are utilized to predict PPB concentrations over a prediction horizon (N_p) and calculate planned control action over a control horizon (N_u) accordingly. This concept is similar to event-based MPC (Pawlowski *et al.* 2014, 2012), where the event is fixed in this work. This concept also allows for sufficient time to determine PPB concentration daily

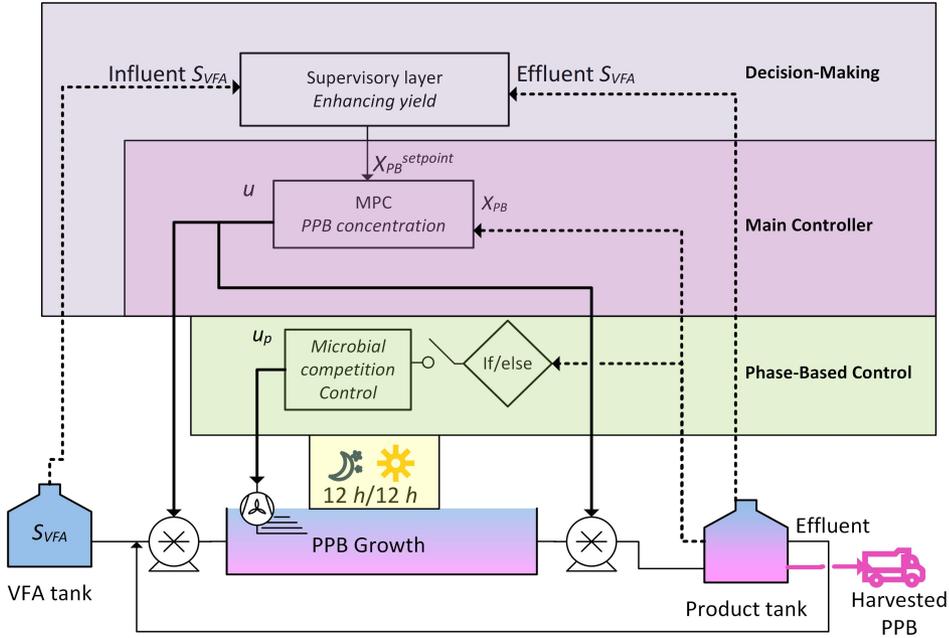


Figure 5.1: MPC-based control system architecture for a raceway reactor. The figure illustrates the three main components of the control system: (i) the phased-based controller for transitioning to the MPC controller, (ii) the main controller for regulating the PPB concentration while handling uncertainties and disturbances, and (iii) the decision-making supervisory layer for assigning the PPB set-point concentration based on a preferred operational strategy. Dashed lines represent variables to be measured (X_{PB} as the primary controlled variable, and influent and effluent VFAs for the supervisory layer). Solid bold lines represent manipulated variables: u , which adjusts the speed of the inflow and outflow pumps (thus controlling the flow rate), and u_p , which controls the paddlewheel.

with an off-line spectroscopic measurement combined with conventional TSS/VSS monitoring if real-time monitoring is not available (Cerruti *et al.* 2020).

5.3.2. ADAPTIVE GPC ALGORITHM: MAIN CONTROLLER

Model predictive control is a model-based control strategy. Although the mechanistic PBM model provides detailed process dynamics, using it as the base model for an MPC controller presents significant challenges. Due to its complex biological characteristics and integrated structure, the PBM model is highly nonlinear. As a result, employing such a large model for MPC controller design leads to a nonlinear non-convex optimization problem that must be solved at each control step, causing computational complexities and a heavy computational burden (Ahmed

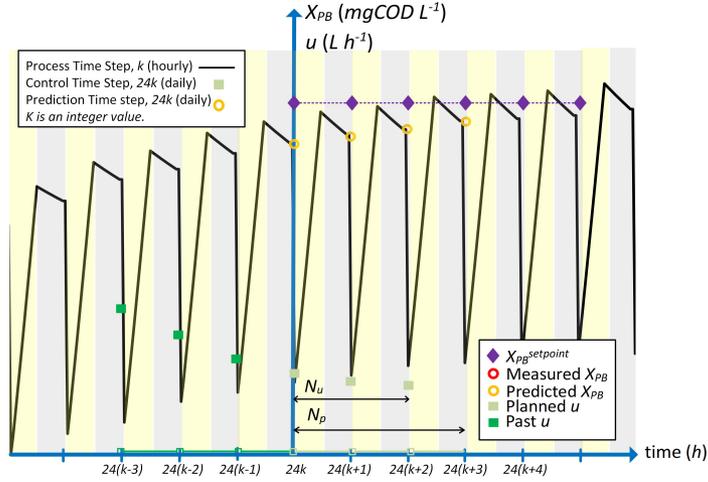


Figure 5.2: Schematic representation of the raceway reactor operation and the integration of model predictive control: hourly process time step vs daily control time step.

and Rodríguez 2020). Furthermore, designing an MPC controller based on this model requires either measurements or estimations of every state variable at each control step. Measuring all the state variables is economically and practically unfeasible (Dochain 2013), while developing a state estimator is also challenging, particularly for PPB states that grow through different pathways ($X_{PB,ph}$, $X_{PB,anc}$, and $X_{PB,aec}$). Additionally, the effectiveness of MPC relies on the accuracy of the model, but the PBM model is susceptible to potential mismatches, such as those related to the parallel PPB growth constant (M_S). Therefore, an input-output model is employed in this work to characterize the relationship between the feeding flow rate and the PPB concentration. To capture the variations of the process such as those related to the nature of the process like parallel growth pathways as well as external disturbances, an adaptive version of the input-output model is employed. This adaptive approach allows the model parameters to be updated based on new sets of observations, ensuring accurateness of predictions as well as robustness of the controller against biological and meteorological variations.

Given the input-output model as the basis of MPC, the generalized model predictive control (GMPC) algorithm can be used as a feedback controller (Clarke *et al.* 1987). In this method, the GMPC controller calculates the control actions over a control horizon (N_u) that minimizes a cost function based on a prediction horizon (N_p). The cost function, J , is defined as follows:

$$J = \sum_{j=1}^{N_p} \delta [\hat{y}(k+j|k) - w(k+j)]^2 + \sum_{j=1}^{N_u} \lambda [\Delta u(k+j-1)]^2, \quad (5.1)$$

where $\hat{y}(k+j)$, $w(k+j)$, and $\Delta u(k+j)$ denote the j -step ahead prediction on

data up to time step k , the future set-point trajectory, and the planned control increments, respectively. Moreover, δ and λ are the controller design parameters representing the error and the control weighting factors. The predicted output, \hat{y} , of the actual output, y , over the prediction horizon N_p is obtained by a single-input single-output discrete time linear model as follows:

$$A(q^{-1})y(k) = B(q^{-1})u(k) + \epsilon(k)/\Delta, \quad (5.2)$$

in which

$$\Delta = 1 - q^{-1}, \quad (5.3)$$

where ϵ denotes zero mean white noise, and $A(q^{-1})$ and $B(q^{-1})$ are the linear models. These linear models are the rational functions of the time shift operator q^{-1} (i.e. $q^{-d}x_k = x_{k-d}$ for $d \in \mathbb{Z}$) that can be written as follows:

$$A(q^{-1}) = 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a}, \quad (5.4a)$$

$$B(q^{-1}) = b_0 + b_1q^{-1} + \dots + b_{n_b}q^{-n_b}, \quad (5.4b)$$

in which n_a and n_b express the order of the system with respect to the outputs and inputs, respectively. Since, the adaptive version of the GMPC controller is considered to tackle with improper future predictions, the parameters of model (6.6) should be updated. If we consider $\theta = [a_1, \dots, a_{n_a}, b_0, \dots, b_{n_b}]^T$ as the vector of the linear model coefficients, the online estimation of this parameter vector at time step k , i.e. $\hat{\theta}(k)$, can be derived using the least-squares method as follows:

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{P(k-1)\phi^T(k)}{1 + \phi^T(k)P(k-1)\phi(k)}(y(k) - \hat{y}(k)), \quad (5.5)$$

where $\phi(k)$ is the augmented vector of past input and output observations, $P(k)$ is the covariance matrix, and $\hat{y}(k)$ is the prediction output. The identification process can be written as follows:

$$\phi(k) = [y(k-1), \dots, y(k-n_a), u(k), \dots, u(k-n_b)]^T, \quad (5.6a)$$

$$P(k) = P(k-1) + \frac{P(k-1)\phi^T(k)\phi(k)P(k-1)}{1 + \phi^T(k)P(k-1)\phi(k)}, \quad (5.6b)$$

$$\hat{y}(k) = \phi^T(k)\hat{\theta}(k-1). \quad (5.6c)$$

Given this adaptive model, the minimization of the cost function J expressed by (6.5) can be explicitly derived, assuming no constraints on the control signals (Camacho *et al.* 2007). Therefore, if \mathbf{f} is defined as the free response of the process, the optimal vector of the planned control actions, i.e. \mathbf{u} can be written as

$$\mathbf{u} = (\mathbf{G}^T\mathbf{G} + \lambda\mathbf{I})^{-1}\mathbf{G}^T(\mathbf{w} - \mathbf{f}), \quad (5.7)$$

where \mathbf{G} , \mathbf{I} , and \mathbf{w} express the step response matrix of the system, the identity matrix, and the vector of the future set-points, respectively. However, the actual control action sent to the process, is the first element of the vector \mathbf{u} that can be written as

$$\Delta u = \mathbf{K}(\mathbf{w} - \mathbf{f}), \quad (5.8)$$

where \mathbf{K} denotes the first row of the matrix $(\mathbf{G}^T \mathbf{G} + \lambda \mathbf{I})^{-1} \mathbf{G}^T$. Now, the first element of Δu is the implemented control action, and the rest of the elements are planned ones that will be re-updated during next control time steps. The adaptation of the system parameters is reflected in the response matrix of the system (\mathbf{G}) and the free response (\mathbf{f}). In other words, to derive \mathbf{G} , we need to find two polynomials E_j and F_j based on the Diophantine equation as follows:

$$1 = E_j(q^{-1})\Delta A(q^{-1}) + q^{-j}F_j(q^{-1}), \quad (5.9)$$

in which the degrees of polynomials E_j and F_j are n_a and $j-1$ (j as in (6.5)), respectively, and they can be derived by dividing 1 by $\Delta A(q^{-1})$ until the remainder is a factor of $q^{-j}F_j(q^{-1})$, and then, the quotient is $E_j(q^{-1})$. Therefore, the polynomial $G_j(q^{-1})$ can be written as

$$G_j(q^{-1}) = E_j(q^{-1})B(q^{-1}). \quad (5.10)$$

The matrix \mathbf{G} is then an $N_u \times N_u$ matrix based on the coefficients of the polynomial $G_j(q^{-1})$ (Camacho *et al.* 2007). The elements of the free response vector \mathbf{f} can also be written as follows:

$$\mathbf{f}_{j+1} = q(1 - \Delta A(q^{-1}))\mathbf{f}_j + B(q^{-1})\Delta u(k-d+j) \quad (5.11)$$

in which $\mathbf{f}_1 = y(k)$. Therefore, as can be seen in equations (5.9), (5.10), and (5.11), these are derived based on the system polynomials of $A(q^{-1})$ and $B(q^{-1})$. Therefore, as these system polynomials are updated at each time step, the response matrix of the system (\mathbf{G}) and the free response (\mathbf{f}) are also updated accordingly.

It should also be highlighted that the only physical constraint that may be taken into account in the actuator limits as follows:

$$u_{min} \leq u \leq u_{max}, \quad (5.12)$$

where u_{min} and u_{max} express the upper and lower bounds of the actuator inputs. Considering the reactor configuration, the lower bound is zero, while the upper bound can be defined based on the volume of the reactor. In case of taking the constraint into account, the optimization problem written by (6.5) subject to inequality constraint of (5.12) has to be solved numerically (Camacho *et al.* 2007).

5.3.3. SUPERVISORY LAYER: DECISION-MAKING OPERATIONAL SCENARIOS

As discussed in Section 5.2.4, maximum productivity of PPB depends on availability of two sources, i.e. the light intensity and the VFA concentration in influent. From a design perspective, the daily product extraction is scheduled before sunrise. This implies that if there is too much VFA in the influent, such that the illumination of one day was insufficient to cultivate maximum productivity, there will be some unconverted VFA in the effluent. Therefore, determining an appropriate value for the desired PPB set-point concentration of the controller (w in the objective function J expressed by (6.5)) enables the control system to operate as efficiently as possible. In addition to this point, variations in each of these two sources highlight the importance of selecting the desired PPB set-point concentration.

In thinking of a suitable value for the desired PPB set-point concentration, it is essential to prioritize the operational strategy based on either “*quality*” or “*quantity*”. In other words, increasing the feeding rate can enhance quantity but may compromise quality, and vice versa. Quantity and quality can be considered as a higher production rate and a higher PPB concentration, respectively. The production rate at control time step k can be defined as

$$Q(k) = X_{PB}(k)u(k), \quad (5.13)$$

where Q [mgCODh⁻¹] denotes the production rate, but since the feeding is configured for only one hour per day, it can be considered as a daily production rate. Hence, increasing u may result in a higher production rate, but reduces the quality, i.e. X_{PB} . The another factor to consider is yield. This comes along as decreasing X_{PB} may lead to unconverted VFA remaining in the effluent. Yield of production, Y , at control time step k can be defined as follows:

$$Y(k) = \frac{Q(k)}{S_{VFA,i}(k-1)u(k-1)}, \quad (5.14)$$

where $S_{VFA,i}$ is the influent VFA concentration. To have some indications towards the factors defined, Table 5.1 provides the steady-state values of the open-loop process simulation. As can be seen, increasing the feeding flow rate (u) leads to a higher daily production (Q) but with less steady-state PPB concentration (X_{PB}). On the other hand, less PPB concentration results in less yield (Y), but some unconverted VFA in the outlet.

Table 5.1: Steady-state values of the inlet and outlet VFA, PPB concentration, yield, and daily production rate under three different operational conditions.

Light intensity 12 h dark/12 h light [Wm ⁻²]	Paddlewheel 12 h dark/12 h light	$S_{VFA,i}$ [mgCODL ⁻¹]	u [Ld ⁻¹]	$S_{VFA,o}$ [mgCODL ⁻¹]	Q [mgCODd ⁻¹]	$Y \times 100$	X_{PB} [mgCODL ⁻¹]
54	on/on	3000	20	≈ 0	17792	29.6	889.6
			25	238.6	23025	30.6	921.0
			30	637.8	25713	28.6	857.1
60	on/on	3000	20	≈ 0	19020	31.7	951.0
			25	127.5	24332	32.4	973.2
			30	534.1	27300	30.3	910.0
60	off/on	2500	20	212.0	14764	29.5	738.2
			25	594.4	17187	27.4	687.5
			30	939.3	18120	24.1	604.0

In addition to the discussion above, it is inevitable that fluctuations will occur in both incoming VFA and light intensity. It also highlights the importance of determining an appropriate desired PPB concentration given the process conditions. Therefore, a supervisory layer is developed in this chapter to overcome this challenge. The supervisory layer is responsible for decision-making considering the current status of the process. Utilizing such a decision-making supervisory layer within the feedback loop enables the control system to update the desired set-point for the PPB concentration. Therefore, a criterion should be designed for quality-driven and

quantity-driven strategies. In this sense, the concentration of VFA in the inlet and outlet plays a crucial role in determining the desired PPB set-point concentration, and thereafter the production rate and the yield. Considering the relations among Q , Y , and X_{PB} the two following operational scenarios can be discussed:

- *Quality as priority*: In this scenario, the PPB set-point should be set as high as possible. A suitable measurement for tracking this trajectory could be the outlet VFA concentration. As this concentration approaches the minimum assigned value, i.e. S_{VFA}^{min} , it indicates that most portion of carbon sources have been consumed and converted to PPB. This indication provides valuable feedback to the controller subject to possible uncertainties, enabling it to calculate control actions even when no additional information is available for the biological and meteorological conditions. Therefore, a stepwise increase (ΔX_{PB}) is implemented for the set-point until the outlet VFA concentration goes below S_{VFA}^{min} . It should be noted that in practice a buffer range, like $S_{VFA,o}^l \leq S_{VFA,o} < S_{VFA,o}^u$ should be taken into account in order to keep the process operation stable. It should also be highlighted that using this scenario contributes to not only PPB output quality, but also to wastewater treatment by reducing output chemical oxygen demand (COD).
- *Quantity as priority*: In this scenario, by reducing the PPB set-point, the production rate will be increased. However, as can be seen by a few examples provided in Table 5.1, the yield is also decreased. To tackle this issue, a novel solution is to recycle the soluble effluent, which primarily consists of unconverted VFA. Yield without recycling can be written as (5.14). Recycling unconverted VFA reduces the amount of VFA required from the VFA tank, thereby increasing the yield to some extent. In other words, it can be written as follows:

$$S_{VFA,i}(k+1)u(k+1) = S_{VFA,i}(k+1)u^*(k+1) + S_{VFA,o}(k)u(k), \quad (5.15)$$

where u^* is the required flow rate of VFA from the fermented stream, which obviously is less than what should be used without circulation. Given (5.14), the new yield based on circulation, Y_{rec} , can be written based on u^* as follows:

$$Y_{rec}(k) = \frac{X_{PB}(k)u(k)}{S_{VFA,i}(k-1)u^*(k-1)}, \quad (5.16)$$

and by substituting (5.15), it gives

$$Y_{rec}(k) = \frac{X_{PB}(k)u(k)}{S_{VFA,i}(k-1)u(k-1) - S_{VFA,o}(k-2)u(k-2)}, \quad (5.17)$$

in which it can be seen that $Y_{rec} > Y$ by comparing (5.17) and (5.14). In contrast to the alternative strategy, we should implement a stepwise reduction (ΔX_{PB}). However, we also require an indication for the set-point reduction. To do so, (5.17) is rewritten as follows:

$$\frac{1}{Y_{rec}(k)} = \frac{S_{VFA,i}(k-1)u(k-1)}{X_{PB}(k)u(k)} - \frac{S_{VFA,o}(k-2)u(k-2)}{X_{PB}(k)u(k)}, \quad (5.18)$$

and

$$\frac{1}{Y_{rec}(k)} = \frac{1}{Y(k)} - \frac{S_{VFA,o}(k-2)}{S_{VFA,i}(k-2)} \times \frac{S_{VFA,i}(k-2)u(k-2)}{X_{PB}(k)u(k)}, \quad (5.19)$$

Now, it can be seen that a potential measurement indication can be $\frac{S_{VFA,o}}{S_{VFA,i}}$. In other words, as the process approaches its steady-state condition considering *recycling effluent*, this factor determines the amount of increase in yield. This strategy not only boosts production rates but also improves yield, which might otherwise decline, but is offset by recirculation. Like the other strategy, a buffer range is also required to be taken into account in practice for operational stability.

Considering these two operational strategies, i.e. *quality-driven* and *quantity-driven*, the decision-making layer can be formulated based on the decision tree given in Figure 5.3. The assignment of the design parameters for this mechanistic decision tree, along with the complementary notes on upper and lower bounds ($S_{VFA,o}^u$, $S_{VFA,o}^l$, α , and β) will be discussed in a simulation study later.

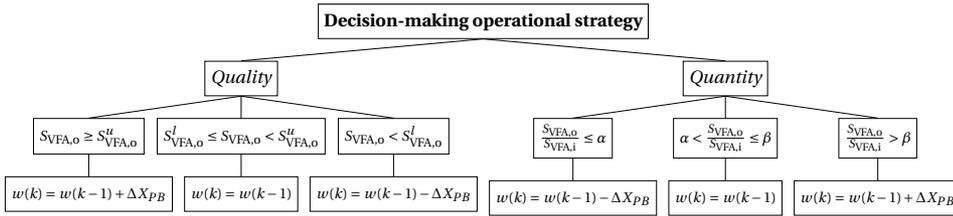


Figure 5.3: Decision-making supervisory layer for assigning a suitable PPB concentration set-point based on either quality- or quantity-driven operational strategy.

5.3.4. OVERRIDE START-UP CONTROL: PHASED-BASED CONTROL

The microbial community considered in the PBM has been divided into three categories: PPB, aerobic bacteria (AEB), and anaerobic bacteria (ANB) (Alloul *et al.* 2023). Furthermore, three different growth pathways, i.e. photoheterotrophic (ph), aerobic chemoheterotrophic (aec), and anaerobic chemoheterotrophic (anc) have been defined for PPB. Since the operational condition is not favorable for anaerobic growth, concentrations of anaerobic bacteria (X_{ANB}) and anaerobic chemoheterotrophic PPB ($X_{PB,anc}$) are very negligible. Therefore, the main competition is between other two types of PPB ($X_{PB,ph}$, $X_{PB,aec}$) with aerobic bacteria (X_{AEB}) only during the start-up operation, as the oxygen supply contributes to the growth of both competitors.

This operational scenario predominantly occurs during the start-up phase, when no microbial biomass dominates and the oxygen concentration is high. Once PPB becomes the dominant species, it enhances its growth accordingly. In other words, the reactor is then a PPB-dominated system, because of availability of light, excess in organic carbon and limited oxygen conditions (Capson-Tojo *et al.* 2023b). Therefore, an override control strategy can be a solution to bring the process from the start-up

phase to the PPB-dominated operation (Chung *et al.* 2006; Sheik *et al.* 2022). A feasible candidate for the control action is regulation of the paddlewheel. As modeled in the PBM, if the paddlewheel is turned on, it results in oxygen depletion from the bulk through accelerating oxygen diffusion for growth, thereby reducing oxygen concentration. Conversely, when the paddlewheel is turned off, oxygen concentration is higher (Alloul *et al.* 2023).

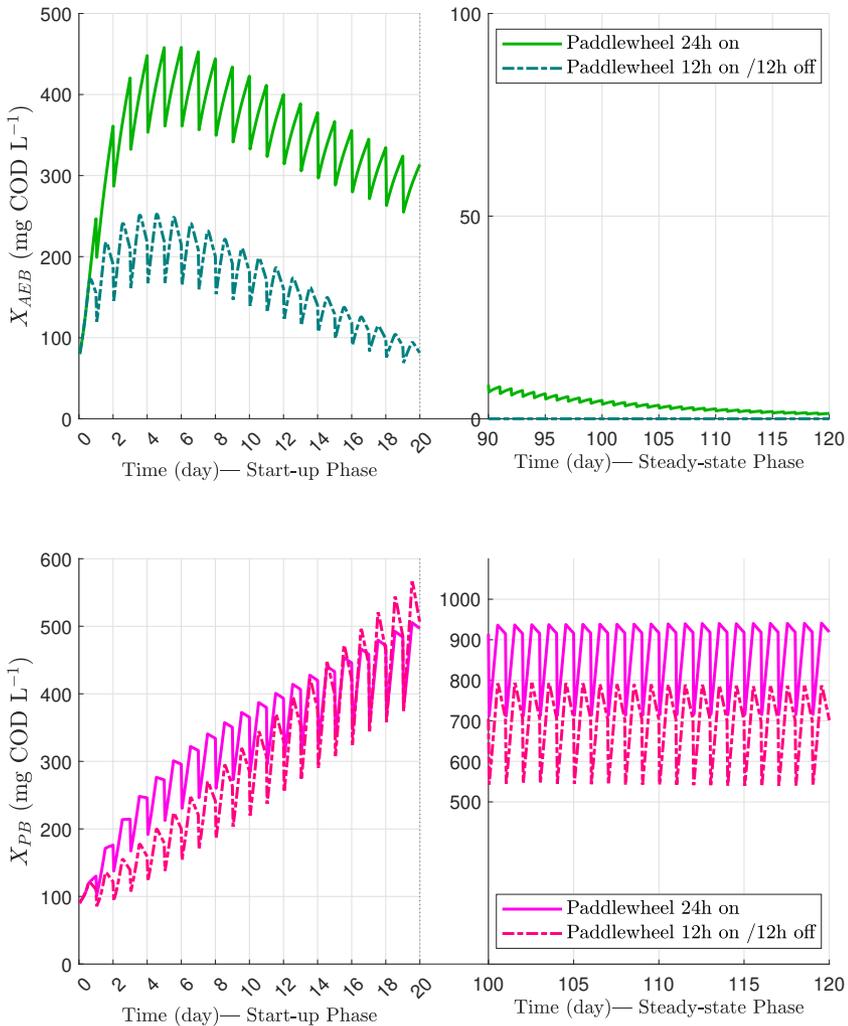


Figure 5.4: Open loop outputs for PPB (X_{PP}) and aerobic bacteria (X_{AEB}) for two phases, i.e. start-up and steady-state under two operational conditions w.r.t. the paddlewheel.

