Doctoral theses at NTNU, 2023:128

Andrea Tuveri

Bayesian Estimators for Bioprocess Monitoring Under Uncertainty

Application and Challenges

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Natural Sciences
Department of Chemical Engineering

NTNU
Norwegian University of
Science and Technology

Andrea Tuveri

Bayesian Estimators for Bioprocess Monitoring Under Uncertainty

Application and Challenges

Thesis for the Degree of Philosophiae Doctor

Trondheim, May 2023

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Chemical Engineering

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Natural Sciences
Department of Chemical Engineering

ISBN 978-82-326-5395-9 (printed ver.)
ISBN 978-82-326-6334-7 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2023:128

Printed by NTNU Grafisk senter

To my parents Anna and Pierpaolo.

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (PhD) at the Norwegian University of Science and Technology (NTNU). The work performed during this PhD was carried out at the Department of Chemical Engineering from November 2018 to December 2022, under the supervision of Prof. Nadav Bar (supervisor, Department of Chemical Engineering) and Prof. Lars Imsland (co-supervisor, Department of Engineering Cybernetichs).

The continuously growing increase in biotechnological processes for production of fine and bulk chemicals, results in the necessity of more competitive solutions to enable producers surviving in the markets. Process control is therefore one of the key enablers for improving performance while maintaining quality. The main motivation of this work is the investigation and application of different state estimators (i.e. soft sensors). This is important to infer real-time information on the unmeasured variables of interest (i.e. metabolite concentrations) by using all the available information (i.e. output measurements), ensuring the possibility of state-feedback and therefore enabling the implementation of control actions, to achieve nearly-optimal performance. The original duration of the project was three years, but then extended to four years for student supervision and technical support in the laboratory.

The project was part of the iFermenter project, which has received funding from the Bio Based Industries Joint Undertaking (JU) under the European Union's Horizon 2020 research and innovation programme. The main purpose of iFermenter was to fully exploit sugar residuals from pulp industry, delivering strains that can handle sugar residual streams mixtures of carbons not commonly consumed by bacteria, and handle the inhibitors contained in the raw materials. Additionally, the project aimed to develop and implement control algorithms to enhance the automation in biotechnological processes, with the goal of creating affordable add-on to bioreactors, easily transferable to other high value products by minimal investments. The consortium consisted of experts in different fields like microbiology, synthetic biology, metabolic engineering, process system engineering, and envir-

onmental engineering.

During my PhD studies I had the privilege to support the development of the CyberGenome LAB at the Department of Chemical Engineering (IKP) at the Norwegian University of Science and Technology (NTNU), transferring knowledge to students. Additionally, I had the opportunity to work together with colleagues to bridge the application of control algorithms in an experimental set-up. This was achieved by a tight cooperation within the Process System Engineering group at the Department of Chemical Engineering (IKP) and together with the Department of Engineering Cybernetics (ITK). The cooperation resulted in the publications which are part of this thesis.

This research work was supported within the iFermenter project from the Bio Based Industries Joint Undertaking (JU) under the European Union's Horizon 2020 research and innovation programme, grant agreement 790507. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the Bio Based Industries Consortium.

Declaration of Compliance

I hereby declare that the thesis is an independent work in agreement with the exam rules and regulations of the Norwegian University of Science and Technology (NTNU). This work is original and my own work, and the sources I used were properly cited and acknowledged.

Øvre Årdal, 12th April 2023
Andrea Tuveri

Acknowledgments

The last four years represented a challenging and interesting journey for me, where I had the possibility to learn and develop. This was not only thanks to me, but also to all the people that I had around me.

Firstly, I would like to thank my supervisor Nadav Bar for giving me the opportunity to pursue this PhD. Nadi has believed in me and challenged me, motivating me to achieve my goals. I have learned a lot from his innovative and creative mindset and his communication skills, which helped me to become a better communicator. I am grateful to him for giving me the freedom to decide my research directions and to make mistakes, supporting me when that happened and making me able to develop as a scientist and as a person.

Secondly, I would like to thank my co-supervisor Lars Imsland. I am honored and grateful to have had the possibility to collaborate with Lars, he has been my compass throughout these years. Lars has taught me to challenge my own ideas, making me gain the confidence I did not have when I have started. His strong passion for his work, his deep knowledge and his selfless service to his academic duties are a true motivation and example for me.

Thirdly, I would like to specially thank Carol, Fernando, Haakon, Johannes, José and Pedro for the enjoyable collaboration, the valuable feedback, the constant support, the fruitful discussions and the nice moments spent together. Additionally, I would like to thank the students that I have had the opportunity to co-supervise during their mater thesis Helene and Kevin, and their 4th year project Anne and Lilian. I have very much enjoyed working with all of you and from each of you I have learned a lot.

The happy, pleasant and welcoming environment in the PSE group had no less importance, this really made every tough moment lighter with a smile, a talk or a hug and this overall experience very memorable. Starting from Sigurd and Johannes which always ensured having social gatherings and activities. Followed by my fantastic colleagues Adriaen, Adriana, Allyne, Alex, Ana, Bahareh, Chris-

topher, Cristina, David, Dinesh, Evren, Fabienne, Halvor, Julian, Lucas B., Lucas C., Rafael, Robert, Simen, Tobias, Zawadi and the others for not simply being my colleagues but also differently participating in my life as close friends and being one of the most important part of my PhD experience. Special thanks also go to my friends Shreya, Martina and Joakim for being part of a lot of fun and exciting moments together and Anacleta for being a dear friend and always making me feel a bit of Italy with me during the coffee breaks.

A very special thank also goes to Kristin Hestetun, because the finalization of this manuscript would have taken much more time without the trust and flexibility I have been given and to my opponents, Berit Floor Lund, Maurício Bezerra de Souza Jr. and Jannike Solsvik, for taking the time to evaluate this thesis.

I would also like to thank the iFermenter members, particularly Christoph, Daniel, Julian, Peter and Thais for the nice and collaborative environment as well as the constant support and discussions. In addition, I would like to extend my thanks to the colleagues and professors at the department of Chemical Engineering at the University of Cagliari for all the knowledge and good memories I carried with me all along to Norway.

Finally, I would like to thank the people that although the distance have been my strength and support all the way before and throughout my PhD experience. My friends Marco and Tomáš for always being there for me even if distant. My parents to whom I am eternally grateful for the education received, being always supportive and letting me follow my dreams. My wonderful brother Luca, for always believing in me. My grandparents, my great aunts together with Cristina and Guglielmo for their support and their active role in my education. Reidun and Solveig, for welcoming home and making me feel part of their family. And my beloved Karsten, for being an example of dedication and strength, being the person I can look up to and learn from, always ready to give me the support and motivation I needed and giving much happiness throughout all these years.

Abstract

In the face of growing competition and soaring energy prices, the biotechnological industry has also identified, as other chemical industries did before, the need for *on-line* process optimization as a strategic solution. However, these solutions are dependent on full-state feedback, which is often not available in an industrial setting. Therefore, it becomes necessary to implement state estimators to combine the mechanistic knowledge, usually obtained through unstructured Monod-like models, and available measurements. This thesis therefore investigates the implementation of three different state estimators to monitor a Fed-batch cultivation of *Corynebacterium glutamicum* using experimental data.

In a real application, one of the limiting factors in the application of state estimators is the lack of full knowledge on the system dynamics, often caused by lack of full knowledge and changes in metabolism. Indeed, the unmodelled dynamics can cause a significant difference between the system equations used for the predictions and the real process (i.e. structural plant-model mismatch). In this thesis, we presented several approaches to handle these uncertainties using experimental data.

Firstly, to handle the system uncertainty, we considered Bayesian state estimators. More specifically, because of the need of avoiding negative concentrations, we considered the constrained form of Extended (EKF) and Unscented (UKF) Kalman filters and the Moving Horizon Estimator (MHE).

Secondly, to properly take the structural plant-model mismatch into account, we implemented the state estimators with a variable and state dependent process noise covariance matrix. Additionally, to handle the changes in metabolism which were not included in the model dynamics, we proposed the possibility to change the tuning parameters for the process parameters in a batch-to-batch fashion. This improved the estimation accuracy, resulting in the possibility of monitoring the unmeasured state continuously.

Thirdly, to further improve the estimation accuracy, we exploited the ability of

MHE to easily handle constraints to implement model adaptation through the use of combined state and parameter estimation. To deal with the ill-posed problem that arises by the addition of the parameters as optimization variables, we implemented and compared two different regularization methods. Although the inclusion of parameter adaptation resulted in improved estimates, it still suffered from limitations due to the missing dynamics.

Fourthly, considering the possibility to use infrequent and delayed measurements to improve the real-time information, we investigated the effect of different approaches to include them. This improved the estimation accuracy, however the results showed that different approaches must be considered depending on the application and quality of measurements.

Lastly, to summarise our results, we discussed them in the context of the available literature. As the results also showed, there is no single solution to address all the challenges, and the different approaches have their advantages and limitations. It is therefore necessary for the practitioner to have a clear understanding of the specific application to then find the most suitable approach to select.

Contents

List of Figures

1 | Introduction

"I believe that no one who is familiar, either with mathematical advances in other fields, or with the range of special biological conditions to be considered, would ever conceive that everything could be summed up in a single mathematical formula, however complex."

RONALD A. FISHER (1890-1962)

This chapter opens the work by introducing the motivation. Consecutively, the contributions and publications related to the research objectives are presented and the overall structure outlined.

1.1 Motivation

Biotechnological processes have acquired large momentum in the last decades, for production of fine and bulk chemicals. Therefore, improving their performance while maintaining quality has become of paramount importance for ensuring competitive industrial applications. Improvements in performance have been traditionally achieved through application of [1–3]:

- **Genetic Engineering**: creating genetically modified organisms (GMO), by altering the DNA makeup of organisms [4].
- **Synthetic Biology**: using the tools developed and applied in genetic engineering to design, or redesign, genetic-circuits in the organisms. This is done to make them produce the desired substances, usually not present in the natural world [1, 3, 5].
- **Metabolic Engineering**: optimizing regulatory processes within the organisms to increase the production/consumption of the desired substances, therefore aiming to the optimal functioning of metabolic pathways [1, 3–5].

Within the biotechnological industry, pharmaceutical companies have seen a solid growth in the last two decades, thanks to the improvements in the aforementioned fields [1, 2]. However, although in 2017 they generated almost 2.2% of the world's gross domestic product, they have been recently affected by economic challenges, associated to the use of outdated development and manufacturing technologies [6]. Therefore, they are seeking for technological modernization, introducing initiatives such as quality-by-design (QbD) and process analytical technology (PAT) [6]. This brought them to identify improvements in digitalization and data analytics as key strategic activities to achieve faster and cheaper product manufacturing, therefore requiring Process Systems Engineering knowledge [2, 7, 8].

Process Systems Engineering (PSE) enables control and optimization at the process level [1], yielding to (nearly-)optimal operations and, by consequence, achieving more profitable processes and tighter product quality specifications [2, 9]. As reported in [8], the scope of PSE is vast and consists in the integration of different disciplines to address problems such as reduction of energy consumption, respect of environmental regulations and economical effectiveness of process operations.

However, the lack of readily available *on-line* measurements of key process variables, one of the core elements of PSE [8], hinders the real-time application of its tools in the bio-industry [1]. Thus, the next generation bio-tech industry, and in particular the pharmaceutical industry, will need to make increasing use of **state estimators** [2, 6, 10–12, 9], because of their ability of maximizing the information obtained by the combination of measurements and models [1], by considering uncertainties. The work behind this thesis therefore aims to contribute to the development of state estimators, trough their application to a fed-batch cultivation process of *Corynebacterium glutamicum*, investigating the possibilities to handle plant-model mismatch when applied to an experimental data-set.

1.2 Thesis Outline

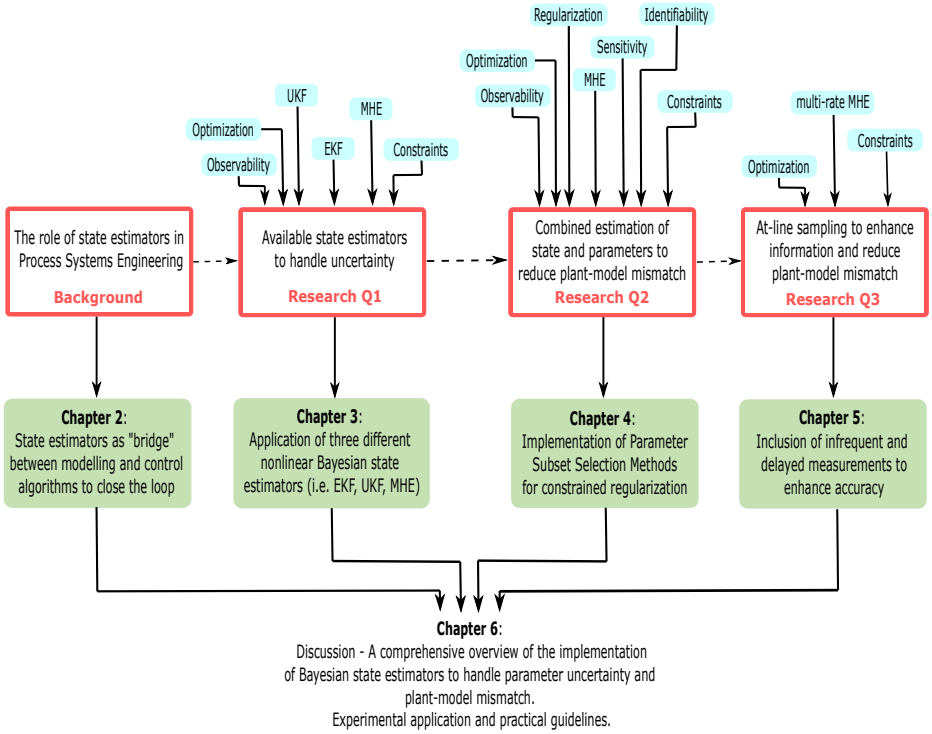


Figure 1.1: Outline of the thesis. The figure shows the summary of the research questions together with the implemented techniques and the chapters that address them.

The thesis is organized, as reported in Fig. 1.1, into six chapters. Chapter 2 introduces the relevant background material for the work and points out the role of state estimators. Thereafter, Chapters 3-5 present the original publications, answering the three research questions. Finally, Chapter 6 discusses the implementation

choices, while Chapter 7 outlines the conclusions and future research directions.

As reported in Sec. 1.1, the importance of PSE for the achievement of more profitable plant operations (i.e. tight product quality specifications) and, more specifically, the key role which state estimators have in the PSE-skill set, by maximizing the return from the investment in model development trough the combination of real-time process information, represents the first step towards the achievement of closed-loop control in practice [6] (further elaborated in this thesis through **Chapter 2**, Fig. 1.1). This led us to the following research questions:

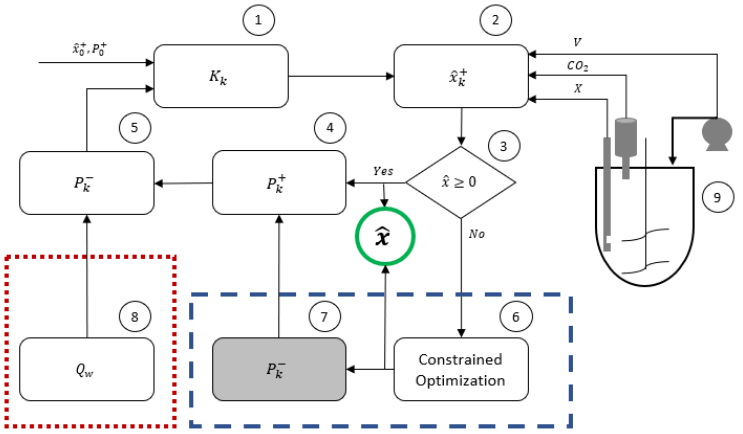


Figure 1.2: Implementation block diagram for EKF and UKF integrated with the bacterial cultivation process (Chapter 3). The recursive filter is represented by the blocks from 1 to 5. After each iteration the Kalman gain is calculated (1) followed by a correction (2) using the available *on-line* measurements from the process (9). The constraints are activated in (6-7) if condition (3) is not satisfied. Consecutively, the error covariance matrix is updated in (4) and then used for the ahead projection (5), where the updated process noise covariance is used (8). The estimated states are represented by the continuous green circle.

Research Question 1. *What are the available state estimator to handle uncertainty in nonlinear biological processes? Can these estimators accurately estimate the unmeasured variables of interest, under plant-model mismatch and high input changes?*

Chapter 3 presents the experimental application of three Bayesian estimators (i.e. Extended and Unscented Kalman Filter and Moving Horizon Estimator). Indeed, as also mentioned in Chapter 2, Bayesian estimators result in an optimal choice to address uncertainty when dealing with nonlinearities [13]. Firstly, to compensate for the absence of a comparison between Extended and Unscented Kalman Filter (EKF and UKF) applied to biological processes, we considered the comparison of

the two state of the art estimators (i.e. EKF and UKF). However, although the UKF presented a slight improvement, the results showed that the two filters performed similarly. Secondly, motivated by avoiding the tedious common practice of trial-and-error tuning of state estimators, found in the most applications in bioprocesses, and inspired by the works of [14–16], we focused on the importance of properly tuning the model uncertainty (i.e. process noise covariance matrix Q) to introduce the noise in a nonlinear fashion, by considering that the noise rightfully enters the model through the uncertain parameters [17] (Fig. 1.2). Thirdly, because the work of [18] was the sole exception implementing constraints to recursive state estimators (i.e. Kalman-like filters) in a bioprocess found in literature, we enforced the importance of using constraints on the states to avoid infeasible (i.e negative) concentrations. This led us to the implementation of a QP-problem [19] for constraining the EKF and the UKF, as well as the investigation of an optimization-based approach, such as the Moving Horizon Estimator (MHE, Fig. 1.3).

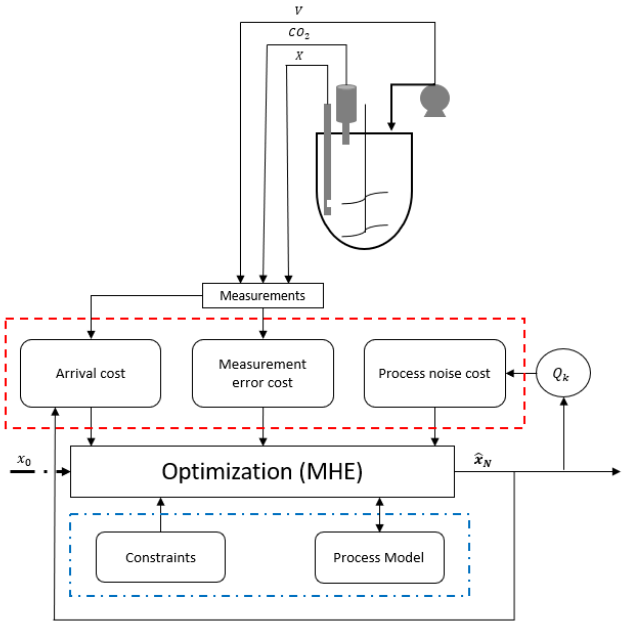


Figure 1.3: Block diagram for the MHE implementation (Chapter 3). The different terms for the cost function are represented by the dashed red line, while the different constraints are reported within the dashed-dotted blue line. The cost function takes the measurements into account in both arrival and measurement error cost, while the process noise is weighted by the error covariance matrix which is recalculated at every iteration.

Research Question 2. *How can we implement on-line model adaptation to further improve the accuracy of the estimated states under plant-model mismatch?*

Chapter 4 presents the implementation of an MHE for combined state and parameter estimation in an experimental application. This was done to improve the estimation accuracy of the variables of interest, considering the time varying nature of parameters [17, 9]. The selection of an MHE was motivated by the works of [20, 21] to exploit its advantage [22] of explicitly incorporating constraints to regularize the problem. This was used to estimate the parameters along with the states, augmenting the system by adding them as states with zero dynamics, as it is usually done when implementing Kalman-like filters [23]. Firstly, to evaluate the possibility to simultaneously estimate states and parameters in absence of full-state feedback, the implementation was done using an experimental data-set. Indeed, previous works [21] implemented an MHE for the same purpose. However, their work presented an *in silico* application under full-state feedback (i.e. the *at-line* measurements were available infrequently with a time delay). Secondly, to deal with the ill-posed problem that arises by the combination of state and parameter estimation, a transformation-based regularization method (Subset Selection by Transformation, SST) [24] was implemented and compared with a state of the art method (Orthogonalization Method, OM) [25]. Additionally, a geometric interpretation of SST [24] was given, to enhance the interoperability of SST in comparison to OM. Indeed, while OM reduces the number of the decision variables, SST [24] reduces their search space (Fig. 1.4). Lastly, to determine the available degrees of freedom for the optimization problem, a stopping criterion based on the structural identifiability of the system [26] was proposed.

Research Question 3. *When infrequent and delayed at-line measurements are available, how can we best include them to enhance the quality of the real-time information?*

Chapter 5, following the works of [27–30] and taking into account the possibility of industrial applications, investigated the implementation of different approaches for the inclusion of *at-line* measurements. This was motivated by the possibility of improving the estimation accuracy by the use of infrequent and delayed measurements in a multi-rate MHE (MMHE). Firstly, to present the challenges encountered by the practitioner in a real application, we implemented the MMHE using an experimetal data set. Secondly, to investigate different possible approaches to obtain more or less smooth estimates, we defined the optimization problem by either considering or not availability of *at-line* measurements at all times. This was done through the use of variable and fixed structure MMHE as defined in [27, 28]. Thirdly, because in real application it is possible that measurements are available with long delays, to avoid the necessity of increasing the estimation horizon (and

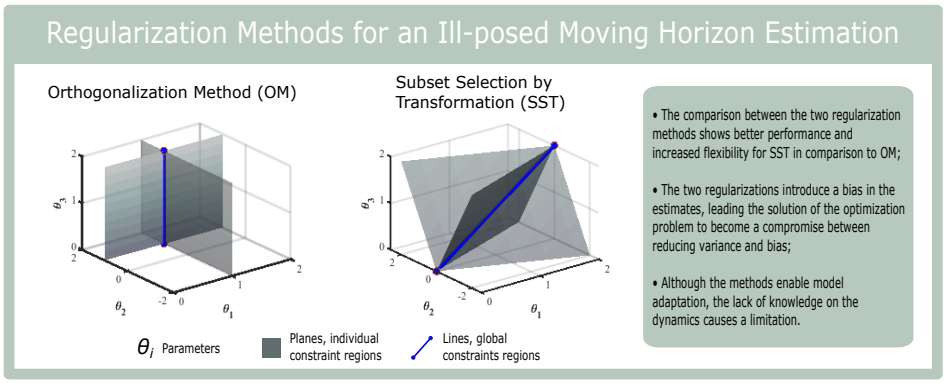


Figure 1.4: Geometric representation of the two regularization methods proposed in Chapter 4. While the Orthogonalization Method (OM) constraints singular parameters θ_i, leaving the other ones as degrees of freedom, Subset Selection by Transformation (SST) constraints linear combinations of different θ_i, leaving them free to vary within the projection of the constrained region along their axes.

therefore the size of the optimization problem), we implemented the MMHE with and without filter recalculation (Fig. 1.5) [30, 31]. Lastly, to limit the impulsive behaviour obtained when considering low uncertainty in the infrequent information from the *at-line* measurements, we highlighted that their integration often results in a trade-off between rejecting disturbances due to the measurement error or the model error (due to plant-model mismatch) [30].

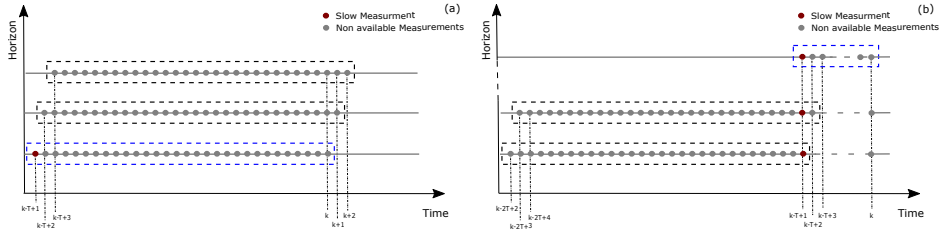


Figure 1.5: Schematic representation with and without filter recalculation. The sample (red) is collected at time $k - T + 1$ and its measurement is available at time k with a delay of 29 minutes ($T = 30$ is the horizon). The black dashed line represents the window horizon. Without filter recalculation (a), the measurement is put at time $k - T + 1$ (i.e. sample time), however, it disappears in the next window shifting. When filter recalculation is performed (b), the optimization problem is recalculated *off-line*, from time $k - 2T + 2$ to time k, resulting in a correction of the estimation from time $k - T + 1$.

1.3 Contributions

Section 1.1 proposes general Process Systems Engineering techniques as the mean for technological modernization, which the pharmaceutical industry is seeking for. However, this span of techniques is wider than our scope. Indeed, our scope has just focused on some specific directions as follows:

- The importance of state estimators, for inferring information on key process variables under absence of measurement devices, for the promotion of feedback control applications, was outlined.
- The application of two state of the art recursive Bayesian estimators (i.e. Extended and Unscented Kalman Filter) was presented and their performance compared. This was done in an experimental application, where high input variation was selected to test the predictive capability of the model in use.
- It was highlighted how, although utilizing a simple Monod-growth model, the limitations of the plant-model mismatch could be counteracted by taking it into account in the model uncertainty.
- The importance of constrained estimation techniques was highlighted, given that this might be necessary when substrates or other metabolites are consumed, since concentrations are strictly positive entities.
- As a response to the necessity of constraints on the state estimates, the Moving Horizon Estimator was tested and its limitation, given the presence of a steady state off-set, discussed.
- To avoid the ill-posed optimization problem, resulting from the addition of the parameters as decision variables, two regularization methods for *on-line* parameter estimation were implemented in an MHE and compared.
- A stopping criterion for the definition of the available degrees of freedom in the optimization problem was proposed, based on the structural identifiability of the system.
- A practical implementation, highlighting the importance of selecting the best approach when including infrequent measurements for improving the estimation accuracy of a fully observable system, was presented using experimental data.

1.4 Publications

The work on this PhD project has resulted in the main authorship of three journal papers, one conference paper, and one co-authorship. The list of papers, together with authors contributions, is given as follows:

1. A research paper published in Journal of Process Control. The paper has been a collaboration between the Department of Chemical Engineering and the Department of Engineering Cybernetics at NTNU:

   Andrea Tuveri, Fernando Pérez-García, Pedro A. Lira-Parada, Lars Imsland, and Nadav Bar. "Sensor fusion based on Extended and Unscented Kalman Filter for bioprocess monitoring". *Journal of Process Control*, 106, 2021, https://doi.org/10.1016/j.jprocont.2021.09.005.

   Andrea Tuveri: Conceptualization, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization.
   Fernando Pérez-García: Validation, Investigation, Writing – review & editing, Supervision.
   Pedro A. Lira-Parada: Data Curation, Writing – review & editing, Visualization.
   Lars Imsland: Conceptualization, Software, Resources, Writing – review & editing, Supervision.
   Nadav Bar: Conceptualization, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

2. A published paper in a conference proceeding (13th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems). The paper has been a collaboration between the Department of Chemical Engineering and the Department of Engineering Cybernetics at NTNU:

   Andrea Tuveri, Haakon Eng Holck, Caroline S.M. Nakama, José Matias, Johannes Jäschke, Lars Imsland and Nadav Bar. "Bioprocess Monitoring: A Moving Horizon Estimation Experimental Application". *IFAC-PapersOnLine*, 55(7):222–227, 2022, https://doi.org/10.1016/j.ifacol.2022.07.448.

   Andrea Tuveri: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing and Visualization.
   Haakon Eng Holck: Software and Writing - Review & Editing.
   Caroline S.M. Nakama: Conceptualization, Writing - Review & Editing, Visualization and Supervision.

José Matias: Conceptualization, Writing - Review & Editing, Visualization and Supervision.
Johannes Jäschke: Writing - Review & Editing, Visualization and Funding acquisition.
Lars Imsland: Conceptualization, Resources, Writing - Review & Editing, Visualization and Supervision.
Nadav Bar: Conceptualization, Writing - Review & Editing, Visualization, Supervision, Project administration and Funding acquisition.

3. A research article published in Computers and Chemical Engineering. The paper has been a collaboration between the Department of Chemical Engineering and the Department of Engineering Cybernetics at NTNU:

   Andrea Tuveri, Caroline S.M. Nakama, José Matias, Haakon Eng Holck, Johannes Jäschke, Lars Imsland and Nadav Bar. "A Regularized Moving Horizon Estimator for Combined State and Parameter Estimation in a Bioprocess Experimental Application". *Computers and Chemical Engineering*, 172, 2023, https://doi.org/10.1016/j.compchemeng.2023.108183.

   Andrea Tuveri: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing and Visualization.
   Caroline S.M. Nakama: Conceptualization, Resources, Writing - Review & Editing, Visualization and Supervision.
   José Matias: Conceptualization, Writing - Review & Editing, Visualization and Supervision.
   Haakon Eng Holck: Software and Writing - Review & Editing.
   Johannes Jäschke: Writing - Review & Editing, Visualization and Funding acquisition.
   Lars Imsland: Conceptualization, Resources, Writing - Review & Editing, Visualization and Supervision.
   Nadav Bar: Conceptualization, Writing - Review & Editing, Visualization, Supervision, Project administration and Funding acquisition.

4. A research paper ready to be submitted. The paper has been a collaboration between the Department of Chemical Engineering and the Department of Engineering Cybernetics at NTNU:

   Andrea Tuveri, Caroline S.M. Nakama, Lars Imsland and Nadav Bar.
   "A practical implementation of Moving Horizon Estimation with Delayed

Measurements in a Bioprocesses Experimental Application". *to be submitted*.

Andrea Tuveri: Conceptualization, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing and Visualization.
Caroline S.M. Nakama: Conceptualization, Writing - Review & Editing, Visualization and Supervision.
Lars Imsland: Conceptualization, Writing - Review & Editing, Visualization and Supervision.
Nadav Bar: Conceptualization, Writing - Review & Editing, Visualization, Supervision, Project administration and Funding acquisition.

1.5 Additional Publications not included in the Thesis

During his PhD the candidate contributed as a co-author in the following journal paper, which is not part of this thesis:

1. A research article published in Biochemical Engineering Journal. The paper has been a collaboration between the Norwegian University of Science and Technology and the Danish Technical University:

   Pedro A. Lira-Parada, **Andrea Tuveri**, Gerd M. Seibold and Nadav Bar. "Comparison of noninvasive, in-situ and external monitoring of microbial growth in fed-batch cultivations in *Corynebacterium glutamicum*". *Biochemical Engineering Journal*, 170, 2021, https://doi.org/10.1016/j.bej.2021.107989.

   Andrea Tuveri: performed part of the experimental work and read and edited the manuscript.

2 | Background

"The most we can know is in terms of probabilities."

RICHARD P. FEYNMAN (1918-1988)

2.1 Process Systems Engineering (PSE)

As recently reviewed in [8], Process Systems Engineering is the discipline related to the development of systematic techniques such as mathematical modelling, data analytics, design, optimization and control, for describing and regulating the behaviour of physico-chemical systems.

However, our intent is not an in-depth analysis of Process Systems Engineering topics, but a brief introduction of its application to biotechnological processes, to highlight the role that state estimators take in it. Therefore, the reader is referred to [8] and references therein for a more detailed overview.

2.2 Bioprocess Systems Engineering (*bio*-PSE)

Bioprocess Systems Engineering represents a sub-group of the more general PSE, aiming at applying PSE techniques to systems where also biological phenomena are considered [1]. Following the works of [1, 2], we define four core components for *bio*-PSE as 1) Modelling, 2) Monitoring, 3) Optimization and 4) Control.

2.2.1 Modelling

Bioprocesses are modelled, most often, using *unstructured* models based on Monod kinetics [32, 17] or their variations (e.g. Haldane, Teissier [33]), considering the growth of the organism on a single limiting substrate. Additionally, they can be easily extended to multiple sugars [34] and substrate inhibition (e.g. Haldane [33, 35]).

However, due to their complexity, limited understanding of all the undergoing phenomena, and limitations in terms of range and validity due to their dependence on specific operating conditions [1, 2], biological systems present several limitations for modellers. It is therefore left to the practitioner to discern between all the available techniques for the most appropriate model type (i.e. structured or unstructured [36]), given the application of interest [37, 38].

By starting from their complexity, the inclusion of knowledge from systems biology and metabolic engineering enables the development of complex high-fidelity models (i.e. *structured* models). Therefore, integrative models can be used to merge mechanistic knowledge with gene expression and cellular function, unravelling the biological mechanisms which regulate them. This allows the connection of phenomena which occur at the bioreactor level (i.e. macro scale) to the ones taking place at the molecular level, including metabolic fluxes (i.e. Flux Balance Analysis, FBA [39]) or genetic circuits (e.g. Hill-Langmuir equation [40]). High-fidelity models are certainly relevant for the investigation of the process behaviour

in a rapid and inexpensive way, by reducing the number of individual experiments. However, although holding excellent predictive capabilities, they can become computationally intractable for *on-line* applications.