As can be seen in Figure 5.4, the maximum capacity for PPB cultivation is attained when the paddlewheel operates continuously throughout the day. This outcome is biologically explainable, as lower levels of oxygen enhance the productivity of photoheterotrophic PPB (Alloul *et al.* 2021; Capson-Tojo *et al.* 2021). Furthermore, during the start-up phase, the high oxygen levels result in the aerobic bacteria's oxygen affinity being at its maximum level, consequently maximizing their production as depicted in Figure 5.4. Therefore, to facilitate a smooth transition from the start-up phase, a heuristic control approach for regulating paddlewheel is implemented in this chapter. This controller can effectively suppress the growth of aerobic bacteria. It is worth noting that utilizing this override control approach, we can switch from this override control to the main MPC controller, when the competition between PPB and non-PPB is minimal. This strategic activation helps prevent any sudden changes that might otherwise destabilize the process. This is chosen because the process stability is crucial, given that the solution of the objective function (6.5) relies mainly on appropriate initialization. In the next section, the implementation of the developed control system and its results will be discussed.

5.4. RESULTS AND DISCUSSIONS

In this section, the proposed control strategy is assessed via a step-by-step simulation study. First, the main controller is evaluated, including an assessment of its effectiveness and robustness against the most critical model mismatch, namely PPB microbial parallel growth constant, and the most probable disturbances, namely light perturbation and incoming VFA concentration. Secondly, the effectiveness of the phase-based override controller is discussed, and finally, the performance of the main controller coupled with the supervisory layer is assessed.

5.4.1. MPC FOR PPB CONCENTRATION: PERFORMANCE ASSESSMENT UNDER DIFFERENT PERTURBATIONS

Given the control structure discussed in Section 5.3.1, to assess the performance of the proposed controller, the process is considered in the PPB-dominant condition with 24h paddlewheel in use. Prediction and control horizons are determined based on the process settling time to the open-loop step response. As a rule of thumb (Seborg *et al.* 2016), the control horizon N_c can be chosen between $\frac{t_s}{3\Delta t} < N_c < \frac{t_s}{2\Delta t}$, in which t_s denotes the settling time that is around 8d in this case for step response w.r.t. the feeding rate, and Δt expresses the sampling time, which is set to 1d as discussed in Section 5.3.1. The prediction horizon is also selected close to the control horizon (Seborg *et al.* 2016). Hence, N_c and N_p are assigned the value of 4d and 5d, respectively. The constraint on the control input can be posed based on the reactor volume. Since the total reactor volume is 100L in the PBM, the upper limit can be physically considered 40 Lh^{-1} , while the lower limit can be zero, i.e. $0 \leq u \leq 40 \text{ L/h}$. In addition, the model expressed by (6.6) is taken into account as the base model of the MPC controller. The order of model is set to $n_a = 1$ and $n_b = 1$ with respect to the output and the input, respectively. Thus, the parameter vector to be updated at each time step is $\theta = [a_1, b_0, b_1]^T$. As increasing the order

did not improve the performance, and the chosen orders provide sufficient control performance, these values are considered fixed for the simulation studies.

Four operating scenarios: (i) set-point tracking without disturbance; (ii) set-point tracking subject to fluctuation in incoming VFA; (iii) set-point tracking under different illumination scenarios; and (iv) set-point tracking with mismatch in the PPB parallel growth constant are considered to assess the controller performance. As depicted in Figure 5.5, the AGMPC controller is able to track the assigned set-points by regulating the feeding flow rate as the control action. As mentioned in Table 5.1, the steady-state equilibrium of the system with initial feeding rate of 25 Lh^{-1} is 921 mgCODL^{-1} , in which by lowering the PPB set-point concentration, the production rate can be increased. Such an observation brought us to design a supervisory layer that. Moreover, it has been observed that assigning a set-point either too high may cause process instability, since required amounts of carbon sources may not be available to convert to PPB, or too low may violate the actuator constraint and drastically decrease the performance and yield. This one also motivates to introduce a supervisory layer to avoid such occurrences.

5

While having a storage tank for VFA produced from an anaerobic fermentor process helps to stabilize the VFA concentration feeding to the raceway reactor, fluctuations in VFA levels are inevitable. To assess the designed AGMPC controller, a potential $\pm 20\%$ disturbances for the nominal incoming VFA concentration is implemented to the process. As shown in Figure 5.6, the controller is able to keep the process stable to the assigned set-point, subject to the incoming VFA disturbances. As mentioned previously, light and VFA are the two main sources for PPB growth. In case of a significant decrease in VFA such that there is no sufficient VFA biologically available to convert and reach the designated PPB set-point concentration, especially in the presence of adequate light intensity, the process may become unstable. This instability arises from the absence of an optimal solution for the control action within the considered actuators constraints. For instance, for a -20% decrease in the nominal inlet VFA, the outlet PPB concentration can be decreased by 100 mgCODL^{-1} without adjusting the feeding rate. Conversely, a significant increase in VFA poses less of a challenge. However, assigning a set-point concentration that is too low can lead to diminished yield and productivity, as significant amounts VFA may remain unconverted. This again highlights the importance of a suitable set-point to be assigned, considering the process status.

To assess the robustness of the proposed control strategy against light intensity, three illumination scenarios are considered, namely (i) controlled (constant) illumination with a constant intensity of 54 Wm^{-2} , (ii) natural illumination with the total intensity equal to the controlled illumination, and (iii) natural illumination with uncertainty that may happen due to meteorological events, like cloud formation (depicted in Figure 5.7 - top figure). As can be seen in Figure 5.7, even if the light distribution is varying, the controller keeps the process stable on the assigned PPB set-point. Switching from controlled illumination to natural light, even though the total intensity remains constant, results in a decrease in the feeding flow rate determined by the controller. This indicates that apart from light intensity, the distribution of light also influences growth (in agreement with Capson-Tojo

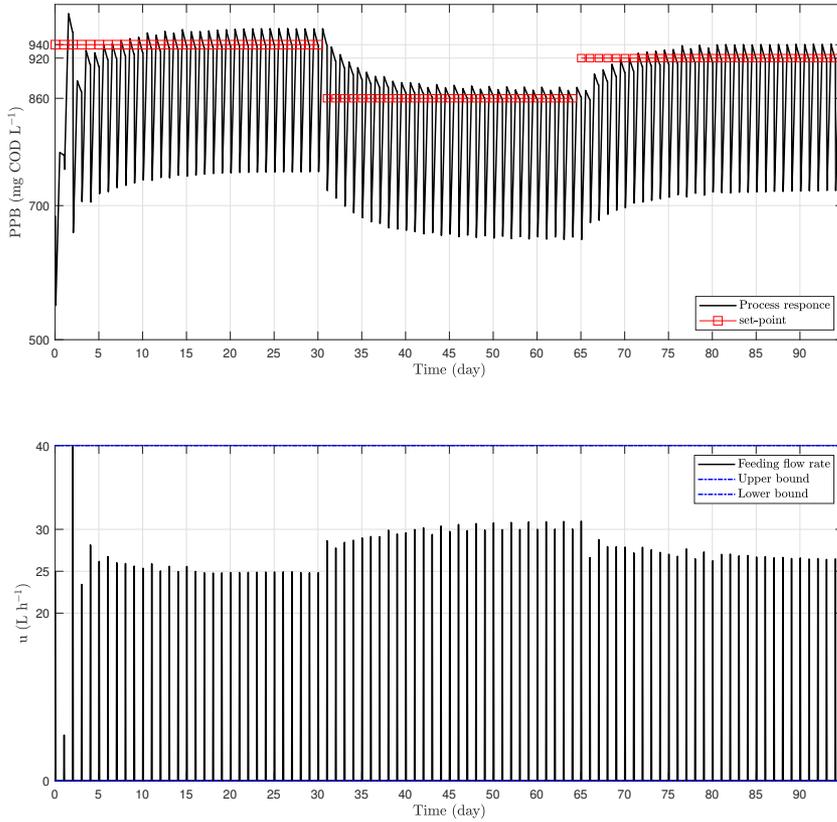


Figure 5.5: AGMPC set-point tracking — process response and control action.

et al. (2023b)). These results indicate the automatic controller can handle these perturbations without light distribution information. As the light intensity and distribution may not be the same for every day, in case of meteorological events that perturbs the planned light intensity, the controller still satisfies the control objective as shown in Figure 5.7. Once again, if the total intensity becomes too low, in case of too high set-point, there is no potential energy source available to convert to PPB, thereby the process becomes unstable, and the importance of assignment of an appropriate set-point is, then, highlighted.

As mentioned in Section 5.2.1, the most important source of the model mismatch is the parallel metabolic growth constant (M_S). Determination of this parameter is experimentally and mathematically complex, as it may change due to changes in species and continuous daily switching between light/dark conditions (Alloul *et al.* 2021). The proposed AGMPC controller is robust against uncertainty and model mismatch due to its reliance on an input-output model that is free of mechanistic relationships. Therefore, this model is daily updated based on observed data to

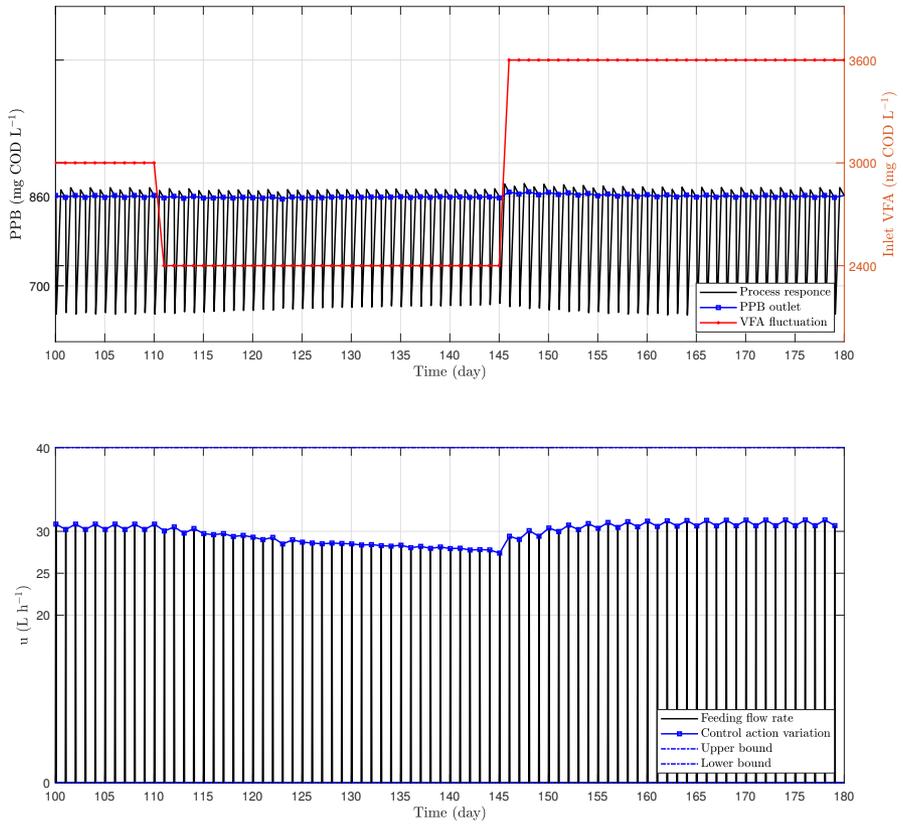
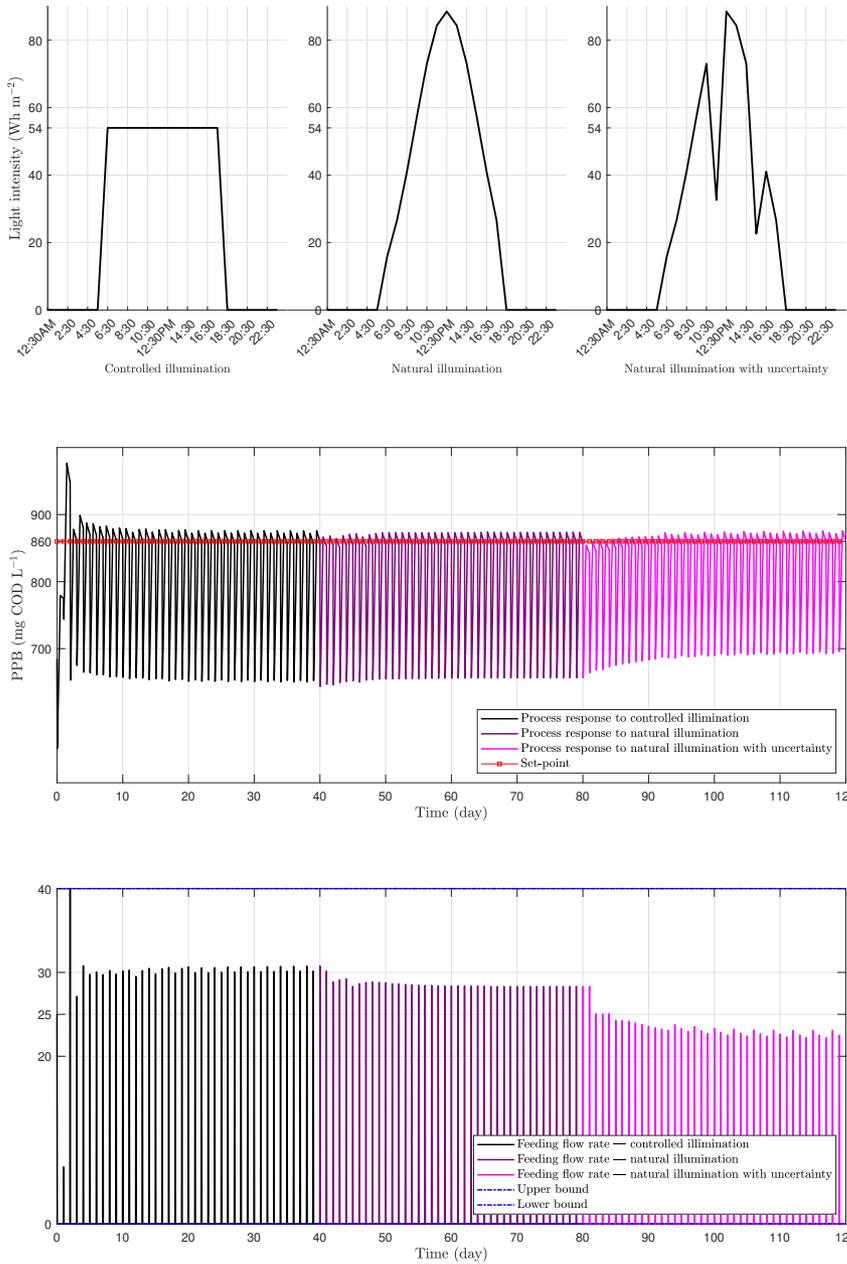


Figure 5.6: AGMPC set-point tracking subject to inlet VFA variation — process response and control action.

accurately capture changes over time. As depicted in Figure 5.8, even though the parallel metabolic growth is changed over time, the controller tracks the assigned set-point by regulating the feeding flow rate. To compare the results, the open loop steady-state values of PPB concentration for the different parallel metabolic growth constants on the last five days for the feeding flow rate of $\approx 25 \text{ L h}^{-1}$ have been also drawn in Figure 5.8. As can be seen, it is an important contributing factor to the PPB concentration, which can drastically change the output concentration without the controller. Thus, the controller keeps the output concentration fixed even when the parallel growth constant changes and no information about these changes is available.

As discussed above, it has been shown that the controller effectively tracks an assigned set-point and can satisfactorily manage two significant potential disturbances: incoming VFA concentration and changes in illumination scenarios, which are the primary resources enabling PPB growth. Moreover, the robustness of



5

Figure 5.7: AGMPC set-point tracking subject to different illumination scenarios — daily illumination intensity, process response, and control action.

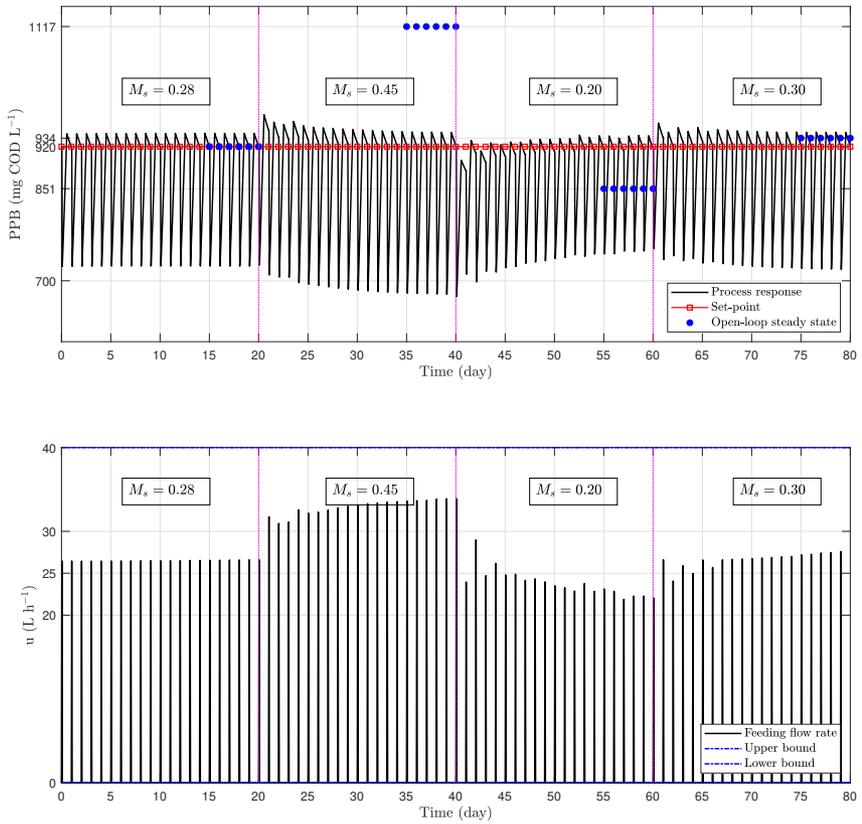


Figure 5.8: AGMPC set-point tracking subject to uncertainty of the PPB metabolic growth constant — process response, and control action.

the controller against model mismatch of the parallel metabolic growth constant has been demonstrated as well. It has been also investigated, whether in some scenarios with severe fluctuations, one might need to assign an appropriate set-point in order to make the best use of available sources to convert to PPB and to avoid process instability. The simulated examples are for the process conditions wherein PPB are the dominant species. In the following section, the discussion focuses on designing a controller to transition the process from the competition phase (start-up phase) to this PPB dominant phase.

5.4.2. OVERRIDE CONTROL FOR MICROBIAL COMPETITION: PHASE-BASED CONTROL

According to the discussion in Section 5.3.4, the paddlewheel is considered as a control regulator for the competition phase during the start-up phase. When the

paddlewheel is activated, it boosts the growth of PPB, if it is the dominant bacteria. Alternatively, turning it off suppresses the growth of non-PPB bacteria. Due to inability of the main controller to find a control input within the search domain dictated by the control input constraint, it is suggested to use an override control with a heuristic approach for the first few days during the start-up phase to prevent non-PPB growth by turning the paddlewheel off, and then switch to MPC controller for PPB concentration control and full-time paddlewheel activation to enhance PPB growth to a maximum extent.

As can be seen in Figure 5.4, the growth of non-PPB is decreased after 4–7 days. Therefore, by suppressing non-PPB growth during these days, by keeping the paddlewheel deactivated, PPB can be the dominant bacteria in a shorter time, then we can switch to the MPC controller and turn the paddlewheel on 24 h to get the maximum growth of PPB. The switching time for the paddlewheel and the activation of the MPC controller depends on the initial condition (in this case concentration in influent) for non-PPB namely, aerobic heterotrophic bacteria, X_{AHB} . As depicted in Figure 5.9, when $X_{AHB} = 10 \text{ mgCODL}^{-1}$, the transition can occur as early as day 2. For $X_{AHB} = 100 \text{ mgCODL}^{-1}$, this transition can take place from day 5 onwards. However, if $X_{AHB} = 200 \text{ mgCODL}^{-1}$, switching to the main controller on day 5 may lead to difficulties for AGMPC in stabilizing the PPB concentration at 860 mgCODL^{-1} as shown. Alternatively, delaying the switch until day 6 can mitigate these control action variations. The process response shown in Figure 5.9 also highlights that when the concentration of aerobic heterotrophic bacteria is lower, or when competition decreases due to a delay in switching, the PPB set-point concentration can be reached more rapidly, while maintaining a higher production rate. Therefore, it can be concluded that such an override control benefits the PPB growth.

5.4.3. SUPERVISORY LAYER: DISCUSSION ON DECISION-MAKING OPERATIONAL STRATEGIES

As discussed above, the proposed AGMPC control system is able to control the assigned output PPB concentration subject to the model mismatch, influent VFA variations, and different illumination scenarios. According to the discussion in Section 5.3.3, the main parameter that affects the process performance is the PPB set-point concentration to be assigned. According to an operational decision expressed by the decision tree given in 5.3, the design parameters can be assigned as follows:

(i): If the priority is *quality* and reaching water treatment criteria, the PPB set-point concentration should be increased to get the maximum potential of available sources, i.e. incoming VFA concentration and light intensity for PPB cultivation. As explained, $S_{VFA,o}$ is a reasonable indication to check how much VFA remains unconverted after one cycle of the process and then to decide for increase/decrease of the PPB set-point trajectory. The cross-checking boundaries, i.e. $S_{VFA,o}^1$ and $S_{VFA,o}^2$ are set to 250 mgCODL^{-1} and 150 mgCODL^{-1} , respectively, as can be seen in the decision chart in Figure 5.3. Therefore, as long as $S_{VFA,o} \geq S_{VFA,o}^1$, the set-point is increased, while if $S_{VFA,o} < S_{VFA,o}^2$, the set-point is decreased. Given that the outlet VFA concentration

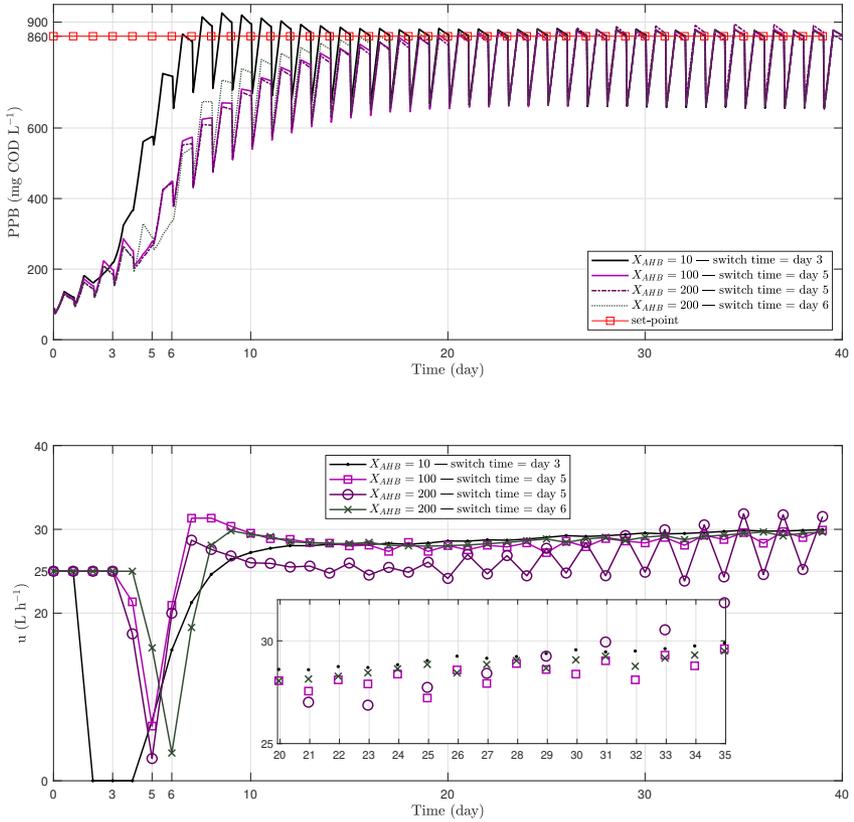


Figure 5.9: Override control for microbial competition during the start-up phase. The override control switches to the MPC controller after a specific day.

is the only indicator, in instances of significant increases in light intensity resulting from meteorological fluctuations, it may be required to lower the set-point. This adjustment is aimed at preventing process instability, as insufficient VFA may be not available for conversion to PPB, leading to a subsequent drop in outlet PPB concentration. Consequently, the lower bound is considered in such cases to address this concern. The buffer range, i.e. $S_{VFA,o}^l \leq S_{VFA,o} < S_{VFA,o}^u$ is also considered keeping the process stable between a specific range of outlet VFA, instead of continuously increasing and decreasing the PPB set-point concentration. Implementing the proposed decision-making layer for *quality* successfully achieves the highest PPB concentration given available sources, as mentioned in Table 5.2. According to Table 5.2, for different operational settings in terms of the required sources for PPB utilization, by checking the outlet VFA, the set-point, and consequently the output PPB concentration, are increased. This decision strategy also contributes to COD removal, as the outlet VFA concentration is decreased by regulating the feeding

flow rate. The outcomes associated with both PPB concentration and COD removal align with the conclusions drawn by Alloul *et al.* (2021, 2023), emphasizing the impact of augmenting hydraulic retention time (HRT), achievable through reducing the feeding flow rate as managed by the controller, on both COD removal and PPB concentration.

Table 5.2: Operational conditions (five operational conditions in terms of available sources, i.e. light intensity and incoming VFA, $S_{VFA,i}$) and performance outcomes: comparative analysis of quality-driven (denoted by #1) and quantity-driven (denoted by #2) approaches in PPB cultivation process.

Operational scenario	Light intensity 12 h dark/12 h light [Wm ⁻²]	$S_{VFA,i}$ [mgCODL ⁻¹]	$S_{VFA,o}$ [mgCODL ⁻¹]	Q [mgCODd ⁻¹]	$Y \times 100$	X_{PB} [mgCODL ⁻¹]	Output u [Ld]	Required input u^* [Ld]
#1	54	3000	170.07	22790.50	31.67	950.00	23.99	23.99
#2			1179.60	28914.18	39.66	741.96	38.97	24.30
#1	60	3000	191.26	25244.49	32.99	989.98	25.50	25.50
#2			1117.80	30631.33	40.96	782.01	39.17	24.93
#1	50	3000	173.61	21298.37	30.71	989.98	23.12	23.12
#2			1153.2	26,382.78	40.73	782.01	36.26	21.59
#1	54	2500	211.97	25695.50	34.00	850.00	30.23	30.23
#2			979.48	28140.02	41.44	629.39	44.71	27.16
#1	54	3500	239.00	21298.37	28.84	1009.70	20.19	20.19
#2			1355.00	26,382.78	38.69	830.00	32.88	20.15

(ii): For the another operational strategy, *quantity* as a production rate is a priority. It can be achieved by reducing the set-point, which increases the feeding flow rate, and thereafter, the output production rate, according to (5.13). While the higher production can be achieved, this operational strategy is not appropriate for COD removal. Therefore, it has been suggested to recycle the soluble materials for the subsequent cycle after separation. It helps to reduce the amount of consumption of the VFA tank from the VFA tank, as denoted by u^* in Table 5.2. As discussed in Section 5.3.3, $\frac{S_{VFA,o}}{S_{VFA,i}}$ can be an appropriate indication to decide for decreasing the PPB set-point concentration. As can be seen in the decision chart in Figure 5.3, α and β are the two boundaries to decide for increase/decrease of the PPB set-point.

These parameters are typically regarded as design parameters that should to be determined by a process expert. For instance, they should not be set at levels where control action becomes saturated. Considering the nominal design provided in the original PBM model, α and β are set to $\frac{1}{3}$ and $\frac{2}{5}$. Since in this example, recirculation is taken into account and a portion of feeding rate includes it, the upper actuator limit is also set to 50 Lh⁻¹. Investigating data given in Table 5.2, while increasing the production rate, the required input from the VFA stream, u^* is also decreased due to the recirculation. Therefore, using this operational decision scenario successfully increases the production rate and decreases the amount of VFA that needs to be provided from the VFA tank. This highlights that the VFA feeding rate in this operational strategy is close to the feeding rate computed by the *quality* decision strategy (see the column of u^* in Table 5.2). Moreover, according to (5.19) and the factor $\frac{S_{VFA,o}}{S_{VFA,i}} \geq \frac{1}{3}$, the yield should exceed $\frac{3}{2}$ of the yield in case of a non-recycling process. In other words, without recirculation, reducing the PPB set-point leads to a corresponding decrease in yield. Conversely, by recycling non-converted VFA, the

yield can be increased by almost $\frac{3}{2}$.

The last assessment includes investigating how the supervisory layer reacts to perturbation and uncertainty during operation. In this regard, the following operational conditions are assumed:

1. Start: A light intensity of 54 Wm^{-2} and incoming VFA concentration of 3000 mgCODL^{-1} as the nominal condition,
2. Event 1: Perturbation on incoming VFA by 20% increase on day 100, i.e. a light intensity of 54 Wm^{-2} and an incoming VFA concentration of 3600 mgCODL^{-1} ,
3. Event 2: Perturbation on light intensity by a 10% increase on day 200, i.e. a light intensity of 60 Wm^{-2} and an incoming VFA concentration of 3600 mgCODL^{-1} ,
4. Event 3: Perturbation on light intensity by a 20% decrease, while considering a mismatch in the parallel growth constant on day 300, i.e. $M_S = 0.32$, a light intensity of 50 Wm^{-2} and an incoming VFA concentration of 3600 mgCODL^{-1} .

5

As can be seen in Figures 5.10 and 5.11, for the both decision strategies, the PPB concentration should be increased upon Events 1 and 2 by the supervisory layer, as the incoming VFA on Event 1 and the light intensity on Event 2 are increased. The basis of such a decision for the quality-driven scenario is keeping the outlet VFA concentration within the specified range as indicated in Figure 5.10 (blue line in top figure), while for the quantity-driven scenario, the criterion is the ratio between the inlet and outlet VFA concentrations, resulting in a higher VFA in the effluent that should be recycled. On Event 3, it is assumed that the parallel growth constant is $M_S = 0.32$ (the nominal constant is $M_S = 0.28$). As discussed in Section 5.4.1 and more specifically the discussion of Figure 5.8, the parallel growth constant is a main source of model mismatch, and therefore, it affects the production, while its value is not known. Therefore, considering this unknown parameter, the supervisory decision layer should determine an appropriate set-point based on the defined criterion. As can be seen in Figures 5.10 and 5.11, however, the light intensity is decreased, but the increase in M_S is the reason of more labor division among different types of PPB (Alloul *et al.* 2021), and consequently the higher PPB concentration in comparison with Event 1. In terms of the calculated control action for the both scenarios, feeding flow rate is adjusted according to the changes (Figures 5.10 and 5.11, bottom ones). On the occurrence of an event, the supervisory layer intervenes to restore stability to the process by addressing the decision criteria that have been violated. Moreover, for the quantity-driven scenario, the flow rate required from the VFA tank (u^* according to Equation (5.15)) is also shown in Figure 5.11 (red line in bottom figure), which is lower than the actual feeding flow rate, as the unconverted VFA is recycled for the next cycle of the process. According to the feeding flow rates depicted in Figures 5.10 and 5.11, the flow rate from the VFA tank is within a similar range ($\approx 20 \text{ Ld}^{-1}$). However, in quantity-driven operational scenarios, the HRT is prolonged due to circulation compared to quality-driven scenarios. This prolonged HRT seems to correspond to an increased yield, aligning with the discussion on PPB aggregation and HRT presented by Blansaer *et al.* (2022).

Another point to be discussed is the adaptation and switching dynamics. As discussed in Section 3.1 and shown in Figure 2, the system parameters are theoretically updated every 24 hours (daily) based on new observations (measurements). However, as long as the process remains within a stable operating domain, meaning no significant perturbations occur, the observations reach a steady state, and the system parameters remain unchanged. This can be clearly seen in Figures 11 and 12 (bottom ones), where the parameter vector θ starts switching when an event occurs, such as on days 100, 200, and 300.

5.4.4. IMPLICATIONS OF THE PROPOSED CONTROL SYSTEM AND FURTHER DEVELOPMENT

The presented control system is the first developed automatic controller for PPB cultivation in a raceway reactor. This reduces the need for skilled labors to supervise the process, not only to ensure process stability against biological perturbations and environmental disturbances, but also to enhance process performance according to the preferred operational strategies mentioned in this work. The best control can be achieved on reliable measurements. Cerruti *et al.* (2020) have discussed a method to measure PPB concentration. Measurement methods based on flow through cell UV-Vis and NIR spectroscopy (Qi *et al.* 2023) allow online measurement if only one species, such as PPB are dominant. If some errors occur due to a lack of precise measurements, developing a mathematical prediction method based on reliable available measurements, such as using either a mechanistic model (Piaggio *et al.* 2024) or an observer (Kemmer *et al.* 2023), would address data availability for the developed control system. It should be also highlighted the control system is based on input-output model, which allows including delay in measurement by increasing time shift operators. Moreover, if any error occurs for PPB measurement, it can be somehow offset via the supervisory layer by cross-checking outlet VFA consecration, for which more reliable and faster measurement is available, which also highlights another advantage of the developed hierarchical control system.

This chapter highlights dealing with the start-up phase for the transition between non-PPB and PPB bacteria communities. This has been addressed by proposing an override control based on mechanistic and heuristic understanding. As mentioned, the capability of the MPC controller to maintain the process on the assigned set-point depends on an appropriate initialization. To include the competition phase into the MPC model, using a simplified mechanistic model instead of such a proposed linear input-output model would address it. Moreover, designing a mechanistic-derived override controller can also be considered as a further development. Finally, a novel supervisory layer based on two proposed operational strategies has been proposed in this work to enhance the process performance. This is based on mechanistic analysis of the process and the simulation model. As a further investigation, application of alternative approaches, namely fuzzy logic system (Ghanavati *et al.* 2021), neural networks (Sadeghassadi *et al.* 2018), and switched systems (Moradvandi *et al.* 2024), can be taken into account.

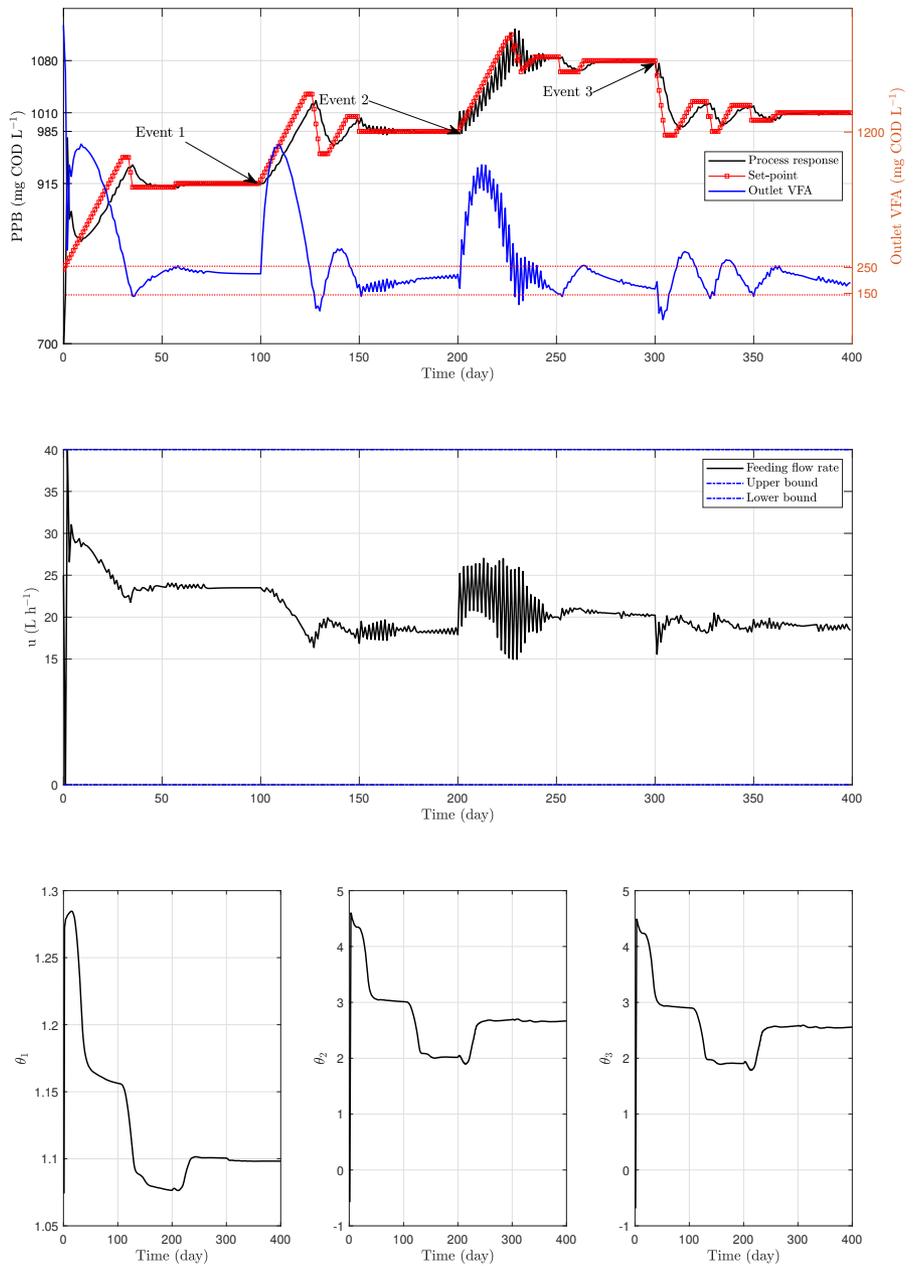


Figure 5.10: Quality-driven scenario: AGMPC set-point tracking integrated with the decision-making supervisory layer to assign the appropriate set-point subject to the operational conditions and fluctuations — process response, determined set-point, outlet VFA, feeding flow rate, and adaptations of parameters.

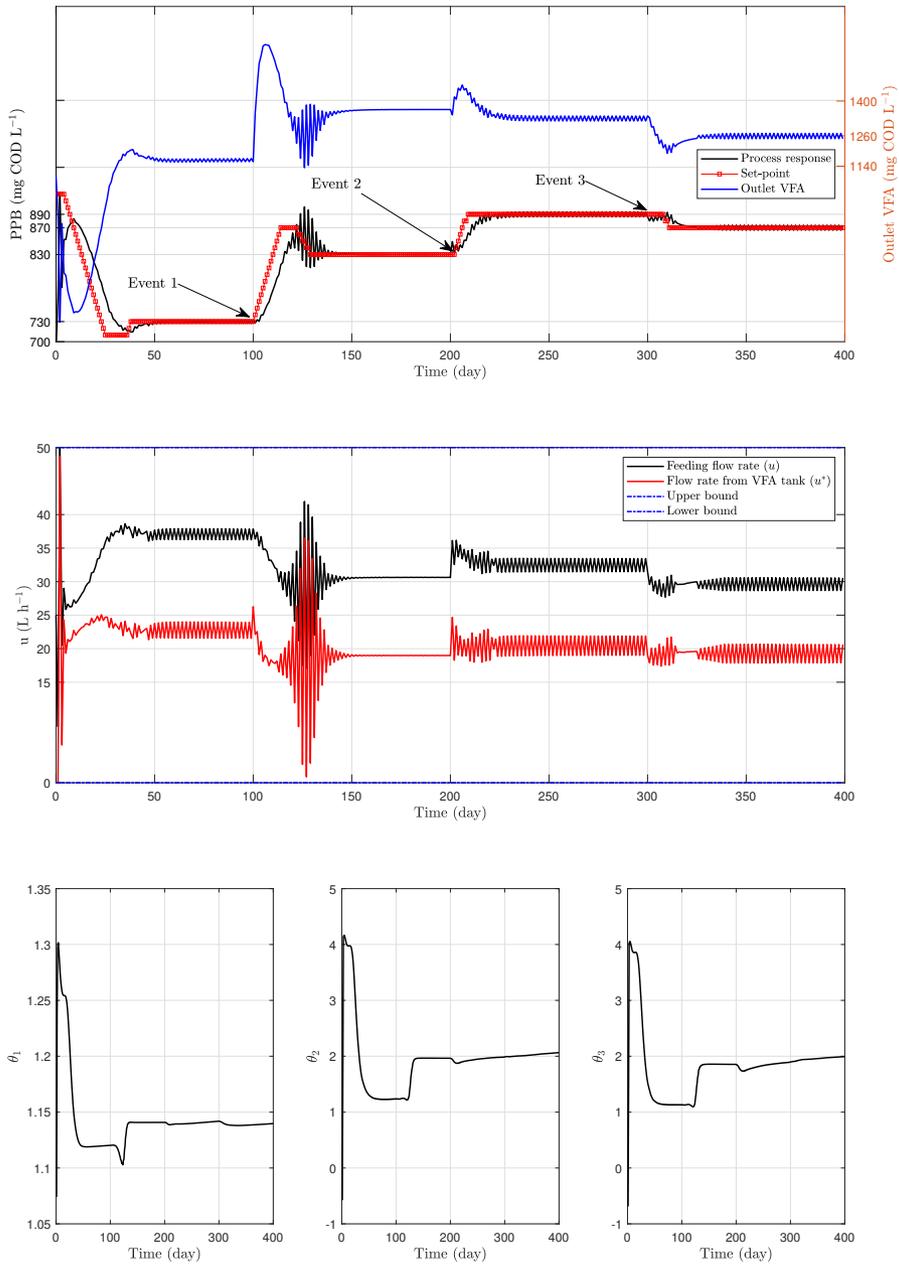


Figure 5.11: Quantity-driven scenario: AGMPC set-point tracking integrated with the decision-making supervisory layer to assign the appropriate set-point subject to the operational conditions and fluctuations — process response, determined set-point, outlet VFA, and feeding flow rate, and adaptations of parameters.

5.5. CONCLUSIONS

In this chapter, a control system on the basis of adaptive generalized model predictive control (AGMPC) for PPB raceway reactors is developed. PPB cultivation in raceway reactors is subject to biological and meteorological fluctuations. The proposed control strategy is able to effectively deal with environmental disturbances. However, significant changes in two essential sources for PPB utilization, namely incoming VFA concentration and light intensity, may lead to process and control inefficiency and instability if the set-point is set either too low or too high. Therefore, the AGMPC controller is integrated to a supervisory layer to assign an appropriate set-point given the process condition. Two operational strategies, namely quality-driven and quantity-driven, are developed for assigning a set-point. In both operational strategies, the hierarchical control system is able to fulfill the process objectives under various perturbations. In the quality-driven scenario, maximizing PPB concentration in each cycle can be achieved by monitoring the outlet VFA concentration. This approach also facilitates COD removal, as it ensures minimal outlet VFA concentrations are attained. In the quantity-driven scenario, decreasing the PPB set-point results in an increased production rate. Simultaneously, the system is configured to recycle unconverted outlet VFA, thereby enhancing yield through the extension of HRT. Moreover, an override control strategy is developed in order to transition the process from the microbial competition phase to the PPB-dominant phase. To achieve this, an investigation is conducted to determine the transition, given an initial condition of the process, the paddlewheel should be deactivated for a few days before switching to full-time activation. The effectiveness of the proposed control framework has been assessed via the PBM model as a benchmark. This automatic control framework can also be used for full-scale plants, even they are supervised by unskilled labors.

REFERENCES

- Ahmed, W. and J. Rodríguez (2020). “A model predictive optimal control system for the practical automatic start-up of anaerobic digesters”. In: *Water Research* 174, p. 115599. ISSN: 0043-1354.
- Alloul, A., M. Cerruti, D. Adamczyk, D. G. Weissbrodt, and S. E. Vlaeminck (2021). “Operational Strategies to Selectively Produce Purple Bacteria for Microbial Protein in Raceway Reactors”. In: *Environmental Science & Technology* 55.12, pp. 8278–8286. ISSN: 1520-5851.
- Alloul, A., A. Moradvandi, D. Puyol, R. Molina, G. Gardella, S. E. Vlaeminck, B. De Schutter, E. Abraham, R. E. Lindeboom, and D. G. Weissbrodt (2023). “A novel mechanistic modelling approach for microbial selection dynamics: Towards improved design and control of raceway reactors for purple bacteria”. In: *Bioresource Technology* 390, p. 129844.
- Alloul, A., S. Wuyts, S. Lebeer, and S. E. Vlaeminck (2019). “Volatile fatty acids impacting phototrophic growth kinetics of purple bacteria: Paving the way for protein production on fermented wastewater”. In: *Water Research* 152, pp. 138–147.
- Blansaer, N., A. Alloul, W. Verstraete, S. E. Vlaeminck, and B. F. Smets (2022). “Aggregation of purple bacteria in an upflow photobioreactor to facilitate solid/liquid separation: Impact of organic loading rate, hydraulic retention time and water composition”. In: *Bioresource Technology* 348, p. 126806.
- Camacho, E. F., C. Bordons, E. F. Camacho, and C. Bordons (2007). *Constrained model predictive control*. Springer.
- Capson-Tojo, G., D. J. Batstone, M. Grassino, S. E. Vlaeminck, D. Puyol, W. Verstraete, R. Kleerebezem, A. Oehmen, A. Ghimire, I. Pikaar, J. M. Lema, and T. Hülsen (2020). “Purple phototrophic bacteria for resource recovery: Challenges and opportunities”. In: *Biotechnology Advances* 43, p. 107567. ISSN: 0734-9750.
- Capson-Tojo, G., D. J. Batstone, and T. Hülsen (2023a). “Expanding mechanistic models to represent purple phototrophic bacteria enriched cultures growing outdoors”. In: *Water Research* 229, p. 119401. ISSN: 0043-1354.
- (2023b). “Expanding mechanistic models to represent purple phototrophic bacteria enriched cultures growing outdoors”. In: *Water Research* 229, p. 119401.
- Capson-Tojo, G., S. Lin, D. J. Batstone, and T. Hülsen (2021). “Purple phototrophic bacteria are outcompeted by aerobic heterotrophs in the presence of oxygen”. In: *Water Research* 194, p. 116941.
- Carreño-Zagarra, J., J. Guzmán, J. Moreno, and R. Villamizar (2019). “Linear active disturbance rejection control for a raceway photobioreactor”. In: *Control Engineering Practice* 85, pp. 271–279.
- Cerruti, M., J.-H. Kim, M. Pabst, M. C. M. Van Loosdrecht, and D. G. Weissbrodt (2022). “Light intensity defines growth and photopigment content of a mixed culture of purple phototrophic bacteria”. In: *Frontiers in Microbiology* 13.
- Cerruti, M., B. Stevens, S. Ebrahimi, A. Alloul, S. E. Vlaeminck, and D. G. Weissbrodt (2020). “Enrichment and Aggregation of Purple Non-sulfur Bacteria in a Mixed-Culture Sequencing-Batch Photobioreactor for Biological Nutrient Removal From Wastewater”. In: *Frontiers in Bioengineering and Biotechnology* 8.
- Chung, Y.-C., I.-L. Chien, and D.-M. Chang (2006). “Multiple-model control strategy for a fed-batch high cell-density culture processing”. In: *Journal of Process Control* 16.1, pp. 9–26.
- Clarke, D., C. Mohtadi, and P. Tuffs (1987). “Generalized predictive control—Part I. The basic algorithm”. In: *Automatica* 23.2, pp. 137–148.
- De Andrade, G., M. Berenguel, J. Guzmán, D. Pagano, and F. Ación (2016). “Optimization of biomass production in outdoor tubular photobioreactors”. In: *Journal of Process Control* 37, pp. 58–69.
- Dochain, D. (2013). *Automatic control of bioprocesses*. John Wiley & Sons.
- Fernández, I., F. Ación, J. Guzmán, M. Berenguel, and J. Mendoza (2016). “Dynamic model of an industrial raceway reactor for microalgae production”. In: *Algal Research* 17, pp. 67–78.
- Ghanavati, M. A., E. Vafa, and M. Shahrokhi (July 2021). “Control of an anaerobic bioreactor using a fuzzy supervisory controller”. In: *Journal of Process Control* 103, pp. 87–99.
- Gupta, N., R. De, H. Kodamana, and S. Bhartiya (2022). “Batch to Batch Adaptive Iterative Learning Control Explicit Model Predictive Control Two-Tier Framework for the Control of Batch Transesterification Process”. In: *ACS Omega* 7.45, pp. 41001–41012.
- Han, H.-G., S.-J. Fu, H.-Y. Sun, and J.-F. Qiao (2021). “Hierarchical nonlinear model predictive control with multi-time-scale for wastewater treatment process”. In: *Journal of Process Control* 108, pp. 125–135.
- Hulsen, T., E. M. Barry, Y. Lu, D. Puyol, J. Keller, and D. J. Batstone (2016). “Domestic wastewater treatment with purple phototrophic bacteria