As stated above, the understanding of all the undergoing phenomena taking place in a bioprocess can often be limited [41]. This, together with the growing interest in data-driven modelling techniques, resulted in the consideration of data-driven models (i.e. black-box or Machine Learning models). Black-box models have the advantage of being able to capture and describe the complex dynamics of all the undergoing processes (i.e. both at the bioreactor and molecular level). Simultaneously, these models suffer because of their dependence on big data-sets, sensitivity to noise and limitations under unseen process conditions, given that they only present descriptive, but not explanatory, capabilities. However, hybrid models (i.e. grey-box models) can present an alternative solution to black-box models, by complementing the use of black-box and mechanistic (i.e. white-box) models. Indeed, hybrid models can combine the mechanistic understanding of white-box models together with the data-based nature of black-box models, therefore combining their explanatory and descriptive capabilities. As mentioned in [2], one of the main challenges of hybrid models is on how to best combine data-based and mechanistic knowledge. As was shown in [42] in the early 90's, one option could be the use of data-based approaches to model single growth-related parameters. However, as recently reviewed in [43], by today several directions can be considered.

Indeed, taking into account limitations in terms of range and validity due to the effect of the inputs to the cell metabolism (i.e. media recipes, supplements, operating conditions) [2], data-driven or hybrid models can constitute a reasonable choice [44], given their capability to learn from data (entirely or partially) the effect of the external inputs [43].

It is therefore important that the practitioner, when discerning between the different approaches and levels of complexity, selects the one more suitable for the application of interest. In doing that, it becomes important to take into account the available mechanistic knowledge, the available data, the time which the practitioner intends to invest and finally, the desired output. Indeed, the latter will depend on the necessity, whether the model is required to be *descriptive* (i.e. visualization), *explanatory* (i.e. process understanding) or *predictive* (i.e. simulation and forecasting). Relevant guidelines can be found in [7, 37, 38, 43].

2.2.2 Monitoring

The availability of real-time information on critical process parameters, also referred here as variables of interest, is crucial for early detection of abnormal behaviour, but also to regulate product quality and yield, by the implementation of control actions [7, 18]. The most straightforward way to measure qualitative and quantitative characteristics is through the use of sensors or measurement devices. Although a variety of possibilities can be found in [45–47], the most widely used, and relevant for our work, are sensors for measurement of pH, dissolved oxygen (DO_2), biomass (e.g. absorbance probe) and different metabolites (e.g. infrared probes such as NIR, MIR and Raman spectroscopy), together with devices to measure off-gas composition and high performance liquid chromatography (HPLC). Similarly to [46, 47], these methods can be divided according to the way the sample is analysed:

1. *in-situ* measurements (i.e. on site), where the analytical test is conducted directly in the bulk material. This is further divided, based on the type of interaction between the sample and the analytical device, in:
    - *in-line*, where the detector has direct interface with the process;
    - *on-line*, when the sample is removed automatically and transported to the detector maintaining sterile conditions, through a by-pass.
2. *ex-situ* measurements (i.e. out site), where the sample is transported or stored and the analytical test is conducted in an external environment. Based on the type of interaction between the sample and the analytical device, they are further divided in:
    - *at-line*, when the detector is located in close proximity and the sample is transported, through automated sample devices, to the analytical site;
    - *off-line*, when the sample is stored and the analysis is conducted with the support of an operator, a posteriori.

In-line and *on-line* measurements can directly provide real-time information, while *at-line* could present a delayed real-time information, depending on the analytical technique employed. The *off-line* measurements instead, can only be used a posteriori, for calibration and validation.

However, when interested in real-time monitoring of a bioprocess, many variables, in practice, are either not measurable, due to the high costs of the instrument, or can only be measured infrequently, using devices that can only perform the analysis *off-line*. In this case, the necessity of inferring information on all the variables of

interest, requires the employment of state estimators. State estimators (also called *soft sensors* [48]), beside inferring information on the unmeasured variables, using the real-time information generated by the analytical devices which are available, present, as additional feature, the ability to reduce the noise and the biases generated by the real-time sensor measurements (Fig. 2.1).

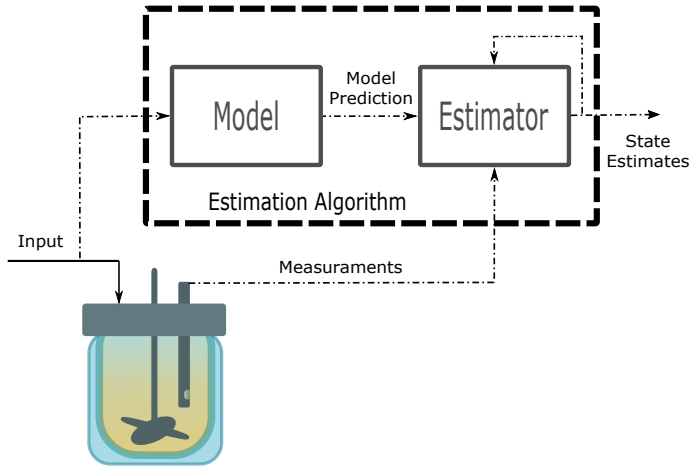


Figure 2.1: State estimation diagram, partially adapted from [9].

To this extent we could say that, by doing so, they create *virtual measurements* [45] by combining the real-time information from the available sensors (i.e. sensor fusion) to the process knowledge contained in the model (which can be either mechanistic, data-based or hybrid as reported in Sec. 2.2.1).

Whereas a more detailed classification can be found in [49], in this work we will classify state estimators applied to bioprocesses, in accordance to [9, 50, 51, 13], into three categories, as reported in Fig. 2.2: 1) deterministic, which assumes that both model and measurements can be fully determined given the operating conditions; 2) Bayesian, assuming that the model and the measurements are affected by random noise; 3) hybrid, which are a combination of the former two. Most importantly, we will consider their nonlinear form, given the nonlinear nature of the models that describe biological systems.

These three categories, as reported in Fig. 2.2, can be further divided [9]. Firstly, deterministic estimator are divided into first-order systems (FOS) based estimators (e.g. extended Luenberger observer, ELO), where the dynamics are given by linearized first-order differential equations and data-based regressions (e.g. partial least squares, PLS), of which application can be found in [18, 52]. Secondly,

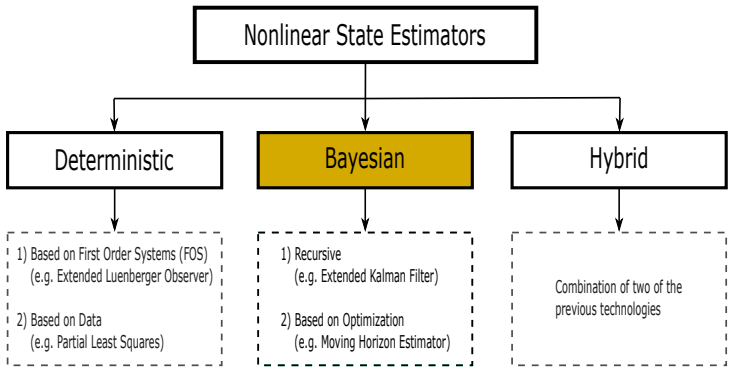


Figure 2.2: Classification of nonlinear state estimators, adapted from [9].

Bayesian estimators (Chapter 3) are divided into recursive (e.g. Extended and Unscented Kalman filters [53]) and optimization based ones (e.g. Moving Horizon Estimator [22]). Lastly, hybrid estimators are obtained by combining deterministic and Bayesian approaches together. An implementation of the latter and its advantages in comparison to a deterministic data-based one is found in [18].

Among all the presented classes, Bayesian estimators (Fig. 2.2), by assuming the intrinsic stochastic (i.e. uncertain) nature of all variables, provide an approach that uses the available real-time data to estimate the variables of interest, based on probability distributions [51]. Therefore allowing the characterization of our expectations, by using a measure of confidence or uncertainty [22]. This class of estimators, which involves Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), Particle Filter (PF), Ensemble Kalman filter (EnKF), Moving Horizon Estimator (MHE) and many others [49, 13], while differing from the deterministic approaches by the way the uncertainty is treated, are the most widely used in bioprocesses [9, 49, 51], being suitable for complex nonlinear systems [9, 49, 13].

2.2.3 Optimization

Optimization of biological processes is of interest for cost reduction, quality enhancement, as well as meeting safety and environmental requirements, therefore representing means for achieving competitive industrial applications (Chapter 1). By being traditionally mostly characterized by batch or semi-batch (i.e. fed-batch) operations [54, 55], the optimization of bioprocesses requires to obtain a dynamic trajectory rather than a constant set-point and therefore necessitates the implementation of dynamic optimization. A general overview of dynamic optimization

applied to biological processes can be found in [54, 55], where the authors covered the available analytical and numerical tools to compute the optimal solution both in the nominal case and in presence of uncertainty.

In a practical situation, given the batch-to-batch (i.e. *run-to-run*) variability caused by uncertain model parameters or plant-model mismatch, disturbances or variations in initial conditions, the solutions based on nominal models can only deliver a sub-optimal response and therefore uncertainty must be taken into account [55]. As reported in [55], in the absence of measurements, robust solutions can be implemented, considering the worst-case scenario. By yielding to a conservative approach, this motivates the use of stochastic (i.e. uncertainty-based) or adaptive (i.e. measurements-based) approaches [56]. In case uncertainty characteristics are changing or unknown (i.e. stochastic case) or not all the variables of interest are measured (i.e. adaptive case), the implementation of a Bayesian state estimator can be of use. Indeed, although they are mostly considered as techniques to estimate the value of the states (i.e. average, mean or expectation), through Bayesian estimation we consider the variables of interest as *random variables* (RV) and obtain a measure of their uncertainty (i.e. how much we can expect them to vary from their mean), using information from the measurements and the knowledge from the model. This approach leads the optimization problem being divided in two subproblems: 1) the estimation problem, where the information from the measurements is used; 2) the optimization problem, where the inputs are updated towards the optimum through the model, while using the states as initial conditions. These implementations come with their limitations, therefore special considerations must be taken, as reported in [55–57]. However, this enhances the importance and the usefulness of state estimators to implement dynamic Real Time Optimization (RTO) [48, 56]. Moreover, it becomes important to mention that state estimators can be directly phrased as least-squares optimization problems (e.g. Moving Horizon Estimator, MHE) [8, 22].

2.2.4 Control

Control is considered an integral aspect of Process Systems Engineering [8]. Indeed it combines all the previously mentioned components to obtain the desired outputs from the process. By considering the entire production plant consisting of several units, following the work of [56, 58], the control system is divided in several layers, based on their time scale, as reported in Fig. 2.3.

Traditionally, process control was mostly employed for regulation purposes. However, with the continuous improvement of advanced control and computational methods, its scope has widened to cover the interaction between all the different unit operations [8], in the so called *plant-wide control* [58]. The typical control

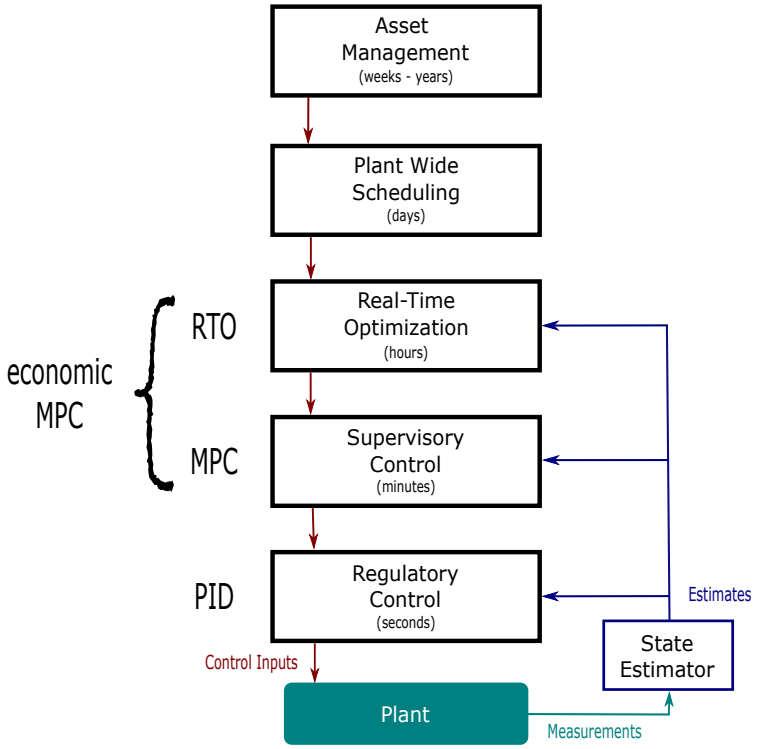


Figure 2.3: Typical control hierarchy. The figure was adapted from figures in [56, 58]. The control inputs coming from the upper layer to the lower layer are reported in red. Measurements (i.e. plant output) from the plant are reported in green. The estimated states (i.e. estimated outputs) are reported in blue.

hierarchy used in *plant-wide control* is presented in Fig. 2.3. For our work, we will limit to define the last three layers reported in Fig. 2.3: 1) Regulatory control; 2) Supervisory control; 3) Real Time Optimization.

Firstly, the role of the regulatory layer (Fig. 2.3) is to stabilize the plant by using, most often, single-input single-output (SISO) PI control loops [58]. In a bioreactor, its task would be to maintain pH, dissolved oxygen (DO_2), temperature and stirring speed at predefined set-points or trajectories by manipulating the flow of acid or base solutions and nutrients, aeration and power used for the heating system and the motor [2]. Therefore, the objective of the regulatory layer (i.e. stabilizing layer) is to control secondary measurements so that expert operators or the upper layer (i.e. supervisory layer) can handle the disturbances on the variables of interest (i.e. primary outputs) [58].

Secondly, the scope of the supervisory layer (i.e. tracking layer [59]) is to track a trajectory, which consists in the execution of a recipe [59]. This layer can be implemented using different advanced control structures [58], however, simple PIDs can encounter difficulties by handling strong nonlinear dynamics. Therefore, the method that obtained the most attention for trajectory tracking is nonlinear model predictive control (NMPC) [56, 59].

Thirdly, the real time optimization layer (i.e. recipe optimization [59]) is the one that determines the *optimal* recipe based on economic considerations and final desired product quality [59], through dynamic optimization as reported in Sec. 2.2.3. In case recipe optimization and tracking optimization are handled in an integrated manner, the problem can be formulated in a single layer approach, removing the time scale separation. Therefore, the recipe optimization can be reformulated, by adding directly the trajectory constraints, obtaining a formulation that maximizes the economic performance of the process, known as economic MPC (eMPC, Fig. 2.3) [56, 59].

At last, it is important to mention that all the discussion above has been done without considering the measurements availability. Indeed, as visible in blue in Fig. 2.3, the implementation of this hierarchical control approach requires the availability of measurements (i.e. state feedback). However, sometimes the evidence necessary for making decisions is not directly available from process measurements [22, 60]. It is therefore here that the relevance of state estimators becomes of key importance. In fact, when coupling process knowledge (Sec. 2.2.1) with secondary observations (i.e. subset of measurable states), it becomes possible to infer information on the unmeasured variables of interest (Sec. 2.2.2).

2.3 Fed-batch Cultivation Case Study

As reported in Sec. 2.2.2, state estimators combine the real-time information to the process knowledge contained in the model. Therefore, to enable the possibility to monitor the the fed-batch bacterial cultivation process of *Corynebacterium glutamicum* used in this work, we present the model and the measurements used by the three Bayesian estimators implemented. The system in Fig. 2.4 is described, following simple Monod-growth kinetics on a single substrate, trough the use of an unstructured mechanistic model. The material balances used to the describe the

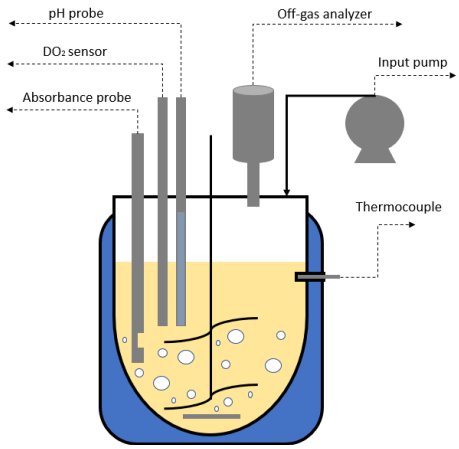


Figure 2.4: The figure reports the different measurement devices available for the experimental set-up.

dynamics of the process are as follows:

$$\dot{V} = F_{in} \tag{2.1a}$$

$$\dot{X} = \mu_{max}\frac{S}{K_S + S}X - X\frac{F_{in}}{V} - k_d X \tag{2.1b}$$

$$\dot{S} = -\mu_{max}\frac{S}{K_S + S}\frac{X}{Y_{XS}} - S\frac{F_{in}}{V} + S_{in}\frac{F_{in}}{V} \tag{2.1c}$$

$$\dot{\mathrm{CO}}_2 = \mu_{max}\frac{S}{K_S + S}\frac{X}{Y_{XCO_2}} - q_{air}\mathrm{CO}_2 \tag{2.1d}$$

Eq. 2.1 reports the model, used throughout the works presented in Chapters 3-5, where the right hand sides define input u (F_{in}, Eq. 2.1a), growth, consumption and formation (i.e. first term in Eq. 2.1b, 2.1c and 2.1d), dilution (i.e. second term in Eq. 2.1b and 2.1c), cell death (i.e. third term in Eq. 2.1b), substrate addition given the feeding concentration S_{in} (i.e. third term in Eq. 2.1c) and CO_2 released in the off-gas (i.e. second term in Eq. 2.1d).

The process was conducted at constant temperature (i.e. temperature in the water jacket), pressure and pH (i.e. by addition of titrating agents, KOH and H_3PO_4) while dissolved oxygen (DO_2) was maintained above 30% by correction of the stirring speed. Through the sensors available in the reactor set-up (Fig. 2.4), measurements for volume (i.e. input pump), biomass (i.e. absorbance probe) and CO_2 (i.e. off-gas measurement device) were available. Therefore, because not all the variables of interest (i.e. states) described by the system dynamics (Eq. 2.1) were directly measured, the implementation of a state estimator was needed to monitor

the glucose composition (Eq. 2.1c).

The system can be compactly described as:

$$\dot{x}(t) = f(x(t), \theta(t), u(t)) + w(t) \tag{2.2a}$$
$$y(t) = h(x(t)) + v(t) \tag{2.2b}$$
$$w(t) \sim \mathcal{N}(0, Q(t)) \tag{2.2c}$$
$$v(t) \sim \mathcal{N}(0, R) \tag{2.2d}$$

where $x = [V, X, S, \mathrm{CO}_2]^T$ is the state vector, $y = [V, X, \mathrm{CO}_2]^T$ the output measurement vector, $\theta = [\mu_{max}, K_S, k_d, Y_{XS}, Y_{X\mathrm{CO}_2}]^T$ the parameters vector and $u = [F_{in}]$ the input, while $w \in \mathbb{R}^{n_x}$ and $v \in \mathbb{R}^{n_y}$ represent the continuous-time noise vectors. The discretized system (Eq. 2.2) becomes:

$$x_k = F_{k-1}(x_{k-1}, \theta, u_{k-1}) + w_{k-1} \tag{2.3a}$$
$$y_k = H_k x_k + v_k \tag{2.3b}$$
$$w_k \sim \mathcal{N}(0, Q_k) \tag{2.3c}$$
$$v_k \sim \mathcal{N}(0, R) \tag{2.3d}$$

with discrete-time process ($w_k \in \mathbb{R}^{n_x}$) and measurement ($v_k \in \mathbb{R}^{n_y}$) noise. Additionally, $x_k \in \mathbb{R}^{n_x}$, $\theta \in \mathbb{R}^{n_\theta}$, $u_k \in \mathbb{R}^{n_u}$ and $y_k \in \mathbb{R}^{n_y}$ are respectively states, parameters, inputs and outputs.

2.3.1 Process Noise Covariance Definition

To define the additive process noise $w_k \sim \mathcal{N}(0, Q_k)$ (Eq. 2.3) with variable process noise covariance matrix Q_k, following the works of [14] and [15], we defined a new noise vector $\omega \sim \mathcal{N}(0, Q_\omega)$ (where $\omega \neq w_k$). The noise ω was considered in the system dynamics as follows:

$$\begin{aligned}
\dot{V} =& F_{in} + \omega_V \\
\dot{X} =& -\frac{F_{in}}{V} X + (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{k_S}) + S} X - (k_d + \omega_{k_d}) X + \omega_X \\
\dot{S} =& \frac{F_{in}}{V}(S_{in} - S) - (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{k_S}) + S} \frac{X}{(Y_{XS} + \omega_{Y_{XS}})} + \omega_S \\
\dot{\mathrm{CO}}_2 =& (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{k_S}) + S} \frac{X}{(Y_{X\mathrm{CO}_2} + \omega_{Y_{X\mathrm{CO}_2}})} - q_{air} CO_2 + \omega_{\mathrm{CO}_2}
\end{aligned} \tag{2.4}$$

defining the noise vector $\omega \in \mathbb{R}^{(n_x + n_\theta)}$ as:

$$\omega = \left[\omega_{\mu_{max}} \ \omega_{k_s} \ \omega_{k_d} \ \omega_{Y_{XS}} \ \omega_{Y_{XCO_2}} \ \omega_V \ \omega_X \ \omega_S \ \omega_{CO_2} \right]^T$$

ω was added to the deterministic model (Eq. 2.1) twofold. In the state dynamics ($\omega_V, \omega_X, \omega_S, \omega_{CO_2}$), in order to address structural plant-model mismatch, and in the parameters ($\omega_{\mu_{max}}, \omega_{K_S}, \omega_{k_d}, \omega_{Y_{XS}}, \omega_{Y_{XCO_2}}$), to address the uncertainty in the model parameters. With this approach, the process noise covariance $Q_k \in \mathbb{R}^{n_x \times n_x}$ becomes a state-dependant matrix, updated at each sampling time k. By compactly rewriting Eq. 2.4 as:

$$\dot{x} = f(x, u, \theta, \omega) \tag{2.5}$$

and defining $G_k \in \mathbb{R}^{n_x \times (n_\theta + n_x)}$ as the Jacobian of Eq. 2.5 with respect to the noise ω:

$$G_k = \frac{\partial f(x, u, \theta, \omega)}{\partial \omega} \tag{2.6}$$

the process noise covariance matrix is obtained as:

$$Q_k = G_k \cdot Q_\omega \cdot {G_k}^T \tag{2.7}$$

where Q_ω is a constant related to the statistics of the parameter uncertainty.

2.4 Bayesian State Estimators

Between the various implementation approaches for state estimators reported in Sec. 2.2.2, the focus will be placed here on Bayesian state estimators, because of their ability to handle uncertainty when dealing with nonlinear processes. The Bayesian approach to state estimation, called Bayesian inference [61], allows us to infer information on the outcome of our future processes. Bayes' rule (or theorem) is the base for it (i.e. Bayesian estimation) and can be stated as follows [62, 61, 13]:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{2.8}$$

where we indicate with x the states and with y the measurements. The conditional probability density function (PDF) $p(x|y)$ and $p(y|x)$ are the PDF of the realization x (or y) to occur given that y (or x) is true, while $p(x)$ and $p(y)$ represent the (prior) PDF of the realization x (or y) without conditions. By defining the set of measurements $Y_k = [y_0, y_i, \ldots, y_k]$, the posterior PDF $p(x_k|Y_k)$ (i.e. the solution of the estimation problem), is obtained from Eq. 2.8 as [62, 61, 13]:

$$p(x_k|Y_k) = \frac{p(y_k|x_k)p(x_k|Y_{k-1})}{p(y_k|Y_{k-1})} \tag{2.9}$$

Eq. 2.9 can be visualized as reported in Fig. 2.5. The application of Bayesian estimation to biological processes can be dated back to the early '80s and '90s, when the pioneering works of [42, 63–68, 23] proposed the application of Extended

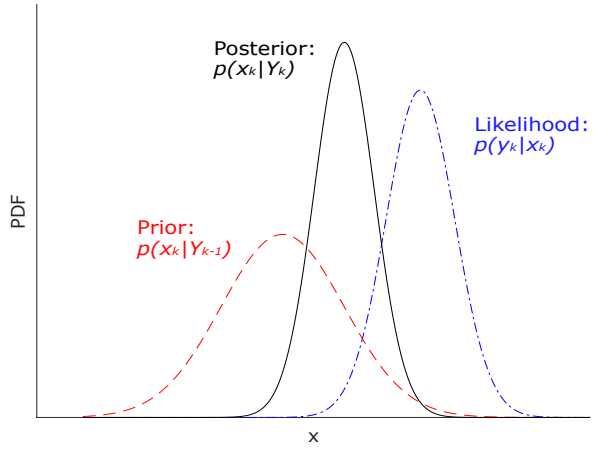


Figure 2.5: Application of Bayes' theorem for the estimation update. The figure reports how the prior PDF $p(x_k|Y_{k-1})$ propagates to $p(x_k|Y_k)$ trough the use of the Likelihood function $p(y_k|x_k)$ based on the the observations (i.e. measurements).

Kalman Filters (EKFs), for considering the random nature of process and measurements, when interested in real-time monitoring of a cultivation process. These works covered the development of high-fidelity first principle models together with the combination of state and parameter estimation [63–66], the implementation of hybrid models to cope with the complex behaviour of biological systems [42] and the use of infrequent or delayed measurements [67, 68, 23].

Further studies have showed the experimental application of the EKF for combination of state and input estimation and the application of advanced measurement techniques (e.g. near infrared and Raman spectroscopy) by coupling *on-line* and *off-line* measurements. However, the linearized nature of EKFs, which could potentially cause numerical challenges [69, 53], coupled with the improvement of computational power, led the attention to nonlinear estimation techniques as Unscented Kalman Filters (UKFs), Particle Filters (PFs) and Moving Horizon Estimators (MHEs).

The most recent experimental applications covered and spanned between applications of 1) EKFs coupling *on-line* and *off-line* measurements [70–72]; 2) PFs implemented with advanced measuring devices (i.e Raman Spectroscopy) [73], considering input uncertainty and complex sugar mixtures [34] and using time-delayed *off-line* measurements [74]; 3) UKFs coupled with NMPC [75], with hy-

brid adaptations [18] and for joint state and parameter estimation [76]; 4) MHEs formulated through a min-max optimization approach to account for parameter uncertainty [77] and coupled with NMPC implementation [78].

In accordance with the aforementioned works, our work intends to address tuning, constraints handling, robustness due to high input variation, multi-rate estimators to include infrequent measurements and adaptive approaches which were given little or no attention before. Indeed, by using experimental data, our work deals with the application of three different Bayesian estimators (i.e. EKF, UKF and MHE) to estimate the concentration of the variables of interest under absence of full state feedback and under high input variation for a fully observable system.

2.4.1 Extended Kalman Filter (EKF)

When applying Bayesian state estimators to nonlinear systems, the most common approach is the Extended Kalman Filter (EKF) [79, 53]. In the EKF, the probability density function (PDF) is propagated, at every iteration, through the linearization of the system around the operating point, inherently causing linearization errors which may not be appropriate for some systems [62, 53].

In this work, we implemented the discrete-time EKF [62, 53], however, as reported in [62], this can be also implemented as continuous-time or hybrid, depending on how the Riccati differential equation is solved. At $k = 0$ the EKF is initialized as:

$$\hat{x}_0^+ = E\left[x_0\right] \tag{2.10a}$$

$$P_0{}^+ = E\left[(x_0 - \hat{x}_0^-)(x_0 - \hat{x}_0^-)^T\right] \tag{2.10b}$$

For $k = 1, \cdots, N$ the estimated states are calculated as follows:
(1) the Jacobians of the process model are calculated using the filter estimate ($\hat{x}_{k-1}$) as the nominal state trajectory:

$$F_{k-1} = \left.\frac{\partial f(x_{k-1}, \theta, u_{k-1})}{\partial x}\right|_{\hat{x}_{k-1}} \tag{2.11a}$$

$$G_{k-1} = \left.\frac{\partial f(x_{k-1}, \theta, u_{k-1})}{\partial w}\right|_{\hat{x}_{k-1}} \tag{2.11b}$$

(2) the predictions for state estimates ($\hat{x}_k^-$) and the estimation error covariance matrix ($P_k{}^- \in \mathbb{R}^{n_x \times n_x}$), where $Q_{k-1} = G_{k-1} Q_\omega {G_{k-1}}^T$ and $Q_\omega \in \mathbb{R}^{n_\omega \times n_\omega}$ ($n_\omega = n_x + n_\theta$, Eq. 2.7) is a tuning parameter, are obtained as:

$$\hat{x}_k^- = F(\hat{x}_{k-1}^+, \theta, u_{k-1}) \tag{2.12a}$$

$$P_k{}^- = F_{k-1} P_{k-1}{}^+ {F_{k-1}}^T + Q_{k-1} \tag{2.12b}$$

(3) the Jacobian of the observation model

$$H_k = \left.\frac{\partial h(x_k)}{\partial x}\right|_{\hat{x}_k} \tag{2.13}$$

(4) through measurement update, incorporating the measurement y_k, the Kalman gain (K_k), the estimate ($\hat{x}_k^+$) and the posterior error covariance matrix (P_k^+) are updated as:

$$K_k = P_k^- {H_k}^T (H_k P_k^- {H_k}^T + R)^{-1} \tag{2.14a}$$
$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - H_k \hat{x}_k^-) \tag{2.14b}$$
$$P_k^+ = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R K_k^T \tag{2.14c}$$

For the linear Kalman Filter (KF), P_k^+ is equal to the covariance of the estimation error and therefore quantifies the uncertainty in the estimates. Differently, in the linearized Kalman filter (i.e. EKF) this is no longer true because of the linearization error. However, if the nonlinearities in the system and measurement functions are not too severe or, if the linearization errors are small, then P_k^+ is a good approximation of the covariance of the estimation error [79, 53].

2.4.2 Unscented Kalman Filter (UKF)

Differently from the EKF, which updates the first order moment (i.e. mean) through the nonlinear function and the second order moment (i.e. covariance) through linearization, the UKF instead approximates the PDF using a minimal set of deterministically chosen weighted sample points (i.e. sigma points), by propagating them through the nonlinear function (i.e. true system) as reported in Fig. 2.6. The sigma points therefore completely capture the true mean and covariance of the prior random variable and, when propagated through the true nonlinear system, enable to obtain the posterior mean and covariance accurately to the second order [80]. Therefore, while the UKF is accurate up to second order moments in the PDF propagation, the EKF is accurate up to first order moment [53].

The sigma points (i.e. weighted samples, $\chi_{k,i}$) are calculated using the mean and square root decomposition, using Cholesky factorization, of the covariance matrix

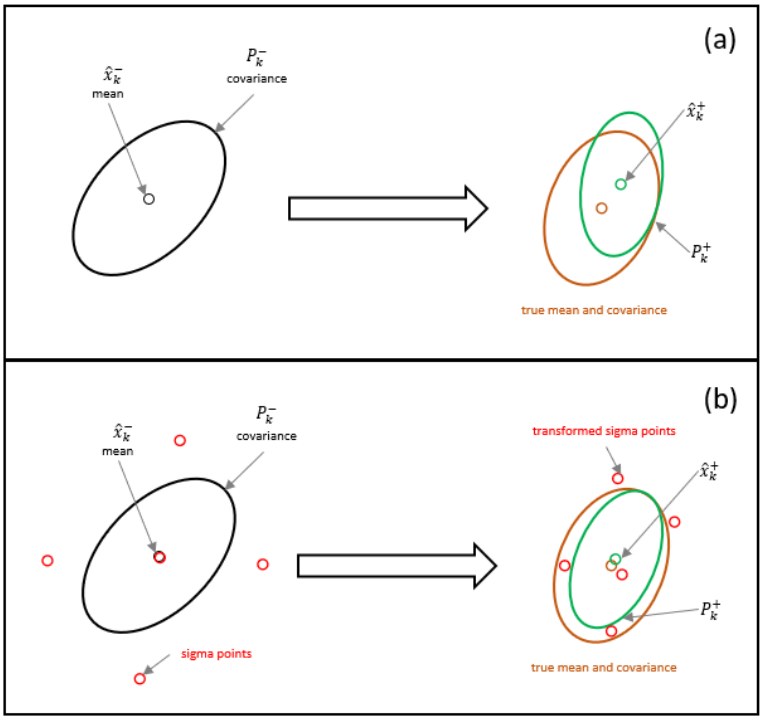


Figure 2.6: The figure, adapted from [80, 53], reports the comparison for mean and covariance propagation for the EKF (a) by linearization and the UKF (b) trough the sigma-points transformation.