- using a novel continuous photo anaerobic membrane bioreactor". In: *Water Research* 100, pp. 486–495. ISSN: 0043-1354.
- Imhoff, J. (2006). "The phototrophic alpha-proteobacteria. Prokaryotes: a handbook on the biology of bacteria, Vol 5: Proteobacteria: alpha and beta subclasses". In: *Springer, New York. doi* 10, pp. 0–387.
- Kemmer, A., N. Fischer, T. Wilms, L. Cai, S. Groß, R. King, P. Neubauer, and M. N. Cruz Bournazou (2023). "Nonlinear state estimation as tool for online monitoring and adaptive feed in high throughput cultivations". In: *Biotechnology and Bioengineering* 120.11, pp. 3261–3275. ISSN: 1097-0290.
- Moradvandi, A., E. Abraham, A. Goudjil, B. De Schutter, and R. E. Lindeboom (2024). "An identification algorithm of switched Box-Jenkins systems in the presence of bounded disturbances: An approach for approximating complex biological wastewater treatment models". In: *Journal of Water Process Engineering* 60, p. 105202. ISSN: 2214-7144.
- Pataro, I. M., J. D. Gil, J. L. Guzmán, M. Berenguel, and J. M. Lemos (2023). "A learning-based model predictive strategy for pH control in raceway photobioreactors with freshwater and wastewater cultivation media". In: *Control Engineering Practice* 138, p. 105619.
- Pawłowski, A., I. Fernández, J. Guzmán, M. Berenguel, F. Ación, and J. Normey-Rico (2014). "Event-based predictive control of pH in tubular photobioreactors". In: *Computers & Chemical Engineering* 65, pp. 28–39.
- Pawłowski, A., J. Guzmán, J. Normey-Rico, and M. Berenguel (2012). "A practical approach for Generalized Predictive Control within an event-based framework". In: *Computers & Chemical Engineering* 41, pp. 52–66.
- Piaggio, A. L., G. Smith, M. K. de Kreuk, and R. E. Lindeboom (2024). "Application of a simplified model for assessing particle removal in dissolved air flotation (DAF) systems: Experimental verification at laboratory and full-scale level". In: *Separation and Purification Technology* 340, p. 126801. ISSN: 1383-5866.
- Puyol, D., E. Barry, T. Hülsen, and D. Batstone (2017). "A mechanistic model for anaerobic phototrophs in domestic wastewater applications: Photo-anaerobic model (PANM)". In: *Water Research* 116, pp. 241–253.
- Qi, X., Y. Lian, L. Xie, Y. Wang, and Z. Lu (2023). "Water quality detection based on UV-Vis and NIR spectroscopy: a review". In: *Applied Spectroscopy Reviews*, pp. 1–25. ISSN: 1520-569X.
- Sadeghassadi, M., C. J. Macnab, B. Gopaluni, and D. Westwick (2018). "Application of neural networks for optimal-setpoint design and MPC control in biological wastewater treatment". In: *Computers & Chemical Engineering* 115, pp. 150–160.
- Seborg, D. E., T. F. Edgar, D. A. Mellichamp, and F. J. Doyle III (2016). *Process dynamics and control*. John Wiley & Sons.
- Sheik, A. G., V. R. K. Machavolu, M. M. Seepana, and S. R. Ambati (2022). "Integrated supervisory and override control strategies for effective biological phosphorus removal and reduced operational costs in wastewater treatment processes". In: *Chemosphere*, p. 132346.

6

ANAEROBIC DIGESTION UNDER METEOROLOGICAL FLUCTUATIONS

Temperature plays a critical role in performance and stability of anaerobic digestion processes, subject to frequent meteorological fluctuations. However, state-of-the-art modeling and process control approaches for anaerobic digestion often do not consider the temporal dynamics of the temperature, which can influence microbial communities, kinetics, and chemical equilibrium, and consequently, biogas production efficiency. Therefore, to account for anaerobic digesters operating under fluctuating meteorological conditions, the Anaerobic Digestion Model no. 1 (ADM1) is mechanistically extended in this chapter to incorporate temporal changes into temperature-dependent parameters. This extension defines inhibition functions for microbial activities using the cardinal temperature model, and accounts for the lag in microbial adaptation to temperature fluctuations using a time-lag adaptation function. Thereafter, given that temperature fluctuations are a significant disturbance, a control framework based on Model Predictive Control (MPC) is developed to regulate the feeding flow rate and to ensure stable production rates despite temperature disturbances without relying on direct temperature control. An adaptive MPC approach is formulated based on a linear input-output model, where the parameters of the linear model are updated online to capture the nonlinear dynamics of the process and frequent changes in the dynamics accurately. In addition, a fuzzy logic system is employed to assign a reference trajectory for the production rate based on the temperature and its rate of change. Integrating this fuzzy logic system with the MPC controller enhances the production rate on warm days and avoids the operational failure in production on cold days. Additionally, to enhance biogas production rates, the feasibility of utilizing a portion of the produced biogas for external heating purposes is also investigated. It is demonstrated that by utilizing the proposed MPC approach, the additional amount of feed for the digester to produce methane required

This chapter is an adapted version of *model predictive control of feed rate for stabilizing and enhancing biogas production in anaerobic digestion under meteorological fluctuations*, Ali Moradvandi, et al. (under review), *Journal of Process Control*.

for a self-consumption biogas-fueled heating system can be calculated according to the meteorological variations. This enhances the process performance and stability. Finally, a thermally optimized dome digester semi-buried in the ground, operating under climate conditions of the Netherlands is considered as a case study to validate the extended model in agreement with biological and physicochemical behaviors of real-world applications, and to demonstrate the effectiveness of the proposed control system in handling temperature changes and enhancing performance.

6.1. INTRODUCTION

Biogas as a type of renewable energy is the primary product of Anaerobic Digestion (AD), an industrial biological process technology that converts wastewater through biochemical and physio-chemical conversions into methane and carbon dioxide (Nguyen *et al.* 2019). This biological process occurs in four stages: first, the hydrolysis of organic matter; followed by acidogenesis and next by acetogenesis of intermediate products; and finally, methanogenesis for biogas production (Batstone *et al.* 2002). These stages are characterized by intricate intracellular and extracellular interactions among microorganisms, as well as soluble and particulate matters. The performance of AD, i.e. the amount of biogas produced, depends on various environmental, biological, and operational parameters. This performance can be negatively impacted by not only inoperative design and inexperienced operators, but also by inevitable perturbations, drawing attention and efforts towards maximizing the efficiency despite disturbances and varying operating conditions.

The operating temperature is among the most important parameters, influencing process performance and possibly causing instability in case of continuous fluctuations and sudden perturbations (Kovalovszki *et al.* 2020). Temperature fluctuations might affect biological processes including microbial activity and growth, and yields and thermodynamics, as well as physicochemical processes. In general, three main temperature domains can be considered for anaerobic digester operations, namely psychrophilic 4–15 °C, mesophilic 20–40 °C, and thermophilic 45–70 °C. The impact of temperature on anaerobic digestion has been explored from various angles in the literature (Nie *et al.* 2021). These range from studies on the role of temperature on microorganisms and microbial bioconversion (Mei *et al.* 2016), to investigations of temperature effects on process optimization and performance (Caposciutti *et al.* 2020), and analyses of the effect of temperature conditions on process sustainability and energy efficiency (Calise *et al.* 2023), all of which show the importance of temperature in this context. However, in many of these investigations, the operating temperature has been assumed to be either constant or varying in a limited range.

Mathematical simulation models provide a valuable virtual benchmark for design, monitoring, and control. This is particularly relevant for assessing the effects of temperature on biogas production, which is crucial for sustainability management, optimal digester design, and effective control. For instance, Pedersen *et al.* (2020) investigated the thermal management of biogas digesters by proposing a non-calibrated heat network model. Similarly, Ahmadi *et al.* (2023) studied a

thermally optimal design for large-scale biogas plants through energy analysis using a heat transfer model. Donoso-Bravo *et al.* (2013) and Kovalovszki *et al.* (2020) proposed two mechanistic models with explicit temperature dependency based on two different anaerobic digestion models, namely AM2 and BioModel, respectively. Donoso-Bravo *et al.* (2013) showed that by using a cardinal temperature model, coupled with a simplified anaerobic digestion model, temperature fluctuations can be reflected in the overall system behaviors, more importantly biogas prediction in the presence of seasonal changes. Kovalovszki *et al.* (2020) also proposed a novel approach to incorporating short-term and long-term temperature changes in BioModel to capture the dynamic temperature dependency on process response and biogas production. However, to simultaneously investigate both thermal analysis and process behaviors within a comprehensive framework, a detailed model that integrates thermal and biological aspects is still required, which has not yet been proposed in the literature.

Since the temperature within the digester can be influenced by factors such as digester design as well as environmental, meteorological, and operational conditions and fluctuations, designing a control system to maintain an optimal temperature to ensure consistent biogas production and efficient energy management is required (Garkoti *et al.* 2024). Therefore, some studies have proposed temperature-controlled digesters by implementing various control strategies, such as internal model control based PID control (Kumar *et al.* 2019), fuzzy-PID control (Anand *et al.* 2021), and adaptive neural network control (Anand *et al.* 2022). However, these methods aim to regulate the temperature and to manage disturbances without incorporating biological parameters. From suitability and energy efficiency points of view, it is also very important to investigate other parameters like feeding flow rate to ensure stable biogas production. Integrating biological parameters can enhance biogas production subject to temperature perturbations. Advanced control strategies, such as model predictive control (MPC), are particularly well-suited for this aim. Furthermore, steering the feeding profile using MPC approach to control biogas production has been proposed for various objectives, e.g. demand-orientated and load-flexible biogas production (Dittmer *et al.* 2022), robust automatic process start-up (Ahmed and Rodríguez 2020), and maximization of methane production rate (Ghanavati *et al.* 2021). However, temperature changes have not been considered as a potential and influential parameter in these studies.

To the best knowledge of the authors, design of an MPC controller for anaerobic digestion has not yet been considered for biogas production management, subject to meteorological perturbations and uncertainties by regulating feeding flow rate. In the current chapter, to address temperature fluctuations and to assess advanced control strategies, a simulation model is developed. Anaerobic Digestion Model no. 1 (ADM1) (Batstone *et al.* 2002), a widely accepted model for anaerobic digestion pilot plants, is extended for accounting how physio-chemical and biological parameters dynamically adapt to the operating temperature by introducing temperature inhibition functions. In addition, the operating temperature is modelled based on a heat network (Pedersen *et al.* 2020) that accounts for ambient conditions and digester configurations. These two models together take several parameters and variables

into account to provide a comprehensive virtual benchmark for analyzing both thermal and biological aspects of an anaerobic digester under varying meteorological conditions. This novel proposed extension is validated against research analyzing temperature fluctuations in anaerobic digestion, demonstrating reliable results. From a process control point of view, the feeding flow rate is regulated by the adaptive MPC controller to deal with the operational temperature fluctuations caused by meteorological variations, without directly controlling the temperature. The proposed MPC approach is based on a linear input-output model. The adaptive version of MPC is adopted in order to capture the highly-nonlinear dynamics of the process and to take into account the temperature fluctuations and potential model mismatches, while providing a computationally effective framework for the MPC controller. In addition to the proposed MPC controller, the enhancement of the process productivity, i.e. biogas production rate, is considered. A fuzzy logic system, acting as an expert process supervisor, is employed to assign a reference trajectory based on the temperature variations. This has been previously used to maximize the methane production rate based on volatile fatty acids (Ghanavati *et al.* 2021; Robles *et al.* 2018), but in this chapter, it is designed according to the digester temperature and the change rate of the temperature. Additionally, a parallel preventive inhibition action is also proposed in order to enhance the process efficiency, while dealing with temperature changes. This includes a mechanism to balance inhibitory effects of increasing ammonia concentrations and lowering pH that are associated with a disbalance between acidogenic and methanogenic conversion rates. Lastly, to maximize the biogas production rate, and to increase the net biogas production, a process management option is also investigated. It concerns the integration of active temperature management in the anaerobic digester with a biogas-fueled heating system, by regulating the extra feeding flow rate required for this heating system. The objective is to avoid severe and seasonal temperature changes, while maximizing net biogas production rate. In summary, the main contributions of the present work can be listed as follows:

- ADM1 is extended temperature-wise and integrated into a heat transfer network, in order to obtain a comprehensive model to investigate the impact of different operating temperature profiles;
- An adaptive MPC framework based on a linear input-output model to control biogas production under varying meteorological conditions by regulation of the feeding flow rate is developed, as opposed to the conventional direct temperature control approach;
- A fuzzy logic system is developed based on the temperature and its rate of change over two consecutive days to assign a change value to the reference trajectory of the production rate, enhancing production when temperature rises and preventing operational failures of zero production when temperature drops due to washout;
- A preventive inhibition approach to maintain pH, integrated with the MPC framework to increase input-to-product conversion (biogas and methane yield)

is proposed, in which the response to temperature fluctuations is incorporated;

- A self-consumption biogas-fueled heating system is integrated with the process in order to enhance the overall process performance; the proposed MPC framework is employed to calculate the required amount of extra feeding flow rate, and it shows a positive net biogas production.

The chapter is organized as follows. In the process model (Section 6.2) the development of the temperature-wise extended ADM1 and its integration with a heat transfer model is discussed. In the process control section (Section 6.3), designing an adaptive MPC controller, assigning a reference trajectory using the fuzzy logic system, the preventive inhibition mechanism, and biogas MPC management using self-consumption biogas-fueled heating system are provided. In the simulation study (Section 6.4), a virtual dome digester operating in the Dutch climate is simulated using the extended model and the proposed control system is then evaluated. In the last section, conclusions are drawn.

6.2. PROCESS MODEL

Biochemical and physicochemical phenomena of anaerobic digestion have been mechanistically modelled at different levels of complexity that can be used for process dynamics analysis, and control. Among those, ADM1 is the one that describes the process in a comprehensive structure through four stages, namely: hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Batstone *et al.* 2002) and it is flexible for further model development. Although the temperature is initially considered as a constant variable within the model, the structure of ADM1 allows mechanistic extension to incorporate timely temperature dynamics. Therefore, in this section, the model extension is discussed. The extended model is considered as a benchmark to simulate the anaerobic digestion under operating temperature fluctuations and to assess the performance of the proposed control system subsequently.

6.2.1. ANAEROBIC DIGESTION MODEL: ADM1

The state variables of ADM1 can be represented as variables for soluble matter, S_i , particulate matter, X_i , gaseous variables, $S_{gas,i}$, and ion variables, S_{i-} , in which i denotes the component name. These state variables are formulated on the basis of mass balances in a set of differential equations considering biochemical and physicochemical interactions (Thamsiriroj and Murphy 2011). The model's stoichiometry is determined according to chemical oxygen demand (COD). Through conversions of main composite and particulate components (total composites, X_c , carbohydrates, X_{ch} , proteins, X_{pr} , and lipids, X_{li}) and soluble components (like S_{su}) to gaseous compounds (methane, $S_{gas,ch4}$, carbon dioxide, $S_{gas,co2}$, and hydrogen, $S_{gas,h2}$), methanogenesis can be inhibited by changes of pH, hydrogen, and free ammonia. These changes mostly affect the biochemical uptake rate, which can be formulated by inhibition functions (I_j). The model is formulated for a

continuous-flow stirred tank (CSTR) with a fixed volume. The model is basically represented for a continuous-flow stirred tank (CSTR) with a fixed volume. A summary of the model is provided in Table 6.1. The model consists of several parameters that can change if the temperature is not constant, which is discussed in the next section.

6.2.2. TEMPERATURE DEPENDENCY OF THE PARAMETERS: INTEGRATION OF DYNAMICAL TEMPERATURE

The ADM1 parameters are mostly dependent on the change of the operating temperature. In this section, the temperature dependency of ADM1 and integration of temporal temperature dynamics into this model are discussed.

physicochemical processes: The liquid-gas transfers and the acid-base reactions are temperature-dependent. Henry's constants of gases, ion/acid dissociation constants, and the partial pressure of water can be corrected by van 't Hoff's equation (Batstone *et al.* 2002). In other words, the temperature dependence of the equilibrium constants at temperature T , can be corrected according to a base temperature T_{base} as follows:

$$\ln \frac{K_{i,T_{\text{base}}}}{K_{i,T}} = \frac{\Delta H^{\circ}}{R} \left(\frac{1}{T} - \frac{1}{T_{\text{base}}} \right) \quad (6.1)$$

where $K_{i,T_{\text{base}}}$ and $K_{i,T}$ denote values of the Henry's constant at T_{base} and T , respectively, and ΔH° and R express enthalpy of volatilization and the universal gas constant. The associated correction to ADM1 parameters is given in Table 6.2. Besides, the partial pressures of each biogas are also related to the operating temperature, which can be written based on the ideal gas law as provided in Table 6.2. Furthermore, as suggested by Lee (2017), the volumetric mass transfer constant, k_{L^a} , can also be corrected based on the 5th order of the base and operating temperature ratio as expressed in Table 6.2.

Biochemical processes: Reaction pathways, thermodynamics, and yields are affected by temperature fluctuations, which subsequently changes microbial kinetics and the microbial population dynamics. Raising the temperature can enhance reaction rates up to their optimal point, but they will be subsequently diminished beyond that optimal temperature (Batstone *et al.* 2002). This works like inhibitory factors. Therefore, temperature inhibition functions are defined in this study to model dynamical effects of temperature changes on some selected biochemical processes. To model the temperature inhibition functions, I_T , the cardinal temperature equation (Rosso *et al.* 1993) is employed as follows:

$$I_T = \frac{(T - T_{\text{max}})(T - T_{\text{min}})^2}{(T_{\text{optimum}} - T_{\text{min}}) \left[(T_{\text{optimum}} - T_{\text{min}})(T - T_{\text{optimum}}) - (T_{\text{optimum}} - T_{\text{max}})(T_{\text{optimum}} - T_{\text{min}} - 2T) \right]} \quad (6.2)$$

in which T , T_{max} , T_{min} , and T_{optimum} denote the operating temperature, the maximum and the minimum temperatures that the growth rate is no longer observed beyond them, and the optimal temperature for maximum growth rates, respectively. Basically, three main operating temperature ranges are defined as psychrophilic 4–15 °C, mesophilic 20–40 °C, and thermophilic 45–70 °C ranges, with the optimal

Table 6.1: Summary of ADM1; processes are divided into biochemical, ρ_i , physiochemical, $\rho_{AB,i}$, and liquid-gas, $\rho_{T,i}$, processes with $v_{i,j}$ as stoichiometric coefficients; q denotes the flow rate; k_{dis} , $k_{hyd,i}$, $k_{m,i}$, $K_{S,i}$, and $k_{dec,i}$ express disintegration, hydrolysis, maximum uptake, half saturation coefficient, and biomass decay rates, respectively; I_i denotes the inhibition function; $K_{a,i}$ and $k_{AB,i}$ denote acid dissociation, and acid-base kinetic constants; q_{gas} and V_{gas} denote gaseous flow rate and volume, while V_{liq} expresses the liquid volume; and $k_L a$ is the volumetric mass transfer rate and $P_{gas,i}$ and K_H denote the partial gas pressure and the Henry's constant.

Process	Variable	Variable description	Rate Expression (ρ)	Mass balance ($\frac{d}{dt}$)
Disintegration	X_c	Composites	$k_{dis}X_c$	Expression (1) ¹
	S_I	Soluble inerts	-	Expression (2) ²
	X_I	Particulate inerts	-	Expression (1) ¹
Hydrolysis	X_{ch}	Carbohydrates		
	X_{pr}	Protein	$k_{hyd,i}X_i$	Expression (1) ¹
	X_{li}	Lipid		
Acidogenesis	S_{su}	Monochacharides (sugars)		
	S_{aa}	Amino acids	$k_{m,i} \frac{S_i}{K_{S,i}+S_i} X_i I_i$	Expression (2) ²
	S_{fa}	Fatty acids		
Acetogenesis	S_{va}	Valerates		
	S_{bu}	Butyrates	$k_{m,i} \frac{S_i}{K_{S,i}+S_i} X_i I_i$	Expression (2) ²
	S_{pro}	Propionates		
Methanogenesis	S_{Ac}	Acetate		
	S_{H2}	Soluble hydrogen	$k_{m,i} \frac{S_i}{K_{S,i}+S_i} X_i I_i$	Expression (2) ²
	S_{CH4}	Soluble methane	-	
Death/Growth	X_{su}	Sugar degraders		
	X_{aa}	Amino acids degraders		
	X_{fa}	Fatty acids degraders		
	X_{c4}	Valerate and butyrate degraders	$k_{Dec,i}X_i$	Expression (1) ¹
	X_{pro}	Propionate degraders		
	X_{ac}	Acetate degraders		
	X_{H2}	Hydrogen degraders		
Acid-base conversion	S_{cat}	Cations		
	S_{an}	Anions		
	S_{va^-}	Valerates ion		
	S_{bu^-}	Butyrates ion		
	S_{pro^-}	Propionates ion	$k_{AB,i^-} [S_i^- (K_{a,i} + S_{H^+}) - K_{a,i} S_i]$	Expression (3) ³
	S_{ac^-}	Acetate ion		
	S_{HCO3^-}	Bicarbonate		
	S_{NH3}	Ammonia		
	S_{IC}	Inorganic carbon		
Liquid-gas transfer	$S_{gas,H2}$	Hydrogen gas		
	$S_{gas,CH4}$	Methane gas	$k_L a (S_i - K_H P_{gas,i})$	Expression (4) ⁴
	$S_{gas,CO2}$	Carbon dioxide gas		

¹ Expression (1): $\frac{q}{V_{liq}} X_{in,i} - \frac{X_i}{t_{res} + V_{liq}/q} + \sum_j \rho_j v_{i,j}$

² Expression (2): $\frac{q}{V_{liq}} + (S_{in,i} - S_i) + \sum_j \rho_j v_{i,j}$

³ Expression (3): $\frac{q}{V_{liq}} + (S_{in,i^-} - S_i^-) - \rho_{AB,i}$

⁴ Expression (4): $-\frac{q_{gas}}{V_{gas}} S_{gas,i} + \frac{V_{liq}}{V_{gas}} \rho_{T,i}$

temperatures of 35°C and 55°C for the mesophilic and the thermophilic ranges, respectively (Nie *et al.* 2021).

This inhibition function is included into ADM1 for hydrolysis kinetic rate constants and maximum uptake rates of acidogenesis, acetogenesis, and methanogenesis as given in Table 6.2. Although each component can have its own inhibition function, for the sake of simplicity and lack of available experimental data, 5 temperature inhibition functions, i.e. hydrolysis of carbohydrates, protein, and lipid ($I_{T,\{ch,pr,li\}}$), acidogenesis of sugars, amino acids, and fatty acids ($I_{T,\{su,aa,fa\}}$), acetogenesis of valerate and butyrate ($I_{T,c4}$), acetogenesis of propionate ($I_{T,pro}$), and methanogenesis of acetate and hydrogen ($I_{T,\{ac,h2\}}$), are calibrated based on the data set given by Bergland *et al.* (2015) and Donoso-Bravo *et al.* (2009). In other words, the corresponding inhibition function for each stage of anaerobic digestion is determined by applying the nonlinear least squares method to identify parameters T_{max} , T_{min} , and $T_{optimum}$ of the cardinal function expressed by (6.2).

Temperature adaptation: The operating temperature can vary substantially day-by-day in some regions, subject to diverse variations in meteorological conditions. Although the response of physicochemical processes is fast, the microbial growth rates as biochemical processes require more time to adapt themselves to new conditions (Kovalovszki *et al.* 2020). As suggested by Kovalovszki *et al.* (2020), an effective temperature ($T_{effective}$) according to the operating temperature and an adaptation constant (τ) can be defined as follows:

$$\frac{dT_{effective}}{dt} = \frac{T_{effective} - T}{\tau} \quad (6.3)$$

Therefore, the effective temperature is used for the correction of not only the methanogenesis pathways, as the most sensitive pathways to temperature perturbations (Prakash *et al.* 2019), but also for other microbial growth rates (Kovalovszki *et al.* 2020). The effective temperature is also considered for the design of the fuzzy logic system, which will come later. In other words, the effective temperature ($T_{effective}$) is the main digester temperature to be taken into account. The operating temperature (T) is only utilized for temperature corrections of dissociation, ionization, and gas-liquid mass transfer.

6.2.3. THERMAL BALANCE MODEL: HEAT TRANSFER NETWORK

The main mathematical model is ADM1 with the dynamical temperature extension as explained above. However, another model is required to simulate the operating temperature over a specific period of time. The operating temperature (T) is the substrate liquid temperature of the digester, which may be influenced by meteorological conditions and the design of the reactor. Therefore, a thermal/energy balance model should be provided to simulate it. This is carried out by a thermal balance digester model, inspired by the heat transfer network introduced by Pedersen *et al.* (2020). A dynamical thermal model for the expression of the state variable, denoted by T , can be written as follows:

$$\rho_{liq} C_{liq} V_{liq} \frac{dT}{dt} = Q_{IRR} + Q_{ADV} + Q_{CON} + Q_{RAD} + Q_{EX}, \quad (6.4)$$

where ρ_{liq} and C_{liq} denote the density of the digester substrate and specific heat capacity of the substrate, respectively. Various forms of heat transfers are

Table 6.2: Temperature dependency of the various ADM1 parameters and corresponding correction expressions according to the operating temperature (T) and the base temperature (T_{base}). Base temperature (T_{base}) and operating temperature (T) should be given on Kelvin.

Parameter description	Notation	Expression	Unit	Correction description	Reference
Ion constant of water	K_w	$10^{-14} \exp \frac{5500}{100R} \left(\frac{1}{T_{\text{base}}} - \frac{1}{T} \right)$	M	van 't Hoff's equation	Batstone <i>et al.</i> 2002
Acid dissociation constant of carbon dioxide	$K_{a,co2}$	$10^{-6.35} \exp \frac{7046}{100R} \left(\frac{1}{T_{\text{base}}} - \frac{1}{T} \right)$	M	van 't Hoff's equation	Batstone <i>et al.</i> 2002
Acid dissociation constant of inorganic nitrogen	$K_{a,IN}$	$10^{-9.25} \exp \frac{31955}{100R} \left(\frac{1}{T_{\text{base}}} - \frac{1}{T} \right)$	M	van 't Hoff's equation	Batstone <i>et al.</i> 2002
Henry's constant of methane	$K_{H,CH4}$	$0.035 \exp \frac{-19410}{100R} \left(\frac{1}{T_{\text{base}}} - \frac{1}{T} \right)$	Mbar ⁻¹	van 't Hoff's equation	Batstone <i>et al.</i> 2002
Henry's constant of carbon dioxide	$K_{H,CO2}$	$0.0014 \exp \frac{-14240}{100R} \left(\frac{1}{T_{\text{base}}} - \frac{1}{T} \right)$	Mbar ⁻¹	van 't Hoff's equation	Batstone <i>et al.</i> 2002
Henry's constant of hydrogen	$K_{H,H2}$	$7.8^{-4} \exp \frac{-4180}{100R} \left(\frac{1}{T_{\text{base}}} - \frac{1}{T} \right)$	Mbar ⁻¹	van 't Hoff's equation	Batstone <i>et al.</i> 2002
Volumetric mass transfer coefficient	k_{L^o}	$k_{L^o} = k_{L^o, \text{base}} \left(\frac{T}{T_{\text{base}}} \right)^5$	d ⁻¹		Lee 2017
Partial gas pressure of water	$P_{\text{gas},H2O}$	$0.0313 \exp 5290 \left(\frac{1}{T_{\text{base}}} - \frac{1}{T} \right)$	bar	van 't Hoff's equation	Batstone <i>et al.</i> 2002
Partial gas pressure of methane	$P_{\text{gas},CH4}$	$S_{\text{gas},CH4} \frac{RT}{64}$	bar	Ideal gas equation	Thamsiriroj and Murphy 2011
Partial gas pressure of carbon dioxide	$P_{\text{gas},CO2}$	$S_{\text{gas},CO2} RT$	bar	Ideal gas equation	Thamsiriroj and Murphy 2011
Partial gas pressure of hydrogen	$P_{\text{gas},H2}$	$S_{\text{gas},CH4} \frac{RT}{16}$	bar	Ideal gas equation	Thamsiriroj and Murphy 2011
Hydrolysis kinetic rate constant	k_{hyd}	$k_{\text{hyd}} I_T$	d ⁻¹	Temperature inhibition	Rosso <i>et al.</i> 1993
Maximum uptake rate of acidogenesis, acetogenesis, and methanogenesis	k_m	$k_m I_T$	d ⁻¹	Temperature inhibition	Rosso <i>et al.</i> 1993

$$I_T = \frac{(T - T_{\text{max}})(T - T_{\text{min}})^2}{(T_{\text{optimum}} - T_{\text{min}})(T - T_{\text{optimum}})(T_{\text{optimum}} - T_{\text{max}})(T_{\text{optimum}} - T_{\text{min}} - 2T)}$$

summarized in Table 6.3.

6.3. PROCESS CONTROL

In this section, we discuss a step-by-step design of an advanced control system for anaerobic digestion under temperature fluctuations. In order to do so, we first define the control problems. As mentioned in the original ADM1 development report (Batstone *et al.* 2002), three operational strategies with respect to temperature can be considered: (i) temperature-controlled digestion with minor ($\pm 3^\circ\text{C}$) changes, (ii) variable-temperature digestion with temperature changes in one specific range, (iii) variable-temperature digestion with temperature changes between mesophilic and thermophilic conditions. In second and third operational conditions, no temperature control is applied. Considering the conventional temperature ranges and realistic meteorological perturbations, the third operational scenario may be seen less often in real-world applications, as a digester working in two distinct temperature ranges may not be efficient as well. Therefore, the first two can be considered for designing a control system configuration. As far as the authors are aware of, most of the proposed control methods in the literature are meant to control the temperature of the digester as a temperature-controlled system, which is not really an anaerobic digestion control problem, while it is a "temperature control" or "temperature regulation" problem. Therefore, in this chapter, we develop a control system for digestion considering the second operating scenario with new control objectives such as maximizing total biogas production (in this case, methane) subject to the varying operational temperature using feeding flow rate regulation. In this scenario, two cases, namely a constant set-point, and a reference trajectory are taken into account for the design of a control system. In the first case, the controller should be able to maintain biogas production at the given set-point, even if the operational temperature varies. In the second case, in addition to the objective in case one, a reference trajectory is assigned according to the operational temperature based on a fuzzy logic system. This approach enhances productivity when the temperature rises and reduces the risk of zero production (washout) when the temperature drops. Furthermore, we define an operational strategy using self-consumption biogas-fueled heating to bring the process to the first operational scenario in order to be able to increase the value of the constant set-point, and consequently improve the productivity during operation. The buildup of the control system is discussed in the next sections.

6.3.1. MODEL PREDICTIVE CONTROL: MAIN CONTROLLER

The primary control objective is to control biogas production, while the operating temperature fluctuates within a specific temperature range, by regulating the feeding flow rate without temperature control. From the control point of view, the feeding flow rate is the most feasible control action in practice. The MPC strategy can be an appropriate candidate, as it anticipates process responses and takes control action accordingly to prevent operation failure (Anand *et al.* 2021). MPC control is model-based. However, the developed mechanistic model is too complex and

Table 6.3: Various forms of heat transfers to integrate ADM1 parameters with a dynamical operating temperature (T) through a heat transfer network.

Heat loss/gain (notation)	Expression	Heat source	Design parameter	Note
Solar irradiation (Q_{RR})	$Q_{solar} \cdot A_T$	Solar transmission	Surface area (A) Digester wall material absorptivity (η)	Q_{solar} can be collected from weather institutes.
Advection (Q_{ADV})	$q \rho_{inf} C_{inf} (T_{inf} - T)$	Influent temperature (T_{inf})	Influent flow rate (q)	ρ_{inf} and C_{inf} denote influent density and heat capacity.
Convection/conduction (Q_{CON})	$AU(T_{ambient} - T)$	Conductive and convective heat transfers at digester walls exposed to the ambient (T_{amb})	Surface area (A) Digester wall material	Conductive and convective series resistances (U) ¹
Radiation (Q_{RAD})	$\frac{\sigma(T_{sky}^4 - T^4)}{\frac{1}{3} + 2 \frac{\epsilon_{surface}}{\epsilon_{sub}} + \frac{\epsilon_{sub}}{\epsilon_{sky}}}$	Radiative heat transfer with sky in which $T_{sky} = 0.0552 T_{ambient}^{3/2}$	Surface radiative emissivity ($\epsilon_{surface}$) Substrate radiative emissivity (ϵ_{sub}) Surface area (A)	The reactor surface functions as a radiation shield of the bulk fluid
External heating ($Q_{heating}$)	Fixed value in Watts $\dot{m}_{recirculation} C_{recirculation} (T_{heater} - T)$	Heat exchanger	Heater capacity Heating temperature (T_{heater}) Recirculation mass flow ($\dot{m}_{recirculation}$)	In case of using external heating

¹ Series resistances (U) between the bulk and the ambient can be written as $U = \frac{1}{\sum_j R_{CNV,j} + \sum_j R_{END,j} + R_{CON,forced}}$, in which

- $R_{CNV,j} = \frac{1}{k_j}$ denoted the convective resistance driven by convective heat transfer coefficient (h_j) between two fluids,
- $R_{END,j} = \frac{\Delta x_j}{k_j}$ denoted conductive resistance driven by thickness of the conducting layer (Δx_j), and the thermal conductivity (k_j),
- $R_{CON,forced}$ denoted the forced convection driven by wind speed.

computationally expensive to be used for MPC control. As an alternative, a linear input-output model can be utilized instead (Ghanavati *et al.* 2021; Moradvandi *et al.* 2024). To properly capture the plant's dynamics with this simpler model, its parameters should be updated at each time step based on new and past observations of the input and output variables. In addition, as the plant is subject to temperature changes, employing an adaptive prediction model can mimic these changes at every time step. Therefore, the adaptive MPC control is utilized to incorporate these aforementioned characteristics of the process. It should be noted that once the parameters are set for a particular time step, they are fixed over the prediction horizon for the calculation of the control actions.

According to the MPC framework, the control action is calculated for a control horizon (N_u) that minimizes a cost function based on a prediction horizon (N_p). The cost function, J , and the process constraint are, then, defined as follows:

$$J = \sum_{j=1}^{N_p} \delta [y(k+j|k) - w(k+j)]^2 + \sum_{j=1}^{N_u} \lambda [\Delta u(k+j-1)]^2 \quad (6.5a)$$

$$\text{s.t. } u_{min} \leq u(k) \leq u_{max}, \quad (6.5b)$$

where $y(k)$ denotes the j -step ahead prediction of the output of the process (methane production rate) at time k , $w(k)$ expresses the future set-point or reference trajectory (i.e. set-point or reference trajectory for methane production rate) at time k , and $\Delta u(k)$ denotes the planned control input increments (i.e. change in feeding flow rate) at time k . Moreover, δ and λ are the controller design parameters representing the error and the control weighting factors. In a simple form, it is assumed that the methane production rate at time step k ($y(k)$) is a function of the feeding flow rate ($u(k)$), which can also be affected by other potential operating conditions like operating temperature. The problem constraint is also given by (6.5b), where u_{min} and u_{max} express the upper and lower bounds of the actuator. To simplify the problem, as suggested above, an input-output model is considered as a basis of the MPC framework. Therefore, the prediction, \hat{y} , of the actual output, y , is replaced in (6.5a) and the model function is expressed by a single-input single-output discretized parameter linear model as follows (Goodwin and Sin 2014):

$$A(q^{-1})y(k) = B(q^{-1})u(k), \quad (6.6)$$

where $A(q^{-1})$ and $B(q^{-1})$ are rational functions of the time operator q^{-1} (i.e. $q^{-z}x_k = x_{k-z}$ for $z \in \mathbb{Z}$), and they can be written as follows:

$$A(q^{-1}) = 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a}, \quad (6.7a)$$

$$B(q^{-1}) = b_0 + b_1q^{-1} + \dots + b_{n_b}q^{-n_b}, \quad (6.7b)$$

in which n_a and n_b express the order of the system with respect to the outputs and the inputs, respectively. Considering $\theta = [a_1, \dots, a_{n_a}, b_0, \dots, b_{n_b}]^T$ as the vector of the linear functions' coefficients, the online estimation of this parameter vector at time step k , i.e. $\hat{\theta}(k)$ can be derived using the least square method as follows:

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{P(k-1)\phi^T(k)}{1 + \phi^T(k)P(k-1)\phi(k)}(y(k) - \hat{y}(k)), \quad (6.8)$$

where $\phi(k)$ is the augmented vector of past input and output observations, $P(k)$ is the covariance matrix, and $\hat{y}(k)$ is the prediction output, which can be written as follows:

$$\phi(k) = [y(k-1), \dots, y(k-n_a), u(k), \dots, u(k-n_b)]^T, \quad (6.9a)$$

$$P(k) = P(k-1) + \frac{P(k-1)\phi^T(k)\phi(k)P(k-1)}{1 + \phi^T(k)P(k-1)\phi(k)}, \quad (6.9b)$$

$$\hat{y}(k) = \phi^T(k)\hat{\theta}(k-1). \quad (6.9c)$$

Using this approach as discussed by Camacho *et al.* (2007), Ghanavati *et al.* (2021), and Goodwin and Sin (2014), since the model is updated at every time step based on new and past observations, this adaptive scheme is an effective approach to deal with disturbances and impreciseness of simplified linear model by updating $\hat{\theta}$ and predicting \hat{y} accordingly.

6.3.2. ASSIGNING A REFERENCE TRAJECTORY BASED ON A FUZZY LOGIC SYSTEM

As discussed, the primary control objective is to maintain the methane production stable even when the operating temperature varies. Therefore, an appropriate set-point (w in (6.5a)) determined, ensuring it is feasible to reach under all possible temperatures. Meteorological conditions may vary drastically, not only seasonally, but also diurnally, affecting the process operation. Although the proposed MPC controller aims to stabilize process production, the set-point can also be adjusted according to the operating temperature at each time step. A decision for the set-point adjustment can be made by an expert, who is aware of the process and its conditions. To design an automatic control system, a fuzzy logic system is employed to obtain an appropriate reference trajectory. The fuzzy logic system works based on IF-THEN rules that are written according to expert knowledge (Mendel 1995). In this chapter, the aim of using a fuzzy logic system is to assign a reference trajectory for the MPC controller according to meteorological changes. As explained, these meteorological variations yield on varying operating and effective temperatures, thereby affecting the process performance and the production rate. In other words, if the temperature rises, the production rate can be increased, and if the temperature decreases, the production rate should be decreased, thereby adjusting the reference trajectory for the controller in order to be matched to the temperature. Since the effective temperature accounts for the temperature adaptation for biological processes, it is selected as a reference to design a fuzzy logic system.

Therefore, the effective temperature ($T_{\text{effective}}(k)$) is one input to the fuzzy logic system. The direction of effective temperature changes based on two consecutive days ($\Delta T_{\text{effective}}(k) = T_{\text{effective}}(k) - T_{\text{effective}}(k-1)$) is also considered as another fuzzy logic system input. The output of the fuzzy logic system is the value of change in trajectory ($\Delta w(k)$). The first input assigns the range of the change in the reference trajectory, and the second input determines the rate of the change. Each fuzzy logic

system consists of four compartments: (i) fuzzification, (ii) inference, (iii) rule base, and (iv) defuzzification. Membership functions for the inputs are trapezoidal for the sides and Gaussian for the middle, while the membership functions for the output are all Gaussian (De Barros *et al.* 2016). Three, five, and seven membership functions are considered for $T_{\text{effective}}$, $\Delta T_{\text{effective}}$, and Δw , respectively. A decision-making rule table that relates inputs to the output is given in Table 6.4. All ranges for the inputs and the output are shaped symmetrically to facilitate easy tuning of the fuzzy logic system. A Mamdani inference system and the center of gravity method are employed to convert linguistic values into a crisp numerical value during defuzzification (Anand *et al.* 2021). The details of the membership functions will be given and discussed for an example in Section (6.4.3).

Table 6.4: Decision-making fuzzy rules for assigning a value for the set-point change (Δw) based on the effective temperature ($T_{\text{effective}}$) and the rate of the change of the effective temperature ($\Delta T_{\text{effective}}$).

$\Delta T_{\text{effective}} \rightarrow$ $T_{\text{effective}} \downarrow$	Big negative	Negative	Zero	Positive	Big positive
Low	Big negative	Medium negative	Zero	Positive	Positive
Medium	Medium negative	Negative	Zero	Positive	Medium positive
High	Negative	Negative	Zero	Medium positive	Big positive

6.3.3. PARALLEL PREVENTIVE INHIBITION MECHANISM

Biogas production and process efficiency can be significantly reduced when temperature is varying (Nie *et al.* 2021). These variations trigger inhibitory factors (Yuan and Zhu 2016), that result in a different inhibitory response by various trophic groups and lead to process acidification due to VFA accumulation (Cysneiros *et al.* 2012). For example, a specific reduction of acidogenic bacteria due to high concentration of free ammonia was observed by (Yenigün and Demirel 2013). To deal with these situations, a few preventive inhibition actions can be taken in order to prevent the operation failure and enhance the performance of the main MPC controller. In other words, in case of either ammonia or pH-induced inhibition, this action can act as a preventive strategy. Since the feeding flow rate is utilized for biogas production within the framework of MPC, a different strategy should be considered. These strategies should be fast in response and preferably should not involve the biological processes to avoid affecting the performance of the main MPC control system. In this regard, pH adjustment by regulation of alkalinity (cations, anions and the charge balance) as a parallel physical fast process (Cysneiros *et al.* 2012; Yuan and Zhu 2016) is a well-known strategy for preventing acidification. Furthermore, this can also be considered in association with physicochemical approaches like chemical additions (Yuan and Zhu 2016). Consequently, a more stable pH during operation prevents inhibition, allowing the main MPC system to regulate biogas production more effectively. Therefore, the efficiency of maintaining a constant pH with the main MPC controller will be studied in the results and discussion section, using the developed temperature-extended ADM1 model.

6.3.4. BIOGAS MPC MANAGEMENT USING SELF-CONSUMPTION BIOGAS-FUELED HEATING

As biogas is a type of organic fuel that can be used for heating, burning a portion of the produced biogas from anaerobic digestion, more specifically methane, to heat up the anaerobic digester, may not only upgrade the overall process efficiency, but also prevent the aforementioned inhibitions by raising the operating temperature (Caposciutti *et al.* 2020). In other words, a biogas-fueled boiler heats up the digester to bring the process to the first operational strategy, i.e. temperature-controlled digestion with minor temperature ($\pm 3^\circ$) changes. Therefore, the digester should be fed with extra feed to produce the required methane for the self-consumption biogas-fueled heating system. It should be noted, that temperature can fluctuate freely within the proposed boundary conditions. Since it is assumed that the fuel of the boiler is provided from produced methane, the amount of methane required for burning, i.e. $\dot{m}_{\text{burned ch}_4}$, can be written as follows:

$$\dot{m}_{\text{burned ch}_4} = \frac{Q_{\text{EX}}}{\eta_{\text{ch}_4} \tau}, \quad (6.10)$$

in which η_{ch_4} denotes the fuel burning efficiency or lower heating value and τ expresses the adaptation constant used in (6.3). Moreover, Q_{EX} represents the external heating used in (6.4), which can be written as follows:

$$Q_{\text{EX}} = \rho_{\text{sub}} C_{\text{sub}} V_{\text{sub}} (T_{\text{heater}} - T). \quad (6.11)$$

Therefore, using the MPC control system presented in Section 6.3.1, feeding flow rate profiles subject to temperature variations as well as different T_{heater} settings can be investigated to manage biogas production effectively, which will be discussed for a dome digester in the next section.

6.4. RESULTS AND DISCUSSIONS: A SIMULATION STUDY

In this section, a case study is defined to assess the temperature-wise extension of ADM1. Historical meteorological changes and perturbations are used as a benchmark to verify the response of the anaerobic digester. The proposed control strategies, including MPC controller with a constant set-point as well as a fuzzy-driven reference trajectory, and the integration of stable pH with the MPC controller are assessed on the defined case study as well. Finally, a discussion on results of biogas MPC management using biogas-fueled heating strategy is also drawn.

6.4.1. A DOME DIGESTER WITH THE METEOROLOGICAL CONDITIONS CORRESPONDING TO THE NETHERLANDS CLIMATE CONDITIONS

To verify the extended model and the proposed control strategies, a full-scale concrete dome anaerobic digester is virtually located in De Bilt, the Netherlands. The dome digester is constructed according to the full-scale thermally optimal design specifications as discussed by Ahmadi *et al.* (2023) and presented in Table 6.5. As Hreiz *et al.* (2017) and Ahmadi *et al.* (2023) discussed, it concerns a

semi-buried dome digester with approximately one-fourth of its surface (mostly dome part) exposed to solar radiation (Figure 6.1). It should be noted that this type of digesters usually do not have a direct control system. Moreover, as they are not currently in use in the Netherlands, this investigation provides its feasibility for future consideration.

Table 6.5: Design parameters and characteristics of a semi-buried dome digester used for simulation based on optimal thermally design discussions by Ahmadi *et al.* (2023).

Parameter description	Expression	Value/Expression	Unit
Operating conditions			
Feeding flow rate (on average)	q	290	$\text{m}^3 \text{ day}^{-1}$
Influent temperature (on average)	T_{inf}	297.15	K
Soil temperature (on average at a depth of 1 m)	T_{soil}	283.15	K
Sky temperature	T_{sky}	$0.0552T_{\text{ambient}}^{3/2}$	K
Digester geometry			
Total volume	V	3761.87	m^3
Liquid/gas volume ratio	-	3.36	-
Dome surface	A_{dome}	536.81	m^2
Wall surface	A_{wall}	471	m^2
Floor surface	A_{floor}	490.62	m^2
Characteristic length (for forced convection)	L_c	13.13	m
Wall thickness	ΔX_{wall}	0.3	m
Insulation thickness	$\Delta X_{\text{insulation}}$	0.02	m

Regarding meteorological and ambient conditions, three influential variables, i.e. daily mean ambient temperature (T_{ambient}), daily mean wind speed, and daily mean solar irradiation (Q_{solar}), have been taken into account due to their significant impacts. The datasets for these variables are obtained based on historical data for a 30-year period (1992-2021) from the Royal Netherlands Meteorological Institute (KNMI). A comprehensive heat network discussed in Section (6.2.3) and schematically presented in Figure 6.1 along with the aforementioned meteorological variables, is considered for the thermal balance and to determine daily changes in the operating temperature. As can be seen, four resistance series, i.e. air-cover-dome-biogas-substrate, soil-wall-biogas-substrate, soil-wall-substrate, and soil-floor-substrate, are defined. The conductive and convective resistances, and all parameters required to simulate the heat network, are provided in Table 6.6. There is also no external heating.