$P \in \mathbb{R}^{n_x \times n_x}$ of the prior, to propagate the system from time $(k-1)^+$ to k^-, as:

$$\chi_{k-1,0} = \hat{x}_{k-1}^+ \tag{2.15a}$$

$$\chi_{k-1,i} = \hat{x}_{k-1}^+ + \tilde{x}_i \qquad i = 1, \cdots, 2n_x \tag{2.15b}$$

$$\tilde{x}_i = \left(\sqrt{(n+\lambda){P_{k-1}}^+}\right)_i^T \qquad i = 1, \cdots, n_x \tag{2.15c}$$

$$\tilde{x}_{n+i} = -\left(\sqrt{(n+\lambda){P_{k-1}}^+}\right)_i^T \qquad i = 1, \cdots, n_x \tag{2.15d}$$

where, λ is defined as:

$$\lambda = \alpha^2(n_x + k) - n_x \tag{2.16}$$

and n_x, α and k are respectively the size of the state vector and two tuning parameters. The parameter α controls the size of the sigma point distribution [53]. The sigma points are than:

(1) transformed using the nonlinear system $F(\cdot)$:

$$\chi_{k,i} = F(\chi_{k-1,i}, \theta, u_k) \tag{2.17}$$

(2) combined to obtain an *a priori* state estimate at the discrete time point k:

$$\hat{x}_k^- = \sum_{i=0}^{2n} {W_m}^{(i)} \chi_{k,i} \tag{2.18}$$

and (3) used to estimate the *a priori* error covariance as:

$$\hat{P}_k^- = \sum_{i=0}^{2n} {W_c}^{(i)} (\chi_{k,i} - \hat{x}_k^-)(\chi_{k,i} - \hat{x}_k^-)^T + Q_k \tag{2.19}$$

The weights W_m and W_c are defined as:

$$ {W_m}^{(0)} = \frac{\lambda}{n+\lambda} \qquad i = 0 \tag{2.20a}$$

$$ {W_c}^{(0)} = \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta) \qquad i = 0 \tag{2.20b}$$

$$ {W_m}^{(i)} = {W_c}^{(i)} = \frac{\lambda}{2(n+\lambda)} \qquad i = 1, \cdots, 2n \tag{2.20c}$$

with β used to incorporate higher order moments of the distribution [53]. For the measurement updates, the sigma points are transformed into $\gamma_{k,i}$:

$$\gamma_{k,i} = H_k \chi_{k,i} \tag{2.21}$$

and further combined to obtain the predicted measurement $\hat{y}_k$ at time point k:

$$\hat{y}_k = \sum_{i=0}^{2n} {W_m}^{(i)} \gamma_{k,i} \tag{2.22}$$

The covariance matrix of the predicted measurements is calculated as:

$$P_y = \sum_{i=0}^{2n} {W_c}^{(i)} (\gamma_{k,i} - \hat{y}_k)(\gamma_{k,i} - \hat{y}_k)^T + R \tag{2.23}$$

Whereas, the cross-covariance between $\hat{x}_k^-$ and $\hat{y}_k$ as:

$$P_{xy} = \sum_{i=0}^{2n} {W_c}^{(i)} (\chi_{k,i} - \hat{x}_k^-)(\gamma_{k,i} - \hat{y}_k)^T \tag{2.24}$$

Through the measurement update (Eqs. (2.25b), (2.25c) and (2.25d)), implemented following the reformulation of [19] to avoid round-off errors, the posterior statistics are calculated (i.e. approximated) as:

$$K_k = P_{xy} {P_y}^{-1} \tag{2.25a}$$

$${\chi_{k,i}}^{+} = {\chi_{k,i}}^{-} + K_k(y_k - \gamma_{k,i}) \tag{2.25b}$$

$$\hat{x}_k^{+} = \sum_{i=0}^{2n} {W_m}^{(i)} \chi_{k,i} \tag{2.25c}$$

$${P_k}^{+} = \sum_{i=0}^{2n} {W_c}^{(i)} (\chi_{k,i} - \hat{x}_k^{-})(\chi_{k,i} - \hat{x}_k^{-})^T \tag{2.25d}$$

2.4.3 Moving Horizon Estimator (MHE)

As an alternative to the recursive state estimators previously presented (i.e. EKF and UKF), the MHE results an alternative approach for state estimation, where the state estimates are obtained through the dependence on a bounded number of past measurements [81]. Being an optimization-based method which uses a moving window of past data (i.e. moving horizon, Fig. 2.7), the MHE has the ability to directly incorporate constraints in both state and parameters, being formulated as an optimization problem. Indeed, as it will be further presented in Chapter 3, while for EKF and UKF we additionally implemented a QP-problem to obtain the constrained forms of both EKF and UKF, this is straightforward in the MHE. Additionally, because of its use of a moving window of past data, the MHE has also the possibility to easily handle multi-rate measurements[82, 22, 83, 84, 49].

The MHE derives from batch least squares estimator (BLS), which was introduced by [61] as the first optimization based estimator [81]. However, to overcome the *curse of dimensionality* that comes with the BLS [85, 82, 81], the MHE, to approximate the BLS, only employs a bounded horizon (i.e. window) of past measurements to reduce the size of the optimization problem and summarizes the measurements which are not considered in the current horizon through the arrival cost Γ [81] (Fig. 2.7).

The MHE problem [20, 86] consists in finding the states and their noise (or additionally also the parameters θ as in Chapter 5) by solving the following constrained

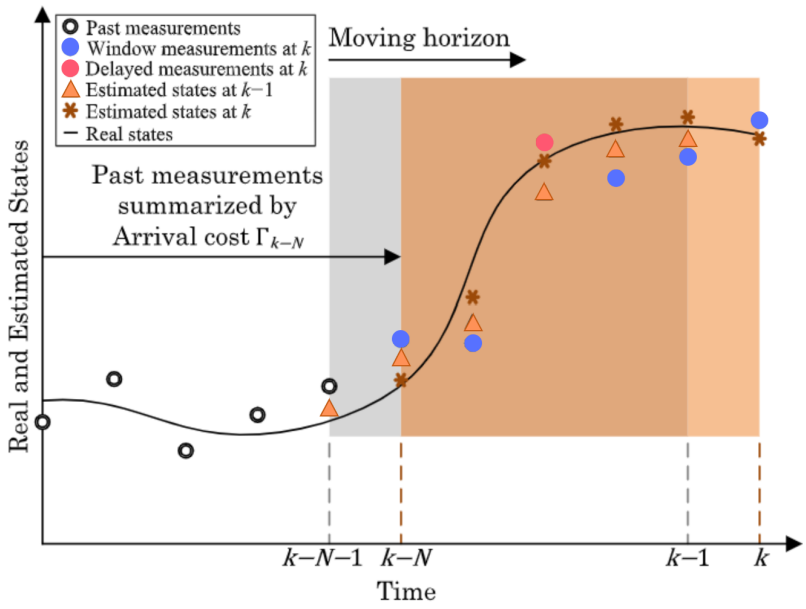


Figure 2.7: The figure reports the characteristic features of the MHE. (1) the moving horizon used at time k (from $k-N$ to k in orange), (2) the arrival cost Γ considering the past measurements up to $k-N$ and (3) the ability to include delayed measurements at time k. The figure is used with permission from [81].

least-squares optimization problem:

$$\min_{x_i, w_i} \quad \left\|\hat{x}_L - x_L\right\|^2_{P_L} + \sum_{i=L}^{N} \|y_i - h(x_i)\|^2_V + \sum_{i=L}^{N-1} \|w_i\|^2_{W_k} \tag{2.26a}$$

$$\text{s.t.} \quad x_{i+1} = F(x_i, \theta, u_i) + w_i \qquad i = L, \ldots, N-1 \tag{2.26b}$$

$$x_i \geq x_{min} \qquad i = L, \ldots, N \tag{2.26c}$$

The cost function (Eq. 2.26a) is given by the summation of three squared Euclidean norms. The first term is the arrival cost (Γ) which summarises past information (up to $L = k - N$, Fig. 2.7). In this work the arrival cost Γ was calculated through the QR-factorization proposed by [20], however, as reviewed in [81], its calculation can be performed through other several approaches (e.g. EKF). The second term is the output noise cost, while the third is the process noise cost. Following the work of [20], the three terms are weighted by:

$$P_L = P^{-1/2}, \qquad V = R^{-1/2}, \qquad W_k = {Q_k}^{-1/2} \tag{2.27}$$

where the notation for the squared norm is $\|b\|^2_B = b^T B^T B b$ [20]. The matrices

$P \in \mathbb{R}^{n_x \times n_x}$, $R \in \mathbb{R}^{n_y \times n_y}$ and $Q_k \in \mathbb{R}^{n_x \times n_x}$ (Eq. 2.27) are respectively estimation error, measurement noise and process noise covariance matrices. The states x_i are constrained with x_{min} as a lower bound, to avoid negative (i.e. infeasible) concentrations, while the term $\hat{x}_L$ (Eq. 2.26a) represents the optimal estimate of x_L.

3 | Experimental Application of Bayesian State Estimators for Bioprocess Monitoring

"The best material model of a cat is another, or preferably the same, cat."

NORBERT WIENER (1894-1964)

Journal of Process Control 106 (2021) 195–207

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Sensor fusion based on Extended and Unscented Kalman Filter for bioprocess monitoring

Andrea Tuveri [a], Fernando Pérez-García [a], Pedro A. Lira-Parada [a], Lars Imsland [b], Nadav Bar [a,*]

[a] Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[b] Department of Engineering Cybernetics, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

ARTICLE INFO

Article history:
Received 23 March 2021
Received in revised form 11 August 2021
Accepted 6 September 2021
Available online 1 October 2021

Keywords:
Sensor fusion
Soft Sensor
Extended Kalman Filter
Unscented Kalman Filter
Observability
Corynebacterium glutamicum

ABSTRACT

Availability of measurements of key process variables is of great importance to allow real time monitoring and control applications. However, in microbial fermentation processes, the unavailability of sensors or their high cost is a major barrier in automated process control applications. In this study, we implemented an Extended Kalman Filter (EKF) and an Unscented Kalman Filter (UKF), both augmented to take state constraints into account, in order to estimate biomass formation, sugar consumption and CO_2 formation in a fed-batch bacterial cultivation process. The filters use a simple monod-growth model combined with an *in-situ* absorbance probe and an infrared off-gas measurements device for monitoring of biomass production and substrate consumption. We tuned the covariance matrices of the filters by a dedicated experiment, and tested their performance on independent set of experiments. Our results demonstrate precise estimation of the biomass and glucose consumption during the batch and the feeding phases, particularly when the process covariance matrix is adapted to the phases to account for model inaccuracies during the feeding phase.

1. Introduction

Corynebacterium glutamicum is an industrial workhorse widely used as a production organism in white biotechnology [1]. *C. glutamicum* has been used for decades in the large-scale production of amino acids, more specifically L-glutamate and L-lysine [1]. However, the product portfolio from *C. glutamicum* bioprocesses is continuously growing and, nowadays, *C. glutamicum* has become a production platform for various bulk and fine chemicals, materials, and biofuels [2]. The wide use of this bacteria requires improvement to enable automatic processes with implementation of feedback controllers to avoid manual and, therefore time consuming, operations. Ideally, all the important process variables can be measured and applied in the feedback control strategy in order to achieve the desired performance of the process. However, the lack of *on-line* sensors is the main barrier for the implementation of feedback controllers. Indeed, high frequency measurements are of great importance in monitoring and control applications, but measurements of all the states in a desirable frequency is not always possible. One way to overcome this challenge is to employ sensor fusion, in which different combinations of measurements are used, together with a state model, to extrapolate information about unmeasured states in a so called Soft Sensor.

In microbial cultivation, one of the most accurate measurement techniques to quantify biomass (X), is through cell dry weight (CDW). This offers a direct measurement of the biomass, but its main disadvantage is the infrequent and delayed measurements (measurements that are analytically time-consuming, and can only be read after a considerable time delay). One alternative is to determine the *on-line* optical density (OD) through specific absorbance wave lengths, in particular *in-situ* near-infrared (NIR) probes [3]. However, correlation between the OD data and the biomass can be inaccurate, especially under low cellular concentrations and under high stirring rates, which generate noise in the NIR signal, emerging from the formation of small bubbles [4]. Measuring the substrate and product concentrations is a more challenging task. Sugars can be measured by High Performance Liquid Chromatography (HPLC) methods, or they can be measured photometrically by using enzymes kits or by spectroscopic data, using appropriate wavelengths with NIR probes [5]. Although instruments as HPLC or large spectral range NIR devices [5] are available, their high costs or their delayed measurements are major drawbacks. The use of simpler and easily available measurements such as CO_2 offered a partial solution to the lack of high frequency measurement devices. As Gudi et al. [6] state, under constant pH and head pressure the amount of CO_2 in the broth

* Corresponding author.
E-mail address: nadi.bar@ntnu.no (N. Bar).

https://doi.org/10.1016/j.jprocont.2021.09.005

can be considered constant due to rapid gas-phase dynamics, thus the gaseous CO_2 outflow can be related to growth [5].

Previous reports include a model-free Extended Kalman Filter (EKF) for parameter and state estimation in a yeast fermentation, using macroscopic and elemental balances, with transfer rates of O_2 and CO_2 in the liquid phase as the only *on-line* measurements [7]. Gudi et al. [6] implemented an EKF in a bacterial culture with the use of delayed measurements and model equations including carbon dioxide evolution rate (CER) and oxygen uptake rate (OUR) as secondary measurements. Further studies have also applied the EKF in presence of input uncertainty in an anaerobic digestion [8] and in a yeast fermentation [3], coupling *on-line* and *off-line* measurements. Goffaux et al. [9] presented an alternative solution for state estimation with scarce *off-line* measurements, where they applied EKF in the time intervals between samples to overcome the scarcity of measurements. These successful pioneering implementations of the EKF in biological processes [6,7], combined with new measurement techniques such as Near Infrared Spectroscopy [4,10,11], brought efficient applications of the EKF as state estimator in microbial fermentation systems. However, the implementation of EKF suffers from numerical challenges due to linearization, particularly when dealing with highly non-linear processes [12].

The numerical challenges associated with nonlinear systems led the attention to nonlinear estimation techniques such as the Unscented Kalman Filters (UKF) [13], and the Particle Filters (PF) [14], that gained momentum due to the improvement of computational power. PFs for instance, have been applied in filamentous fungi to estimate the states during penicillin production process using CO_2 and O_2 *on-line* measurements and time delayed measurements of biomass [15]. In a similar process, the PF was used to estimate the penicillin using off-gas measurements combined with Raman Spectroscopy [16]. Later, PFs have been implemented in bacterial cultivations [17] under input uncertainty, using complex substrate mixtures (e.g. xylose, mannose and glucose). Although the PFs reported above provided satisfying results, their implementation can turn challenging in time-critical applications [15]. The UKFs in contrast, are less computationally demanding than the PFs and are comparable to computation efforts of EKFs [12,14,18]. The UKF approach appears to be promising, as reported in few implementations in bioprocesses [5,19]. One is the work of Dewasme et al. [19], which presents the implementation of a Nonlinear Model Predictive Controller (NMPC) coupled with an UKF. The other is the work of Kramer et al. [5], applying a sigma-point Kalman Filter (SPKF) with NIR spectroscopy, coupled with partial least squares (PLS) modelling for state estimation in a yeast fermentation. Furthermore, to the best of our knowledge, only their work applied state constraints in a real bioprocess application. The implementation of state constraints is important in bioprocess, for instance, during substrate depletion in batch processes, the state estimator can calculate (infeasible) negative values of sugar concentrations.

Here, we present an implementation of UKF and compare its performances with an EKF. Different from the work of [5], we implemented the UKF without the use of *off-line* measurements and sophisticated NIR probes [5] (e.g. large spectral range). Furthermore, unlike the approach of [5] that is based on the projection of the sigma-points in the constrained area without accounting for the new measurements, our approach is implemented as a QP-problem (quadratic programming problem) to recalculate and correct the sigma-points based on the measurements, knowledge on their noise, and the predicted covariance [18]. For a fair comparison, we applied the same constraint handling also to the EKF. The tuning, and then the testing of the estimators were conducted by specially designed sets of experiments (Exp. I for tuning and II for testing). The implementation of the filters was carried out after the experiments were performed, in a true online fashion.

The two different state estimators could improve the prediction of biomass and substrate concentration *on-line* during cultivation accurately. The filters are able to estimate all the model states also in the complete absence of direct sugar measurements and in spite of model inaccuracies that emerge from variances in the parameters and missing dynamics. The accuracy of the estimation of the EKF and the UKF was comparable. The state estimators were able to use measurements of CO_2 and biomass to obtain information of the glucose. Our results, however showing limitations under unmodelled dynamics, achieved good estimation results without the need of any *off-line* measurement or measurement devices such as HPLC, NIR spectroscopy and Raman Spectroscopy, hence simplifying laboratory set-ups, allowing a fully automatic monitoring of the states.

2. Material and methods

2.1. Platform microorganism

The Gram-positive, biotin auxotroph, and rod-shaped soil bacterium *Corynebacterium glutamicum* is an industrial workhorse widely used as a production organism in industrial biotechnology [1]. *C. glutamicum* ATCC13032 [20] was used in this work.

2.2. Experimental setup

The cells of *C. glutamicum* were harvested from a 200 mL overnight shake flasks pre-culture, in complex medium 2YT (16 g of tryptone, 10 g yeast extract and 5 g of NaCl per litre) before inoculation. The fermentation was performed in a 2.7 L baffled stirred tank reactor Labfors5 (Infors AG, Switzerland). The initial working volume of 1.5 L was inoculated to an OD_{600} of 1 approximately. The reactors were equipped with two six-bladed Rushton impellers, with a distance from the bottom of the reactor of 6 cm and 12 cm. The feed of 500 mL with 100 g/L of glucose was added when the dissolved oxygen increased above 60% and remained above it. An absorbance probe for biomass monitoring (ASD12-N Absorption Probe, Optek GmbH) was used to measure absorbance in the culture broth (wavelength 840–910 nm) and an infrared off-gas analyzer for offgas composition (BlueInOne Ferm, BlueSens GmbH). Dissolved oxygen was controlled above 30% by stirrer speed (200–1100 rpm), while the reactor was kept at 1 bar and was aerated with 2 NL/min pressurized air. The temperature was kept at 30 °C and the pH was maintained at 7 by addition of KOH (184 g/L) and H_3PO_4 (188 g/L). Antifoam (Antifoam 204, Sigma) was added manually when necessary. The processes were considered finished when the dissolved oxygen value raised from 30% to 60%. *C. glutamicum* was grown using the minimal medium CGXII and glucose as sole-carbon source (20 g/L initial concentration). The composition of CGXII per litre is as follows: 10 g $(NH_4)2SO_4$, 5 g urea, 0.26 g KH_2PO_4, 0.53 g K_2HPO_4, 0.01325 g $CaCl_2 \times 2H_2O$, 0.25g $MgSO_4 \times 7\ H_2O$, 0.2mg biotin, 1mg $FeSO_4 \times 7H_2O$, 1 mg $MnSO_4 \times H_2O$, 0.1mg $ZnSO_4 \times 7H_2O$, 0.02mg $CuSO_4$, and 0.002mg $NiCl_2 \times\ 6H_2O$ [21].

2.3. Analytical procedures

Cell supernatants were automatically drawn and diluted 1 : 10, filtered through membrane filters (0.45 μm pore size), and stored at 4 °C by the NUMERA system (Securecell, Switzerland). A high-pressure liquid chromatography system (UltiMate 3000 series, Thermo Scientific, U.S.) was used for the quantification of extracellular glucose. The separation of sugars was done using a prewarm at 80 °C in a 300 × 7.8mm NUCLEOGEL SUGAR 810 Pb column (Macherey-Nagel, Germany), followed by a refractive index detector (RefractoMax 520, Thermo Scientific). Deionized

water was used at 0.4 mL/min as mobile phase. Biomass was determined through filtering 3 mL of culture broth in pre weighted filters of 0.2 μm. Each sample was rinsed with water and dried at 60 °C for 72 h.

3. State estimation

This section is structured as follows: firstly, we introduce the system (Section 3.1) and assess its structural identifiability (Section 3.1.1). Secondly, we present the available measurement signals (Section 3.2) and, given them, we analyse the observability of the system (Section 3.3). Lastly, we propose our implementation approach (Section 3.4) and apply it to the EKF and the UKF.

3.1. System model

The system is modelled using Monod-like kinetics for growth on a single sugar, with linear cell death and considering dilution when the feeding is added, as follows [22]:

$$\begin{cases} \frac{dV}{dt} = F_{in} \\ \frac{dX}{dt} = -\frac{F_{in}}{V}X + \mu_{max}\frac{S}{K_s+S}X - k_d X \\ \frac{dS}{dt} = \frac{F_{in}}{V}(S_{in}-S) - \mu_{max}\frac{S}{K_s+S}\frac{X}{Y_{XS}} \\ \frac{d\mathrm{CO_2}}{dt} = \mu_{max}\frac{S}{K_s+S}\frac{X}{Y_{X\mathrm{CO_2}}} - q_{air}\mathrm{CO_2} \end{cases} \tag{1}$$

with state and input vector defined as:

$$x = [V,\ X,\ S,\ \mathrm{CO_2}]^T \quad \text{and} \quad u = [F_{in}]$$

where V, X, S and CO_2 denote the states, volume, biomass, substrate and carbon dioxide, respectively. F_{in} is the flow in the reactor, S_{in} is the concentration of substrate in the input, μ_{max}, K_S, k_d, Y_{XS} and $Y_{X\mathrm{CO_2}}$ are the time invariant parameters (Table 1) and q_{air} is the air inflow. The equation for CO_2 is adopted from [5]. The parameters were obtained from one experimental data set of biomass, glucose and CO_2 values from the *off-gas* analyzer by using a nonlinear least-squares data fitting algorithm (lsqnonlin, Matlab).

Table 1
Model parameters of Equation (1) with estimated values, units, and standard deviation.

Parameter	Description	Value	Unit	Std. Dev.
μ_{max}	Maximum growth rate	0.19445	$[h^{-1}]$	$3.25 \cdot 10^{-6}$
K_S	Monod growth constant	0.007	$[g\,L^{-1}]$	$3.92 \cdot 10^{-6}$
k_d	Death rate constant	0.006	$[h^{-1}]$	$4.49 \cdot 10^{-6}$
Y_{XS}	S from X yield	0.42042	$[g\,g^{-1}]$	$3.58 \cdot 10^{-6}$
$Y_{X\mathrm{CO_2}}$	CO_2 from X yield	0.54308	$[g\,g^{-1}]$	$2.22 \cdot 10^{-6}$

The parameters standard deviation was calculated using the Fisher Information Matrix (FIM). Indeed the FIM is a lower bound for the parameters covariance matrix [23]. The FIM is approximated by the number of samples (N), the weighted residual sum of squares (RSS) and the Jacobian matrix J for the set of optimal parameters ($p = \hat{p}$, Table 1) as follows [23]:

$$FIM_p = N(RSS)^{-1}JJ^T|_{p=\hat{p}} \tag{2}$$

The parameter covariance matrix is then calculated as:

$$cov(\hat{p}) = FIM_p{}^{-1} \tag{3}$$

3.1.1. Structural identifiability

Structural identifiability is essential when conducting parameter estimation. However, structural identifiability only assess if it possible to uniquely determine the process parameters given the outputs (measurements) and the dynamics of the system. Indeed, structural identifiability does not imply practical identifiability, since this will be dependent on the quantity and quality of the experimental data and on the accuracy with which the model can describe the dynamics of the system (minimum model mismatch) [24]. For the structural identifiability analysis the GenSSI toolbox was used in this work [25]. The method computes the Lie derivatives of the system, and the Jacobian of the series coefficient with respect to the parameters [26]. The computation determined that the system in Eq. (1) is structurally identifiable.

3.2. Signal processing

The signals are collected every 60 s from the different measurement devices and handled through the Process Information Management System Lucullus (Securecell, Switzerland). Signals from the absorbance probe (wavelength range 840–910 nm) are calibrated with the regression curves from experimental data to set them in CDW. The absorbance probe is an invasive probe (ASD12-N Absorption Probe, Optek GmbH) that measures absorbance in the culture broth in a range of $0.05 - 4$ concentration units (CU). The calibration curve from CU to cell dry weight (CDW) is obtained by linear regression from experimental data for two separate ranges.

For $CU \geqslant 0.9$:

$$CDW_{CU} = 22.187 \cdot CU - 5.0991 \tag{4}$$

and for $CU < 0.9$:

$$CDW_{CU} = 11.124 \cdot CU + 0.66116 \tag{5}$$

Signals from the *off-gas* analyzer are instead scaled with respect to the volume of the broth, to obtain a relative value for the CO_2. The *off-gas* analyzer is a non-invasive infrared (IR) measurement device (BlueInOne Ferm, Blue-Sens GmbH) that measures the concentration of CO_2 in the outflow in a range between 0%–25%. The signal for the volume is obtained by integrating the signals from the pumps and the volume withdrawn with the *off-line* samples.

3.3. Local observability

The observability of a system defines when it is possible to infer the internal states given the history of measurements and the corresponding inputs [14]. However, when dealing with nonlinear systems, global observability is not applicable, and we have to consider local observability as defined in [27]. We consider the system as:

$$\dot{x} = f(x) + g(x) \cdot u \tag{6}$$

$$y = h(x) \tag{7}$$

We can find the observation space of the system by calculating the map O, therefore we calculate the Lie derivatives as follows:

$$L_1 = h(x) \tag{8}$$

$$L_2 = L_f h = \frac{\partial h}{\partial x} f \tag{9}$$

$$L_3 = L_g h = \frac{\partial h}{\partial x} g \tag{10}$$

$$L_4 = L_f(L_f h) = \frac{\partial (L_f h)}{\partial x} f \tag{11}$$

$$L_5 = L_g(L_f h) = \frac{\partial (L_f h)}{\partial x} g \tag{12}$$

$$L_6 = L_f(L_g h) = \frac{\partial (L_g h)}{\partial x} f \tag{13}$$

$$L_7 = L_g(L_g h) = \frac{\partial (L_g h)}{\partial x} g \tag{14}$$

From this we obtain the observation space as:

$$O = \{L_1, L_2, L_3, L_4, L_5, L_6, L_7\} \tag{15}$$

We then find the observability codistribution dO:

$$dO = span\{d(L_1), d(L_2), d(L_3), d(L_4), d(L_5), d(L_6), d(L_7)\} \quad (16)$$

This gives the dimension of the observability codistribution. When $dO(x)$ has the dimension of the state vector (n), the system is locally observable. This can be used to identify, given the system, the necessary measurements for the system to be locally observable. As reported in Appendix, the observability analysis gives the possibility to evaluate, given the process dynamics and the measurement function, which measurements are needed to ensure the observability of the system at every time [27].

3.4. Proposed approach

The two different state estimators are implemented as in Fig. 1. When glucose reaches negative values, a quadratic function subject to linear constraints is optimized (Quadratic Programming, QP) to obtain feasible estimates (block 3, Fig. 1). The implementation of the QP-problem differs in the two cases since it constraints directly the states in the EKF (Eq. (31)) and the sigma points in the UKF (Eq. (50)). Indeed, constraining the sigma points has a direct impact on the error covariance matrix (block 7, Fig. 1) [18], which is therefore updated in the UKF. To adapt the filter to the case of abruptly changing dynamics, as it happens under a sudden increase in sugar concentration, the parameter uncertainty is increased to correct the filter trust in the measurements (block 8, Fig. 1).

3.4.1. Extended Kalman filter

The Extended Kalman Filter (EKF) is implemented as in [12, 14]. The system is given as follows, where f (Eq. (17)) is obtained by discretizing the system described in Eq. (1) by explicit Euler method.

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1}) \quad (17)$$

$$y_k = h(x_k) + v_k \quad (18)$$

$$w_k \sim \mathcal{N}(0, Q_k) \quad (19)$$

$$v_k \sim \mathcal{N}(0, R) \quad (20)$$

with discrete-time process (w_k) and measurement (v_k) noise, respectively with covariance Q_k and R. At $k = 0$ the EKF is initialized as:

$$\hat{x}_0^+ = E[x_0] \quad (21)$$

$$P_o{}^+ = E\left[(x_0 - \hat{x}_0^-)(x_0 - \hat{x}_0^-)^T\right] \quad (22)$$

For $k = 1, \cdots, N_{end}$ estimated states are calculated as follows:

(a) the Jacobians of the process model

$$F_{k-1} = \left.\frac{\partial f(x_{k-1}, u_{k-1}, 0)}{\partial x}\right|_{\hat{x}_{k-1}}$$

$$G_{k-1} = \left.\frac{\partial f(x_{k-1}, u_{k-1}, 0)}{\partial w}\right|_{\hat{x}_{k-1}}$$

(b) the predictions for state and covariance

$$\hat{x}_k^- = f(\hat{x}_{k-1}^+, u_{k-1}, 0) \quad (23)$$

$$P_k{}^- = F_{k-1}P_{k-1}{}^+F_{k-1}{}^T + G_{k-1}Q_{k-1}G_{k-1}{}^T \quad (24)$$

(c) the Jacobian of the observation model

$$H_k = \left.\frac{\partial h(x_k)}{\partial x}\right|_{\hat{x}_k} \quad (25)$$

(d) the measurement update, incorporating the measurement y_k

$$K_k = P_k^- H_k{}^T(H_k P_k^- H_k{}^T + R)^{-1} \quad (26)$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - h(\hat{x}_k^-)) \quad (27)$$

$$P_k^+ = (I - K_k H_k)P_k^-(I - K_k H_k)^T + K_k R K_k^T \quad (28)$$

We constrained the EKF using the QP-problem proposed by [18] for constraining the sigma-points in the UKF, differently from the constrained approaches proposed in [28].

The QP-problem defined as:

$$\min_{\hat{x}_k^+} \quad J \quad \text{s.t.} \quad D_k\hat{x}_k^+ \le d_k \quad (29)$$

is solved by minimizing the objective function

$$J = \hat{x}_k^{+T}(D_k{}^T R_k{}^{-1} D_k + (P_k{}^-)^{-1})\hat{x}_k^+ + \\ -2(y_k{}^T R_k{}^{-1} D_k + (\hat{x}_k^-)^T(P_k{}^-)^{-1})\hat{x}_k^+ \quad (30)$$

over $\hat{x}_k^+$, subject to the constraints:

$$D_k\hat{x}_k^+ \le d_k \quad (31)$$

3.4.2. Unscented Kalman filter

The system is as for the EKF, given in Eqs. (17)–(20). The state estimate and the covariance for the Unscented Kalman Filter are implemented as follows [12,14,18], with additive noise. To propagate the system from time $(k-1)^+$ to k^-, we choose the sigma points as:

$$\chi_{k-1,0} = \hat{x}_{k-1}^+$$

$$\chi_{k-1,i} = \hat{x}_{k-1}^+ + \tilde{x}_i \qquad i = 1, \cdots, 2n$$

$$\tilde{x}_i = \left(\sqrt{(n+\lambda)P_{k-1}{}^+}\right)_i^T \qquad i = 1, \cdots, n$$

$$\tilde{x}_{n+i} = -\left(\sqrt{(n+\lambda)P_{k-1}{}^+}\right)_i^T \qquad i = 1, \cdots, n \quad (32)$$

Here, λ is given by

$$\lambda = \alpha^2(n+k) - n \quad (33)$$

where n is the size of the state vector and α and k are tuning parameters. The parameter α controls the size of the sigma point distribution [12]. The matrix square root was calculated by using Cholesky factorization. We transformed the sigma points using the nonlinear system $f(\cdot)$:

$$\chi_{k,i} = f(\chi_{k-1,i}, u_k, 0) \quad (34)$$

and combined them to obtain an *a priori* state estimate at the discrete time point k:

$$\hat{x}_k^- = \sum_{i=0}^{2n} W_m{}^{(i)}\chi_{k,i} \quad (35)$$

and estimated the *a priori* error covariance as:

$$\hat{P}_k^- = \sum_{i=0}^{2n} W_c{}^{(i)}(\chi_{k,i} - \hat{x}_k^-)(\chi_{k,i} - \hat{x}_k^-)^T + Q_k \quad (36)$$

The weights W_m and W_c are defined as:

$$W_m{}^{(0)} = \frac{\lambda}{n+\lambda} \qquad i = 0 \quad (37)$$

$$W_c{}^{(0)} = \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta) \qquad i = 0 \quad (38)$$

$$W_m{}^{(i)} = W_c{}^{(i)} = \frac{\lambda}{2(n+\lambda)} \qquad i = 1, \cdots, 2n \quad (39)$$

with β used to incorporate higher order moments of the distribution [12]. For the measurement updates, we transform the sigma points into $\gamma_{k,i}$:

$$\gamma_{k,i} = h(\chi_{k,i}) \quad (40)$$

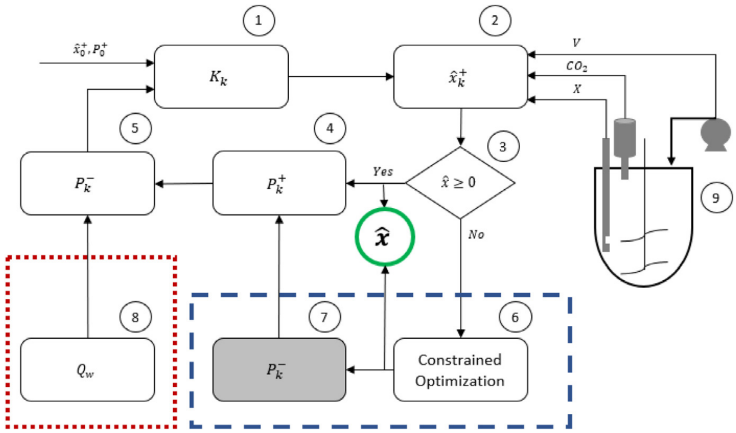


Fig. 1. Implementation Procedure. The figure reports the proposed implementation for the EKF and UKF integrated with the bacterial cultivation process. The recursive filter is represented by the blocks from 1 to 5. Where 1 represents the Kalman gain calculation after each loop, 2 the correction term, where the newly available outputs from the *on-line* process 9 (volume, biomass and CO_2) are used to estimate the states. The condition to activate or not the optimization is represented by step 3. Indeed if the concentration of glucose is non negative the error covariance matrix is updated in 4 and then used for the projection ahead in 5. In case the condition in 3 is not satisfied (negative glucose concentrations are estimated), the optimization is solved (6, blue dashed area). In the case of the EKF only the step 6 is implemented, while for the UKF also step 7, since the constrained estimation of the sigma points has a direct impact on the calculation of the error covariance matrix. The estimated states are represented by the continuous green circle. Step 8 (red dotted area) represents the entering parameters covariance, which is increased once the glucose is fed again to the system (changing systems dynamics) and kept constant afterwards. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and combined them to obtain the predicted measurement at time point k:

$$\hat{y}_k = \sum_{i=0}^{2n} W_m{}^{(i)} \gamma_{k,i} \tag{41}$$

We then estimated the covariance of the predicted measurements as:

$$P_y = \sum_{i=0}^{2n} W_c{}^{(i)} (\gamma_{k,i} - \hat{y}_k)(\gamma_{k,i} - \hat{y}_k)^T + R \tag{42}$$

The cross-covariance between $\hat{x}_k^-$ and $\hat{y}_k$ as:

$$P_{xy} = \sum_{i=0}^{2n} W_c{}^{(i)} (\chi_{k,i} - \hat{x}_k^-)(\gamma_{k,i} - \hat{y}_k)^T \tag{43}$$

The measurement update is:

$$K_k = P_{xy} P_y{}^{-1} \tag{44}$$

$$\chi_{k,i}{}^+ = \chi_{k,i}{}^- + K_k(y_k - \gamma_{k,i}) \tag{45}$$

$$\hat{x}_k = \sum_{i=0}^{2n} W_m{}^{(i)} \chi_{k,i} \tag{46}$$

$$P_k{}^+ = \sum_{i=0}^{2n} W_c{}^{(i)} (\chi_{k,i} - \hat{x}_k^-)(\chi_{k,i} - \hat{x}_k^-)^T \tag{47}$$

The measurement update equations (Eqs. (45), (46) and (47)) are implemented following the reformulation of [18] to avoid round-off errors. The constraints implementation is done by solving a QP-problem as shown in [18] when the sugar estimate goes below zero.