According to the given conditions and specifications of the dome digester, the defined heat network model is simulated, and the operating and effective temperature variations over a year are depicted in Figure 6.2. As previously mentioned, the operating temperature will be used to adjust the temperature for the physicochemical parameters, while the effective temperature will be utilized to correct the biological parameters and to assign a reference trajectory based on the fuzzy logic system.

Table 6.6: Thermal parameters used for simulation.

Parameter description	Notation	Value/Expression	Unit	Reference
Thermal conductivity				
Wall (plain concrete walls surrounded by moist earth)	k_{wall}	1.5	$\text{W m}^{-1} \text{K}^{-1}$	Metcalf <i>et al.</i> (1991)
Dome (plain concrete with air space plus brick facing)	k_{dome}	1.2	$\text{W m}^{-1} \text{K}^{-1}$	Metcalf <i>et al.</i> (1991)
Insulation (fiberglass)	k_{ins}	0.04	$\text{W m}^{-1} \text{K}^{-1}$	Metcalf <i>et al.</i> (1991)
Air	k_{air}	0.026	$\text{W m}^{-1} \text{K}^{-1}$	Metcalf <i>et al.</i> (1991)
Convection				
Dry wall–biogas coefficient	$h_{\text{dw-b}}$	2.15	$\text{W m}^{-1} \text{K}^{-1}$	Pedersen <i>et al.</i> (2020)
Wet wall–biogas coefficient	$h_{\text{ww-b}}$	2.70	$\text{W m}^{-1} \text{K}^{-1}$	Pedersen <i>et al.</i> (2020)
Biogas–substrate coefficient	$h_{\text{bs-s}}$	2.20	$\text{W m}^{-1} \text{K}^{-1}$	Pedersen <i>et al.</i> (2020)
Wet wall–substrate coefficient	$h_{\text{ww-s}}$	177.25	$\text{W m}^{-1} \text{K}^{-1}$	Pedersen <i>et al.</i> (2020)
Floor–substrate coefficient	h_{fs}	244.15	$\text{W m}^{-1} \text{K}^{-1}$	Pedersen <i>et al.</i> (2020)
Forced (Air) convection				
Thermal conductivity	k_{air}	0.026	$\text{W m}^{-1} \text{K}^{-1}$	Cengel and Ghajar (2011)
Dynamic viscosity	ν_{air}	1.82×10^{-5}	Pa s	Cengel and Ghajar (2011)
Density	ρ_{air}	1.205	Kg m^{-3}	Cengel and Ghajar (2011)
Reynolds number	Re	$\frac{\rho_{\text{air}} v_{\text{wind}} L_c}{\mu_{\text{air}}}$	–	Cengel and Ghajar (2011)
Nusselt number	Nu	$0.037 Re^{4/5} Pr^{1/3}$	–	Cengel and Ghajar (2011)
Air–insulation convection coefficient	h_{air}	$\frac{Nu_{\text{air}} k_{\text{ins}}}{L_c}$	$\text{W m}^{-1} \text{K}^{-1}$	Cengel and Ghajar (2011)
Substrate/influent thermal coefficients				
Heat capacity	C_{inf}	4.179×10^3	$\text{J Kg}^{-1} \text{K}^{-1}$	Pedersen <i>et al.</i> (2020)
Density	ρ_{inf}	1×10^3	Kg m^{-3}	Pedersen <i>et al.</i> (2020)
Radiative parameters				
Stefan-Boltzmann constant	σ	5.67×10^{-8}	$\text{W m}^{-2} \text{K}^{-4}$	Mohr and Taylor (2005)
Wall emissivity	ϵ_{wall}	0.75	–	Cengel and Ghajar (2011)
Substrate emissivity	ϵ_{sub}	0.67	–	Pedersen <i>et al.</i> (2020)

6.4.2. SIMULATION OF A DOME DIGESTER SYSTEM: ASSESSMENT OF THE PROPOSED EXTENDED ADM1 MODEL

In this section, the temperature-wise extended ADM1 proposed in Section 6.2.1 is simulated, considering the dynamical operating and effective temperatures, reactor configurations, and other aforementioned operational conditions of the defined case study. The initial conditions are set to the default initial conditions of the original ADM1 (Batstone *et al.* 2002), in which the substrate is protein-rich, thereby ammonia accumulation may be inevitable. Variations of the temperature inhibition functions are shown in Figure 6.3. These functions are calibrated based on experimental data given by Bergland *et al.* (2015) and Donoso-Bravo *et al.* (2009), and inhibit the maximum growth rates of hydrolysis, acidogenesis, acetogenesis, and methanogenesis of the corresponding compounds in case of a deviation from the optimal temperature. Therefore, as can be seen, they inhibit growth rates more on cold days and less when the temperature is close to the optimal level. Consequently, this is reflected in the methane production, which is depicted in Figure 6.4. Thus, the methane production rate is higher in a range when the temperature is relatively higher, in line with the season-based production discussed by Hreiz *et al.* (2017). Accumulation of fatty acids ($S_{fa} = 1.3 \text{ gL}^{-1}$ at the lowest temperature) and volatile fatty acids ($S_{va+bu+pro+ac} = 1.4 \text{ gL}^{-1}$ at the lowest temperature) occur as the temperature drops. The correlation between the fatty acid accumulation and methane production rate with temperature aligns with the findings from experimental studies conducted by Erdirencelebi and Ebrahimi (2022).

While the effective temperature peaks (as shown in Figure 6.2 for the effective temperature), methane production does not peak (as shown in Figure 6.4). Based on the model output, it is anticipated that a temperature-induced increase in

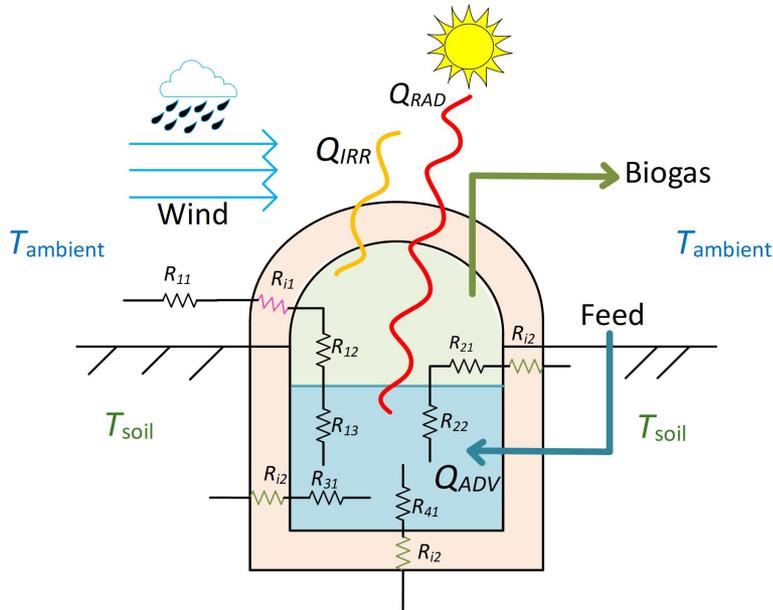


Figure 6.1: Schematization of the heat map of a dome digester: besides heat transfers by radiation Q_{RAD} , solar irradiation Q_{IRR} , and advection Q_{ADV} through the influent feeding to the digester, three resistance series are considered: R_{11} , R_{12} , R_{13} and R_{22} , R_{21} , R_{31} , and R_{41} are the convection resistors of air–cover, dry wall–biogas, biogas–substrate, wet wall–biogas, wet wall–substrate and floor–substrate, respectively. Furthermore, R_{i1} and R_{i2} express the conduction resistors for dry and wet walls.

6

the hydrolysis rate constant of protein increases the free ammonia concentration and its inhibitory effect (as shown in Figure 6.5). This then counteracts the temperature-induced increase in methane production rate (Nie *et al.* 2021; Yuan and Zhu 2016). Therefore, although the temperature increase suffices for more methane production, it is suppressed by free ammonia inhibition. In other words, due to the disproportionately accelerated process rates at elevated temperatures, the total concentration of free ammonia increases more than the shift in the acid-base equilibrium, and acetoclastic methanogenesis remains inhibited. It then takes some time for the methanogens to overcome the free ammonia induced inhibition, although it does not lead to significant VFA accumulation. Temperature and methane production stabilize then up till September, after which the drop in effective temperature results in significant VFA accumulation, that leads to acidification, further suppression of the growth rate and ultimately wash out (as shown in Figure 6.5), in agreement with research conducted by Xing and Wang (2021).

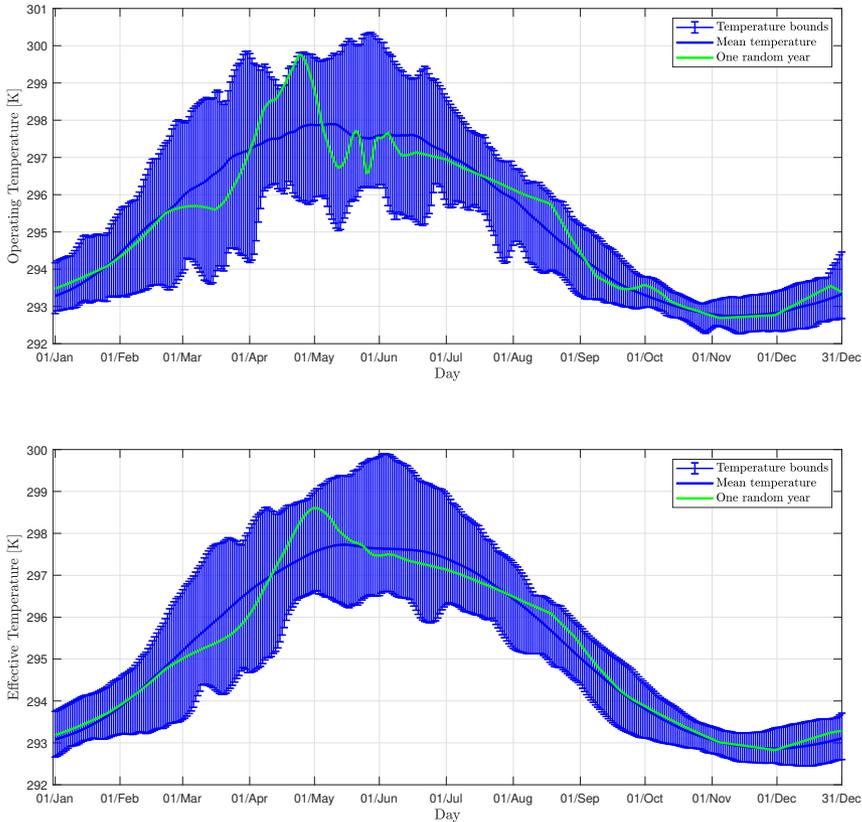


Figure 6.2: Operating temperature (T) and effective temperature ($T_{\text{effective}}$) based on 30 years historical meteorological data given from KNMI and simulated based on the heat network thermal model.

6.4.3. CONTROL SYSTEM ASSESSMENT: DISCUSSION ON MPC CONTROLLER

To assess the performance of control configuration presented in Section 6.3, the dome digester coupled with the proposed MPC controller is simulated in a closed loop. The objective is to maintain the methane (or biogas) production subject to meteorological fluctuations, i.e. varying operating and effective temperatures. Therefore, the performance of the proposed control strategy is evaluated based on its ability to track the assigned set-points accurately. In this regard, three yearly random selected temperature scenarios from the historical data are considered, along with three different set-points for each year: constant set-points for year one ($2600 \text{ m}^3 \text{ d}^{-1}$) and year two ($2300 \text{ m}^3 \text{ d}^{-1}$), and step changes in the set-points for year three. The prediction and control horizons are set to 8 and 5 days, respectively.

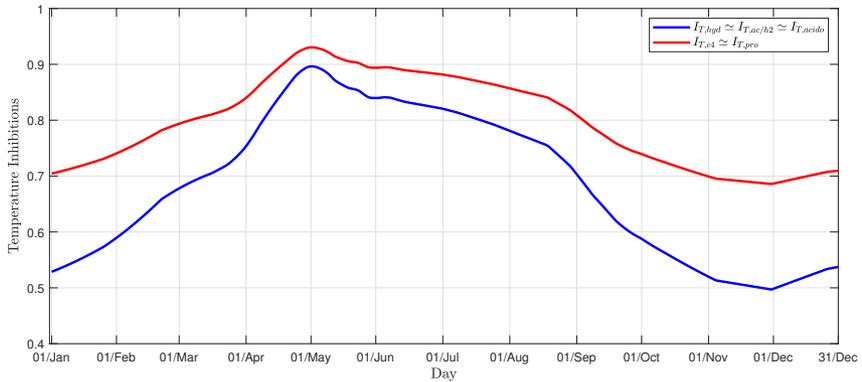


Figure 6.3: Variations of the temperature inhibition functions over a year.

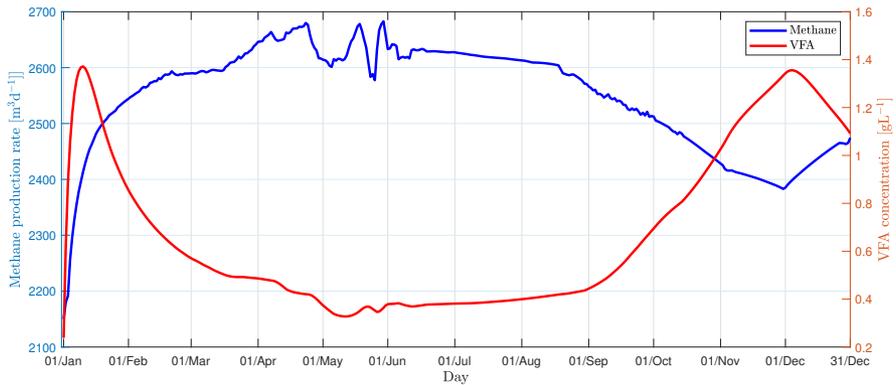


Figure 6.4: Variations of methane production rate and VFA concentration over a year according to the temperature fluctuations.

As shown in Figure 6.6, despite varying temperatures, the MPC controller is able to track the assigned set-points using the first order of internal model expressed by (6.7a) and (6.7b). However, when the set-point is too high (as in year one), the controller encounters spikes at the moments of considerable changes in temperature. In contrast, with a moderate set-point (as in year two) or step-wise set-points across different seasons (as in year three), the methane production rate is smoother. As explained earlier, the motivation of the current study is to use the feeding flow rate as a biological manipulator to control methane production in response to varying temperatures, instead of controlling the temperature directly. Therefore, as shown in Figure 6.6, the control action varies in relation to the temperature (the higher the temperature, the lower the control action required). This outcome was expected, as temperature inhibition reduces methane production when the temperature is

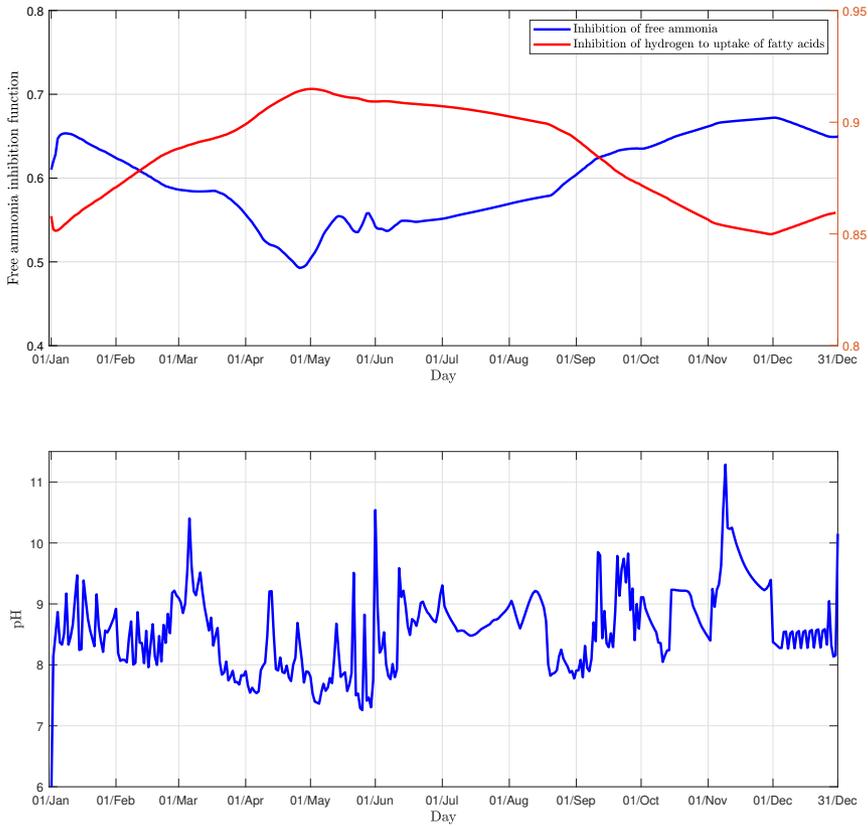


Figure 6.5: Variations of free ammonia inhibition function and inhibition function of hydrogen to uptake of fatty acids (as an example to show the trend of inhibition functions of hydrogen to the uptakes), and pH variation over a year.

low. Consequently, the feeding flow rate should be increased to offset the lower growth rate by supplying more influent. However, increasing the feeding flow rate reduces the hydraulic retention time, which may increase the risk of acidification and washout (Khan *et al.* 2016). On the other hand, as shown and discussed in the previous section, when the temperature is low, there is also a risk of acidification and free ammonia accumulation.

In addition to the proposed MPC controller with either a constant set-point or a reference trajectory, as discussed, the stabilized pH during operation can offset VFA and free ammonia accumulations, and consequently enhance process performance, and robustness (Yuan and Zhu 2016). A constant pH value can be regulated by manipulating the physicochemical processes and hydrogen ions, and adding chemical components that do not affect the biological processes. Therefore, by

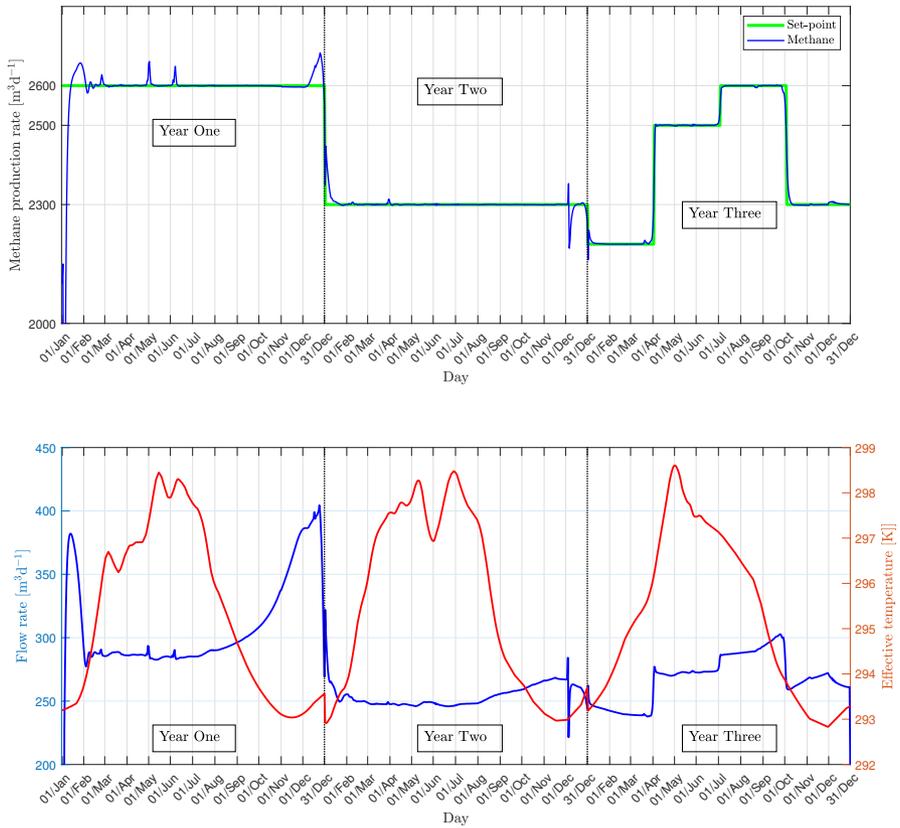


Figure 6.6: MPC set-point tracking subject to varying temperature — process response, and control action.

assuming a constant hydrogen ion concentration of 8×10^{-8} gCOD L^{-1} (i.e. pH = 7.1), the defined digester is simulated over year one to compare the results with the MPC controller that is not integrated with a constant pH. The results are shown in Figure 6.7. As can be seen, the inhibitory effect of free ammonia, which is a major factor preventing an increase in methane production at high temperatures, is drastically reduced. It also reduces the VFA concentration overall, particularly at low temperatures, which diminishes the risk of acidification. Consequently, the feeding flow rate required to maintain methane production (with the set-point assigned to $2600 \text{ m}^3 L^{-1}$) is reduced by almost 20%. This shows the advantages of a constant pH with the proposed MPC strategy.

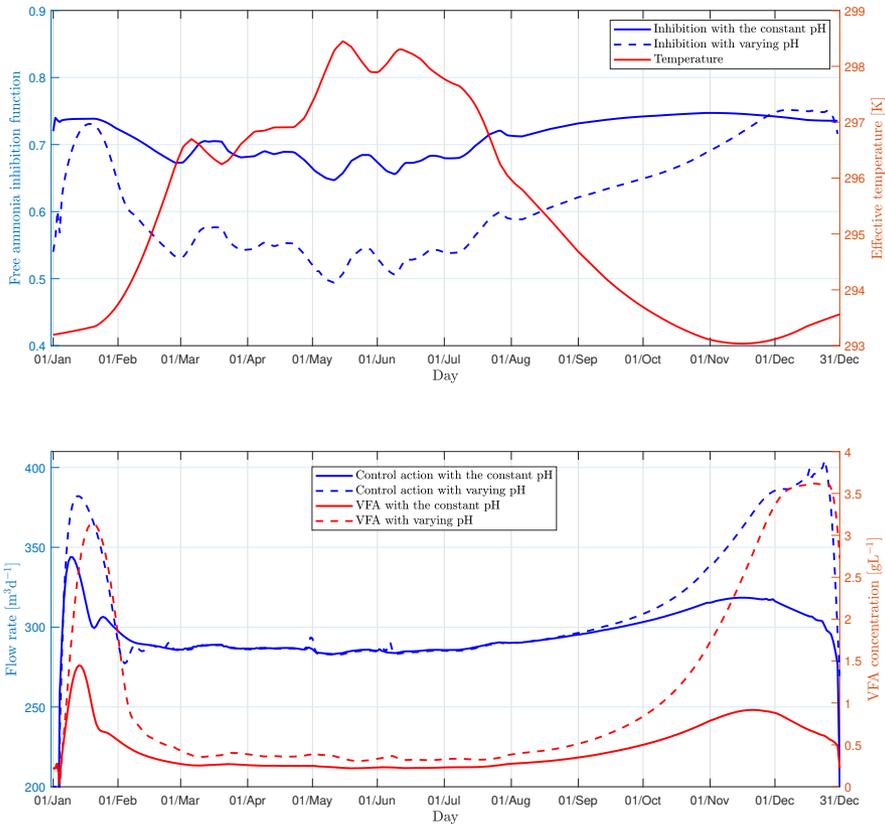


Figure 6.7: MPC with the constant pH and varying pH during operation — comparison of control action, VFA concentration, and free ammonia inhibition.