The QP-problem defined as:

$$\min_{\chi_{k,i}} \quad J \quad \text{s.t.} \quad D_k \chi_{k,i} \leq d_k \tag{48}$$

is solved by minimizing the objective function

$$J = \chi_{k,i}{}^T (D_k{}^T R_k{}^{-1} D_k + (P_k{}^-)^{-1}) \chi_{k,i} + \\ -2(y_k{}^T R_k{}^{-1} D_k + (\chi_{k,i}{}^-)^T (P_k{}^-)^{-1}) \chi_{k,i} \tag{49}$$

for $\chi_{k,i}$, subject to the constraints

$$D_k \chi_{k,i} \leq d_k \tag{50}$$

4. Results

The results are structured as follows: we first assess the observability of the system in Section 4.1, then we propose in Section 4.2 the tuning parameters for the EKF and UKF and evaluate the performance of the filters (Section 4.3) using two different experiments.

4.1. Observability analysis

Here we conducted local observability analysis using Lie derivatives (Section 3.3). For the system to be observable the dimension of the observability co-distribution (dO) must be equal to the dimension of the state vector ($n = 4$) at any time. This was done by simulation of different available measurement scenarios Appendix. The calculation of the Lie derivatives was done in Matlab by symbolic differentiation and this was feasible as the problem is of small size [29]. The results showed that the system is locally observable when measuring volume, biomass and CO_2 (Fig. A.8). Measuring only the volume and the biomass was critical for local observability (Fig. A.8). However, the combination of measurements of volume and CO_2 or biomass and CO_2 showed that the system can be locally observable under specific process conditions, *i.e.* in absence of glucose depletion or under continuous feeding Appendix.

4.2. Filter tuning

The filter receives the signals from the sensors every 60 s. Both the EKF and the UKF filters were tuned using Experiment I only.

Table 2
Noise, description and numerical value.

Noise	Description	Value
$\sigma^2_{\mu_{max}}$	μ_{max} variance	$1.05 \cdot 10^{-11}$
$\sigma^2_{K_S}$	K_S variance	$1.54 \cdot 10^{-11}$
$\sigma^2_{k_d}$	k_d variance	$2.02 \cdot 10^{-11}$
$\sigma^2_{Y_{XS}}$	Y_{XS} variance	$1.28 \cdot 10^{-11}$
$\sigma^2_{Y_{XCO_2}}$	Y_{CO_2} variance	$4.91 \cdot 10^{-12}$
n_V	Additive noise V	$1 \cdot 10^{-2}$
n_X	Additive noise X	$1 \cdot 10^{-2}$
n_S	Additive noise S	$1 \cdot 10^{-2}$
n_{CO_2}	Additive noise CO_2	$1 \cdot 10^{-4}$

To ensure a fair comparison, the tuning parameters are the same for the EKF and the UKF. The initial conditions are given as:

$$x_0 = \begin{bmatrix} 1.5 & 1.2 & 20 & 0 \end{bmatrix}^T$$

$$\hat{x}_0^+ = x_0$$

$$P_0{}^+ = \begin{bmatrix} 2.09 \cdot 10^{-8} & 0 & 0 & 0 \\ 0 & 1.10 \cdot 10^{-5} & 0 & 0 \\ 0 & 0 & 1.09 \cdot 10^{-4} & 0 \\ 0 & 0 & 0 & 2.17 \cdot 10^{-5} \end{bmatrix}$$

The measurement noise covariance matrix R was tuned as follows

$$R = diag\begin{bmatrix} 10^{-2} & 1 & 10^{-3} \end{bmatrix} \tag{51}$$

while the process noise covariance matrix Q_k was tuned as in [3, 30]:

$$Q_k = G_k \cdot Q_w \cdot G_k{}^T \tag{52}$$

where the Jacobian G_k is defined as:

$$G_k = \frac{\partial f(x, u, w)}{\partial w}$$

Here, w represents the noise vector

$$w = \begin{bmatrix} \sigma^2{}_{\mu max} & \sigma^2{}_{ks} & \sigma^2{}_{kd} & \sigma^2{}_{Y_{XS}} & \sigma^2{}_{Y_{XCO_2}} & n_V & n_X & n_S & n_{CO_2} \end{bmatrix}^T$$

with description and variance values in Table 2 obtained as explained in Section 3.

The matrix Q_w is a diagonal matrix of the variances elements of w:

$$Q_w = diag\begin{bmatrix} \sigma^2{}_{\mu max} & \sigma^2{}_{ks} & \sigma^2{}_{kd} & \sigma^2{}_{Y_{XS}} & \sigma^2{}_{Y_{XCO_2}} & n_V & n_X & n_S & n_{CO_2} \end{bmatrix}$$

The process noise is then modelled with the variance values reported in Table 2 obtained from the parameters standard deviation (Table 1) and with an additive noise (n_i) in order to avoid Q_k to become zero or indefinite when the substrate S is depleted. The value of Q_k then varies at every iteration (k). Because the dynamics of the process is significantly different between the first and the second batch phase, we tested two different tuning cases for Q_w: (1) Constant parameter uncertainty (CU, Table 2), and (2) modified parameter uncertainty (MU), increasing, during the second batch phase (from 21 h), the variance of $\sigma^2{}_{K_S}$ and $\sigma^2{}_{Y_{XCO_2}}$ to $3.38 \cdot 10^{-2}$ and $4.91 \cdot 10^{-2}$ respectively. Indeed, the parameter uncertainty, and therefore the covariance, is increased once the glucose is fed again to the system (second batch) and kept to that value once it changes. The rational behind the modified process noise is the increased uncertainty due to unmodelled dynamics of the glucose uptake by our simple model during the second batch phase, caused by utilization of byproducts accumulated under overflow metabolism (cf. discussion in Section 5).

The system measurement function $h(x)$ is:

$$h(x) = \begin{bmatrix} V & X & CO_2 \end{bmatrix}^T \tag{53}$$

Table 3
RMSE values for the tuning experiment (Exp. I). The values show how the filters perform similarly and always better than the model and also how the RMSE improved with modified parameter uncertainty (MU) compared to constant uncertainty (CU) between first and second batch phase.

RMSE	Biomass			Glucose		
	Model	EKF	UKF	Model	EKF	UKF
CU	2.45	1.15	1.15	1.78	1.64	1.53
MU	2.45	1.15	1.15	1.78	1.04	0.92

Table 4
RMSE values for the validation experiment (Exp. II). The values confirm the improvement obtained by the use of the filters and the similar performance of the two filters and also how the RMSE improved with modified parameter uncertainty (MU) compared to constant uncertainty (CU) between first and second batch phase.

RMSE	Biomass			Glucose		
	Model	EKF	UKF	Model	EKF	UKF
CU	2.47	1.10	1.11	2.22	1.34	1.28
MU	2.47	1.10	1.11	2.22	0.84	0.86

The linearized measurement matrix (H, Eq. (25)) is then:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{54}$$

The available measurements in our process are the volume, the value from the absorption probe and the carbon dioxide ratio in the off-gas:

$$y = \begin{bmatrix} V & CDW & \frac{CO_2}{V} \end{bmatrix}^T \tag{55}$$

The other tuning parameters for the UKF (defined in Section 3.4.2) were tuned as follows:

$$\alpha = 0.001$$

$$k = 1$$

$$\beta = 2$$

According to [12] α is chosen heuristically between 0 and 1 and should be kept as low as possible. Since we assume that the noise follows a Gaussian distribution, the value of β is optimal at 2 [12]. For the tuning of the UKF in the case of modified parameter uncertainty (MU) between first and second batch, the parameter α was selected equal to 1 to avoid divergence of the filter when the uncertainty of the estimates is high.

4.3. Experimental results

Tuning Experiment (Exp. I). The model prediction of biomass and cell death was accurate for the first batch, but exhibited a large discrepancy with the reference measurements (CDW) during the second batch (Fig. 2). In comparison, both the EKF and the UKF estimated the biomass formation with good accuracy compared to the reference measurements (RMSE 1.15, Table 3 and Fig. 2).

Increasing parameter uncertainty (MU, Section 4.2), because of unmodelled dynamics in the second batch, did not affect the biomass estimation (Table 3 and Fig. 2 bottom row).

The glucose prediction of the model and estimated values of the EKF and the UKF were accurate during the first batch phase when compared to the *off-line* measurements (Fig. 3). When maintaining the parameter uncertainty constant during the entire process (CU, Section 4.2), the values estimated from the EKF and the UKF showed the inability of the filters to correctly estimate the sugar consumption after feeding (Fig. 3 top row), likely due to unmodelled utilization of accumulated byproducts

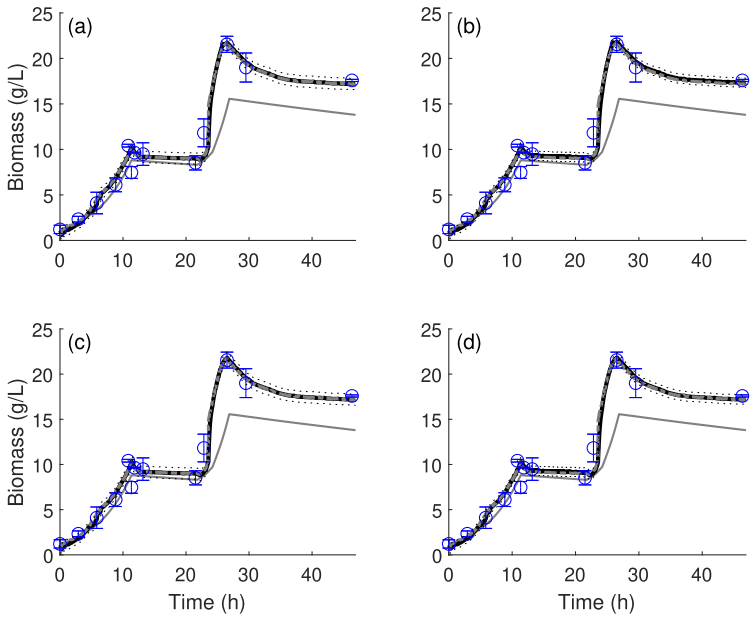


Fig. 2. Biomass growth-tuning experiment (Exp. I). This experiment was used to tune the filters. The experimental results correspond to the *off-line* measurement (blue circles with error bars), and the sensor measurement (dashed–dotted grey line for the absorption signal). The figures report the results for the two filters with same (a and b) and with different parameter uncertainty between first and second batch (c and d). The model (thin grey line), the EKF (Fig. a and c, black continuous line) and the UKF (Fig. b and d, black continuous line) can follow the measurements in the batch phase (0–21 h). However, the model simulations present discrepancies from the experimental data after the feeding (from 21 h), which are eliminated by the application of the EKF and the UKF.

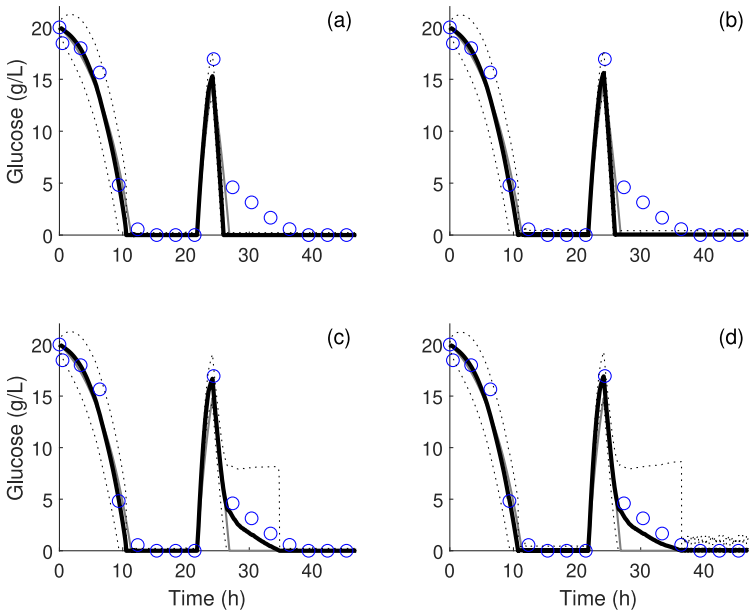


Fig. 3. Glucose concentration-tuning experiment (Exp. I). The experimental results correspond to the *off-line* measurement (blue circles). The figures report the results for the two filters with same (a and b) and with different parameter uncertainty between batch and feeding phase (c and d). The model (thin grey line), the EKF (Fig. a, black continuous line) and the UKF (Fig. b, black continuous line) can follow the measurements in the first batch phase and during the feeding (0–26 h). However, the model simulations present discrepancies from the experimental data after the feeding (a and b, from 26 h), which are not eliminated by the application of the EKF and the UKF. The dotted lines represent the confidence intervals ($2-\sigma$ intervals) of the estimated values. The results indicate that the variance is similar for both filters under dynamic conditions, but it differs under steady state between the EKF (a) and the UKF (b). However, when considering different parameter uncertainty between the first and second batch phase, both the EKF (c) and the UKF (d) are able to follow the measurements.

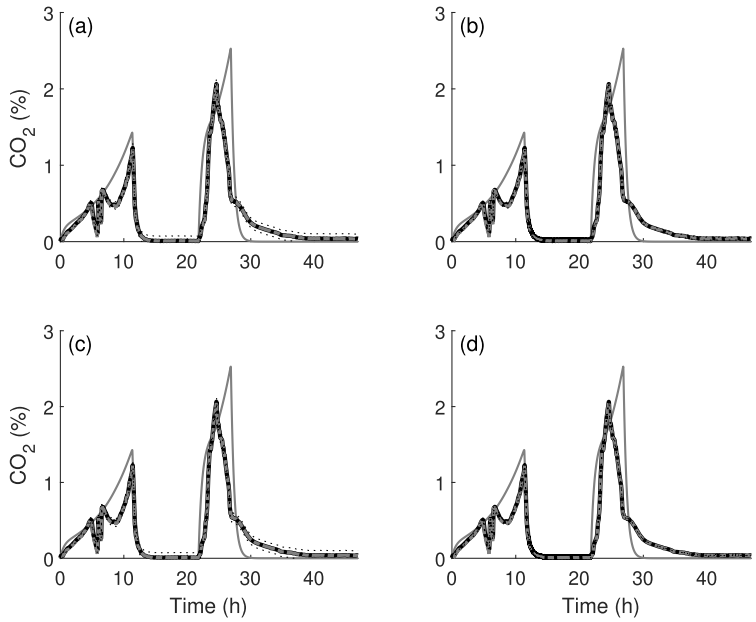


Fig. 4. CO_2 formation-tuning experiment (Exp. I). The figure reports the percentage of CO_2 in the off-gas for the two filters with same (a and b) and with different parameter uncertainty between first and second phase (c and d). The model (grey line) follows the measurements (dashed–dotted line) form the off-gas analyser in the batch phase, with the exception of noise between 5 and 7 h, while shows higher discrepancy during and after feeding. The $2-\sigma$ intervals (dotted lines) show that the EKF (a and c) shows a reduced uncertainty compared to the UKF (b and d).

(see Section 5). Most importantly, when increasing the parameter uncertainty during the second batch phase, to account for this unmodelled dynamics (MU, Section 4.2), both filters exhibited an improvement in the estimation of glucose consumption, with values closer to the *off-line* measurements (Table 3 and Fig. 3 bottom row). Interestingly, the confidence intervals of the EKF ($2-\sigma$ bounds, Fig. 3 bottom left) increased under sugar depletion, reflecting the uncertainty of the prediction, to then decrease at the end of the fed-batch. Indeed, however the estimated values of the UKF were slightly better than the ones of the EKF for the CU and MU cases (7% and 12%, respectively, Fig. 3 bottom right), the confidence intervals ($2-\sigma$ bounds) of the UKF were larger than the ones of the EKF during the second batch. This is likely because of the large discrepancy between the model prediction and measurements. This is mostly visible for the CO_2 (Fig. 4), reflecting the fact that the model uncertainty is large (see Section 5). Indeed, the model predictions of CO_2 formation followed the measurements of the off-gas during the first batch (Fig. 4) but were reduced rapidly to zero in the second one. This is in contrast to the measurements which indicate CO_2 formation emerges from catabolism of byproducts (see Section 5).

The open-loop model was less accurate compared to the two filters, exhibiting nearly double RMSE (2.45 for biomass and 1.78 for glucose, Table 3).

Note that the *off-line* measurements of the biomass (CDW) and the sugar concentration (HPLC measurements) were only used to evaluate the estimated values but were not used by the filter algorithms during the calculations.

Validation Experiment (Exp. II). We tested the performance of the model and the filters in a dedicated validation experiment. The tuning of the filters was kept unchanged, *i.e.* they were not adapted to the validation experiment (Figs. 5–7). For completeness of this work, we present the results for the CU case also in the validation experiment. Indeed, these results show that under constant parameter uncertainty the filters are unable to correctly estimate the drop in glucose consumption after feeding.

Firstly, we observed that the model had a large discrepancy compared to the *off-line* measurements (Fig. 5), both during the first and the second batch. Secondly, the error of the model is in magnitude of 2-fold compared to the filters (Table 4). The model predictions are particularly erroneous in the first exponential growth phase of the Experiment II (Figs. 5–7), discrepancy that emerged likely due to experimental and cell-to-cell variations.

Both the filters estimated the biomass accurately by using the *on-line* measurements (both for CU and MU case) with respect to the reference values (Fig. 5). The RMSE values of the filters were 55% lower than the value of the model (Table 4).

The glucose predicted by the model in the validation experiment was inaccurate for both the first and the second batch (Fig. 6), with a slower consumption than the reference measurements during the first batch and faster consumption after feeding. The EKF and the UKF showed accurate estimations of the glucose during the first and second batch (until 25 h), but during glucose depletion (26–30 h), the drop in the sugar consumption rate was not captured by the model nor by the filters in the CU case (total RMSEs 1.34 and 1.28, for the EKF and the UKF respectively, Fig. 6 top row). Most importantly, the filters with modified parameter uncertainty (MU, Section 4.2) successfully captured the slower sugar consumption after feeding (Fig. 6 bottom row), although the confidence intervals of the UKF exhibited chattering (Fig. 6 bottom right). Indeed, the case with modified parameter uncertainty (MU) between the two phases showed an improvement of the estimates at the cost of an higher uncertainty (Fig. 6).

The modelled CO_2 prediction generally captured the dynamics of the CO_2 formation in the validation experiments, but was delayed in the first batch, and with larger error during and after feeding (Fig. 7). The filters however, with the information from the measurements, captured the CO_2 formation most accurately.

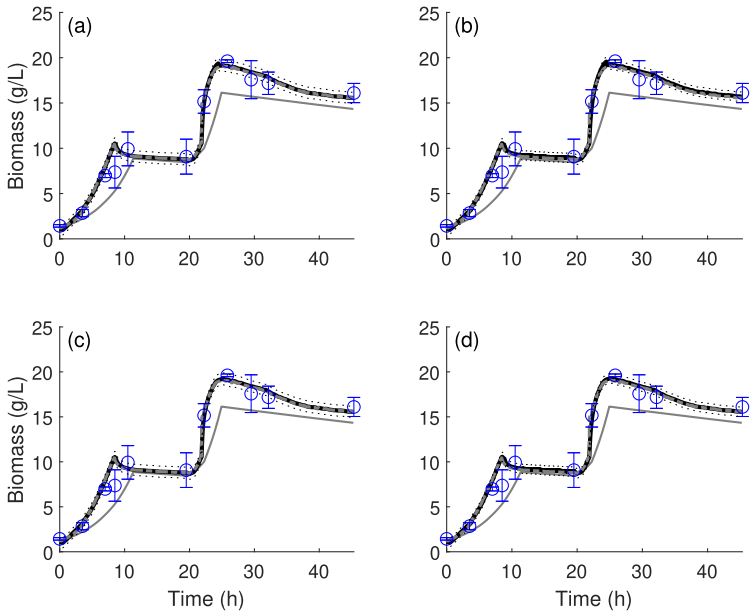


Fig. 5. Biomass growth-validation experiment (Exp. II). These experimental results were used to validate, test and compare the model performance against the EKF and the UKF. The figures report the results for the two filters with same (a and b) and with different parameter uncertainty between first and second batch (c and d). The experimental results correspond to the *off-line* measurement (blue circles with error bars), and the sensor measurement (dashed–dotted grey line for the absorption signal). The model (thin grey line), presents discrepancies from the experimental data, showing a slower growth during the whole fed-batch. However, the EKF (Fig. a and c, black continuous line) and the UKF (Fig. b and d, black continuous line) can follow the measurements.

5. Discussion

This work presents the implementation of two Kalman-like filters to estimate biomass, glucose and CO_2 during a fed-batch fermentation process. The filters were based on a process model and updated their estimates using only two measurement devices, *i.e.* an *in-situ* absorbance probe to measure biomass and an off-gas sensor to measure CO_2. The performance and accuracy of the filters were evaluated using a separate set of control measurements, *i.e.* biomass and glucose (Section 2.3), and this information was not used by the filters at any time point during the experiments.

One important novelty of this work is to be able to predict the unmeasured glucose by two filters under excess of substrate, using only an *in-situ* absorbance probe and an off-gas analyzer (Section 2). We also compared and evaluated the performance of the two filters by training–testing sets of experiments. Previous works did present successful implementations of state estimators in microbial cultivation processes [5,9,16,19], but using slow and accurate *off-line* sampling and/or advanced measurement devices, for instance using a NIR with a large spectral range (950–1750 nm) that is able to detect sugars, biomass and ethanol [5]. In fact, successful estimation of sugars based on manual biomass sampling and/or advanced NIR spectroscopy and CO_2 measurements was already demonstrated in [5], but only under substrate limitation during feeding. In contrast, excess of substrate usually leads to changes in metabolism, which influence the dynamics of substrate consumption and CO_2 formation. Because this dynamics are not modelled, the filters, which only rely on the model and the measurements, can lead to erroneous estimates of the sugar (see below). In this work, we resolved this problem by increasing the parameter uncertainty in Q_w during the second batch under excess of substrate (MU), forcing the filters to rely more on the measurements than on the model. This led both filters to accurately estimate the glucose concentration in the complete absence of direct sugar measurements and under varying metabolic conditions. Importantly, we demonstrated accurate predictions using exclusively the low cost off-gas analyser and a simple and low cost absorbance probe (Section 2.2), potentially eliminating the need to measure glucose with more expensive and *off-line* devices such as HPLC.

Not surprisingly for a simple Monod model, its prediction of a fed-batch process was biased for the biomass, glucose and CO_2, particularly under feeding. These results were consistent in both experiments, and the slight variations in the values are likely due to variations in experimental conditions (e.g. inoculation, cell culture and variations in media composition) which were not included in the model dynamics. The large discrepancy of the model compared to the reference measurements after feeding, particularly for glucose consumption, emerges most likely due to unmodelled dynamics. For instance, different authors discussed the formation of organic acids under oxygen limitation in *C. glutamicum* [31–35], such as acetate and lactate that become available for the cells. Indeed, this intermediates are released by the cells during cultivation phases with reduced growth and available substrate and subsequently utilized [35] as alternative substrates, when the cells enter a state of prolonged substrate limitation [36]. This is more likely to occur under high glucose feeding rates and low dissolved oxygen [33,34] and is not part of our model. Our simple open-loop Monod model, with its parameter estimated by a dedicated experiment (Section 3), although maybe suitable for the first batch phase, did not performed reliably in the second batch phase, for both subsequent experiments we tested. The performance of the two filter strategies was similar for the biomass, but the UKF predicted the overall glucose consumption more accurately than the EKF for both the

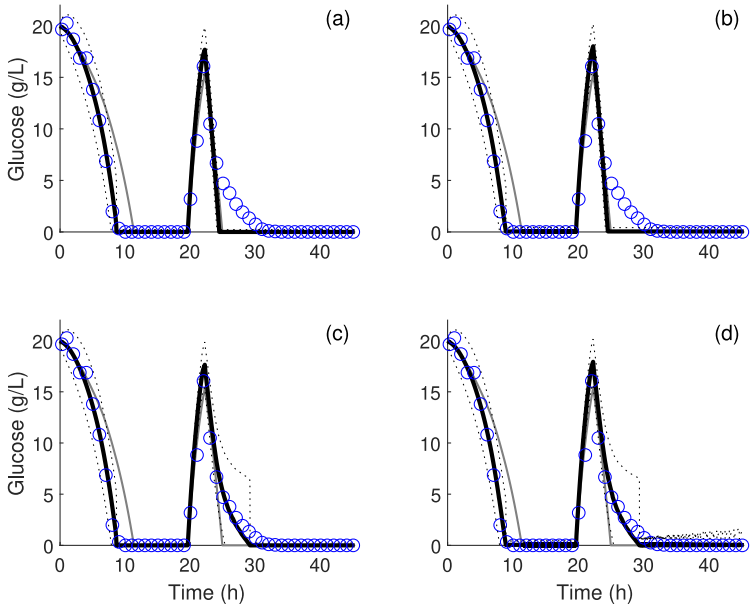


Fig. 6. Glucose concentration-validation experiment (Exp. II). The experimental results correspond to the *off-line* measurements (blue circles). The figures report the results for the two filters with same (a and b) and with different parameter uncertainty between first and second batch phase (c and d). The model (thin grey line) presents a slower consumption rate during the first batch, however the EKF (Fig. a, black continuous line) and the UKF (Fig. b, black continuous line) can follow the measurements in the first batch and during feeding (0–21 h). The model simulations present discrepancies from the experimental data after the feeding phase (from 26 h) with a faster consumption rate, which are not eliminated by the application of the EKF and the UKF. The dotted lines represent the confidence intervals ($2-\sigma$ intervals) of the estimated values. The results indicate that the variance is similar for both filters under dynamic conditions, but it differs under steady state between the EKF (a) and the UKF (b). The filters (a and b) successfully eliminated the discrepancies between model simulations and experimental data, being also able to estimate the time at which the glucose is depleted (7 h). However, when applying a different parameter uncertainty between the two phases, both the EKF (c) and the UKF (d) followed the measurements.

training set and the testing experiment. Most importantly, we demonstrated in this work that in the absence of full state measurements, both Kalman-like filters required different parameter uncertainty (Q_w) for the first and the second batch (MU). This is in order to account for the unmodelled dynamics, particularly of the byproducts utilization during the second batch phase, which decreases the glucose consumption rate. Previous works [3,5,8] tuned and employed a constant parameter uncertainty Q_w for the entire fed-batch process. Interestingly, they report accurate results for the feeding phase, likely of two important reasons: (1) all the states were measured or (2) substrate is fed with a low feeding rate, presenting a case of frequent and abrupt sugar starvation and subsequently rapid utilization of any potential byproducts. Adaption of the values in the process noise matrix Q_w that correspond to the unmodelled uncertainty, in our case the CO_2 from the biomass yield (Y_{CO2}) and the monod growth constant (K_s), were highly effective and greatly improved both filters performance (over 50% more accurate, see Tables 3 and 4). Notably, the estimators exhibited significant increase in the confidence regions ($2-\sigma$ bounds) of the sugar estimate after the feeding phase. This indicates that although the filter successfully captured the states (indicated by the lower RMSE in both experiments), the estimates had high uncertainty bounds. This is likely because in the complete absence of direct sugar measurements, the large discrepancy between the measurement and model predictions increases the predicted covariance matrix P.

In the MU case, the tuning of the UKF was highly sensitive to the spread of the sigma points (denoted by the parameter α) and therefore we needed to be more conservative to avoid large covariance increase. A possible reason the UKF exhibits larger covariance in the MU case is likely due to the design of the noise covariance matrix Q_k. A previous work presented the possibility to add an extra positive definite matrix to Q_k, seeing it as a trade-off between stability and accuracy [37]. Indeed, when this additive term is set too large, the filter can exhibit significant deviation, and the state covariance matrix P (Eq. (47)) can present high condition number, causing the filter to misinterpret the residual information and therefore lead to divergence issues [38]. Under increased covariance the UKF also exhibits chattering. This instability can be a consequence of the constraint handling implementation (QP-problem, Sections 3.4.1 and 3.4.2). While both implementations used an identical QP-problem, the states of the UKF were constrained by optimization of the sigma points, which may often trigger hard saturation constraints. Due to the different nature of the filters, a large covariance matrix is not encountered in the EKF, which being just a first order approximation, will not see the nonlinear effects [12]. The increase of the covariance matrix in the UKF will be further investigated and it will be part of future works.

We suggest several manners to improve the predictions of the filters we presented: (1) Apply full state measurements or improve the measurement devices, on the expense of increased costs and instrumentation complexity. In industrial bio-processes this may not be a feasible alternative, especially in production of low-value products. The error in the estimation can be also corrected by using a few scarce sugar measurements, particularly at the transient parts of the bio-process. This was demonstrated by [5], where the authors increased the state estimate accuracy by adding sugar measurements to their SPKF. (2) Improve the model

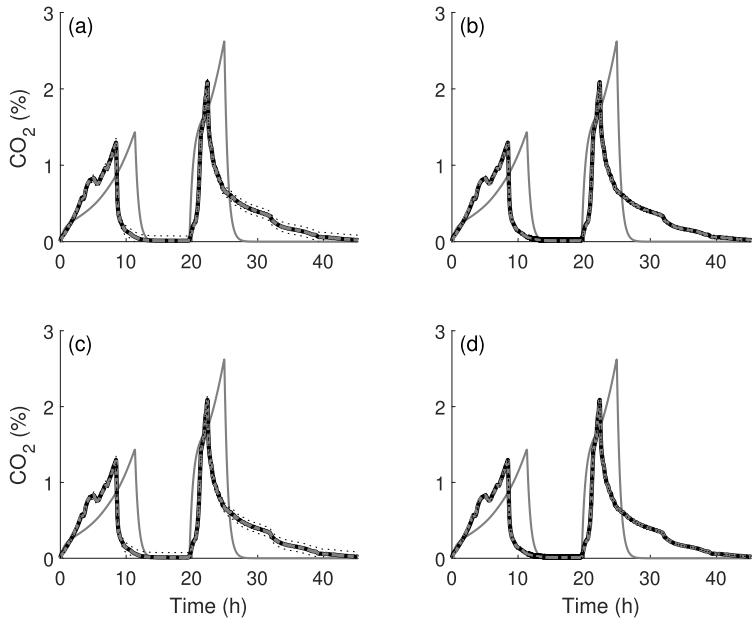


Fig. 7. CO_2 formation-validation experiment (Exp. II). The figure reports the percentage of CO_2 in the off-gas for the two filters with same (a and b) and with different parameter uncertainty between the first and the second (c and d). The model (grey line) shows a more similar dynamic behaviour with the measurements (dashed–dotted line) in the first batch compared to the tuning experiment (Exp. I) with a less accentuate noise between 4 and 6 h but with a slower CO_2 production. However, the model still shows higher discrepancy during and after the feeding phase. The $2-\sigma$ intervals (dotted lines) show that the EKF (a and d) presents a reduced uncertainty compared to the UKF (b and d).

to include intracellular processes in the cells as byproduct formation/consumption and oxygen limitation. Although modelling these processes have the potential to increase the accuracy of the filters, they are increasing the model complexity, subjected to an increase in model uncertainty and parameter over-fitting. (3) Different model dynamics can be adopted for the batch phase and consecutive feeding phases, for instance by separate sets of equation for the sugar consumption, or a separate model parameter set for the batch and the feeding phases, particularly the yield coefficient Y_{XS}. (4) Implementation of adaptive estimation (i.e. adaptive UKF) as presented in [39]. As the authors presented in [39], the measurement noise covariance (R) can be adaptively tuned utilizing the available outputs. However, this would be mostly effective with an high confidence in the process model [40], while, in the presented case a possibility is to tune the process noise covariance (Q) when anomalies are detected in the dynamics of the bacterial cultivation process, allowing an adaptive model uncertainty.

Implementation of a fully *on-line*, low cost state estimator is important for feedback control of microbial cultivations. Although it is possible to estimate sugar concentrations by advanced devices, these are either high-cost (wide range NIR, FTIR) or incorporate major challenges in *on-line* implementation, for instance due to time-demanding measurements or a full integration of the device to the system (HPLC). An estimator that integrates low-cost measurement devices with an accurate model have a great potential to promote a closed feedback control loop in microbial processes and increase their performance.

6. Conclusions

This paper presents an implementation of an UKF and an EKF to estimate biomass and sugar concentrations during a fed-batch process, in the absence of direct sugar measurements. We demonstrated that two simple, low-cost *on-line* measurements could describe the dynamic behaviour of the entire process and lead to an improved accuracy of the estimates, particularly when the process noise variance values were adapted to the phase of the process (MU). Our results indicated that the UKF was more accurate than the EKF. The UKF- and EKF-based soft sensors can therefore provide a solution for sensor fusion that can be used to achieve an holistic feedback control in a real fermentation process.

CRediT authorship contribution statement

Andrea Tuveri: Conceptualization, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Fernando Pérez-García:** Validation, Investigation, Writing – review & editing, Supervision. **Pedro A. Lira-Parada:** Data Curation, Writing – review & editing, Visualization. **Lars Imsland:** Conceptualization, Software, Resources, Writing – review & editing, Supervision. **Nadav Bar:** Conceptualization, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to Prof. Edmund Brekke and the colleagues Ana CSR Dias, José Matias, Lucas Bernardino for the

discussions. The project was funded by the Bio Based Industries Joint Undertaking (JU) under the European Union's Horizon 2020 research and innovation programme under grant agreement N^o 790507.

Appendix. Observability analysis

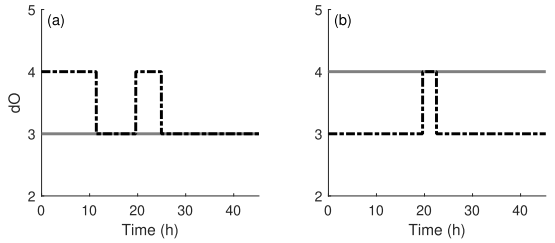


Fig. A.8. The simulation results report the dimension of the observability space *dO* for the different measurement combinations (I–IV). In the left are reported the results for Case I (continuous grey) and II (dashed–dotted black). While in the right for Case III (dashed–dotted black) and IV (continuous grey).

Observability test was conducted considering different measurement functions as follows:

Case I — Volume and biomass

$$h(x) = \begin{bmatrix} V \\ X \end{bmatrix} \tag{56}$$

Case II — Volume and CO_2

$$h(x) = \begin{bmatrix} V \\ CO_2 \end{bmatrix} \tag{57}$$

Case III — Biomass and CO_2

$$h(x) = \begin{bmatrix} X \\ CO_2 \end{bmatrix} \tag{58}$$

Case IV — Volume, biomass and CO_2

$$h(x) = \begin{bmatrix} V \\ X \\ CO_2 \end{bmatrix} \tag{59}$$

The results of the observability analysis are reported in Fig. A.8. The simulation test showed that the system is observable when information about volume, biomass and CO_2 is available (Case IV), with $dO = 4$. On the contrary, it showed that in the other considered cases the system into consideration was not locally observable (Case I, II and III). Indeed, in case of availability of only volume and biomass (Case I) the dimension of the observability space *dO* was equal to 3. Measuring instead volume and CO_2 (Case II) its dimension was 4 except when glucose is depleted, while when measuring biomass and CO_2 (Case III) it showed that the dimension of the observability space was 3 except during feeding.

This results show that to ensure local observability of the system, measurements of volume, biomass and CO_2 are necessary and that only measurements of volume and biomass are not sufficient. However, they also show that, under specific conditions, the sole measurements of volume and CO_2 (Case II) or biomass and CO_2 (Case III) ensure local observability. Indeed, in absence of glucose depletion volume and CO_2 ensure local observability, while, under continuous feeding, also biomass and CO_2 alone are sufficient.

References

[1] V.F. Wendisch, J.M. Jorge, F. Pérez-García, E. Sgobba, Updates on industrial production of amino acids using *Corynebacterium glutamicum*, 2016, http://dx.doi.org/10.1007/s11274-016-2060-1.

[2] J. Becker, C.M. Rohles, C. Wittmann, Metabolically engineered corynebacterium glutamicum for bio-based production of chemicals, fuels, materials, and healthcare products, Metab. Eng. (ISSN: 10967176) 50 (2018) 122–141, http://dx.doi.org/10.1016/j.ymben.2018.07.008.

[3] D. Krämer, R. King, On-line monitoring of substrates and biomass using near-infrared spectroscopy and model-based state estimation for enzyme production by *S. cerevisiae*, IFAC-PapersOnLine (ISSN: 24058963) 49 (7) (2016) 609–614, http://dx.doi.org/10.1016/j.ifacol.2016.07.235.

[4] M. Scarff, S.A. Arnold, L.M. Harvey, B. McNeil, Near infrared spectroscopy for bioprocess monitoring and control: Current status and future trends, Crit. Rev. Biotechnol. (ISSN: 0738-8551) 26 (1) (2006) 17–39, http://dx.doi.org/10.1080/07388550500513677.

[5] D. Krämer, R. King, A hybrid approach for bioprocess state estimation using NIR spectroscopy and a sigma-point Kalman filter, J. Process Control (ISSN: 09591524) 82 (2019) 91–104, http://dx.doi.org/10.1016/j.jprocont.2017.11.008.

[6] R.D. Gudi, S.L. Shah, M.R. Gray, Multirate state and parameter estimation in an antibiotic fermentation with delayed measurements, Biotechnol. Bioeng. (ISSN: 10970290) 44 (11) (1994) 1271–1278, http://dx.doi.org/10.1002/bit.260441102.

[7] G. Stephanopoulos, K.-Y.-Y. San, Studies on on-line bioreactor identification. i. theory, Biotechnol. Bioeng. (ISSN: 10970290) 26 (10) (1984) 1176–1188, http://dx.doi.org/10.1002/bit.260261006.

[8] L. Dewasme, M. Sbarciog, E. Rocha-Cózatl, F. Haugen, A. Vande Wouwer, State and unknown input estimation of an anaerobic digestion reactor with experimental validation, Control Eng. Pract. (ISSN: 09670661) 85 (2019) 280–289, http://dx.doi.org/10.1016/j.conengprac.2019.02.003.

[9] G. Goffaux, A. Vande Wouwer, O. Bernard, Improving continuous–discrete interval observers with application to microalgae-based bioprocesses, J. Process Control (ISSN: 09591524) 19 (7) (2009) 1182–1190, http://dx.doi.org/10.1016/j.jprocont.2009.03.009.

[10] P.C. Lopez, I.A. Udugama, S.T. Thomsen, C. Roslander, H. Junicke, M. Mauricio-Iglesias, K.V. Gernaey, Towards a digital twin: a hybrid data-driven and mechanistic digital shadow to forecast the evolution of lignocellulosic fermentation, Biofuels, Bioprod. Biorefin. (ISSN: 1932-104X) 14 (5) (2020) 1046–1060, http://dx.doi.org/10.1002/bbb.2108.

[11] A. Klimkiewicz, P.P. Mortensen, C.B. Zachariassen, F.W. van den Berg, Monitoring an enzyme purification process using on-line and in-line NIR measurements, Chemometr. Intell. Lab. Syst. (ISSN: 01697439) 132 (2014) 30–38, http://dx.doi.org/10.1016/j.chemolab.2014.01.002.

[12] R. Kandepu, B. Foss, L. Imsland, Applying the unscented Kalman filter for nonlinear state estimation, J. Process Control (ISSN: 09591524) 18 (7–8) (2008) 753–768, http://dx.doi.org/10.1016/j.jprocont.2007.11.004.

[13] S.J. Julier, J.K. Uhlmann, H.F. Durrant-Whyte, A new approach for filtering nonlinear systems, in: Proceedings of American Control Conference - ACC, 1995, pp. 1628–1632, http://dx.doi.org/10.1109/ACC.1995.529783.

[14] D. Simon, Optimal State Estimation, John Wiley & Sons, Inc., Hoboken, NJ, USA, ISBN: 9780470045343, 2006, http://dx.doi.org/10.1002/0470045345.

[15] J. Kager, C. Herwig, I.V. Stelzer, State estimation for a penicillin fed-batch process combining particle filtering methods with online and time delayed offline measurements, Chem. Eng. Sci. (ISSN: 00092509) 177 (2018) 234–244, http://dx.doi.org/10.1016/j.ces.2017.11.049.

[16] A. Golabgir, C. Herwig, Combining mechanistic modeling and Raman spectroscopy for real-time monitoring of fed-batch penicillin production, Chem. Ing. Tech. (ISSN: 0009286X) 88 (6) (2016) 764–776, http://dx.doi.org/10.1002/cite.201500101.

[17] P. Sinner, M. Stiegler, C. Herwig, J. Kager, Noninvasive online monitoring of *Corynebacterium glutamicum* fed-batch bioprocesses subject to spent sulfite liquor raw material uncertainty, Bioresour. Technol. (ISSN: 09608524) (2020) 124395, http://dx.doi.org/10.1016/j.biortech.2020.124395.

[18] S. Kolås, B. Foss, T. Schei, Constrained nonlinear state estimation based on the UKf approach, Comput. Chem. Eng. (ISSN: 00981354) 33 (8) (2009a) 1386–1401, http://dx.doi.org/10.1016/j.compchemeng.2009.01.012.

[19] L. Dewasme, S. Fernandes, Z. Amribt, L.O. Santos, P. Bogaerts, A. Vande Wouwer, State estimation and predictive control of fed-batch cultures of hybridoma cells, J. Process Control (ISSN: 09591524) 30 (2015) 50–57, http://dx.doi.org/10.1016/j.jprocont.2014.12.006.

[20] S. Abe, K.-I. Takayama, S. Kinoshita, Taxonomical studies on glutamic acid-producing bacteria, J. Gen. Appl. Microbiol. (ISSN: 1349-8037) 13 (3) (1967) 279–301, http://dx.doi.org/10.2323/jgam.13.279.

[21] L. Eggeling, M. Bott, Handbook of *Corynebacterium Glutamicum*, CRC Press, ISBN: 1420039695, 2005.

[22] P. Sinner, J. Kager, S. Daume, C. Herwig, Model-based analysis and optimisation of a continuous *Corynebacterium glutamicum* bioprocess utilizing lignocellulosic waste, in: IFAC-PapersOnLine, vol. 52, Elsevier B.V., (ISSN: 24058963) 2019, pp. 181–186, http://dx.doi.org/10.1016/j.ifacol.2019.12.255.

[23] van Riel Natal, Parameter estimation in non-equidistantly sampled non-linear state space models;, Eindh. Univ. Technol. 1 (1) (2006) 16, http://dx.doi.org/10.13140/2.1.3335.5846.

[24] P.A. Lira-Parada, E. Pettersen, L.T. Biegler, N. Bar, Implications of dimensional analysis in bioreactor models: Parameter estimation and identifiability, Chem. Eng. J. 417 (2021) 129220, http://dx.doi.org/10.1016/j.cej.2021.129220.
[25] T.S. Ligon, F. Fröhlich, O.T. Chiş, J.R. Banga, E. Balsa-Canto, J. Hasenauer, S.I. GenS, 2.0: multi-experiment structural identifiability analysis of SBML models, Bioinformatics (ISSN: 1367-4803) 34 (8) (2018) 1421–1423, http://dx.doi.org/10.1093/bioinformatics/btx735.
[26] E. Balsa-Canto, A.A. Alonso, J.R. Banga, An iterative identification procedure for dynamic modeling of biochemical networks, BMC Syst. Biol. (ISSN: 1752-0509) 4 (1) (2010) 11, http://dx.doi.org/10.1186/1752-0509-4-11.
[27] H. Nijmeijer, A. van der Schaft, Nonlinear Dynamical Control Systems, vol. 175, Springer New York, New York, NY, ISBN: 978-1-4757-2102-7, 1990, http://dx.doi.org/10.1007/978-1-4757-2101-0.
[28] D. Simon, Kalman filtering with state constraints: a survey of linear and nonlinear algorithms, IET Control Theory Appl. (ISSN: 1751-8644) 4 (8) (2010) 1303–1318, http://dx.doi.org/10.1049/iet-cta.2009.0032.
[29] K. Röbenack, Computation of multiple Lie derivatives by algorithmic differentiation, J. Comput. Appl. Math. (ISSN: 03770427) 213 (2) (2008) 454–464, http://dx.doi.org/10.1016/j.cam.2007.01.036.
[30] S. Kolås, B.A. Foss, T.S. Schei, Noise modeling concepts in nonlinear state estimation, J. Process Control (ISSN: 09591524) 19 (7) (2009b) 1111–1125, http://dx.doi.org/10.1016/j.jprocont.2009.03.002.
[31] M. Inui, S. Murakami, S. Okino, H. Kawaguchi, A.A. Vertès, H. Yukawa, Metabolic analysis of *Corynebacterium glutamicum* during lactate and succinate productions under oxygen deprivation conditions., J. Mol. Microbiol. Biotechnol. 7 (4) (2004) 182–196, http://dx.doi.org/10.1159/000079827.
[32] M. Inui, M. Suda, S. Okino, H. Nonaka, L.G. Puskás, A.A. Vertès, H. Yukawa, Transcriptional profiling of *Corynebacterium glutamicum* metabolism during organic acid production under oxygen deprivation conditions, Microbiology (Reading, England) (2007) 2491–2504, http://dx.doi.org/10.1099/mic.0.2006/005587-0.
[33] V.F. Wendisch, M. Spies, D.J. Reinscheid, S. Schnicke, H. Sahm, B.J. Eikmanns, Regulation of acetate metabolism in *Corynebacterium glutamicum*: Transcriptional control of the isocitrate lyase and malate synthase genes, Arch. Microbiol. (ISSN: 03028933) 168 (4) (1997) 262–269, http://dx.doi.org/10.1007/s002030050497.
[34] F. Pérez-Garcí, P. Peters-Wendisch, V.F. Wendisch, Engineering *Corynebacterium glutamicum* for fast production of l-lysine and l-pipecolic acid, Appl. Microbiol. Biotechnol. (ISSN: 14320614) 100 (18) (2016) http://dx.doi.org/10.1007/s00253-016-7682-6.
[35] J. Becker, C. Wittmann, Bio-based production of chemicals, materials and fuels - *Corynebacterium glutamicum* as versatile cell factory, Curr. Opin. Biotechnol. (ISSN: 09581669) 23 (4) (2012) 631–640, http://dx.doi.org/10.1016/j.copbio.2011.11.012.
[36] N. Paczia, A. Nilgen, T. Lehmann, J. Gätgens, W. Wiechert, S. Noack, Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms, Microb. Cell Factories (ISSN: 14752859) 11 (2012) 1–14, http://dx.doi.org/10.1186/1475-2859-11-122.
[37] K. Xiong, H.Y. Zhang, C.W. Chan, Performance evaluation of UKf-based nonlinear filtering, Automatica (ISSN: 00051098) 42 (2) (2006) 261–270, http://dx.doi.org/10.1016/j.automatica.2005.10.004.
[38] L. Perea, J. How, L. Breger, P. Elosegui, Nonlinearity in sensor fusion: Divergence issues in EKF, modified truncated SOF, and UKf, in: Collection of Technical Papers - AIAA Guidance, Navigation, and Control Conference 2007, vol. 2, ISBN: 1563479044, 2007, pp. 1863–1878, http://dx.doi.org/10.2514/6.2007-6514.
[39] Z. Mahmoudi, K. Nørgaard, N.K. Poulsen, H. Madsen, J.B. Jørgensen, Fault and meal detection by redundant continuous glucose monitors and the unscented Kalman filter, Biomed. Signal Process. Control 38 (2017a) 86–99, http://dx.doi.org/10.1016/j.bspc.2017.05.004.
[40] Z. Mahmoudi, N.K. Poulsen, H. Madsen, J.B. Jørgensen, Adaptive unscented Kalman filter using maximum likelihood estimation, IFAC-PapersOnLine (ISSN: 24058963) 50 (1) (2017b) 3859–3864, http://dx.doi.org/10.1016/j.ifacol.2017.08.356.

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 55-7 (2022) 222–227

IFAC Papers Online CONFERENCE PAPER ARCHIVE

Bioprocess Monitoring: A Moving Horizon Estimation Experimental Application ⋆

Andrea Tuveri * **Haakon Eng Holck** * **Caroline S.M. Nakama** * **José Matias** * **Johannes Jäschke** * **Lars Imsland** ** **Nadav Bar** *

* *Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway (e-mail: nadi.bar@ntnu.no)*

** *Department of Engineering Cybernetics, Norwegian University of Science and Technology (NTNU), Trondheim, Norway*

Abstract: Microbial fermentation processes are most often described by nonlinear time-varying dynamics, which require the implementation of nonlinear state estimators to infer unmeasured metabolites in the cultivation broth. Among the various nonlinear available estimator strategies, the Moving Horizon Estimator (MHE) is an *on-line* optimization approach that easily allows to enforce hard constraints, an important feature that helps to avoid unfeasible concentrations. In this work we implemented an MHE by using experimental data from a fed-batch cultivation process of *Corynebacterium glutamicum*. Available real-time measurements of biomass and CO_2 formation were used to infer sugar concentrations by combining the available measurements with a simple Monod model. We found that the MHE was able to estimate all the three variables of interest, including the unmeasured sugar concentrations, during the entire fed-batch cultivation process. Moreover, we show that the estimates are accurate in comparison to the reference *off-line* samples. This work demonstrates the benefits of MHE as a soft sensor that can monitor bioprocesses in real-time.

Keywords: Estimation and control in biological systems, Parameter and state estimation, Monitoring, Optimization

1. INTRODUCTION

Automation of biological processes is limited due to the unavailability of *on-line* measurement devices that can quantify variables of interest. The lack of sensors can be circumvented by implementing state estimators that enable to monitor the process in real-time. These state estimators are dependent on mathematical models (Rao, 2000) that represent the system in a simplified manner (i.e. Monod growth model). Different applications of nonlinear estimation techniques, including Extended Kalman Filters (EKFs), Particle Filters (PFs) and Unscented Kalman Filters (UKFs), are available, as reported in Tuveri et al. (2021). However, those applications are all based on recursive Bayesian estimators, which approximate the posterior conditional probability density function (*pdf*) using measurements available at the current sampling instant.

Differently from them, optimization-based methods such as the Moving Horizon Estimator (MHE) use a moving window of past data (Robertson et al., 1996; Rawlings and Bakshi, 2006; Bavdekar et al., 2013; Ali et al., 2015). Most importantly, from a practical point of view and in contrast to the recursive Bayesian estimators, the MHE has:

- the ability to explicitly incorporate bound constraints in states and parameters;
- the possibility to easily handle multi-rate measurements;
- higher computational times (CPU times).

Indeed, MHE has the ability to explicitly incorporate physical constraints on states and parameters, and disturbances in the form of time-varying parameters can be added as extra degrees of freedom in the optimization (Robertson et al., 1996; Rao et al., 2003; Kühl et al., 2011). On the other hand, the EKFs or UKFs methods require strategies like clipping (Haseltine and Rawlings, 2005) or other optimization-based methods (Kolås et al., 2009; Tuveri et al., 2021) to avoid the estimation of negative concentrations. Additionally, time-varying parameters and disturbances have to be included as additional states, without the possibility to constrain them within predefined boundaries. Moreover, many real-time systems incorporate measurement devices with various sampling rates and times. For instance, whereas some absorbance probes sample every 10 seconds, sugar measurements by high performance liquid chromatography (HPLC) can be acquired every 30-60 minutes. The MHE, by considering a window of past measurements, is of interest in the case of such multi-rate measurements (Elsheikh et al., 2021), since it easily allows to place them in an adequate position within the time horizon. For those reasons, the MHE is a promising approach within bioprocesses.

⋆ The project was funded by the Bio Based Industries Joint Undertaking (JU) under the EU's Horizon 2020 research and innovation programme - N^o 790507. J. Matias acknowledges financial support from the Norwegian Research Council, SUBPRO, grant number: 237893. C. S. M. Nakama acknowledges financial support from the Norwegian Research Council through FRIPRO Project SensPATH.

10.1016/j.ifacol.2022.07.448

The continuously increasing interest in real-time optimal control has brought advances in *nonlinear* Model Predictive Control (NMPC) algorithms and consequently in MHE methods (Findeisen et al., 2007). Indeed, even though fewer results are available in the literature about MHE, it can be seen as the dual of Model Predictive Control (MPC), since they both share moving horizon approach and dynamic optimization. The pioneering work of Kühl et al. (2011) in fact, transferred a fast real-time iteration approach developed for NMPC to MHE, contributing to its feasibility in real-time.

Although some *in-silico* applications within bioprocess monitoring are available (Raïssi et al., 2005; Valipour and Ricardez-Sandoval, 2021; Bae et al., 2021; Elsheikh et al., 2021), the implementation of MHE combined with experimental data is almost non-existent and, to the best of our knowledge, it was only presented by Goffaux and Wouwer (2008). In Goffaux and Wouwer (2008) the authors present a robust receding horizon approach in the case of uncertain parameters, by selecting the worst parameter realization in a min-max optimization approach. To reduce the high computational demand, model linearization and monotonicity assumptions are required. Differently from the implementation in Goffaux and Wouwer (2008), our approach does not require linearisation or monotonicity of the model, parameter realisation is considered nominal and model uncertainty is a free variable minimized in the cost function. In order to demonstrate the benefits of MHE in bioprocess monitoring, we present an implementation of MHE using real experimental data of a fed-batch bacterial cultivation of *Corynebacterium glutamicum*, previously presented in Tuveri et al. (2021). The estimation performance obtained here by the MHE is accurate with respect to the *off-line* samples. Moreover, we demonstrate the incorporation of hard state constraints directly in the optimization formulation. This is an important advantage when the practitioner needs to avoid nonphysical estimates (i.e. negative concentrations).

2. INPUT AND OUTPUTS

The *on-line* output measurements were collected every 60 seconds and used by the estimator to measure biomass, volume and CO_2 respectively and to infer the unmeasured glucose composition. Signals from the absorbance probe were obtained in concentration units (0.05 - 4 CU) and than converted to g/L (cell dry weight, CDW) using a calibration curve as follows.

CU $\geq$ 0.9:

$$CDW_{CU} = 22.187 \cdot CU - 5.0991$$

CU $<$ 0.9:

$$CDW_{CU} = 11.124 \cdot CU + 0.66116$$

On-line CO_2 signals were obtained as measure of the composition in the outflow (0%–25%). The volume was calculated by integrating *on-line* the signals from the pumps, taking into account also the amount of volume taken for the *off-line* samples (8 mL/sample). The feeding profile is reported in Fig. 1. From here on we will define the two different phases as batch (from zero to the start of the feeding) and fed-batch or second batch (from the feeding on). More information about measurements and experimental setup are reported in Tuveri et al. (2021).

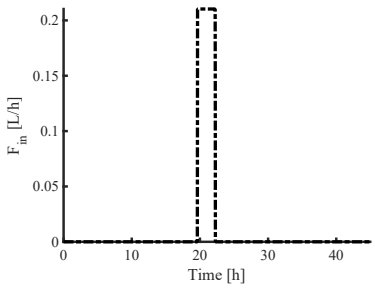


Fig. 1. Feeding profile.

3. SYSTEM MODEL

The dynamics are modelled by using simple Monod kinetics for growth on glucose as fed-batch process (Eq. 1), with the addition of linear cell death (Tuveri et al., 2021), where the state vector is defined by $x(t) = [V(t),\ X(t),\ S(t),\ CO_2(t)]^T$ and the input vector by $u = [F_{in}(t)]$. From here on we will drop the time-dependent notation for simplicity. The states V, X, S and CO_2 are respectively volume, biomass, substrate and carbon dioxide. F_{in} is the inflow of glucose with concentration S_{in} $(100g/L)$, while q_{air} is the inflow of air (2 NL/min).

$$\begin{cases} \frac{dV}{dt} = F_{in} \\ \frac{dX}{dt} = -\frac{F_{in}}{V}X + \mu_{max}\frac{S}{K_s+S}X - k_dX \\ \frac{dS}{dt} = \frac{F_{in}}{V}(S_{in} - S) - \mu_{max}\frac{S}{K_s+S}\frac{X}{Y_{XS}} \\ \frac{dCO_2}{dt} = \mu_{max}\frac{S}{K_s+S}\frac{X}{Y_{XCO_2}} - q_{air}CO_2 \end{cases} \tag{1}$$

The parameters are μ_{max}, K_S, k_d, Y_{XS} and Y_{XCO_2} (Table 1). Those parameters were obtained using a nonlinear least-squares data fitting algorithm (lsqnonlin, Matlab) by a dedicated experiment (Tuveri et al., 2021).

Table 1. Values of model parameters (Eq. 1) with unit and standard deviations.

Parameter	Description	Value	Unit	Std. Dev.
μ_{max}	Maximum growth rate	0.19445	$[h^{-1}]$	$3.25 \cdot 10^{-6}$
K_S	Monod growth constant	0.007	$[g \cdot L^{-1}]$	$3.92 \cdot 10^{-6}$
k_d	Death rate constant	0.006	$[h^{-1}]$	$4.49 \cdot 10^{-6}$
Y_{XS}	S from X yield	0.42042	$[g \cdot g^{-1}]$	$3.58 \cdot 10^{-6}$
Y_{XCO_2}	CO_2 from X yield	0.54308	$[g \cdot g^{-1}]$	$2.22 \cdot 10^{-6}$

The covariance matrix for the parameters is calculated through the Fisher Information Matrix (FIM) as in Tuveri et al. (2021), where also structural identifiability and local observability for the system were positively assessed.

4. MOVING HORIZON ESTIMATION

The MHE estimates the states using past measurements at specific time points in the horizon $T = t_N - t_L$, where t_N represents the current time and t_L the starting point of the horizon. The time horizon is then discretized according to the sampling rate.

The dynamics of the process (Eq. 1) are described by a set of ordinary differential equations (ODEs):

$$\dot{x} = f(x, u, w) \tag{2}$$

$$y = h(x) \tag{3}$$

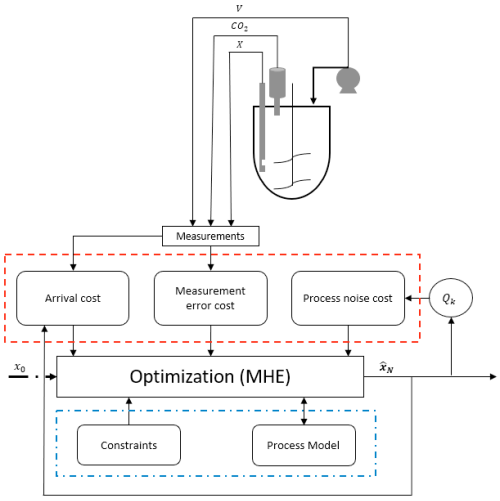


Fig. 2. The figure reports the MHE implementation, where the red dashed line represents the cost function (Eq. 6) and the blue dashed-dotted line its constraints. Measurements are given to both arrival and measurement error cost, while the measurement error covariance matrix is calculated at every iteration to weight for the model noise.

By discretizing it, we obtain:

$$x_{k+1} = F(x_k, u_k, w_k) \tag{4}$$
$$y_k = h(x_k) \tag{5}$$

where k denotes the sampling time t_k and w_k is a random variable.

The MHE problem (Kühl et al., 2011; Andersson et al., 2016) consists in finding the states and their noise obtained by solving the following constrained least-squares optimization problem:

$$\begin{aligned} \min_{x_i, w_i} \quad & \left(\|\hat{x}_L - x_L\|^2_{P_L} + \sum_{i=L}^{N} \|y_i - h(x_i)\|^2_V + \sum_{i=L}^{N-1} \|w_i\|^2_{W_k} \right) \\ \text{s.t.} \quad & x_{i+1} = F(x_i, u_i, w_i) \qquad i = L, \cdots, N-1 \\ & x_i \geq x_{min} \qquad i = L, \cdots, N \end{aligned} \tag{6}$$

The states x_i are constrained with $x_{min} = [0, 0, 0, 0]$ as a lower bound, to avoid negative (unfeasible) concentrations. Following the work of Kühl et al. (2011), we define:

$$P_L = P^{-1/2}, \qquad V = R^{-1/2}, \qquad W_k = {Q_k}^{-1/2} \tag{7}$$

and consider $\|b\|^2_B = b^T B^T B b$. The first term of the optimization (Eq. 6) is the arrival cost (Γ), which summarises the effect of measurements previous to the estimation horizon (up to t_L) and it is updated by single QR-factorization as in Kühl et al. (2011). The term $\hat{x}_L$ (Eq. 6) represents the optimal estimate of x_L.

The matrices P, R and Q_k (Eq. 7) are defined respectively as error, measurement noise and process noise covariances (Tuveri et al., 2021). Moreover, to take into account different process dynamics (Elsheikh et al., 2021), the process noise covariance Q_k is updated as in Tuveri et al. (2021). However since MHE, differently from EKF and UKF, takes into account several past measurements within the horizon T, the corresponding process noise covariance matrix Q_k is used for each sampling time (k) within the horizon. Along the optimization horizon (30 min), states (x_i) and process noise (w_i) are optimized using the information from the mechanistic model and the outputs (1 min sampling rate). The implementation of the MHE is illustrated in Fig. 2. To transform the continuous time model, we apply three point Legendre collocation on finite elements. The Nonlinear Programming (NLP) problem was solved using IPOPT (Wächter and Biegler, 2006) embedded in CasADi (Andersson et al., 2019).

4.1 Arrival Cost Update

Several approaches can be used to calculate the arrival cost. An interesting review on different arrival cost schemes and how they can effect the stability of the MHE is presented in Elsheikh et al. (2021). Here we describe the QR-factorization approach employed in this work, which approximates the arrival cost to a quadratic term that is updated before each new horizon (Kühl et al., 2011). Altough the QR-factorisation is a linearised technique as the EKF, it also holds all the numerical properties of the square-root Kalman Filter and the influence of past information can only grow within the limits of the process noise covariance Q_k (Kühl et al., 2011).