6.4.4. TRAJECTORY ASSIGNMENT: DISCUSSION ON THE FUZZY LOGIC SYSTEM

As proposed, we can use a reference trajectory assigned by the designed fuzzy logic system instead of a fixed set-point to adjust the production rate according to the effective temperature at each time step to enhance the production rate during warm days and to prevent the risk of washout during cold days. The fuzzy rules, detailed in Table 6.4, are designed in alignment with the biological understanding of microbial activities. In other words, an increase in effective temperature leads to an increase in microbial activity, which in turn enhances biogas production, and vice versa, following a symmetric pattern. Therefore, this can be easily transferred to other similar technologies. Since the system is tuned offline, it relies on the initial set-point for determining the trajectory. In other words, the fuzzy logic system

calculates the adjustment (Δw) to be applied to the next set-point. Consequently, the initial set-point is crucial in shaping the overall trajectory. To assess the process behavior under different initial set-points, open-loop simulations are required. This ensures that the method is independent of skilled operators, if the initial set-point is assigned appropriately.

For this case study, the effective temperature over a 30-year period is illustrated in Figure 6.2. Based on the observed maximum and minimum effective temperatures, the reference temperature (T_{ref}) is defined within a range of 290 to 303 K. To accommodate values beyond these boundaries, trapezoidal membership functions are employed for the Low and High fuzzy rules. The possible range of ΔT_{ref} is determined to be between -0.15 and 0.15, with trapezoidal membership functions extending beyond these limits. Similarly, the range of Δw is set between -50 and 50 $\text{m}^3 \text{d}^{-1}$, as larger changes between two consecutive days are considered impractical. The assigned ranges are symmetrically divided for their respective rules, ensuring balanced coverage. The specifications for the corresponding membership functions are summarized in Table 6.7. As depicted in Figure 6.8, the reference trajectory properly changes at each time step based on the effective temperature and its rate of change. As expected, at higher temperatures, the assigned production rate to be tracked is higher, and vice versa. The controller action also changes accordingly to follow the assigned reference trajectory and to offset the temperature variations. Integrating the fuzzy logic system with the adaptive MPC framework to assign the reference trajectory allows us to reduce the yearly feed by 5%, while still producing nearly the same amount of methane annually.

6

Table 6.7: Degree of membership for the fuzzy logic system aiming to assign the change in the reference trajectory according to the operating temperature.

Fuzzy set	Type	Specification
$T_{\text{effective}}$ ([290 303])		
Low	Trapezoidal	[290 291 293 295]
Medium	Gaussian	[1.5 296.5]
High	Trapezoidal	[298 300 301 303]
$\Delta T_{\text{effective}}$ ([-0.15 0.15])		
Big negative	Trapezoidal	[-0.15 -0.12 -0.125 -0.075]
Negative	Gaussian	[0.02 -0.075]
Zero	Gaussian	[0.04 0]
Positive	Gaussian	[0.02 0.075]
Big positive	Trapezoidal	[0.075 0.125 0.13 0.15]
Δw ([-50 50])		
Big negative	Gaussian	[5 -37.5]
Medium negative	Gaussian	[5 -25]
Negative	Gaussian	[5 -12.5]
Zero	Gaussian	[5 0]
Positive	Gaussian	[5 12.5]
Medium positive	Gaussian	[5 25]
Big positive	Gaussian	[5 37.5]

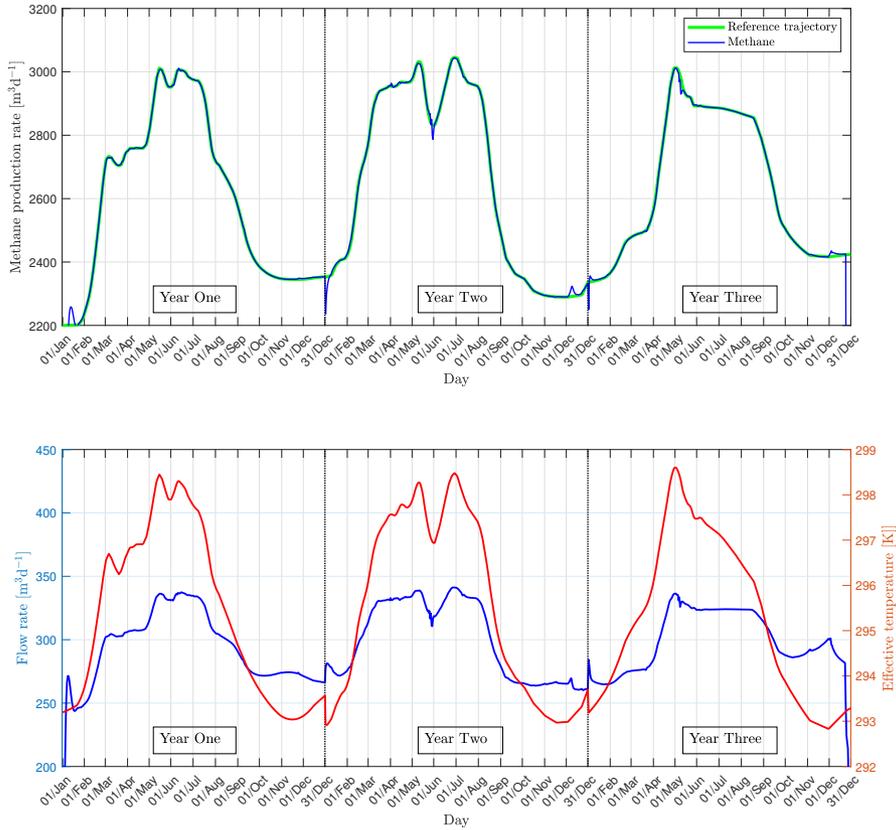


Figure 6.8: MPC with a reference trajectory assigned by the fuzzy logic system according to temperature variations.

6.4.5. BIOGAS MANAGEMENT: DISCUSSION ON BIOGAS-FUELED HEATING STRATEGY

As discussed above, the proposed MPC strategy can handle variations in meteorological conditions by adaptively updating the parameters of the internal linear input-output model. This approach effectively regulates the feeding flow rate to maintain a constant methane production and off-set temperature-change induced inhibition, throughout the year, despite varying operating and effective temperatures. On the other hand, using this control framework can help to estimate the amount of extra feed required to be fed to the digester to produce extra methane and use it to heat the digester. In other words, similar to the reference trajectory assigned by the fuzzy logic system discussed above, a different reference trajectory can be obtained based on the amount of external heating, as given in (6.10). Therefore, the control action calculated is based on this reference trajectory, provides a yearly overview of

the additional influent required for feeding the digester according to temperature fluctuations. In this regard, the reference trajectory in the control framework (w) is the amount of external heating (6.10) at each time step. The exact amount of external heating is dependent on the design of the heating system, which specifies the heating efficiency. However, for the sake of simplification, in this case, it is assumed that the total produced amount of methane can be burned for heating to calculate only the amount of the extra feed.

According to the historical meteorological data and the corresponding simulated operating and effective temperatures depicted in Figure 6.2, the MPC strategy determines the amount of methane required for burning in an external heating system based on $T_{\text{heater}} = 308.15 \text{ K}$ (Ahmadi *et al.* 2023; Batstone *et al.* 2002). The methane requirements for the lower and upper temperature bounds, as well as for a random year are then calculated to assign a reference trajectory based on (6.10). Given the controller calculation, the required feeding flow rate to be fed to the digester to provide the required methane for heating is shown in Figure 6.9. Considering the external heating from extra methane production, the achievable effective temperature is calculated using the heat model and is also depicted in Figure 6.9. It shows that the daily temperature increases by 5° in average, while the temperature range (the difference between the coldest and warmest temperatures) decreases by 1.5° , which brings the digester to the temperature-controlled condition with minor ($\pm 3^\circ$) changes.

6

This self-consumption method can improve process performance and stability in three aspects: (i) yearly feed, (ii) total COD conversion, and (iii) methane production rate. As summarized in Table 6.8 for the simulation of the dome digester and the developed MPC controller with and without an external heating system, using external heating reduces the total yearly required feed for the production of $2600 \text{ m}^3 \text{ d}^{-1}$, even though a portion of the produced biogas is burned for heating. This is due to an increase in COD conversion, as the effective temperature rises with external heating (Figure 6.9), leading to more biogas production and less VFA accumulation. On the other hand, with the increase in the effective temperature, the methane production can also be enhanced by increasing the feeding flow rate without any concern about acidification, free ammonia accumulation, and subsequent washout as can be seen in Table 6.8 (the potential daily methane production rate column). However, the trade-off between COD conversion and enhanced biogas production through increased feeding flow rate should also be considered (Khan *et al.* 2016), although it is beyond the scope of this study.

Thanks to the proposed MPC framework, a comparison of different heating systems in terms of heater temperature can also be investigated. As summarized in Table 6.9, the total required feed for a process with a heater at 308.15 K is almost 4600 m^3 higher compared to a process with a heater at 298.15 K to produce a daily methane production rate of $3000 \text{ m}^3 \text{ d}^{-1}$. However, this extra feeding enhances the process over a year in two ways: (i) improving the conversion of feed to product (with less VFA in the outputs), and (ii) increasing process stability, allowing daily production to increase to $3800 \text{ m}^3 \text{ d}^{-1}$.

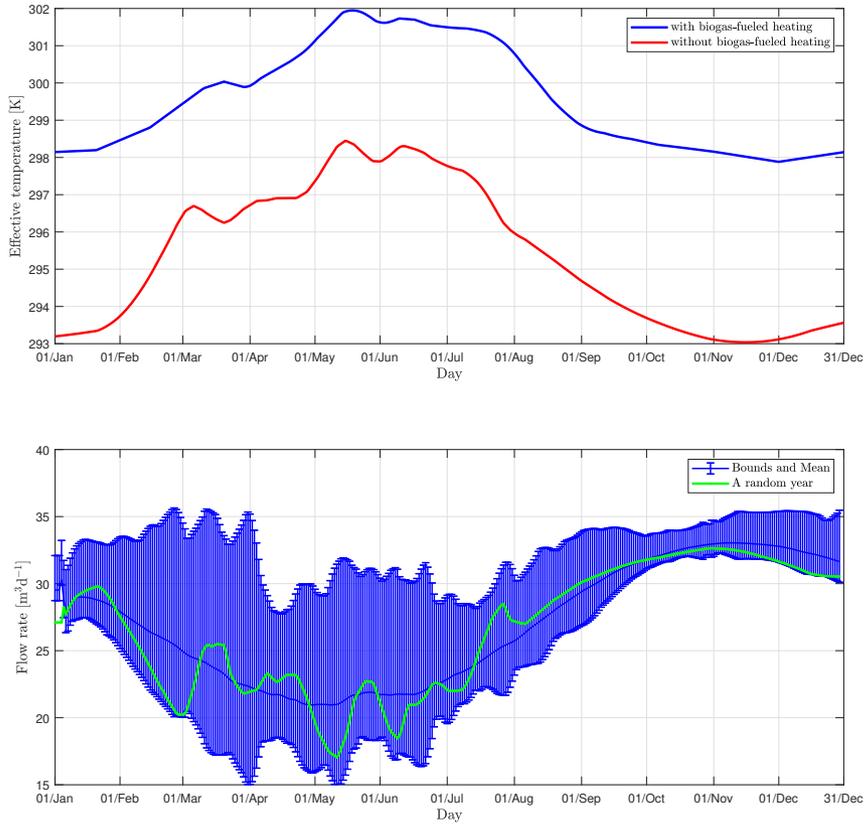


Figure 6.9: MPC with a reference trajectory based in the required external heating to be used as a self biogas-fueled heating system.

Table 6.8: Comparison of the proposed MPC approach for calculating the required feed and the maximum methane production rate with and without an external heating system for a random year; the daily potential methane production can be increased by 46%, with the almost the same amount of yearly feed.

External heating	Yearly extra feed (m ³)	Yearly feeding for daily production of 2600 m ³ d ⁻¹ (m ³)	Total required feed (m ³)	Potential daily methane production rate (m ³ d ⁻¹)
Applied	9.5616e3	1.0156e5	1.1112e5	≤ 3800
Not applied	-	1.1159e5	1.1159e5	≤ 2600

6.4.6. IMPACTS AND IMPLICATIONS OF THE PROPOSED METHODOLOGY

Using the proposed MPC approach integrated with either a fuzzy logic system for assigning a reference trajectory or a self-sustaining biogas-fueled heating system,

Table 6.9: Comparison of the proposed MPC approach for calculating the required feed and the maximum methane production rate with external heating for a random year for two different heater temperatures; increasing heater temperature increases the daily potential of methane production by 26% with less than 3% of increase in the yearly feed.

Heater temperature (K)	Yearly extra feed (m ³)	Yearly feeding for daily production of 3000 m ³ d ⁻¹ (m ³)	Total required feed (m ³)	Potential daily methane production rate (m ³ d ⁻¹)
308.15	9.5616e3	1.2018e5	1.2970e5	≤ 3800
298.15	1.9139e3	1.2385e5	1.2508e5	≤ 3000

not only can address temperature fluctuations, but also enhance the production efficiency. It then highlights the importance and the efficiency of an integrated approach that combines MPC control of the flow rate without an external temperature control approach for AD designers and process supervisors. Such an automatic control system can also be used for processes supervised by unskilled operators. By online measurements, the control system can adapt its parameters to varying conditions and regulate the control action to stabilize and enhance the production rate. It is then proposed to apply and validate this novel temperature-wise extended model as well as the integrated automatic control framework to full-scale plants and in different geographic locations or with different digester geometries, even if they are supervised by unskilled operators as a topic for future work. It should also be pointed out that this investigation depends on meteorological conditions, which determines operating and effective temperature profiles, and therefore needs to be investigated for each specific scenario individually. In this regard, the proposed control scheme relies on expert knowledge only for the pre-analysis of historical meteorological data, initializing the fuzzy logic system, and determining an appropriate temperature for the self-heater. However, during operation, it is fully automatic, utilizing the adaptive manner to handle variations and disturbances, which ensures that the process can be effectively managed.

6

6.5. CONCLUSIONS

In this chapter, the anaerobic digestion model no.1 (ADM1) has been mechanistically extended in order to incorporate temporal temperature variations caused by meteorological fluctuations. The extended model demonstrates reliable outcomes in general agreement with experimental studies. On the other hand, a feeding flow rate control strategy based on MPC approach has also been proposed to deal with varying meteorological conditions and to maintain a stable methane production rate. This method can be employed in the absence of any external heating system, as it regulates the feeding flow rate to compensate for changes in the temperature. To enhance the productivity of the process under these conditions, a fuzzy logic system has been employed to assign a reference trajectory for the methane production rate for the MPC controller. This fuzzy logic system can adjust the reference trajectory to increase the production rate when the temperature rises and to reduce it when

the temperature drops, thereby enhancing the process performance and avoiding operational failures. Thanks to the extended model and proposed control strategy, it has been also demonstrated that the production rate can be increased, if the pH value is fixed to deal with free ammonia and VFA accumulation. This strategy shows improvements in conversion by reducing the required feeding flow rate. Additionally, the adaptive MPC framework enables the calculation of the required extra feed to produce more methane to be used for a self-consumption biogas-fueled heating system in order to increase process performance and stability for a fixed set-point. The effectiveness of the proposed control framework has been assessed using a defined dome digester under climate conditions in the Netherlands.

REFERENCES

- Ahmadi, A., M. Avila, and L. Barna (Sept. 2023). "Pathways for the thermally optimal design and practice of anaerobic digestion in large-scale biogas plants: Heat transfer modeling and energy analysis". In: *Chemical Engineering Research and Design* 197, pp. 884–907. ISSN: 0263-8762.
- Ahmed, W. and J. Rodríguez (2020). "A model predictive optimal control system for the practical automatic start-up of anaerobic digesters". In: *Water Research* 174, p. 115599. ISSN: 0043-1354.
- Anand, K., A. P. Mittal, and B. Kumar (2021). "Modelling and simulation of dual heating of substrate with centralized temperature control for anaerobic digestion process". In: *Journal of Cleaner Production* 325, p. 129235. ISSN: 0959-6526.
- (2022). "ANN-based sensorless adaptive temperature control system to improve methane yield in an anaerobic digester". In: *Biomass Conversion and Biorefinery* 13.8, pp. 7265–7285. ISSN: 2190-6823.
- Batstone, D., J. Keller, I. Angelidaki, S. Kalyuzhnyi, S. Pavlostathis, A. Rozzi, W. Sanders, H. Siegrist, and V. Vavilin (2002). "The IWA Anaerobic Digestion Model No 1 (ADM1)". In: *Water Science and Technology* 45.10, pp. 65–73.
- Bergland, W. H., C. Dinamarca, and R. Bakke (2015). "Temperature Effects in Anaerobic Digestion Modeling". In: *Linköping Electronic Conference Proceedings*. Linköping University Electronic Press.
- Calise, F., F. L. Cappiello, L. Cimmino, M. Dentice d'Accadia, and M. Vicidomini (2023). "Dynamic analysis and investigation of the thermal transient effects in a CSTR reactor producing biogas". In: *Energy* 263, p. 126010. ISSN: 0360-5442.
- Camacho, E. F., C. Bordons, E. F. Camacho, and C. Bordons (2007). *Constrained model predictive control*. Springer.
- Caposciutti, G., A. Baccioli, L. Ferrari, and U. Desideri (2020). "Biogas from Anaerobic Digestion: Power Generation or Biomethane Production?" In: *Energies* 13.3, p. 743. ISSN: 1996-1073.
- Cengel, Y. A. and A. Ghajar (2011). "Heat and mass transfer (a practical approach, SI version)". In: *McGraw-670 Hill Education* 671.52, p. 964.
- Cysneiros, D., C. J. Banks, S. Heaven, and K.-A. G. Karatzas (2012). "The effect of pH control and 'hydraulic flush' on hydrolysis and Volatile Fatty Acids (VFA) production and profile in anaerobic leach bed reactors digesting a high solids content substrate". In: *Bioresource Technology* 123, pp. 263–271. ISSN: 0960-8524.
- De Barros, L. C., R. C. Bassanezi, and W. A. Lodwick (2016). *First Course in Fuzzy Logic, Fuzzy Dynamical Systems, and Biomathematics*. Springer.
- Dittmer, C., B. Ohnmacht, J. Krümpel, and A. Lemmer (2022). "Model Predictive Control: Demand-Orientated, Load-Flexible, Full-Scale Biogas Production". In: *Microorganisms* 10.4, p. 804. ISSN: 2076-2607.
- Donoso-Bravo, A., W. Bandara, H. Satoh, and G. Ruiz-Filippi (2013). "Explicit temperature-based model for anaerobic digestion: Application in domestic wastewater treatment in a UASB reactor". In: *Bioresource Technology* 133, pp. 437–442. ISSN: 0960-8524.
- Donoso-Bravo, A., C. Retamal, M. Carballa, G. Ruiz-Filippi, and R. Chamy (2009). "Influence of temperature on the hydrolysis, acidogenesis and methanogenesis in mesophilic anaerobic digestion: parameter identification and modeling application". In: *Water Science and Technology* 60.1, pp. 9–17.
- Erdirencelebi, D. and G. M. Ebrahimi (2022). "Enhanced sewage sludge treatment via parallel anaerobic digestion at the upper mesophilic level". In: *Journal of Environmental Management* 320, p. 115850. ISSN: 0301-4797.
- Garkoti, P., J.-Q. Ni, and S. K. Thengane (2024). "Energy management for maintaining anaerobic digestion temperature in biogas plants". In: *Renewable and Sustainable Energy Reviews* 199, p. 114430. ISSN: 1364-0321.
- Ghanavati, M. A., E. Vafa, and M. Shahrokhi (July 2021). "Control of an anaerobic bioreactor using a fuzzy supervisory controller". In: *Journal of Process Control* 103, pp. 87–99.
- Goodwin, G. C. and K. S. Sin (2014). *Adaptive filtering prediction and control*. Courier Corporation.
- Hreiz, R., N. Adouani, Y. Jannot, and M.-N. Pons (2017). "Modeling and simulation of heat transfer phenomena in a semi-buried anaerobic digester". In: *Chemical Engineering Research and Design* 119, pp. 101–116. ISSN: 0263-8762.
- Khan, M., H. Ngo, W. Guo, Y. Liu, L. Nghiem, F. Hai, L. Deng, J. Wang, and Y. Wu (2016). "Optimization of process parameters for production of volatile fatty acid, biohydrogen and methane from anaerobic digestion". In:

- Bioresource Technology* 219, pp. 738–748. ISSN: 0960-8524.
- Kovalovszki, A., L. Treu, L. Ellegaard, G. Luo, and I. Angelidaki (2020). “Modeling temperature response in bioenergy production: Novel solution to a common challenge of anaerobic digestion”. In: *Applied Energy* 263, p. 114646.
- Kumar, M., D. Prasad, B. S. Giri, and R. S. Singh (2019). “Temperature control of fermentation bioreactor for ethanol production using IMC-PID controller”. In: *Biotechnology Reports* 22, e00319. ISSN: 2215-017X.
- Lee, J. (2017). “Development of a model to determine mass transfer coefficient and oxygen solubility in bioreactors”. In: *Heliyon* 3.2, e00248.
- Mei, R., T. Narihiro, M. Nobu, and W.-T. Liu (2016). “Effects of heat shocks on microbial community structure and microbial activity of a methanogenic enrichment degrading benzoate”. In: *Letters in Applied Microbiology* 63.5, pp. 356–362. ISSN: 0266-8254.
- Mendel, J. M. (1995). “Fuzzy logic systems for engineering: a tutorial”. In: *Proceedings of the IEEE* 83.3, pp. 345–377.
- Metcalf, L., H. P. Eddy, and G. Tchobanoglous (1991). *Wastewater engineering: treatment, disposal, and reuse*. Vol. 4. McGraw-Hill New York.
- Mohr, P. J. and B. N. Taylor (2005). “CODATA recommended values of the fundamental physical constants: 2002”. In: *Reviews of modern physics* 77.1, p. 1.
- Moradvandi, A., E. Abraham, A. Goudjil, B. De Schutter, and R. E. Lindeboom (2024). “An identification algorithm of switched Box-Jenkins systems in the presence of bounded disturbances: An approach for approximating complex biological wastewater treatment models”. In: *Journal of Water Process Engineering* 60, p. 105202. ISSN: 2214-7144.
- Nguyen, D., S. Nitayavardhana, C. Sawatdeenarunat, K. Surendra, and S. K. Khanal (2019). “Biogas Production by Anaerobic Digestion: Status and Perspectives”. In: *Biofuels: Alternative Feedstocks and Conversion Processes for the Production of Liquid and Gaseous Biofuels*. Elsevier, pp. 763–778. ISBN: 9780128168561.
- Nie, E., P. He, H. Zhang, L. Hao, L. Shao, and F. Lü (2021). “How does temperature regulate anaerobic digestion?” In: *Renewable and Sustainable Energy Reviews* 150, p. 111453.
- Pedersen, S. V., J. Martí-Herrero, A. Singh, S. Sommer, and S. Hafner (2020). “Management and design of biogas digesters: A non-calibrated heat transfer model”. In: *Bioresource Technology* 296, p. 122264.
- Prakash, D., S. S. Chauhan, and J. G. Ferry (2019). “Life on the thermodynamic edge: Respiratory growth of an acetotrophic methanogen”. In: *Science Advances* 5.8.
- Robles, A., G. Capson-Tojo, M. Ruano, E. Latrille, and J.-P. Steyer (2018). “Development and pilot-scale validation of a fuzzy-logic control system for optimization of methane production in fixed-bed reactors”. In: *Journal of Process Control* 68, pp. 96–104. ISSN: 0959-1524.
- Rosso, L., J. Lobry, and J. Flandrois (1993). “An Unexpected Correlation between Cardinal Temperatures of Microbial Growth Highlighted by a New Model”. In: *Journal of Theoretical Biology* 162.4, pp. 447–463.
- Thamsiriroj, T. and J. Murphy (2011). “Modelling mono-digestion of grass silage in a 2-stage CSTR anaerobic digester using ADM1”. In: *Bioresource Technology* 102.2, pp. 948–959.
- Xing, B.-S. and X. C. Wang (2021). “High-rate mesophilic co-digestion with food waste and waste activated sludge through a low-magnitude increasing loading regime: Performance and microorganism characteristics”. In: *Science of The Total Environment* 777, p. 146210. ISSN: 0048-9697.
- Yenigün, O. and B. Demirel (2013). “Ammonia inhibition in anaerobic digestion: A review”. In: *Process Biochemistry* 48.5–6, pp. 901–911. ISSN: 1359-5113.
- Yuan, H. and N. Zhu (2016). “Progress in inhibition mechanisms and process control of intermediates and by-products in sewage sludge anaerobic digestion”. In: *Renewable and Sustainable Energy Reviews* 58, pp. 429–438. ISSN: 1364-0321.