When we shift the horizon to a new start point at t_{L+1}, the arrival cost would ideally be defined as:

$$\begin{aligned} \Gamma(x_{L+1}) = \min_{x_L} \quad & \left(\|\hat{x}_L - x_L\|^2_{P_L} + \|y_L - h(x_L)\|^2_V + \|w_L\|^2_{W_L} \right) \\ \text{s.t.} \quad & x_{L+1} = F(x_L, u_L, w_L) \end{aligned} \tag{8}$$

However, since x_{L+1} is described by a nonlinear function, we do not have an analytical expression for the ideal arrival cost Γ. For obtaining an explicit solution of Eq. 8, some approximations are carried out. First, we define the term $x(t_{L+1}|x_L)$ as the solution of the ODEs (Eq. 4) in the interval from $t \in [t_L, t_{L+1}]$ with x_L as initial value. By linearizing $x(t_{L+1}|x_L)$ around the best available estimate x^*, we obtain:

$$\begin{aligned} x(t_{L+1}|x_L) &\approx x(t_{L+1}|x^*) + A \cdot (x_L - x^*) \\ &\approx \tilde{x} + A x_L \end{aligned} \tag{9}$$

where $\tilde{x} := x(t_{L+1}|x^*) - Ax^*$ and matrix A is the derivative of $x(t_{L+1}|x_L)$ with respect to x_L:

$$A = \left. \frac{\partial F(x_L, u_L)}{\partial x_L} \right|_{x^*}$$

Since in this case $h(x_L)$ is linear, we can represent it as $h(x_L) = Hx_L$, where H is a selector matrix. This way it becomes possible to solve, analytically, Eq. 8 by rewriting it as:

$$\min_{x_L} \left\| \begin{matrix} P_L(\hat{x}_L - x_L) \\ V(y_L - Hx_L) \\ W_L(x_{L+1} - \tilde{x} - Ax_L) \end{matrix} \right\|^2_2 \tag{10}$$

and transforming it using QR-factorization:

$$\begin{pmatrix} P_L & 0 \\ -VH & 0 \\ -W_L A & W_L \end{pmatrix} = \mathcal{Q} \begin{pmatrix} \mathcal{R}_1 & \mathcal{R}_{12} \\ 0 & \mathcal{R}_2 \\ 0 & 0 \end{pmatrix} \tag{11}$$

The QR-factorization decomposes the matrix in the objective function (Eq. 10) into the product of an orthogonal matrix $\mathcal{Q}$ and an upper triangular matrix $\mathcal{R}$ (Elsheikh et al., 2021). From Eq. 11, we then obtain an equivalent problem of the form:

$$\min_{x_L} \left\| \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} + \begin{pmatrix} \mathcal{R}_1 & \mathcal{R}_{12} \\ 0 & \mathcal{R}_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_L \\ x_{L+1} \end{pmatrix} \right\|_2^2 \tag{12}$$

The analytical solution of this problem (Eq. 12) results in the approximated quadratic expression for the arrival cost:

$$\Gamma^{'}(x_{L+1}) = \|\gamma_3\|_2^2 + \|\gamma_2 + \mathcal{R}_2 x_{L+1}\|_2^2 \tag{13}$$

Since the first term of Eq. 13 is given, the arrival cost updates are given by:

$$\hat{x}_{L+1} = -\mathcal{R}_2^{-1}\gamma_2, \qquad P_{L+1} = \mathcal{R}_2 \tag{14}$$

4.2 Moving Horizon Estimation Setup

The filter receives the signals from the sensors every 60 seconds. The states and the input of the system are defined in Section 3, while the measured outputs are:

$$y = [V \ CDW \ \mathrm{CO_2}]^T \tag{15}$$

The initial states are given by:

$$x_0 = [1.5 \ 1.2 \ 20 \ 0]^T$$

with initial covariance matrix:

$$P_0{}^+ = \begin{bmatrix} 2.09{\cdot}10^{-8} & 0 & 0 & 0 \\ 0 & 1.10{\cdot}10^{-5} & 0 & 0 \\ 0 & 0 & 1.09{\cdot}10^{-4} & 0 \\ 0 & 0 & 0 & 2.17{\cdot}10^{-5} \end{bmatrix}$$

and measurement noise covariance matrix R:

$$R = \begin{bmatrix} 10^{-2} & 0 & 0 \\ 0 & 10^{-1} & 0 \\ 0 & 0 & 10^{-3} \end{bmatrix}$$

The process noise covariance matrix Q_k was tuned as in Tuveri et al. (2021), where G_k is the Jacobian with respect to the model noise vector w:

$$G_k = \frac{\partial f(x,u,w)}{\partial w}$$

and Q_k is obtained as:

$$Q_k = G_k \cdot Q_w \cdot {G_k}^T \tag{16}$$

With Q_w defined as the covariance matrix of the noise w:

$$Q_w = diag\left[\sigma^2_{\mu_{max}} \ \sigma^2_{k_s} \ \sigma^2_{k_d} \ \sigma^2_{Y_{XS}} \ \sigma^2_{Y_{XCO_2}} \ \sigma^2_V \ \sigma^2_X \ \sigma^2_S \ \sigma^2_{CO_2}\right]$$

The values of Q_w are reported in Table 2. The values are kept equal to those reported in Tuveri et al. (2021), except the values for σ^2_V and $\sigma^2_{\mathrm{CO_2}}$, which have been modified as they are considered tuning parameters. The additive noise terms on the states are added to prevent the process noise covariance Q_k from being zero or indefinite whenever the substrate is depleted. To compensate for unmodelled dynamics, the values of $\sigma^2_{K_S}$ and $\sigma^2_{Y_{\mathrm{XCO_2}}}$ (Table 2) are increased, once the feeding phase is started and then kept constant thereafter.

Table 2. Variances (σ_i^2) of additive noise (w_i) in parameters and states. The parameters variance is obtained from Table 1. These values are kept constant until the second batch phase (Fed-batch values), when the values of $\sigma^2_{K_S}$ and $\sigma^2_{Y_{\mathrm{XCO_2}}}$ are increased to compensate for the unmodelled dynamics.

Variance	Additive noise	Batch	Fed-batch
$\sigma^2_{\mu_{max}}$	in μ_{max}	$1.05 \cdot 10^{-11}$	-
$\sigma^2_{K_S}$	in K_S	$1.54 \cdot 10^{-11}$	$3.38 \cdot 10^{-2}$
$\sigma^2_{k_d}$	in k_d	$2.02 \cdot 10^{-11}$	-
$\sigma^2_{Y_{XS}}$	in Y_{XS}	$1.28 \cdot 10^{-11}$	-
$\sigma^2_{Y_{\mathrm{XCO_2}}}$	in $Y_{\mathrm{XCO_2}}$	$4.91 \cdot 10^{-12}$	$4.91 \cdot 10^{-2}$
σ^2_V	in V	$1 \cdot 10^{-1}$	-
σ^2_X	in X	$1 \cdot 10^{-2}$	-
σ^2_S	in S	$1 \cdot 10^{-2}$	-
$\sigma^2_{\mathrm{CO_2}}$	in $\mathrm{CO_2}$	$1 \cdot 10^{-1}$	-

5. RESULTS

In this section, the estimation results (Fig. 3) for the system presented in Sec. 3 obtained by the application of the MHE with incorporation of state constraints (Sec. 4) are presented. The parameters were tuned as was presented in Sec. 4.2, while the time horizon was 30 minutes. As was reported in Tuveri et al. (2021), intracellular metabolic changes (e.g byproduct formation) during the batch and the feeding phases was not accounted for by the simple Monod model (because these changes are not entirely mapped), resulting in discrepancy between the model and the real dynamics. The results we obtained here (Fig. 3) are accurate with respect to the *off-line* measurements. The state estimator accurately follows the sugar consumption after adaptation of the tuning to the changes in metabolism due to high glucose feeding.

5.1 State Estimates

Using the MHE described in Sec. 4, we estimated with good accuracy the unmeasured glucose, following the changes in its concentrations also under high model mismatch (Fig. 3). It can be seen that the model predictions present a discrepancy compared to the *off-line* values. This is visible in both the first and the second (after feeding) batch. Despite these poor model predictions, the MHE improves the estimates by using the available information on the measurements.

Firstly, the *off-line* measurements of biomass are based on cell dry-weight, and are usually inaccurate due to manual sampling that exhibit high variance, e.g. at time $t = 8$ (Fig. 3a). The MHE, relying on the *on-line* OD high frequency measurements, estimated the biomass accurately compared to these *off-line* measurements, and at the same time corrected the biased biomass model prediction. Secondly, the consumption of the unmeasured sugars was well captured by the estimator, compared to the highly accurate sugar measurements (HPLC), even under high model-mismatch. Notably, the estimator could capture the change of sugar consumption rate after the feeding phase when we increased the parameters variance in the process noise covariance matrix Q_k. Thirdly, the model predictions for the CO_2 present the largest error. The CO_2 model predictions present a delay in the first

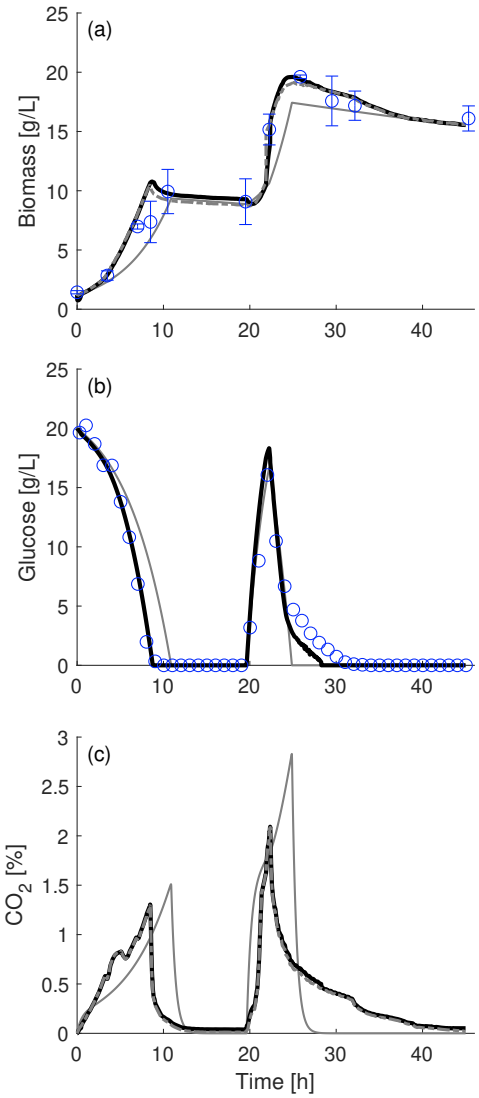


Fig. 3. Estimation results with 1 minute sampling interval. The blue circles (*off-line* measurements) and the dashed-dotted grey lines (sensor measurements) correspond to the experimental results. The solid grey is the open-loop model prediction and the solid black is the estimated value. The figures present the results for biomass (a), glucose concentrations (b) and CO_2 output (c). The MHE accurately estimated the states, and improved the estimate of the glucose compared to the model prediction.

batch and a high model mismatch in the second batch (feeding phase). However, this discrepancy is compensated by the MHE with the information on the measurements. Fourthly, it remains worth mentioning that similar results can be achieved by the application of recursive Bayesian state estimators (i.e. EKF and UKF) as shown in Tuveri et al. (2021). However, here we want to present the MHE

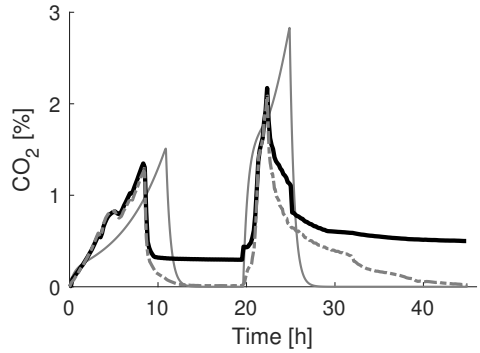


Fig. 4. Estimation results for the CO_2 output with a sampling interval of 10 minutes. The dashed-dotted grey lines (sensor measurements) correspond to the experimental results. The solid grey is the open-loop model prediction and the solid black is the estimated value. The estimated value presents a steady state offset.

as an alternative framework where constraints can be incorporated explicitly. Moreover, the reader must take into account that in this application the initial state vector x_0 was initialized by using the true states of the system at the initial time, given a certain initial error covariance matrix P_0 as in Tuveri et al. (2021). In the case of uncertain initial state vector x_0, the error covariance matrix would be initialized to a different value and the MHE would show slower convergence to correct the states. Our implementation of the MHE was dependent on the sampling frequency. The estimates of biomass X and CO_2 exhibited a steady state offset (Fig. 4) when the sampling rate was lower than 60 seconds. There are several explanations to that, and we will discuss two of them that we think are more relevant:

1) The tuning of the parameters R and Q_k weights differently the contribution of measurements and model information in the optimization problem. Indeed, tuning needs to be done after the frequency of the output measurements is defined. In our case we selected a sampling rate of 60 seconds as in Tuveri et al. (2021) and a relatively short horizon (30 minutes). We found that a short sampling interval combined with a sufficient horizon length improves the performance of the MHE, consistent with what reported in Schei (2008), where the author states that it is desirable to choose short sampling times intervals and a long data window. This is also consistent with what was reported in Haseltine and Rawlings (2005). Here the authors state that for short time horizons there is the possibility that the data within the horizon can not overcome the biasing of the arrival cost approximation. The tuning of the MHE presents therefore a compromise between performance and computational requirement (Schei, 2008), since longer time horizons imply a bigger optimization problem. However, the time it took to solve our optimization problem was in the interval 0.05 to 1 seconds, well within the time update interval (60 seconds).

2) Model mismatch can deteriorate the optimal solution, leading to steady-state offsets (Kühl et al., 2011). As it was presented previously (Elsheikh et al., 2021), due

to the complexity of biological systems, model mismatch is often encountered in bioprocesses and it can not be simply handled by a proper design of Q_k. Moreover, as already presented in Tuveri et al. (2021), the system under consideration presents a high model mismatch due to sudden metabolic changes after an exposure of the bacteria to high glucose concentration following an oxygen deprivation period. There are two possible remedies to this problem. One is to consider the system with unknown non-Gaussian uncertainties (through the use of a Gaussian mixture model) as was recently presented in Valipour and Ricardez-Sandoval (2022). Indeed the authors show how their approach is effective in cases involving either unexpected process or measurement (e.g. sensor drift) noises. The other is to include model parameters as decision variables in the optimization problem (Bae et al., 2021). Interestingly, the optimization problem will adapt the model (by adapting the parameter values) during the metabolic change periods. The inclusion of parameters as decision variables in the optimization problem may not only reduce the time for the practitioner to manually tune the MHE by trial and error, but also reveal possible changes in metabolic behaviors, leading to more robust models. However, if this is not coupled with a proper choice of parameter selection to optimize, it may lead to an ill-conditioned problem with over-fitted parameters, as previously discussed in Bae et al. (2021). This will be however part of further investigations.

6. CONCLUSION

This work presents the implementation of an MHE for the estimation of biomass, glucose concentrations and CO_2 formation in a fed-batch cultivation process. We reported here the efficacy of the MHE as an alternative state estimator in bioprocesses, demonstrating its advantage under necessity of hard state constraints. We showed that although the results were accurate with respect to the *off-line* measurements, simple tuning could not fully compensate for unmodelled dynamics. As a future work, the MHE has the potential to serve as a powerful tool that can both estimate the states in real-time and allow an adaptive parameter estimation. This will enable the detection of changes in metabolic behaviours and, as a consequence, the basis for more robust model predictions.

REFERENCES

Ali, J.M., Hoang, N.H., Hussain, M.A., and Dochain, D. (2015). Review and classification of recent observers applied in chemical process systems. *Computers & Chemical Engineering*, 76, 27–41.

Andersson, J.A.E., Gillis, J., Horn, G., Rawlings, J.B., and Diehl, M. (2019). CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1), 1–36.

Andersson, L.E., Scibilia, F., and Imsland, L. (2016). An estimation-forecast set-up for iceberg drift prediction. *Cold Regions Science and Technology*, 131, 88–107.

Bae, J., Kim, Y., and Lee, J.M. (2021). Multirate moving horizon estimation combined with parameter subset selection. *Computers & Chemical Engineering*, 147, 107253.

Bavdekar, V.A., Gopaluni, R.B., and Shah, S.L. (2013). A comparison of moving horizon and Bayesian state estimators with an application to a pH process. *IFAC Proceedings Volumes*, 46(32), 160–165.

Elsheikh, M., Hille, R., Tatulea-Codrean, A., and Krämer, S. (2021). A Comparative Review of Multi-Rate Moving Horizon Estimation Schemes for Bioprocess Applications. *Computers & Chemical Engineering*, 107219.

Findeisen, R., Allgöwer, F., and Biegler, L.T. (2007). *Assessment and future directions of nonlinear model predictive control*, volume 358. Springer.

Goffaux, G. and Wouwer, A.V. (2008). Design of a robust nonlinear receding-horizon observer-Application to a biological system. *IFAC Proceedings Volumes*, 41(2), 15553–15558.

Haseltine, E.L. and Rawlings, J.B. (2005). Critical evaluation of extended Kalman filtering and moving-horizon estimation. *Industrial & engineering chemistry research*, 44(8), 2451–2460.

Kolås, S., Foss, B., and Schei, T. (2009). Constrained nonlinear state estimation based on the UKF approach. *Computers & Chemical Engineering*, 33(8), 1386–1401.

Kühl, P., Diehl, M., Kraus, T., Schlöder, J.P., and Bock, H.G. (2011). A real-time algorithm for moving horizon state and parameter estimation. *Computers & chemical engineering*, 35(1), 71–83.

Raïssi, T., Ramdani, N., and Candau, Y. (2005). Bounded error moving horizon state estimator for non-linear continuous-time systems: application to a bioprocess system. *Journal of Process control*, 15(5), 537–545.

Rao, C.V. (2000). *Moving horizon strategies for the constrained monitoring and control of nonlinear discrete-time systems*. The University of Wisconsin-Madison.

Rao, C.V., Rawlings, J.B., and Mayne, D.Q. (2003). Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE transactions on automatic control*, 48(2), 246–258.

Rawlings, J.B. and Bakshi, B.R. (2006). Particle filtering and moving horizon estimation. *Computers & chemical engineering*, 30(10-12), 1529–1541.

Robertson, D.G., Lee, J.H., and Rawlings, J.B. (1996). A moving horizon-based approach for least-squares estimation. *AIChE Journal*, 42(8), 2209–2224.

Schei, T.S. (2008). On-line estimation for process control and optimization applications. *Journal of Process Control*, 18(9), 821–828.

Tuveri, A., Pérez-García, F., Lira-Parada, P.A., Imsland, L., and Bar, N. (2021). Sensor fusion based on Extended and Unscented Kalman Filter for bioprocess monitoring. *Journal of Process Control*, 106, 195–207.

Valipour, M. and Ricardez-Sandoval, L.A. (2021). Assessing the Impact of EKF as the Arrival Cost in the Moving Horizon Estimation under Nonlinear Model Predictive Control. *Industrial & Engineering Chemistry Research*, 60(7), 2994–3012.

Valipour, M. and Ricardez-Sandoval, L.A. (2022). A robust moving horizon estimation under unknown distributions of process or measurement noises. *Computers & Chemical Engineering*, 157, 107620.

Wächter, A. and Biegler, L.T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1), 25–57.

4 | A Regularized Moving Horizon Estimator for Combined State and Parameter Estimation

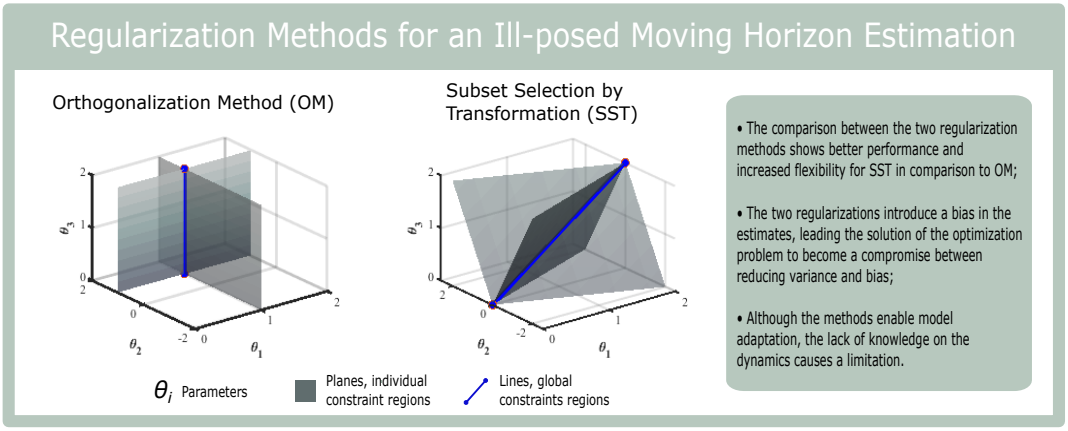
"Among all of the mathematical disciplines, the theory of differential equations is the most important. It furnishes the explanation of all those elementary manifestations of nature which involve time."

MARIUS SOPHUS LIE (1842-1899)

Graphical Abstract

A Regularized Moving Horizon Estimator for Combined State and Parameter Estimation in a Bioprocess Experimental Application

Andrea Tuveri, Caroline S.M. Nakama, José Matias, Haakon Eng Holck, Johannes Jäschke, Lars Imsland, Nadav Bar



Highlights

A Regularized Moving Horizon Estimator for Combined State and Parameter Estimation in a Bioprocess Experimental Application

Andrea Tuveri, Caroline S.M. Nakama, José Matias, Haakon Eng Holck, Johannes Jäschke, Lars Imsland, Nadav Bar

- Online parameter estimation for model adaptation.
- Comparison of two different regularization methods.
- Stopping criterion for regularization methods based on structural identifiability.
- Combination of basic concepts of linear and Lie algebra for constraints selection.
- Implementation of a Moving Horizon Estimator using regularization methods with experimental data.

A Regularized Moving Horizon Estimator for Combined State and Parameter Estimation in a Bioprocess Experimental Application

Andrea Tuveri[a], Caroline S.M. Nakama[a], José Matias[a], Haakon Eng Holck[a], Johannes Jäschke[a], Lars Imsland[b], Nadav Bar[a,*]

[a]*Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway*
[b]*Department of Engineering Cybernetichs, Norwegian University of Science and Technology (NTNU), Trondheim, Norway*

Abstract

Due to the lack or high costs of measurement devices to monitor and control metabolites in microbial cultivation processes, state estimators are often required. These estimators depend on available *on-line* measurements and model dynamics. However, they are often characterised by simple models due to the lack of full knowledge on the process dynamics and high variability in the cell metabolism. This causes uncertainty in the model parameters and therefore the necessity of *on-line* model adaptation, for instance through simultaneous state and parameter estimation. However, these estimation problems are often ill conditioned. The Moving Horizon Estimator (MHE) is a good candidate in this context, since it easily allows enforcing hard constraints as well as regularization to address the ill-posedeness. In this work, we present a method for simultaneous state and parameter estimation in the absence of full state measurements, with the aid of two regularization methods, in a microbial fed-batch cultivation.

Keywords: State Estimation, Parameter Estimation, Moving Horizon Estimator, Regularization, Optimization, Bioprocess

1. Introduction

Real-time monitoring is of paramount importance in control of bioprocesses. Indeed, it is critical for quality assessment and therefore feedback implementations. However, in many applications this is a great challenge, since not all the variables of interest can be directly measured *on-line*. In the case that key variables are unmeasured, state estimators can be used to infer information from the available measurements to compensate for it (Doyle III, 1998; Dochain, 2003). State estimators, also called Soft Sensors (Dochain, 2003), depending on mathematical models (Rao, 2000) that usually represent the real system in an oversimplified manner (e.g. Monod growth model, Monod (1949)), can yield to poor estimates, due to the high model uncertainty, mostly caused by biological variability (Jabarivelisdeh et al., 2020).

Schei (2008) and Mohd Ali et al. (2015) reviewed the implementation of state estimators in chemical processes, discussing design issues, as well as general guidelines for selecting the proper technique in specific applications. Among the various state estimators reported, optimization-based methods, such as the Moving Horizon Estimator (MHE), have two main advantages: (1) they can handle multi-rate measurements easily (Rao, 2000), often encountered in bioprocesses (Elsheikh et al., 2021); (2) they can explicitly incorporate equality or inequality constraints in both states and parameters that may vary in the model (Rao, 2000; Rao et al., 2003), enabling the possibility to handle uncertainty (Doyle III, 1998; Dochain, 2003).

Whereas the capability of handling constraints in combination with multi-rate measurements has been extensively exploited for biological applications, to the best of our knowledge, constraints handling methods have been mostly applied to either state or parameter estimation, not both. Our work aims to account for model uncertainty by *simultaneously* performing state and parameter estimation in absence of full state feedback. Although not applied to a biological process, the work of Liu et al. (2021) also recently presented simultaneous state an parameter estimation in an MHE. However, differently from the work of Bae et al. (2021) and ours, they used a penalty quadratic term for the calculation of the arrival cost (i.e. to take the past information into account).

The majority of the up to date MHE applications in bioprocesses, reconstruct both states and key parameters using multi-rate full state feedback (Gatzke and Doyle III, 2002; Küpper et al., 2009, 2010; Vercammen et al., 2016; Jabarivelisdeh et al., 2020; Elsheikh et al., 2021; Bae et al., 2021; Hernández Rodríguez et al., 2021; Valipour and Ricardez-Sandoval, 2021; Valipour and Ricardez-Sandoval, 2022; Valipour and Ricardez-Sandoval, 2022). Some of the reported applications (Gatzke and Doyle III, 2002; Küpper et al., 2009, 2010; Vercammen et al., 2016; Hernández Ro-

*Corresponding author
Email address: nadi.bar@ntnu.no (Nadav Bar)

dríguez et al., 2021) have shown successful reconstruction of critical model parameters (i.e. model update or adaptation). Elsheikh et al. (2021) presented a review on the utilization of multi-rate measurements and the effect of different arrival cost updates. Valipour and Ricardez-Sandoval (2021) and Jabarivelisdeh et al. (2020) showed the implementation of an MHE along with a nonlinear Model Predictive Controller (NMPC), while Valipour and Ricardez-Sandoval (2022) and Valipour and Ricardez-Sandoval (2022) focused on the improvement of the estimation of unexpected process and measurement noises, by the adoption of non-Gaussian distributions. Bae et al. (2021) presented instead an implementation that attenuates the ill-posed problem arising from the estimation of all parameters, by reducing the number of decision variables (i.e. regularization method). Fewer and less recent works focused on inferring information on the unmeasured states (Raïssi et al., 2005; Goffaux and Wouwer, 2008) without taking into account the possibility to estimate the model parameters. The work of Raïssi et al. (2005) presented an implementation where the solution of the optimization is a bounded set consistent with model, measurements and errors. Goffaux and Wouwer (2008), on the other hand, implemented a min-max optimization problem to deal with parameter uncertainty.

The concept of expressing the uncertain parameters as bounded variables can be dated back to the work of Grossmann and Sargent (1978) and later addressed by Bonvin et al. (2001). Considering that MHE has the ability to explicitly incorporate physical constraints on the parameters, and that disturbances in the form of time-varying parameters can be added as extra degrees of freedom in the optimization problem (Robertson et al., 1996; Rao et al., 2003; Kühl et al., 2011), MHE results a good candidate for performing combined state and parameter estimation. However, uncertainty in the model parameters can potentially cause a large bias in the estimates of the unmeasured states (Dochain, 2003). Moreover, the inclusion of process parameters as decision variables, can lead to an ill-posed optimization problem. Therefore, to obtain a unique local solution, regularization methods are required.

Several approaches have been proposed for selecting parameters subsets (i.e. regularization methods) to attenuate ill-posed parameter estimation problems. McLean and McAuley (2012) reported and divided them in different categories: (1) methods based on the correlation and collinearity indexes (López C. et al., 2015; Anane et al., 2019), (2) Orthogonalization Method (Yao et al., 2003; Lund and Foss, 2008; Thompson et al., 2009; Bae et al., 2021), (3) methods based on the Fisher Information Matrix (FIM) characteristics (Balsa-Canto et al., 2007), and (4) methods based on Principal Component Analysis (PCA) and eigenvalue-eigenvector decomposition (Vajda et al., 1989; Kim and Lee, 2019; Nakama et al., 2020; Chen et al., 2022). In addition to these methods, common practice is also to allow the variation of only one poorly known parameter per measurement value (Dochain, 2003). Further details on regularization techniques can also be found in Kravaris et al. (2013). Most of the regularization methods are dependent on cut-off or threshold values for the selection of the number of constraints. However, some authors already presented alternatives, to avoid cut-off values, based on the minimisation of the mean squared error (MSE) to improve model prediction (Chu et al., 2009; Wu et al., 2011).

To capture the changing or missing dynamics caused by the plant-model mismatch Psichogios and Ungar (1992); Jabarivelisdeh et al. (2020); Bae et al. (2021), our work exploits the ability of MHE to explicitly incorporate physical constraints on the parameters, to continuously update them.

This is done by adding them as single degrees of freedom to the optimization problem and, as also reported in Bae et al. (2021), will result in parameter drifts. To overcome the ill-conditionedness of the problem that arises from that, we implemented and compared two different regularization methods, coupled with a stopping criterion based on structural identifiability. The latter allowed to select, at each time point, the number of active constraints necessary to regularize the problem, based on the available *on-line* information. Firstly, to reduce the search space of the decision variables, we adopted the regularization approach (Subset Selection by Transfomation, SST) as proposed by Kim and Lee (2019) and implemented it in an MHE. This was done using real experimental data of a fed-batch bacterial cultivation of *Corynebacterium glutamicum* (Tuveri et al., 2021). Secondly, to evaluate the properties of SST and show its advantages, we compared it to the known Orthogonalization Method (OM) (Lund and Foss, 2008; Bae et al., 2021). Finally, to avoid using a stopping criterion based on a threshold value, for the selection of the number of the active constraints on the parameters, we proposed a strategy based on the structural identifiability of the system (Villaverde, 2019), in both OM and SST. The results, validated using experimental data from a real bioprocess, present accurate state estimates with respect to the reference values also allowing simultaneous parameter adaptation.

2. Experimental Setup

The experiment was conducted using the *C. glutamicum* ATCC13032 strain. The cells were harvested in a shake flask pre-culture over night in 2YT complex medium and then inoculated in a 2.7 L baffled stirred tank reactor Labfors5 (Infors AG, Switzerland). The experiment was conducted with initial volume of 1.5 L and initial OD_{600} of 1. The 500 mL of glucose feeding (100 g/L) were added once the dissolved oxygen stabilized above 60 %. Temperature, pressure and pH were kept respectively to 30 °C, 1 bar and 7. The reactor was aerated with 2 NL/min pressurized air and the dissolved oxygen was controlled above 30% by modifying the stirrer speed (200–1100 rpm). *C. glutamicum* was cultured on CGXII minimal medium and glucose used as carbon source. *On-line* measurements

for absorbance (840–910 nm, ASD12-N Absorption Probe, Optek GmbH), volume and *off-gas* composition (BlueInOne Ferm, BlueSens GmbH) were available every 10 seconds trough the process information management system Lucullus (Securecell, Switzerland). *Off-line* samples for sugars were instead collected and stored at 4°C by the NUMERA system (Securecell, Switzerland). Glucose samples were analyzed using an high-pressure liquid chromatography system (UltiMate 3000 series, Thermo Scientific, U.S.). The *off-line* biomass was instead evaluated by measuring cell dry weight. The reader is referred to Tuveri et al. (2021) for more details on cultivation process and analytical procedures.

3. Bioprocess description

This section starts by introducing the system dynamics (Sec. 3.1) and the noise structure (Sec. 3.1.1). Following this, *on-line* (i.e. outputs) and *off-line* measurements (i.e. reference values) are described (Sec. 3.1.2) together with the signal processing (Sec. 3.1.3).

3.1. System Model

The process describes an aerobic bacterial cultivation, which is performed in two phases:

- A batch phase, up to complete sugar depletion and stabilization of dissolved oxygen above 60%;
- A fed-batch phase, followed a short period of starvation.

The system dynamics are described using first order Monod-like kinetics (Tuveri et al., 2021). In addition, cell death is considered to be linear and dilution is due to the addition of feeding. The model equations are the following:

$$\begin{aligned}\dot{V} =& F_{in}\\ \dot{X} =& -\frac{F_{in}}{V}X + \mu_{max}\frac{S}{K_s+S}X - k_dX\\ \dot{S} =& \frac{F_{in}}{V}(S_{in}-S) - \mu_{max}\frac{S}{K_s+S}\frac{X}{Y_{XS}}\\ \dot{\mathrm{CO}}_2 =& \mu_{max}\frac{S}{K_s+S}\frac{X}{Y_{X\mathrm{CO}_2}} - q_{air}CO_2\end{aligned} \tag{1}$$

The states V, X, S and CO_2 in Eq. 1 represent the concentrations of volume, biomass, glucose (or substrate, sugar) and carbon dioxide (CO_2), respectively. F_{in} indicates the feeding flow to the reactor, given a substrate concentration S_{in}. In addition, q_{air} represents a constant gas inflow. From herein, we define:

- State vector $x \in \mathbb{R}^{n_x}$ and input $u \in \mathbb{R}^{n_u}$:

$$x = [V,\ X,\ S,\ CO_2]^T \quad \text{and} \quad u = [F_{in}] \tag{2}$$

- Measured outputs $y \in \mathbb{R}^{n_y}$:

$$y = [V,\ X,\ CO_2]^T \tag{3}$$

- Parameters $\theta \in \mathbb{R}^{n_\theta}$:

$$\theta = [\mu_{max}, K_s, k_d, Y_{XS}, Y_{X\mathrm{CO}_2}]^T \tag{4}$$

This allows us to write the model in Eq. 1 as $\dot{x} = f(x, u, \theta)$. Additionally, by adding the parameters as state variables with zero dynamics ($\dot{\theta}_i = 0$), the system in Eq. 1 can be written as:

$$\dot{x}_a = \bar{f}(x_a, u) \tag{5}$$

where $x_a \in \mathbb{R}^{n_x+n_\theta}$ is defined as $x_a = [x, \theta]$.

Table 1: Model parameters $\theta \in \mathbb{R}^{n_\theta}$ for the system in Eq. 1 given values, units and standard deviation (Tuveri et al., 2021).