7

SUMMARY OF FINDINGS AND FUTURE DIRECTIONS

7.1. OVERALL OBJECTIVE AND RESEARCH SCHEME

EVALUATION

This thesis addresses the central research question: *What tools and approaches can be employed to advance resource recovery from wastewater, and how do these tools contribute to achieving the desired outcome for PPB raceway reactors and anaerobic digesters?* To answer this question, we have explored two key technologies: anaerobic digestion (a well-established process) and PPB raceway reactors (a promising emerging technology). Through the application of three main approaches—mechanistic modeling, hybrid system identification, and adaptive predictive control—we have shown how these tools contribute to advancing resource recovery. Mechanistic modeling has provided a detailed and foundational understanding of the processes, establishing the groundwork for further exploration. Hybrid system identification has demonstrated how simpler, low-order models can effectively forecast system behavior and support dynamic adjustments. Finally, adaptive predictive control has shown its ability to ensure robust and efficient operation, optimizing performance despite operational disturbances. Together, these approaches form an integrated framework that advances the efficiency and sustainability of wastewater resource recovery, offering practical solutions for both current and future challenges.

A research scheme comprising four subprojects, namely mechanistic model, prediction model, control configuration, and MPC Model, has been also defined. This approach, which ranges from modeling to control, establishes a proven research trajectory applicable to various processes in resource recovery from wastewater. In summary, if one has a process, they should first develop a mechanistic model to understand its dynamics and behavior, providing a virtual benchmark for simulation. Then, this understanding can be used to develop a supervisory layer that manages the performance of the MPC model based on the process characteristics. Additionally, for tasks such as prediction, data reconciliation, or model-based control, one will need to employ a simplified input-output model. The novelty and contributions of this work lie both in developing new approaches and in adapting existing methods to the specific cases under consideration. In the following sections, the findings for each research direction, as well as potential further explorations for ADs and PPB raceway reactors and resource recovery technologies in general, are summarized.

7

7.2. MECHANISTIC MODELING

7.2.1. PURPLE BACTERIA MODEL

The Purple Bacteria Model (PBM), which has been developed in this PhD thesis, is the third model proposed for describing PPB dynamics, but the first one specifically for raceway reactors. Besides the model structure, the main novelty of the PBM is its mechanistic-metabolic growth contribution modeling, which introduces an empirical parallel growth constant for PPB dynamical selection. This constant accounts for the contribution of other growth pathways in addition to the dominant one, resulting in more accurate predictions for PPB. This is the first time a method has been proposed for including parallel growth, opening up new possibilities for alternative approaches

for inclusion of the microbial parallel growth pathways and the microbial community complexity. The developed PBM model in a mechanistic manner provides clear understanding of microbial mechanisms and process interactions, yielding a reliable virtual model for assessment of process control approaches. The structure of the model allows for controlling the complexity of the model as needed.

As a topic for future work, one alternative for inclusion of parallel growth pathways could involve defining an inhibition function, although identifying an appropriate inhibitor indicator may be challenging. The concept of dividing a state variable based on different growth pathways and integrating them into a more precise mechanistic representation in the model can also be validated for other microbial communities and processes. However, experimental analysis and verification may not be straightforward. This idea can enhance the understanding of process behaviors. When adding additional state variables to enhance process understanding, the model becomes more complex. In the current version of the PBM developed in this PhD thesis, including other species could be considered, while the inclusion of dynamical pH and temperature can also be taken into account in future work, specifically due to their importance from the process control point of view. On the other hand, since the PBM is highly nonlinear and complex, it can only be considered for dynamic analysis and process simulation. Simplifying it, either via the use of PCA and model order reduction methods or in a mechanistic manner (first-principles-based), may yield a simplified version with 5 or 6 state variables. Such simpler models could be used for model-based control design, as it would be less complex while computationally feasible.

7.2.2. EXTENDED TEMPERATURE-WISE ADM1

From both process understanding and process control perspectives, the original ADM1 model has been mechanistically extended to account for temperature variability. This has been achieved by introducing a temperature inhibition function based on the Cardinal model. The model has also been integrated into a heat transfer network, thereby reflecting varying meteorological conditions into the operational temperature. This extended temperature-wise ADM1, integrated with a heat network, can predict product production rates and yields with satisfactory accuracy in line with known biological responses, and provides mechanistic overviews on inhibition factors that affect performance.

As a topic for future work, the Arrhenius model could also serve as an alternative for defining the novel temperature inhibition function. The main challenge in properly defining a temperature inhibition function, whether using the Cardinal or Arrhenius model, is the calibration and verification through dedicated experiments for various microbial communities. Therefore, through sensitivity analysis and parameter identification, required experiments can be defined and conducted to properly calibrate the model.

7.3. HYBRID SYSTEM IDENTIFICATION

7.3.1. STATE-OF-THE-ART HYBRID SYSTEM IDENTIFICATION

A survey has been conducted as a part of this PhD thesis, revealing state-of-the-art in hybrid system identification research. This systematic survey highlights two aspects of hybrid system identification, i.e., parametrizations (how to model) and methods (how to identify). In terms of general contributions to this field, there are numerous potential options that our review highlighted, while application-based contributions are relatively scarce. This systematic survey outlines which parameterizations and effective associated parameter identification methods have been studied, highlighting areas for further advancement and improvement. By focusing on application-based hybrid system identification, both the parameterization approaches and the corresponding parameter identification approaches can be specifically defined and developed to meet practical requirements.

As a topic for future work, a survey on hybrid system modeling—combining mechanistic components with input-output modeling for parts that are challenging to represent mechanistically—can be conducted, particularly in the context of biological processes. This approach can be applied to model dynamical pH in the PBM. It should be highlighted that describing physico-chemical processes is more straightforward with input-output modeling than it is for biological processes, where data can be collected more easily through experiments for the identification procedure.

7.3.2. APPROXIMATING BIOPROCESS USING SYSTEM IDENTIFICATION

Identification of switched Box-Jenkins systems in the presence of bounded disturbances, which has been discussed in this PhD thesis, can be considered as an application-based identification problem. Representing a complex process using a switched Box-Jenkins model is advantageous for several reasons. Firstly, it allows for the consideration of multiple linear subsystems, enabling the system to switch between them and potentially increase prediction accuracy, while mitigating model complexity. Secondly, it accommodates different dynamics for the relationships between input-to-output and disturbance-to-output, which can be biologically explainable. Lastly, it assumes bounded disturbances, a more practical and reliable assumption on disturbances, as stochastic information on disturbances is not always readily available for biological and environmental factors. This approach increases prediction accuracy. However, depending on the specific application—such as data assimilation, prediction, or model-based control—simpler switched systems may be sufficient to meet the requirements, and these models can be derived from the switched Box-Jenkins model.

As a topic for future work, further exploration in terms of modeling switching rules can be taken into account. Considering methods that divide the operational space in a biologically meaningful way, such as piecewise affine systems, can also be investigated. However, identifying a comprehensive operational space for this purpose requires further research in both experimental and modeling aspects. In other words, although a bioprocess might be complex and involve different biological mechanisms in a hybrid manner by its inherent nature, this complexity may not always be reflected in the numerical data used to model the process.

7.4. ADAPTIVE PREDICTIVE CONTROL

7.4.1. CONTROL OF PPB RACEWAY REACTORS

The developed control configuration based on the MPC control approach is the first automatic control system for PPB cultivation, and particularly PPB in raceway reactors. Due to the specific features of the raceway reactor configuration and operation, as well as the characteristics of PPB, the MPC control strategy is integrated with a supervisory layer. This supervisory layer accounts for an operational strategy based on the given requirements, whether it would be production rate or yield. This combination of MPC control and a supervisory decision-making layer optimizes the use of available resources, such as light and carbon sources, to convert them to PPB. Even with potential biological and meteorological fluctuations and incomplete knowledge of PPB growth pathway contributions, this control system configuration ensures effective resource utilization and stable process operation. The model used for the MPC controller is a linear input-output model, which is updated adaptively to capture accurate dynamics and perturbations during operation.

As a topic for future work, instead of adaptive MPC control, nonlinear MPC based on a simplified nonlinear model of the process (as mentioned in Section 7.2.1) can also be investigated. This approach allows the inclusion of resource availabilities into the MPC objective function, as their dynamics can be modelled. In such a case, designing an observer to estimate state variables for which online measurements are not available, may also be necessary. On the other hand, investigating different approaches to drive a decision-making layer, not only in terms of design but also by using methods such as fuzzy logic and neural networks, can be considered as further improvement for this system.

7.4.2. CONTROL OF ANAEROBIC DIGESTION SUBJECT TO METEOROLOGICAL FLUCTUATIONS

The proposed adaptive MPC controller in this thesis can effectively handle meteorological fluctuations and maintain the production rate. Beside the robust operation by using the MPC controller, the production rate can also be adjusted according to temperature changes in order to enhance the production rate when the temperature rises, and to ensure the stable production and avoid any operational failures when the temperature drops. Therefore, an appropriate reference trajectory should be implemented according to the expert knowledge. In this regard, a fuzzy logic system is employed to mimic the expert knowledge based on the temperature and its rate of change to assign a value for a change in the reference trajectory. Moreover, a self-consumption biogas-fueled external heating system has been proposed in order to enhance the overall production and performance. The required extra feed for this external heating can be calculated based on the proposed MPC framework. Finally, the effect of a fixed value for pH has been also studied and according to the enhancement of the production, the integration of MPC controller with a physical preventive inhibition mechanism has been suggested.

As a topic for further development, the mechanism for stabilizing pH during operation can be formulated as a pH control problem that an effective control method can be then designed for. Moreover, although the two aforementioned approaches—namely assign-

ing an appropriate reference trajectory and integration of a self-consumption biogas-fueled heating system—have shown improvement for the process performance and robustness, the thermally optimal digester design can also be investigated by defining a multi-objective optimization problem. In addition, a double-layer fuzzy logic system to take the volatile fatty acid concentration into account for a more reliable supervisory layer to avoid the process washout can be considered for further development as well.

7.5. ADVANCING RESOURCE RECOVERY FROM WASTEWATER VIA MODELING AND CONTROL APPROACHES

As discussed in this PhD thesis, resource recovery from wastewater can be enhanced through modeling and control approaches. This encourages us to focus on improving existing technologies rather than proposing entirely new solutions. A promising topic for future research could be the development of a comprehensive benchmark that integrates various reactors within a single plant and simulates control strategies for them, known as plant-wide modeling and control. Such a virtual simulation benchmark facilitates the investigation of potential perturbations in processes that might compromise operational efficiency and expected outcomes. This research idea can be explored for the integration of PPB and anaerobic digestion based on modeling and control system configurations developed in this thesis, while other resource recovery technologies suffering from inefficiency can be considered as well.

ACKNOWLEDGEMENTS

As my PhD journey comes to an end, I cannot help but reflect on the incredible mix of surprises, ups and downs, adventures, and life-changing milestones I have experienced along the way. And when I say life-changing, I mean much more than academic achievements. The most precious things I have gained during this journey are the soft skills I have developed, which have truly shaped who I am today. It is said that when one is on a long road, it is easy to forget to enjoy the ride because they are so focused on reaching the destination. The PhD journey is much the same—stress and pressure can often overshadow the value of the moment as it is happening. But if there is one lesson I have learned, it is the importance of enjoying and valuing every moment along the way, as that has been the greatest reward of this journey—more meaningful than any degree ever could be. So, to keep a few special moments and names forever in my heart, and to say a big thank you to those who were around me on this PhD journey, I have written some words here.

First and foremost, my gratitude goes to my supervisory team—these incredible individuals who gave me the chance to step out of my comfort zone, make mistakes, and grow. Before applying for a PhD abroad, my former supervisor advised me on two key things: "the weather of the destination and the attitude of the prospective supervisor—everything else can be managed, but these two factors have a huge impact on your life." After almost four years, I can now understand what he said. I was very lucky to work with **Ralph**, **Edo**, and **Bart**; **Ralph**, I could not have imagined a more supportive supervisor than you; **Edo**, your kindness in every situation is something I will always remember; and **Bart**, your attention to details in everything taught me the invaluable lesson of being meticulous in my work. I would say when your supervisors understand you, stand by you, and are transparent and easy-going, you do not just get through your PhD—you truly live it.

Then, it comes to people around me in Water Management department. Let me start with the fantastic 4.59 office—where my life at TU Delft truly began. It all started when I moved there and met **Diana**; we always had something to talk about. Then one day, **Asif** came to decide whether to move to our office—that was his best decision ever he made in his life; becoming a real "friend" to me. Then, **Diego** joined us, bringing his life-changing philosophies. **Zhaxu** came next as a guest professor, managing to travel around Europe in one year, more than I did in four. After returning from last year's New Year holiday, **Jing** along with her sociable husband, and **Greg**, a real "Liverpool" fan who brought a football vibe to the office, were my new-year surprises. I also count **Arash** as part of our office crew, as he often joined us with his endless passion and discussions about machine learning. With these incredible people, I had great moments, nice coffee and lunch chats, life-research-work discussions, and so many cultural exchanges.

This department is filled with wonderful people. Before I even arrived, **Rielle** and

Mariska went out of their way to ensure everything went smoothly—**Mariska** even saved me from getting lost by booking a taxi at the last moment. **Bahareh** and **Aashna**, were the only ones who replied to my introduction emails, making me feel welcome from the start. **Hamed, Saqr, Pamela, Antonella**, and **Henry** advised me to navigate the unknown PhD culture. As an introvert, I tend to keep to myself, so friendships find me rather than the other way around. Making friendship was even harder as it was the Covid time and everyone was working from home, and no one was around. But I cherish the good chats with **Alex, Iosif, Job, Rifki, Bilal, Sumit, Ali**, and later with **David, Suham, Fatima, Bugra, Derya, Ibrahim**, and **Ja-Ho**. I also enjoyed the days with **Casper, Sjoerd** working on their Master projects. But, a turning point for me was when **Alireza** came, an honest friend—when Edo told me that he had hired an Iranian PhD, I never imagined that we would become so close with so many things to talk about.

Working on a European-Indian project was a nice experience. Our bi-yearly consortiums let me travel to Kolkata, Mumbai, Perth, and Sevilla. I also had an opportunity to know a bunch of great and professional people in this project; **Markus**, a humble project leader; Professor **Makarand, Atul**, and **Manikanta**, great people of Pilot 5; Professor **Anju**, and Dr **Isabel**, kindest people in our project; **Giorgio, Abbas, David, Siegfried**, professional collaborators; and **Andre, Ivar, Tenno, Hanson, Tim**, great people to hang out with after our long days.

Next, I want to acknowledge my friends, who we have shared time and thoughts, laughed and argued, learned and grown together. I am so blessed to have had such people in my life. **Alireza**, my brilliant companion, has been an honest, supportive, and scientific-constructive friend, even from afar—I will never forget the night he and **Amir-mohammad** surprised me at the airport before I left Iran. **Hossein**, who always called to check how I was doing. **Ali, Reza, Hasan, Rasool**, and **Amir**, kept me connected to Iran with their calls and messages. I should also mention **Mohammad Amin**, who generously shared his experiences of working on a similar project with me, which significantly shaped the direction of my thesis. Once in Delft, I tried to form a group of like-minded friends for deep conversations. **Hasan, Mahdi, Arash**, and **Mohammad** were among the first. Then, out of nowhere, **Ali** came into my life and introduced me to a weekly event where I had good times and met some kind people—**Ali, Ali, Mohammad, Ali, Ali, Hossein, Ashkan, Mahdi, Ehsan**, and a few more kind individuals (that is funny how many friends I have with the same name as me!). Particularly with **Ali** and **Ehsan**, I spent most of my time here and had a bittersweet mix of memories.

Last but not least, it all comes down to family. Family means everything, and being far from them has only made that more clear. **Maman, Pedar**—my parents—**Adeleh, Saeideh**—my sisters—and little **Adrina**, my niece, you are the world to me. **Maman**, you are in every breath and every thought—I know you are always praying for me. And **Pedar**, I cannot breathe without you for even a moment—you always said I should have gotten my PhD and here it is. My sisters, **Adeleh** and **Saeideh**, you have always been so supportive, and constantly been pushing me forward; very grateful to have you. I love you all so much, and no matter what life brings our way, I hope we can reunite one day.

As last words; I might be asked about the differences between a life **with PhD** and **without PhD**. Will a title of “Doctor” change my life? Some may have contributed valuable work at the cutting edge of science. Others might take pride in their achievements

and publications in prestigious journals. And some might like this title itself. But I am grateful for the broader worldview I have gained, for stepping out of my comfort zone, for new experiences, and for the people I have met along the way. "Soar to see, not to be seen"; this PhD journey has helped me rise to gain a clearer perspective in my life. I am not perfect, and neither is this dissertation. I have come to understand that perfection is an ideal, not always attainable, no matter how much we strive for it. Yet, I have no regrets because this experience has taught me invaluable lessons. One of the most profound lessons is captured in a poem, which also serves as a complete response to the first question. To truly feel its depth, you need to know Farsi, as no translation can fully capture its profound meaning. But there is one couplet:

"Put aside yourself, and you shall see,
that reality holds more to be,
Lose the ego, and in that state,
many truths shall then await."

تا شدم بی خبر از خویش، خبرها دارم
بی خبر شو که خبرهاست در این بی خبری
تا شدم بی اثر، از ناله اثرها دیدم
بی اثر شو که اثرهاست در این بی اثری
تا زدم لاف هنر خواجه به هیچم نخرید
بی هنر شو که هنرهاست در این بی هنری
سرو آزاد شد آن دم که ثمر هیچ نداد
بی ثمر شو که ثمرهاست در این بی ثمری
تا سر خود نسپردیم به خاک در دوست
خاطر آسوده نگشتیم از این دربه دری

Ali
Autumn 2024—Delft

CURRICULUM VITÆ



Ali Moradvandi was born on December 12, 1991, in Tehran, Iran. He earned his Bachelor's degree in Chemical Engineering from Amirkabir University of Technology. With a curiosity about Systems & Control, he pursued a Master of Science at Sharif University of Technology, specializing in Modeling, Simulation, and Control, with a focus on adaptive control in the Department of Chemical Engineering. Ali continued his academic journey by pursuing a PhD at Delft University of Technology, where his research centered on applied control systems focused on advancing resource recovery technologies through mechanistic modeling, hybrid system identification, and adaptive predictive control. This multidisciplinary project

provided him with an invaluable opportunity to explore both the fields of Chemical Engineering and Systems & Control. His research interest includes model predictive control, adaptive control, Lyapunov stability, system identification, and their applications to biological and chemical processes.

LIST OF PUBLICATIONS

Peer-reviewed publications inside the thesis scope

- **Moradvandi A.**, De Schutter B., Abraham E., Lindeboom R.E.F., (under review), Model predictive control of purple bacteria in raceway reactors: Handling microbial competition, disturbances, and performance, *Computers and Chemical Engineering*.
- **Moradvandi A.**, Heegstra S., Ceron-Chafla P., De Schutter B., Abraham E., Lindeboom R.E.F., (under review), Model predictive control of feed rate for stabilizing and enhancing biogas production in anaerobic digestion under meteorological fluctuations, *Journal of Process Control*.
- **Moradvandi A.**, Abraham E., Goudjil, A., De Schutter B., Lindeboom R.E.F., (2024), An identification algorithm of switched Box-Jenkins systems in the presence of bounded disturbances: An approach for approximating complex biological wastewater treatment models, *Journal of Water Process Engineering*.
- Alloul A.¹ **Moradvandi A.**¹, Puyol D., Molina R., Gardella G., Vlaeminck S.E., De Schutter B., Abraham E., Lindeboom R.E.F., Weissbrodt D., (2023), A novel mechanistic modelling approach for microbial selection dynamics: Towards improved design and control of raceway reactors for purple bacteria, *Bioresource Technology*.
¹Equally first authorship.
- **Moradvandi A.**, Lindeboom R.E.F., Abraham E., De Schutter B., (2023), Models and methods for hybrid system identification: A systematic survey, *IFAC-PapersOnLine*.

Peer-reviewed publications outside the thesis scope

- Koo J., **Moradvandi A.**, Jonoski A., Solomatine D.P., Abraham E., (under review), Flood control of reservoir systems: Learning-based explicit and switched model predictive control approaches.
- Ghoreishee S.A., Shahrokhi M., Vafa E., **Moradvandi A.**, (2023), Fixed-time fault-tolerant control of uncertain MIMO switched interconnected systems subject to state delay, unknown control directions, and quantized nonlinear inputs, *International Journal of Adaptive Control and Signal Processing*.
- Sabeti F., Shahrokhi M., **Moradvandi A.**, (2022), Adaptive asymptotic tracking control of uncertain fractional-order nonlinear systems with unknown quantized input and control directions subject to actuator failures, *Journal of Vibration and Control*.
- **Moradvandi A.**, Malek S.A., Shahrokhi M., (2020), Adaptive finite-time fault-tolerant controller for a class of uncertain MIMO nonlinear switched systems subject to output constraints and unknown input nonlinearities, *Nonlinear Analysis: Hybrid Systems*.
- **Moradvandi A.**, Shahrokhi M., Malek S.A., (2019), Adaptive fuzzy decentralized control for a class of MIMO large-scale nonlinear state delay systems with unmodeled dynamics subject to unknown input saturation and infinite number of actuator failures, *Information Sciences*.

- Malek S.A., Shahrokhi M., Vafa E., **Moradvandi A.**, (2018), Adaptive prescribed performance control of switched MIMO uncertain nonlinear systems subject to unmodeled dynamics and input nonlinearities, *International Journal of Robust and Nonlinear Control*.

Conference and workshop abstracts

- **Moradvandi A.**, De Schutter B., Abraham E., Lindeboom R.E.F., Automatic control of purple bacteria in raceway-pond reactors: Simulation study of supervisory model predictive control, *International Conference on Novel PhotoRefineries for Resource Recovery—Photobiorefinery 2024*.
- **Moradvandi A.**, De Schutter B., Abraham E., Lindeboom R.E.F., Control of purple bacteria in raceway-pond reactors: How to deal with lack of knowledge of operators using model predictive control, *IWA Young Water Professionals Benelux 2024*.
- **Moradvandi A.**, Abraham E., De Schutter B., Lindeboom R.E.F., Approximating anaerobic digestion models by switched Box-Jenkin models using an identification algorithm for predictive modeling, *IWA World Conference on Anaerobic Digestion 2024*.
- **Moradvandi A.**, Koerts C., Alloul A., Kleerebezem R., De Schutter B., Abraham E., Weissbrodt D., Lindeboom R.E.F., Photoheterotrophic resource recovery bioreactors: model-based reliability assessment to support process design and control, *IWA International Conference on Water and Wastewater Management 2023*.
- **Moradvandi A.**, Alloul A., Weissbrodt D., Lindeboom R.E.F., Importance of modeling the metabolic switch mechanisms to predict PPB growth, *Applied Physics to PPB-based Environmental Biotechnology—Purplegain 2023*.