Parameter	Description	Value	Unit	Std. Dev.
μ_{max}	Maximum growth rate	0.19445	$[\mathrm{h}^{-1}]$	$3.25 \cdot 10^{-6}$
K_S	Monod growth constant	0.007	$[\mathrm{g} \cdot \mathrm{L}^{-1}]$	$3.92 \cdot 10^{-6}$
k_d	Death rate constant	0.006	$[\mathrm{h}^{-1}]$	$4.49 \cdot 10^{-6}$
Y_{XS}	S from X yield	0.42042	$[\mathrm{g} \cdot \mathrm{g}^{-1}]$	$3.58 \cdot 10^{-6}$
$Y_{X\mathrm{CO}_2}$	CO_2 from X yield	0.54308	$[\mathrm{g} \cdot \mathrm{g}^{-1}]$	$2.22 \cdot 10^{-6}$

3.1.1. Noise Structure Tuning

Following the works of Leu and Baratti (2000) and Kolås et al. (2009), the process noise was added to the deterministic plant model (Eq. 1) twofold. In the states, in order to address unmodelled dynamics, and in the parameters, to address parameter uncertainty (standard deviations are reported in Table 1). Indeed, the idea is that the model uncertainty (Leu and Baratti, 2000; Kolås et al., 2009) arises mostly from the uncertainty in the model parameters ($\omega_{\mu_{max}}, \omega_{K_S}, \omega_{k_d}, \omega_{Y_{XS}}, \omega_{Y_{XCO_2}}$). Moreover, in our case, to prevent the noise covariance matrix from becoming singular when the substrate S is depleted, the noise is added to the state dynamics ($\omega_V, \omega_X, \omega_S, \omega_{CO_2}$). The noise enters the material balances (Eq. 1) as:

$$\begin{aligned}\dot{V} =& F_{in} + \omega_V\\ \dot{X} =& -\frac{F_{in}}{V}X + (\mu_{max}+\omega_{\mu_{max}})\frac{S}{(K_s+\omega_{k_S})+S}X - (k_d+\omega_{k_d})X + \omega_X\\ \dot{S} =& \frac{F_{in}}{V}(S_{in}-S) - (\mu_{max}+\omega_{\mu_{max}})\frac{S}{(K_s+\omega_{k_S})+S}\frac{X}{(Y_{XS}+\omega_{Y_{XS}})} + \omega_S\\ \dot{\mathrm{CO}}_2 =& (\mu_{max}+\omega_{\mu_{max}})\frac{S}{(K_s+\omega_{k_S})+S}\frac{X}{(Y_{X\mathrm{CO}_2}+\omega_{Y_{X\mathrm{CO}_2}})} - q_{air}CO_2 + \omega_{\mathrm{CO}_2}\end{aligned} \tag{6}$$

given the noise vector $\omega \in \mathbb{R}^{(n_x+n_\theta)}$:

$$\omega = \left[\ \omega_{\mu_{max}}\ \ \omega_{k_S}\ \ \omega_{k_d}\ \ \omega_{Y_{XS}}\ \ \omega_{Y_{XCO_2}}\ \ \omega_V\ \ \omega_X\ \ \omega_S\ \ \omega_{CO_2}\ \right]^T$$

with $\omega \sim \mathcal{N}(0, Q_\omega)$, where the diagonal elements of the covariance matrix $Q_\omega \in \mathbb{R}^{(n_x+n_\theta)\times(n_x+n_\theta)}$ are reported in Table 2. For the batch phase, the values of Q_ω related to the parameters θ are tuned by setting them equal to the value of the variance, obtained by parameter estimation on a single experiment, while the ones related to the states x are defined as tuning parameters and therefore selected manually. For the fed-batch phase, only the values of $\sigma^2_{K_S}$ and $\sigma^2_{Y_{X\mathrm{CO}_2}}$ are increased to compensate for unmodelled dynamics (Tuveri et al., 2021). The model with noise (Eq. 6) can be compactly written as:

$$\dot{x} = f(x, u, \theta, \omega) \tag{7}$$

Table 2: Diagonal elements of the covariance matrix Q_ω. The values are kept constant until the fed-batch phase, when the values of $\sigma^2_{K_S}$ and $\sigma^2_{Y_{XCO_2}}$ are increased to compensate for unmodelled dynamics.

Variance	Additive noise	Batch	Fed-batch	Unit
$\sigma^2_{\mu_{max}}$	in μ_{max}	$1.05 \cdot 10^{-11}$	-	$[h^{-2}]$
$\sigma^2_{K_S}$	in K_S	$1.54 \cdot 10^{-11}$	$3.38 \cdot 10^{-2}$	$[g^2.L^{-2}]$
$\sigma^2_{k_d}$	in k_d	$2.02 \cdot 10^{-11}$	-	$[h^{-2}]$
$\sigma^2_{Y_{XS}}$	in Y_{XS}	$1.28 \cdot 10^{-11}$	-	$[g^2.g^{-2}]$
$\sigma^2_{Y_{XCO_2}}$	in Y_{XCO_2}	$4.91 \cdot 10^{-12}$	$4.91 \cdot 10^{-2}$	$[g^2.g^{-2}]$
σ^2_V	in V	$1 \cdot 10^{-1}$	-	$[L^2.h^{-2}]$
σ^2_X	in X	$1 \cdot 10^{-2}$	-	$[g^2.h^{-2}]$
σ^2_S	in S	$1 \cdot 10^{-2}$	-	$[g^2.h^{-2}]$
$\sigma^2_{CO_2}$	in CO_2	$1 \cdot 10^{-1}$	-	$[h^{-2}]$

The process noise covariance $Q_k \in \mathbb{R}^{n_x \times n_x}$ for the system described as in Eq. 6, is updated at each sampling time k as (Tuveri et al., 2021):

$$Q_k = G_k \cdot Q_\omega \cdot {G_k}^T \tag{8}$$

Eq. 8 therefore allows to have a state-dependant varying covariance Q_k, where Q_ω is a constant related to the statistics of the parameter uncertainty. $G_k \in \mathbb{R}^{n_x \times (n_\theta + n_x)}$ is the Jacobian of Eq. 7 with respect to the noise ω:

$$G_k = \frac{\partial f(x, u, \theta, \omega)}{\partial \omega} \tag{9}$$

Additionally, to estimate the parameters θ together with the states x, we define them as additional state variables and the augmented state vector becomes $x_a \in \mathbb{R}^{n_x + n_\theta}$. In accordance with the work of Grossmann and Sargent (1978), the parameters are considered as bounded variables, given that probability distribution functions for the parameters are available. This is done to reflect the drifting characteristics of the model parameters used to describe the system, as was previously showed in Bae et al. (2021). Therefore, by approximating the probability distribution functions of the parameters to be normal, with mean value θ_0 and variance ${\sigma_\theta}^2$ (Table 2), we will treat them as drifting bounded variables, allowing them to vary within their bounds at each iteration. The augmented model (Eq. 10) can therefore be written as:

$$\begin{aligned}
\dot{V} =& F_{in} + \omega_V \\
\dot{X} =& -\frac{F_{in}}{V} X + (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{k_S}) + S} X - (k_d + \omega_{k_d}) X + \omega_X \\
\dot{S} =& \frac{F_{in}}{V}(S_{in} - S) - (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{k_S}) + S} \frac{X}{(Y_{XS} + \omega_{Y_{XS}})} + \omega_S \\
\dot{CO}_2 =& (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{k_S}) + S} \frac{X}{(Y_{XCO_2} + \omega_{Y_{XCO_2}})} - q_{air} CO_2 + \omega_{CO_2} \\
\dot{\mu}_{max} =& \omega_{\mu_{max}} \\
\dot{K}_s =& \omega_{K_s} \\
\dot{k}_d =& \omega_{k_d} \\
\dot{Y}_{XS} =& \omega_{Y_{XS}} \\
\dot{Y}_{XCO_2} =& \omega_{Y_{XCO_2}}
\end{aligned} \tag{10}$$

Eq. 10 can be written compactly as $\dot{x}_a = \bar{f}(x_a, u, \omega)$. The state-dependant covariance matrix for the augmented system $\bar{Q}_k \in \mathbb{R}^{(n_x + n_\theta) \times (n_x + n_\theta)}$ is updated at each sampling time k as:

$$\bar{Q}_k = \begin{bmatrix} Q_k & 0 \\ 0 & Q_\theta \end{bmatrix} \tag{11}$$

Where the state-dependant submatrix Q_k (Eq. 8) is updated at each sampling time k, while $Q_\theta \in \mathbb{R}^{n_\theta \times n_\theta}$ is a constant diagonal matrix with variances ${\sigma_\theta}^2$ defined in Table 2. The latter is done to allow the parameters to vary within the bounds at each iteration k and it is a necessary assumption since the real dynamics of the parameters are unknown.

3.1.2. On-line and Off-line measurements

The *on-line* output measurements y were available with a frequency of 10 seconds and used by the estimator to monitor biomass, volume and CO_2 and to infer the unmeasured glucose composition. Additionally, to have reference values for evaluating the estimation performance, *off-line* samples of X and S were taken with a lower sample frequency and not used at any time by the estimator. The samples for biomass X were collected manually approximately every 3 hours, whereas for glucose S they were collected automatically by the NUMERA auto-sampler every hour. Additional information is available in Tuveri et al. (2021).

3.1.3. Signal Processing

The *on-line* output measurements y were used by the estimator every 60 seconds. Signals from the absorbance probe were obtained in concentration units (0.05 - 4 CU) and than converted to g/L of biomass X (cell dry weight, CDW) using a calibration curve as follows:

$$CDW_{CU} = \begin{cases} 22.187 \cdot CU - 5.0991 & CU \geq 0.9 \\ 11.124 \cdot CU + 0.66116 & CU < 0.9 \end{cases}$$

On-line CO_2 signals were obtained by measurements of the composition in the outflow (0%–25%). The volume measurement V was calculated by integration of the *on-line* flow signals every 10 seconds. Moreover, to consider the amount of volume taken for the *off-line* samples (8 mL/sample), these values were iteratively integrated with the pump signals. Although available every 10 seconds, each of the measurements was only used by the estimator every 60 seconds. This is done to not increase the size of the optimization problem.

4. Background theory

This section presents some important concepts utilized to develop the proposed approaches presented in the following section. We first introduce the sensitivity matrix (Sec. 4.1) which is the base for the regularization methods implemented. Than, briefly discuss the local observability and identifiability of the system (Sec. 4.2), for the further introduction of the stopping criterion used in the regularization methods.

4.1. Sensitivity Matrix

The sensitivity matrix of model outputs is typically defined as the Jacobian of the outputs (y) with respect to the parameters θ (Vajda et al., 1985; Yao et al., 2003; Lund and Foss, 2008; Thompson et al., 2009; Bae et al., 2021; Chen et al., 2022). However, because in this work we are interested in the effect of the parameters on the states, following the work of Bae et al. (2021), we define it as the Jacobian of the state variables (x) with respect to the parameters (θ):

$$S_{\theta,mn} = \left.\frac{\partial x_m}{\partial \theta_n}\right|_{\hat{x},\hat{\theta}} \tag{12}$$

where:

$$m = 1, \ldots, n_x$$
$$n = 1, \ldots, n_\theta$$

and $\hat{x}$ and $\hat{\theta}$ are respectively the estimates of states and parameters. The idea behind using the state variables (x) instead of the output measurements (y) is to identify the model parameters (θ) which have a significant effect on the state variables (Bae et al., 2021). This allows us to identify time-varying values for the parameters (θ) so that the estimated state variables ($\hat{x}$) can better match the *off-line* samples (y^*) and therefore the underlying true process behaviour (i.e. obtain lower RMSE). The sensitivity matrix therefore summarizes locally the influence of the parameters on the state variables. It is important to carefully scale S_θ since this will directly affect the information obtained for the parameter estimation method (Vajda et al., 1989; Yao et al., 2003; Lund and Foss, 2008; Thompson et al., 2009; McLean and McAuley, 2012; Chis et al., 2016). Therefore, to address the problem of different orders of magnitudes in the parameters, their logarithm ($log(\theta)$) is used to avoid scaling issues due to differences in their orders of magnitude (Vajda et al., 1989; Chis et al., 2016). Additionally, to take into account the characteristics of the data and the available variation ranges, we will refer to the same scaling method applied in Thompson et al. (2009), McLean and McAuley (2012) and Bae et al. (2021), by using the state and parameter variance obtained from the covariance matrix. The scaled sensitivity matrix $Z \in \mathbb{R}^{n_x \times n_\theta}$ is defined as:

$$Z_{m,n} = \left.\frac{\partial x_m}{\partial log(\theta_n)} \frac{log(\sigma_{\theta_n})}{\sigma_{x_m}}\right|_{\hat{x},\hat{\theta}} \tag{13}$$

In the following, unless differently stated, we will refer to the sensitivity as defined in Eq. 13.

4.2. Local Observability and Identifiability

Observability of nonlinear systems based on *Lie* algebra is a structural and local property (Hermann and Krener, 1977; Isidori, 1985; Nijmeijer and van der Schaft, 1990; Powel and Morgansen, 2015; Villaverde, 2019). The local observability of the system in Eq. 1 was already analyzed and discussed by Tuveri et al. (2021). However, the focus there was limited to assess the possibility to infer information about the unmeasured variable of interest (S). Here instead, we take into account also the process parameters (θ) and therefore need to evaluate structural identifiability (Villaverde, 2019). By assuming that the only input is F_{in} (previously defined in Eq. 2), being S_{in} and q_{air} constant over time (Sec. 2), the system in Eq. 5 can be written as an input-affine system, with the terms only dependent on the states ($\bar{f}_x$) and the ones dependent on the input ($\bar{f}_u$) such that:

$$\dot{x}_a = \bar{f}_x(x_a) + \bar{f}_u(x_a) \cdot u \tag{14}$$
$$y = h(x) \tag{15}$$

If we now consider that $\bar{f}_u(x_a)$ does not carry any information about the parameters (see Eq. 1), we can define the map O, which represents the observation space of the system (Eq. 14), without considering the term $\bar{f}_u(x_a) \cdot u$, as:

$$O = \{O_1, O_2, O_3\} = \{L^0_{\bar{f}_x} h, L^1_{\bar{f}_x} h, L^2_{\bar{f}_x} h\} \tag{16}$$

Where:

$$L^0_{\bar{f}_x} h = h, \qquad L^1_{\bar{f}_x} h = \frac{\partial h}{\partial x_a} \bar{f}_x \qquad L^2_{\bar{f}_x} h = \frac{\partial (L^1_{\bar{f}_x} h)}{\partial x_a} \bar{f}_x$$

And obtain the codistribution dO:

$$dO = span\{d(O_1), d(O_2), d(O_3)\} \tag{17}$$

The dimension of the codistribution dO defines if the system is locally observable ($dO = n_x$). Moreover, when considering structural identifiability, if dO has the same size as the augmented vector ($n_x + n_\theta$), the system is said to be locally observable with identifiable parameters (Villaverde, 2019). The system in consideration (Eq. 1) is locally observable, since $dO = n_x$ when considering only the states x, but not all the parameters are locally identifiable at every iteration, since $dO < n_x + n_\theta$ when considering the augmented vector x_a. Therefore indicating that the available information in the latter is not enough to estimate all states and parameters together.

5. Proposed approach

Simultaneous estimation of both states and parameters can lead to an ill-posed optimization problem. To attenuate the ill-posed problem that arises by the addition of the parameters as decision variables, we consider two different approaches, namely the Orthogonalization Method (OM) (Sec. 5.2.1) and the Subset Selection by Transfomation (SST) (Sec. 5.2.2). Both methods are based on the sensitivity matrix and reduce the parameter space, but while OM keeps a subset of parameters constant, SST fixes linear combinations of parameters (i.e. clusters of parameters). The SST method can be therefore regarded as a more flexible approach, since it constrains linear combinations and not single parameters. This enables to vary more

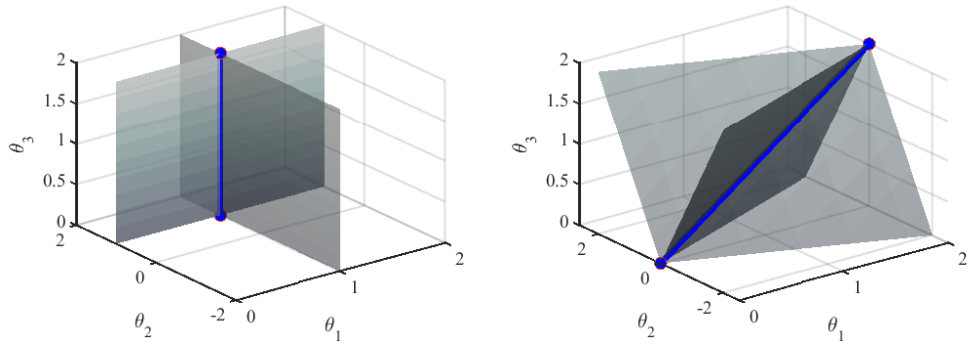


Figure 1: Geometric representation of the constraint implementation. Orthogonalization Method (OM, left) and Subset Selection by Transformation (SST, right). Each plane represents an individual constrained region. When more than one constraint is selected, the constrained region is represented by their intersection (blue line). In OM (left), constraints are used for singular parameters (planes) and the degrees of freedom are the unconstrained parameters (blue line). In SST (right), constraints are implemented by linear combinations of different θ_i and therefore the planes are not required to be normal to an axis, leaving more freedom to the individual parameters. Here, both θ_1 and θ_3 (right) can vary within the projection of the constrained region (blue line) along their axes. The values reported here are only for illustration purposes.

parameters simultaneously. However, a stopping criterion for the decision on how many degrees of freedom can be used for the estimation of the parameters is necessary for both methods (OM and SST). In this work we propose a stopping criterion based on the structural identifiability (Sec. 5.1).

This section will first present the stopping criterion proposed (Sec. 5.1), explaining how to practically use identifiability as a decision making criterion for constraints selection. Secondly, it introduces the two regularization approaches (i.e. OM and SST) together with the application of the proposed stopping criterion (Sec. 5.2). Thirdly, it shows the implementation of the aforementioned methods in a MHE formulation (Sec. 5.3), for the combined estimation of states and parameters.

5.1. Lie-based Constraints Selection

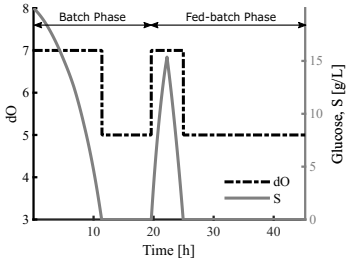


Figure 2: Dimension of the observability space (dO) for the open loop model and Glucose (S) dynamics. dO is calculated without taking into account the terms related to the input (u, Eq. 14). The decrease in dO ($dO = 5$) is a consequence of the sugar depletion.

The intention of this work is to present a consistent and simple method to select the constraints for regularizing the optimization problem (Fig. 1). Although different regularization methods have been presented (McLean and McAuley, 2012), their dependence on a cut-off or threshold value might hinder their results (Chu et al., 2009; Wu et al., 2011; Kim and Lee, 2019; Nakama et al., 2020; Chen et al., 2022). We therefore want to use the information obtained by the local identifiability (structural property) as a baseline for the decision method of constraints selection. That is, at any time, we select the available degrees of freedom given the dimension of the observability co-distribution dO for the augmented system. The open loop test in Fig. 2 reports the dimension of the observability co-distribution dO for the system in Eq. 1, calculated as in Sec. 4.2. The system is locally observable ($dO > 4$, Fig. 2), but not all parameters are structurally identifiable ($dO < 7$ under sugar depletion, Fig. 2). However some information can still be retrieved also when the substrate is depleted. During the dynamic phases of the process (0-11 hours and 20-25 hours) the available information from the data is higher. Both under absence ($dO = 7$) and presence ($dO = 5$) of sugar depletion, it is possible to infer information of some of the parameters without having an ill-posed problem.

To introduce the method, we start by defining the dimension of the observability co-distribution dO. This is equivalent to the maximum number of degrees of freedom (i.e. decision variables of the optimization problem):

$$\mathrm{DOF}_{max} = dO \tag{18}$$

Eq. 18 indicates that dO, being a local property, gives us information about the maximum number of states and parameters that can be estimated, given the available information at each time. The maximum DOF (DOF_{max}) is given by $n_x + n_y$, which is 7 for this problem. Four of these degrees of freedom ($\mathrm{DOF}_x = 4$) are used for the estimation of the state variables (n_x). Following the previous

considerations, the degrees of freedom (DOF_θ) available for estimating the parameters are given by:

$$\mathrm{DOF}_\theta = \mathrm{DOF}_{max} - \mathrm{DOF}_x \tag{19}$$

Therefore, the number of constraints (n_{θ_s}) that we need to impose on the estimation problem to achieve identifiability is:

$$n_{\theta_s} = n_\theta - \mathrm{DOF}_\theta \tag{20}$$

That is, whenever $\mathrm{DOF}_\theta < n_\theta$, we will add linear constraints on the parameters to attenuate the ill-posed estimation problem. However, since in real microbial cultivations the substrate is not entirely consumed ($S \neq 0$), some precautions must be taken. For numerical reasons, small values of S (in this work $< 10^{-3}$ g/L) should be considered $S = 0$ during the *on-line* calculation of dO.

5.2. Regularization Methods

This section describes the two regularization methods adopted (i.e. OM and SST) (Sec. 5.2.1 and Sec. 5.2.2), together with their geometrical interpretation (Sec. 5.2.3).

5.2.1. Orthogonalization Method

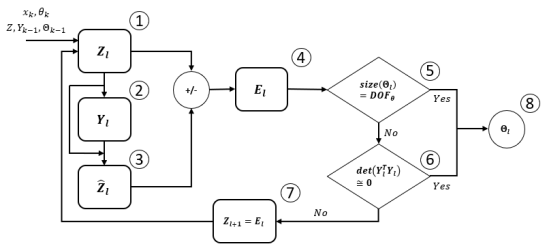


Figure 3: Orthogonalization Method implementation block diagram. Initial values are given and the scaled sensitivity matrix is obtained in (1). Then the scaled parameter with the biggest norm is added to vector Y_l (2) and the matrix projection is calculated in (3), given (1) and (2). Consecutively, the residual error is obtained in (4). Finally, the stopping criteria are verified in (5) and (6). If one of these conditions is satisfied, the vector of constrained parameters is obtained in (8). More specifically, condition (6) is set with a threshold of 10^{-8}. If not the method goes to (7) and starts again to add an extra parameter to be constrained.

Among the various Parameter Subset Selection (PSS) methods, the Orthogonalization Method is the heuristic method most commonly used due to its simplicity (Yao et al., 2003; Lund and Foss, 2008; Kim and Lee, 2019; Bae et al., 2021). In the context of MHE, this method is appealing due to its small computational loads (Bae et al., 2021). However, an unsuitable selection of the cut-off value for the stopping criterion, might result in poor estimates. To avoid that, we present here its implementation together with the stopping criterion previously presented (Sec. 5.1). For that purpose, we start with the scaled sensitivity matrix Z (Eq. 13):

$$Z = [z_1,\ z_2, \ldots,\ z_{n_\theta}] \tag{21}$$

The elements z_i of the Eq. 21 represent the column vectors of Z. Those vectors are sorted in decreasing order, by calculating the Euclidean norm. We take the one with the largest norm z_1, representing the main direction, and normalise it:

$$q_1 = \frac{z_1}{\|z_1\|} \tag{22}$$

At the first step ($l = 1$), $Z_l = Z$, $Y_{l-1} = \varnothing$ and $\Theta_{l-1} = \varnothing$. The parameter θ_l, corresponding to the column vector z_l, is than appended to the vector $\Theta_l = [\Theta_{l-1}, \theta_l]$. Similarly, the matrix Y_l is defined by appending q_l. Given Y_l, we calculate the matrix projection of Z_l onto it:

$$\hat{Z}_l = Y_l (Y_l^T Y_l)^{-1} Y_l^T Z_l \tag{23}$$

If we now subtract the projection $\hat{Z}_l$ (Eq. 23) we obtain the residual error (Eq. 24) which is orthogonal to Y_l:

$$E_l = Z_l - \hat{Z}_l \tag{24}$$

The method stops, differently from Bae et al. (2021), if the following condition, based on the stopping criterion defined in Sec. 5.1, is satisfied:

$$size(\Theta_l) = \mathrm{DOF}_\theta \tag{25}$$

where DOF_θ is obtained as in Eq. 19. If this condition is not satisfied, we set $Z_{l+1} = E_l$ until all the parameters are ranked or until $Y_l^T Y_l$ is nearly singular (Yao et al., 2003; Bae et al., 2021). The reader is referred to Strang (2016); Yao et al. (2003); Lund and Foss (2008); Thompson et al. (2009); McLean and McAuley (2012); Kim and Lee (2019); Bae et al. (2021) for more detailed information.

5.2.2. Subset Selection by Transformation

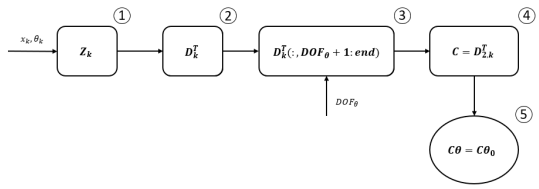


Figure 4: Subset Selection by Transformation block diagram. This method requires the calculation of the right eigenvector matrix through singular values decomposition (SVD) in (2), given the scaled sensitivity matrix in (1). Then the submatrix of the right eigenvectors corresponding to the smallest eigenvalues is selected according to the DOF_θ at the k_{th} iteration (3) and set equal to the constraint coefficient matrix C (4). The constraints are selected in (5) and the right hand side defined based on the initial parameters θ_0.

Subset Selection by Transformation (SST), was firstly introduced by Kim and Lee (2019) and further applied in Kim et al. (2019) and Chen et al. (2022). As observed in Nakama et al. (2020), and reported in Chen et al. (2022), Subset Selection by Transformation (SST) induces the

same regularization effect as Principal Component Regression (PCR). Also known as Truncated Singular Values Decomposition (TSVD) (Kim and Lee, 2019), PCR is an eigenvalue-eigenvector based analysis that reveals existing dependencies among the parameters and that can reduce the parameter space by using the eigenvectors of the reduced Hessian matrix (Nakama et al., 2020). However, while in PCR the parameters are projected onto the subspace of the eigenvectors associated with the larger eigenvalues, in SST the parameter space is restricted by adding constraints in the directions of the eigenvectors associated with the smaller eigenvalues. Following the works of Kim and Lee (2019), Kim et al. (2019) and Chen et al. (2022), we linearized the model in Eq. 1 around $\theta = \theta_0$:

$$\begin{aligned} x = &\ x_0 + S_\theta(\theta - \theta_0) \\ = &\ S_\theta\theta + c \end{aligned} \tag{26}$$

where x_0 and θ_0 represent the initial values, S_θ is the sensitivity matrix (Eq. 12) and $c = x_0 - S_\theta\theta_0$ is a constant. By adopting the Hessian approximation ($\mathcal{H} \approx S_\theta^T S_\theta$), as reported in Vajda et al. (1985), we can analyze the changes in the state variables x with respect to the parameters θ. Based on the eigenvalue-eigenvector decomposition, the Hessian approximation $\mathcal{H} \approx S_\theta^T S_\theta$, can be represented as:

$$S_\theta^T S_\theta = D\Sigma D^T \tag{27}$$

with $D \in \mathbb{R}^{n_\theta \times n_\theta}$ as the orthogonal eigenvectors matrix. Moreover, given that D is an orthonormal matrix ($D^T D = DD^T = \mathbb{I}$), we can rewrite Eq. 26 (Kim and Lee, 2019; Chen et al., 2022):

$$\begin{aligned} x = &\ S_\theta DD^T\theta + c \\ = &\ M\alpha + c \end{aligned} \tag{28}$$

where $M = S_\theta D$ and $\alpha = D^T\theta$. In this way we have obtained a vector $\alpha \in \mathbb{R}^{n_\theta}$ of transformed parameters. Note that the eigenvalue-eigenvector decomposition of the approximated Hessian based on either S_θ or Z (Eq. 13), for the calculation of Eq. 28, does not change the direction of the eigenvectors (D), but just the magnitude of the eigenvalues (Ω). Eq. 27 can then be rewritten with respect to Eq. 13 as:

$$Z^T Z = D\Omega D^T \tag{29}$$

Consequently, based on the stopping criterion defined in Sec. 5.1, the number of active constraints (n_{θ_s}) (Eq. 20) is defined. The vector α (Eq. 28) can therefore be further divided in two subvectors α_1 and α_2 as:

$$x = \ M_1\alpha_1 + M_2\alpha_2 + c \tag{30}$$

where the magnitude of the eigenvalues sorts them in descending order. This two subvectors indicate the clusters of parameters corresponding to the larger eigenvalues ($\alpha_1 = D_1^T\theta$) which are estimated, and the ones corresponding to the smaller eigenvalues ($\alpha_2 = D_2^T\theta$) which are fixed (i.e. constraints). Given this decomposition, it is possible to find a constraint coefficient matrix $C \in \mathbb{R}^{n_{\theta S} \times n_\theta}$:

$$C\theta = r \tag{31}$$

where $C = D_2^T$ is a submatrix of D^T (Eq. 29) and $r \in \mathbb{R}^{n_{\theta S}}$ is a given constraint right-hand side vector. By defining the constraint right-hand side vector as $r = C\theta_0$, Eq. 31 becomes:

$$C\theta - C\theta_0 = 0 \tag{32}$$

Eq. 32 will therefore constrain, at each iteration k, the newly estimated parameters based on the nominal conditions (θ_0), which represent the local optimal solution (Table 1) previously obtained as discussed in Tuveri et al. (2021). Our contribution to the method, in addition to what previously presented in Kim and Lee (2019), consists in its geometric interpretation. Indeed, while the work of Kim and Lee (2019) looks at the problem trough the transformation of the parameters (i.e. $\theta = D\alpha$, by rotating the axis trough directions of the principal components) and constrains the unselected transformed parameters to their nominal values (i.e. $\alpha_2 = \alpha_{2,0}$), in our work we look at it from the constraints point of view (Nakama et al., 2020). Therefore showing that Eq. 32, which mathematically is equivalent to the method of Kim and Lee (2019), geometrically implies fixing the constraints in the directions of the smaller eigenvalues (Fig. 1) while maintaining the direction of the axes of the original parameters θ. This has the advantage of enhancing the interpretability of SST, which, as previously stated in Kim and Lee (2019), can complement OM and PCR. Indeed, our interpretation shows that: (1) by maintaining the same axes (Fig. 1), SST can be directly compared with the state of the art OM (Yao et al., 2003; Lund and Foss, 2008), as will also be further explained in Sec. 5.2.3; (2) by looking at the problem from the constraints point of view, SST has the same regularization effect as PCR.

5.2.3. Interpretation of Regularization Methods

Given the two regularization methods, OM (Sec. 5.2.1) and SST (Sec. 5.2.2), we want to show how the implementation of constraints differs. This concept is geometrically represented (for three dimensions) in Fig. 1, and mathematically translates, for OM, to:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{33}$$

and, for SST, to:

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{34}$$

where Eq.s 33 and 34 are equivalent to Eq. 31. Indeed, while in OM we constrain single parameters θ at every iteration (Eq. 33), in SST we constrain clusters of them

(Eq. 34). The main reasoning for using SST is to allow the change of the parameters along the directions with higher certainty, while restricting them along the less sensitive directions. These directions represent a linear combination (i.e. cluster) of parameters with a fixed relationship that can provide a more flexible approach than constraining single parameters (Fig. 1).

5.3. Moving Horizon Estimator

Consider the system dynamics, described by a set of ordinary differential equations (ODEs), which is given by the system in Eq. 1 augmented by considering the parameters as additional states (Eq. 5) together with the measurement function $h(x) = Hx$ (Eq. 35b). The implementation of one of the regularization methods previously presented OM (Sec. 5.2.1) and SST (Sec. 5.2.2) consists in adding a set of algebraic equations ($g(\theta) = 0$) to enforce equality constraints on the parameters (Eq. 35c). We than obtain a system of differential algebraic equations (DAEs):

$$\dot{x}_a = \bar{f}(x_a, u) + w \tag{35a}$$
$$y = h(x) \tag{35b}$$
$$0 = g(\theta) \tag{35c}$$

where $x \in \mathbb{R}^{n_x}$ is the state vector, $y \in \mathbb{R}^{n_y}$ the output vector, $u \in \mathbb{R}^{n_u}$ the input vector, $\theta \in \mathbb{R}^{n_\theta}$ the parameters vector, $x_a \in \mathbb{R}^{(n_x+n_\theta)}$ the augmented vector and $w \in \mathbb{R}^{(n_x+n_\theta)}$ is the process noise random variable. Here w is different from $\omega \in \mathbb{R}^{(n_x+n_\theta)}$ which is solely a tuning parameter for $\bar{Q}_k$, indeed while $\omega \sim \mathcal{N}(0, Q_\omega)$, $w \sim \mathcal{N}(0, \bar{Q}_k)$. The discretization of Eq. 35, given the sampling time t_k, yields to:

$$x_{a,k+1} = \bar{F}(x_{a,k}, u_k) + w_k \tag{36a}$$
$$y_k = h(x_k) \tag{36b}$$
$$0 = g(\theta_k) \tag{36c}$$

$\bar{F} : \mathbb{R}^{(n_x+n_\theta)} \times \mathbb{R}^{n_u} \to \mathbb{R}^{(n_x+n_\theta)}$ is obtained by discretization of $\bar{f}$. The solution to the MHE problem (Kühl et al., 2011; Andersson et al., 2016) consists of finding parameters, states and their noise within a finite-time horizon $T = t_N - t_L$. Where t_L and t_N are respectively the initial and final times of the horizon. This is done by solving the following constrained least-squares optimization problem every 60 seconds (i.e. when new measurements are available), along the horizon T (30 min):

$$\min_{x_i,\theta_i,w_i} \quad \left\| \begin{matrix} \hat{x}_L - x_L \\ \hat{\theta}_L - \theta_L \end{matrix} \right\|^2_{P_L} + \sum_{i=L}^{N} \|y_i - h(x_i)\|^2_V + \sum_{i=L}^{N-1} \|w_i\|^2_{W_k}$$

$$\text{s.t.} \quad x_{a,i+1} = F(x_{a,i}, u_i) + w_i \qquad i = L, \dots, N-1 \tag{37a}$$
$$g(\theta_i) = 0 \qquad i = L, \dots, N \tag{37b}$$
$$x_i \geq x_{min} \qquad i = L, \dots, N \tag{37c}$$
$$\theta_{min} \leq \theta_i \leq \theta_{max} \qquad i = L, \dots, N \tag{37d}$$

The aim is to obtain states (x_i), parameters (θ_i) and process noise (w_i) using the available information from the model (Eq. 1) and the outputs (1 min sampling rate). The constraints in Eq. 37 are defined as:

- equality constraints on the state variables based on the process dynamics (Eq. 37a) and on the process parameters, based on the regularization method selected in Eq. 37b (Sec.s 5.2.1 and 5.2.2);
- inequality constraints based on physical limitations of the process variables ($x_{min} = [0, 0, 0, 0]$, Eq. 37c) and the parameters uncertainty (given the parameter bounds θ_{min} and θ_{max}, Eq. 37d). The latter allows to maintain the newly estimated parameters close to the nominal values, given that, under presence of structural plant-model mismatch, parameter adaptation might not necessarily lead to an improved model (Marchetti et al., 2008).

The cost function (Eq. 37) is given by the summation of three Euclidean norms. The first term is the arrival cost which summarises past information (its calculation is reported in S1). The second is the output noise cost, and the third is the process noise cost. The three terms are weighted by (Kühl et al., 2011):

$$P_L = P^{-1/2}, \qquad V = R^{-1/2}, \qquad W_k = \bar{Q}_k^{-1/2} \tag{38}$$

where the notation for the squared norm is $\|b\|^2_B = b^T B^T B b$ (Kühl et al., 2011). $P \in \mathbb{R}^{(n_x+n_\theta)\times(n_x+n_\theta)}$, $R \in \mathbb{R}^{n_y \times n_y}$ and $\bar{Q}_k \in \mathbb{R}^{(n_x+n_\theta)\times(n_x+n_\theta)}$ are respectively error, measurement noise and process noise (for the augmented system) covariance matrices. The terms $\hat{x}_L$ and $\hat{\theta}_L$ (Eq. 37) represent, instead, the optimal estimates of x_L and θ_L.

The optimal solution of the nonlinear programming (NLP) was obtained using IPOPT (Wächter and Biegler, 2006) embedded in CasADi (Andersson et al., 2019), where the continuous time model was discretized by a three point Legendre collocation on finite elements.

6. Results

In the present section we first describe the implementation and the tuning parameters (Sec. 6.1). Then we present the results obtained. Firstly, to show how the presence of structural plant-model mismatch and the absence of full state feedback can hinder the estimation results, we present the possibility to use additional constraints on the parameters to find a trade-off between state and parameter estimates (Sec. 6.2). This is shown by using the Subset Selection by Transfomation (SST) method, since one of the novelties of the work is to present SST as a more flexible regularization method in a MHE formulation. Consecutively, to show the advantages of SST, we compare the results obtained by its implementation to the ones obtained by the use of OM. To evaluate the accuracy of the estimates of biomass X and glucose S, and compare the results obtained

trough the different methods, we calculate the root mean squared error (RMSE) with respect to the reference *off-line* measurements, which were not used at any time by the estimator. In addition, to avoid that the zero values for the glucose could average out the RMSE, we calculate it separately for both batch (0-8 hours) and fed-batch phase (20-30 hours), without taking the zero values into account.

6.1. Implementation

The sampling rate was set to 1 min together with an estimation horizon T = 30 min. For the starting phase of the MHE problem, when the estimation window is smaller than the number of available measurements (i.e. $k < T$), the initial state vector was used (Kühl et al., 2011). The parameters θ, to be estimated along with the states, were appended to the augmented state vector x_a at each sampling interval. The MHE was initialized given the real values for the states and the nominal parameters (θ_0) reported in Table 1 as initial conditions:

$$x_{a,0} = \begin{bmatrix} 1.5 & 1.2 & 20 & 0 & 0.19445 & 0.007 & 0.006 & 0.42042 & 0.54308 \end{bmatrix}^T$$

The initial error covariance matrix was selected as:

$$P_0{}^+ = \begin{bmatrix} {P_{x,0}}^+ & 0 \\ 0 & {P_{\theta,0}}^+ \end{bmatrix}$$

with ${P_{x,0}}^+$ obtained, according to Schneider and Georgakis (2013), as $P_0 = diag((\hat{x}_0 - x_0)^T(\hat{x}_0 - x_0))$:

$${P_{x,0}}^+ = \begin{bmatrix} 2.09{\cdot}10^{-8} & 0 & 0 & 0 \\ 0 & 1.10{\cdot}10^{-5} & 0 & 0 \\ 0 & 0 & 1.09{\cdot}10^{-4} & 0 \\ 0 & 0 & 0 & 2.17{\cdot}10^{-5} \end{bmatrix}$$

and

$${P_{\theta,0}}^+ = \begin{bmatrix} 1.05{\cdot}10^{-11} & 0 & 0 & 0 & 0 \\ 0 & 1.54{\cdot}10^{-11} & 0 & 0 & 0 \\ 0 & 0 & 2.02{\cdot}10^{-11} & 0 & 0 \\ 0 & 0 & 0 & 1.28{\cdot}10^{-11} & 0 \\ 0 & 0 & 0 & 0 & 4.95{\cdot}10^{-12} \end{bmatrix}$$

The measurement noise covariance matrix R:

$$R = \begin{bmatrix} 10^{-2} & 0 & 0 \\ 0 & 10^{-1} & 0 \\ 0 & 0 & 10^{-3} \end{bmatrix}$$

And the process noise covariance matrix for the augmented system $\bar{Q}_k$ was tuned as in Section 3.1.1 (Eq. 11), maintaining Q_ω and Q_θ unchanged troughout the different case studies. Moreover, for the arrival cost calculation (Sec. S1), to take into account the variation between the past information and the new horizon T, a weighting matrix $\bar{W}_L \in \mathbb{R}^{(n_x+n_\theta)\times(n_x+n_\theta)}$ (Sec. S1) is defined as follows:

$$\bar{W}_L = \begin{bmatrix} W_L & 0 \\ 0 & \mathbb{I} \end{bmatrix}$$

Where $W_L \in \mathbb{R}^{n_x \times n_x}$ is defined as $W_k = Q_k^{-1/2}$ for $k = L$ and $\mathbb{I}^{n_\theta \times n_\theta}$ is the identity matrix. $\mathbb{I}$ was selected as the identity matrix so that, being always smaller than W_k, changes with respect to the past information in the states are penalized more than the ones in the parameters.

6.2. Trade-off between state and parameter estimates

To present the possibility to find a trade-off between model update and estimation accuracy, we present how the implementation of additional bounds on the parameters (i.e. inequality constraints, Eq. 37d), differently leaves the possibility to adapt the model. In this extent, we present in this section the estimated states and the effects of the bounds on the parameters (i.e. inequality constraints) using three different cases. The aim is to show how the bounds selected can influence the performance of model prediction and estimation accuracy. Table 3 reports the bounds for the three presented cases.

6.2.1. Bounds Selection

In *Case 1* the bounds on the parameters were selected to demonstrate the possibility to find a trade-off between model update and estimation accuracy (Table 3). This was achieved by carefully relaxing the standard deviations σ_{θ_i} reported in Table 1, therefore defining bounds on the parameters that leave them enough possibility to vary and adapt the model. Moreover, the bounds for Y_{XS} were kept narrower than for the other parameters because of its only dependence on the glucose (S).

Table 3: Lower (θ_{min}) and upper (θ_{max}) parameter bounds.

Case	$\theta_{min}/\theta_{max}$	μ_{max}	K_s	k_d	Y_{XS}	Y_{XCO_2}
Case 1	θ_{min}	0.15	$5 \cdot 10^{-3}$	$1 \cdot 10^{-5}$	0.3846	0.10
	θ_{max}	0.30	$8 \cdot 10^{-3}$	$7 \cdot 10^{-3}$	0.4562	2
Case 2	θ_{min}	0.1941	$6.6 \cdot 10^{-3}$	$5.6 \cdot 10^{-3}$	0.4201	0.5429
	θ_{max}	0.1948	$7.4 \cdot 10^{-3}$	$6.4 \cdot 10^{-3}$	0.4208	0.5433
Case 3	θ_{min}	0.10	$5 \cdot 10^{-3}$	$1 \cdot 10^{-5}$	0.3846	0.10
	θ_{max}	0.40	$3 \cdot 10^{-2}$	$9 \cdot 10^{-3}$	0.4562	2

Following the idea that an improvement in the model prediction can compromise the accuracy of the estimates, in *Case 2* we selected tighter bounds on the parameters (Table 3). The main idea is to allow the parameters to only vary within their uncertainty range as also reported in Kim et al. (2019), therefore reducing the penalization on the accuracy of the estimates. These bounds were also selected based on the standard deviations σ_{θ_i} reported in Table 1. However, considering that the standard deviation for the parameters was obtained by a single experiment, the values were relaxed by increasing the order of magnitude of the variance either by two times what reported in Table 1 (i.e. $\theta_0 \pm 10^2 \cdot \sigma_{\theta_i}$). This was done to take only batch-to-batch variations into account.

To show the influence and necessity of bounds to further regularize the optimization problem, in *Case 3* we selected higher variation margins for the parameters. Indeed, the bounds were defined to leave more freedom to the parameters to adapt, however, considering their necessity to avoid the deterioration of the glucose estimates. The only parameter maintained with tight bounds was Y_{XS}, due to its only dependence on S. Moreover, the reader is here warned that lower values were selected as non zero because of the presence of the logarithm in the calculation of the sensitivities (Eq. 13).

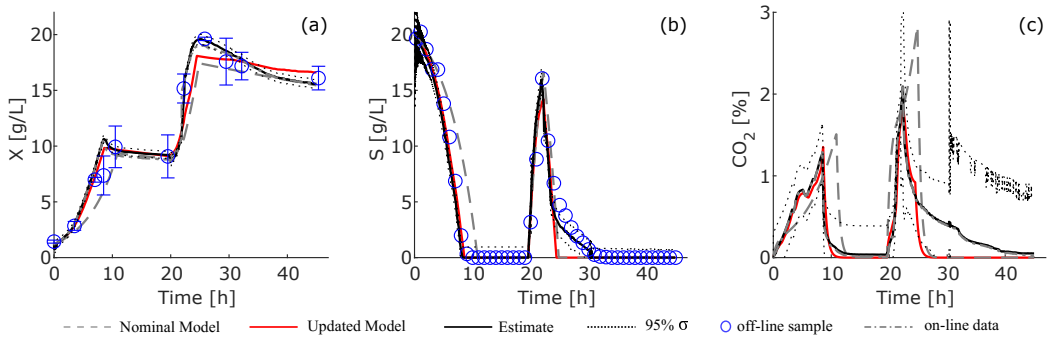


Figure 5: Subset Selection by Transformation, *Case 1* - Biomass (a), glucose (b) and CO_2 (c) compositions and their confidence intervals (95% σ). An improvement in the updated model is visible in (a) during the batch phase (0-20 hours) and during the fed-batch phase (over 20 hours). The updated model captures the sugar and the CO2 dynamics correctly until the end of the fed-batch phase (after 25 hours). *Off-line* samples are only reported as reference, but not used by the estimator.

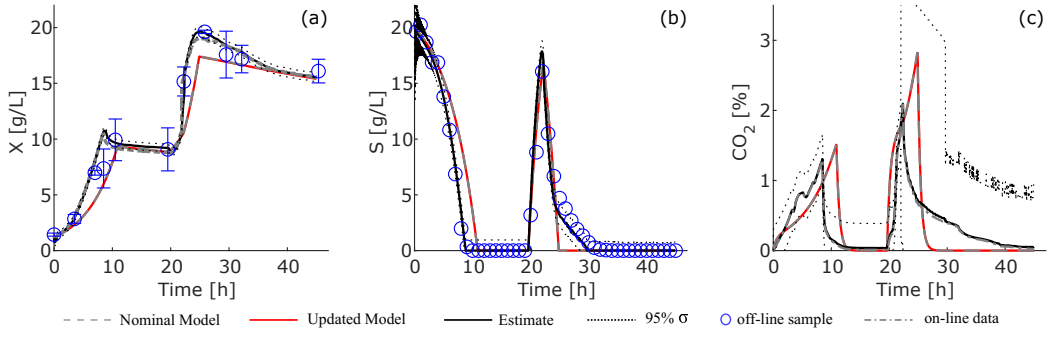


Figure 6: Subset Selection by Transformation, *Case 2* - Biomass (a), glucose (b) and CO_2 (c) compositions and their confidence intervals (95% σ). While improvement in the updated model is moderate, RMSE values show a good performance for the glucose estimate, with a value of 0.93 (b). *Off-line* samples are only reported as reference, but not used by the estimator.

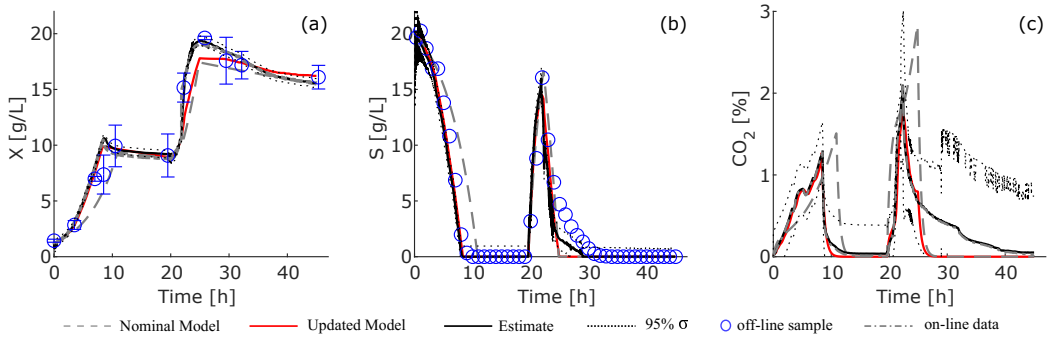


Figure 7: Subset Selection by Transformation, *Case 3* - Biomass (a), glucose (b) and CO_2 (c) compositions and confidence intervals (95% σ). These results show the dependence of the methods on the bounds, to avoid deterioration of both state and parameter estimates. In (b) it is visible how the performance of the SST method is less influenced (in comparison to OM) by the choice of the bounds. *Off-line* samples are only reported as reference, but not used by the estimator.

6.2.2. Estimation Results using Constrained Regularization

By solving the optimization problem, defined in Eq. 37, the estimates are:

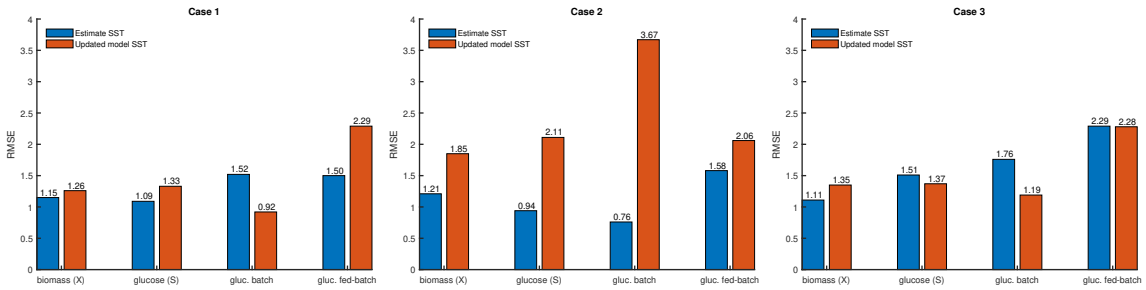


Figure 8: Subset Selection by Transformation - RMSE values calculated for both the updated model and the estimated states with respect to the reference *off-line* values of biomass (X) and glucose (S). RMSE for glucose is also calculated for the single phases (batch and fed-batch) without taking the zero values into account.

- the *estimated states* ($\hat{x}$) presented in Fig. 5, 6 and 7;
- the *estimated parameters* ($\hat{\theta}$, reported in S2), considered as additional states and estimated along with the states.

In addition, the results presented in Fig. 5, 6 and 7 report:

- the *on-line data* (y) for X (y_X) and CO_2 (y_{CO_2}) which are the *on-line* measurements used by the estimator;
- *off-line samples* (y^*) for X ($y_X{}^*$) and S ($y_S{}^*$) which are only reported as reference values to validate the accuracy of the estimates and used to calculate the root mean squared error (RMSE), but not used at any time by the estimator;
- *confidence intervals* for the estimates (95% σ), obtained from the diagonal elements of the error covariance matrix P.

This results are obtained by the implementation of the MHE with SST as regularization method. To evaluate their accuracy, the RMSE (Fig. 8) is calculated with respect to the *off-line* measurements (y^*) of biomass X ($y_X{}^*$) and substrate S ($y_S{}^*$), reported by blue dots in Fig. 5a and b, as follows:

$$RMSE_{y^*} = \sqrt{\frac{1}{n_{y^*}} \sum_{j=1}^{n_{y^*}} (y^* - \hat{x})^2} \tag{39}$$

Additionally, to show the limitation of the model on describing the dynamic behaviour of the state variables (x), due to the plant-model mismatch (particularly visible during the fed-batch phase), we report in Fig. 5, 6 and 7 the model simulations for:

- the *nominal model* (i.e. open loop model), simulated by using the nominal parameters (θ_0), reported in Table 1;
- the *updated model*, simulated a posteriori, using the estimated parameters ($\hat{\theta}$), reported in S2.

Firstly, to give the reader a reference of the accuracy on the estimates obtained by simultaneously esimating states and parameters, we present the RMSE obtained by solely estimating the states, as previously presented in Tuveri et al. (2022) (Table 4).

Table 4: RMSE values calculated for both the nominal model and the estimated states with respect to the reference *off-line* values of biomass (X) and glucose (S) based on the results presented in Tuveri et al. (2022). RMSE for glucose is also calculated for the single phases (batch and fed-batch) without taking the zero values into account. The values of the parameters for the nominal model are reported in Table 1.

	Biomass	**Glucose**		
RMSE	tot.	tot.	Batch	Fed-batch
Estimate	1.23	1.01	0.78	1.71
Nom. Model	1.83	2.11	3.68	2.05

Relatively accurate and comparable estimates were found for the biomass, with RMSE of 1.15 (*Case 1*, Fig. 5a), 1.21 (*Case 2*, Fig. 6a) and 1.11 (*Case 3*, Fig. 7a), presenting respectively 5%, 3% and 10% improvement with respect to solely estimating the states (Table 4). The estimates of glucose present instead a bigger difference, resulting in a RMSE of 1.09 for *Case 1*, 0.94 for *Case 2* and 1.53 for *Case 3*. This therefore results in the sole improvement of *Case 2*, which improves the estimate of the glucose by 7% with respect to solely estimating the states (Table 4), while *Case 1* and *Case 3* penalize it by 8% and 51% respectively. Moreover, the RMSE of the transient phases only (duration in which sugar values are not zero) shows that the error varies between 1.50 and 1.52 for *Case 1*, 0.76 to 1.58 for *Case 2* and 1.76 to 2.36 for *Case 3* (Fig. 8). Therefore, while *Case 2* shows an overall higher accuracy on the glucose estimates, *Case 1* presents a more balanced error between batch and fed-batch phase and *Case 3* presents the highest RMSE.

Effect on the unmeasured states. The results show that *Case 2*, by presenting tighter bounds, yields to more accurate estimate for the glucose (Fig. 6b). Due to the fact that both states and parameters are estimated in absence of direct glucose measurements, it is reasonable that the estimates will be a compromise between improved model prediction capabilities and accurate estimates. Indeed, pa-

rameters related to the glucose dynamics (i.e. μ_{max}, K_s and Y_{XS}, Eq. 1) are dependant on the available information on the glucose. More specifically, Y_{XS}, which is only dependent on the glucose, is the most affected. Therefore, tighter bounds on these parameters lead to more accurate glucose estimates.

Trade-off between variance and bias. When comparing the results obtained in the three different cases, it is possible to see that the unmeasured state (i.e. glucose, S) presents an higher bias in *Case 1* and *Case 3* (Fig. 5b and 7b) with respect to *Case 2* (Fig. 6b). However, as mostly visible for the CO_2, *Case 2* presents higher variance in the estimates (Fig. 6c) with respect to the other two. Therefore, this interestingly shows how the combined estimation of state and parameters presents the necessity to find a trade-off between reducing variance and bias.

Effect of structural model mismatch. The updated model, calculated with the newly estimated parameters, yields to good predictions in *Case 1* ($RMSE_X = 1.26$ and $RMSE_S = 1.33$) and *Case 3* ($RMSE_X = 1.35$ and $RMSE_S = 1.38$), especially for the batch phase, presenting respectively an RMSE of 0.92 (Fig. 5b) and 1.18 (Fig. 7b) for the glucose. Differently, *Case 2*, having tighter parameters bounds, does not presents such improvement in the predictions as the other two cases ($RMSE_X = 1.84$ and $RMSE_S = 2.11$, Fig. 6). Instead it presents a slight penalization of the model prediction of less than 1% (Table 4). However, the improvement is limited in all the cases, due to the lack of knowledge on the process dynamics (i.e. structural plant-model mismatch, 25-30 hours Figures 5,6 and 7). It is sensible to understand that the improvement can not overpass the limitations of the model in use. In fact, the RMSE for the updated model, shows an increase of almost 40% between batch and fed-batch phases (Fig. 8), due to the higher structural plant-model mismatch in the latter. The updated parameters values are reported in Sec. S2.

6.3. Subset Selection by Transformation as an alternative to the Orthogonalization Method

To show the advantages of implementing SST in an MHE formulation, we compare the results previously presented (Sec. 6.2.2) to the ones obtained by the implementation of OM (Figures 9, 10 and 11).

Case 1. The estimates for biomass obtained by SST and OM are comparable, with RMSE of 1.15 for SST (Fig. 8) and 1.16 for OM (Fig. 12). The estimates of glucose present instead a difference, with a 18% increase in RMSE for OM (1.29) with respect to SST (RMSE 1.09). Moreover, the difference increases when looking only at the transient phases (1.66-1.96 for OM and 1.50-1.52 for SST). Additionally, also the updated model presents a 30% improvement in the predictions by SST (RMSE 0.92) with respect to OM (RMSE 1.26).

Case 2. The RMSE values (Figures 8 and 12) show similar results, especially for the glucose estimates (RMSE 0.94 for SST and 0.91 for OM). These results show that, when adopting tighter bounds on the parameters, the performance of the two methods is similar. However, it is important to note that while OM presents a slightly better improved RMSE for the glucose estimate (0.91) while presenting an RMSE for the biomass estimates of 1.22, the SST presents a more pronounced improvement in both of them (0.94 and 1.21) with respect to the case when only estimating the states (Table 4).

Case 3. The results obtained in this case (Fig. 11) show that OM is more influenced than SST (Fig. 7) by the selection of the bounds. Indeed, the RMSE calculated for the glucose estimates and the updated model in the batch phase are respectively 15 and 20% higher for OM (Fig. 12).

These results show us how both methods necessitate bounds to limit the possibility of deteriorating both state and parameter estimates. However, they also show that SST, by yielding lower RMSE values under looser bounds, presents less dependence on the bounds selection.

7. Discussion

In this work we present the use of two different methods, namely Orthogonalization Method (OM) and Subset Selection by Transformation (SST), for combined state and parameter estimation in a MHE framework, where the estimates were obtained based on the knowledge of the *on-line* measurements of volume (V), biomass (X) and carbon dioxide (CO_2).

One of the main contributions of this work is to apply MHE as a state estimator, to an experimental data-set, to estimate, in addition to the states, the model parameters. This is done to account for model uncertainty in the experimental data-set. However, as visible when comparing the RMSE values (Fig. 8), the more uncertain the model parameters are (i.e. model uncertainty), the larger is the estimation bias of the unmeasured states. This is consistent with what is stated in Dochain (2003). Moreover, when including all the parameters, an ill-posed problem arises. To alleviate that, we implemented and compared two regularization methods (i.e. OM and SST). Therefore taking practical identifiability into consideration (McLean and McAuley (2012); Chis et al. (2016); Kim and Lee (2019)). The advantage of this implementation is to enable the selection of different subsets of parameters during the process, to avoid overfitting when adapting them. Another novelty of this work, therefore consists in the implementation of the SST method in an MHE and its comparison to the performance with a state of the art method as the OM. Additionally, an important aspect of this work is also the introduction of a stopping criterion based on the structural identifiability of the system (Villaverde, 2019). This bypasses the need to define an heuristic threshold value for the selection of the number of active constraints, therefore leading to a more consistent model adaptation. Moreover, the choice of this stopping criterion allows both methods to have a comparable way to select the available degrees of freedom. Additional stopping criteria are available in

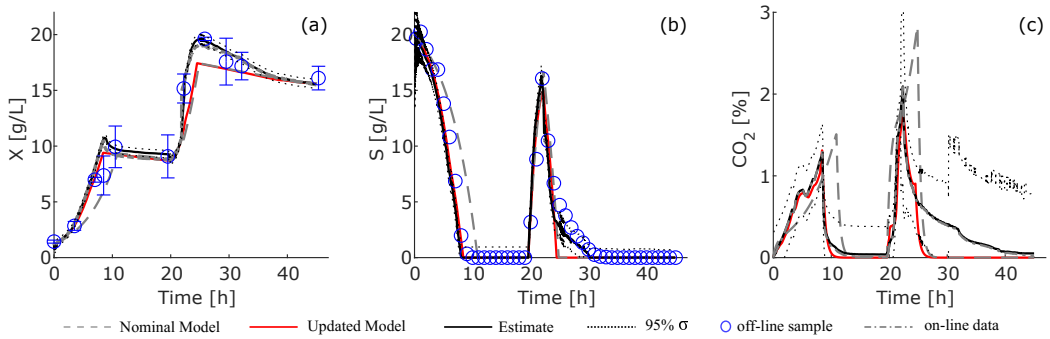


Figure 9: Orthogonalization Method, Case 1 - Biomass (a), glucose (b) and CO_2 (c) compositions, and their confidence intervals (95% σ). An improvement in the updated model is visible in (a) in the batch (0-20 hours) while less pronounced in the feeding phase (25-30 hours). Similarly, the parameter adaptation could not yield to a glucose model (b) better than the nominal during the feeding phase. In contrast, the updated model for the CO_2 (c) could compensate better for the first 25 hours. *Off-line* samples are only reported as reference, but not used by the estimator.

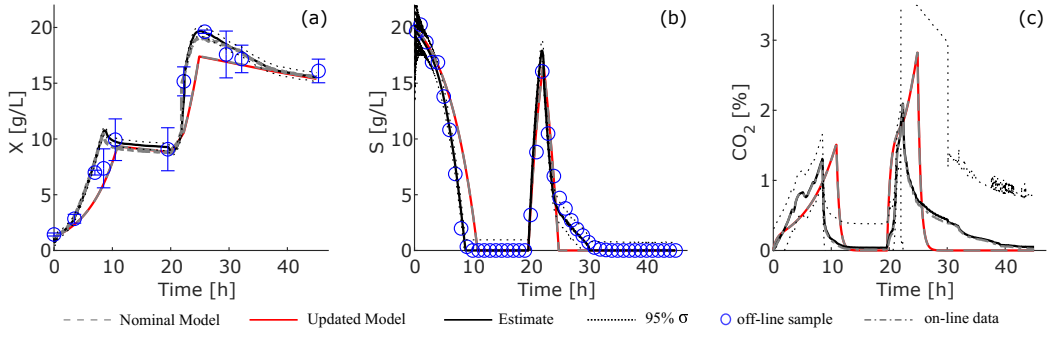


Figure 10: Orthogonalization Method, Case 2 - Biomass (a), glucose (b) and CO_2 (c) compositions and their confidence intervals (95% σ). While the improvement in the updated model is moderate, RMSE values show a good performance for the glucose estimate, with a value of 0.96 (b). *Off-line* samples are only reported as reference, but not used by the estimator.

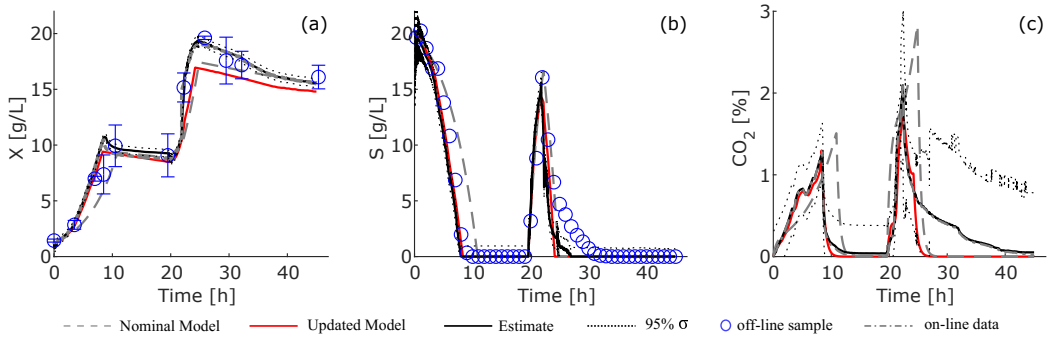


Figure 11: Orthogonalization Method, Case 3 - Biomass (a), glucose (b) and CO_2 (c) compositions and their confidence intervals (95% σ). These results show the dependence of the methods on bounds to avoid deterioration of both states and parameter estimates. This is mostly noticeable in (b). *Off-line* samples are only reported as reference, but not used by the estimator.

literature, like the definition of a threshold as in Nakama et al. (2020) and Bae et al. (2021), or the minimization of the Mean Squared Error (MSE) proposed by Chu et al. (2009); Wu et al. (2011) and applied in Kim and Lee (2019)

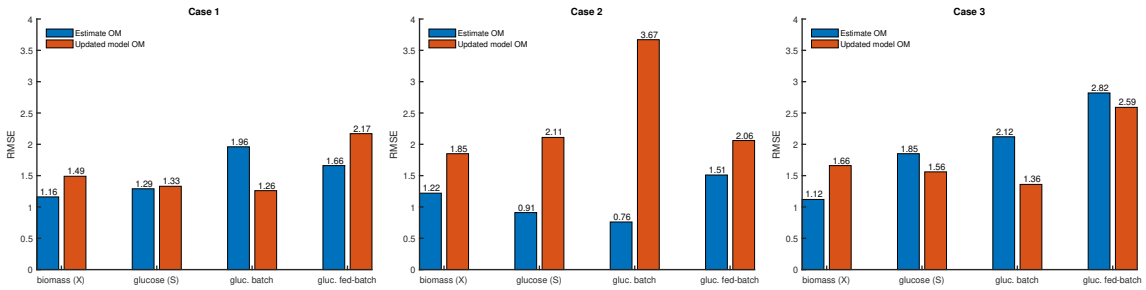


Figure 12: Orthogonalization Method - RMSE values calculated for both the updated model and the estimated states with respect to the reference *off-line* values of biomass (X) and glucose (S). RMSE for glucose is also calculated for the single phases (batch and fed-batch) without taking the zero values into account.

and Chen et al. (2022). Indeed, the comparison between the two regularization methods, when the stopping criteria is a threshold, would be purely based on trial and error tuning of this threshold, and will not necessarily tell us about how effective the two regularizations are. We did not compare the proposed criteria with the minimization of the MSE because, due to the absence of full-state feedback (i.e. measurements of the substrate S is missing) and therefore the lack of information on all the state variables, an *on-line* selection of the degrees of freedom would not be possible.

Our work presents the possibility to estimate glucose by *on-line* measurements of other states (i.e. volume, biomass and CO_2), whereas the estimation of the internal states (i.e. parameters) was not the main focus, the parameters were estimated in order to adapt the model uncertainty. Indeed, the idea is that parameters should adapt in order to consider batch-to-batch variation, aiming to obtain more accurate state estimates (i.e. lower RMSE with respect to the off-line samples y^*). Consistent with Bae et al. (2021), simultaneous estimation of states and parameters improves the estimates of the variables of interest.

The results obtained for *Case 1* in Fig. 5 and Fig. 9 indicate that the SST method yields improved accuracy over the OM method, for both glucose estimate and model prediction capabilities. This improved accuracy (see Sec. 6.3) is a consequence of the different parameter estimation strategies. Indeed, in all the cases OM allows to vary a maximum of three parameters per iteration (see Sec. S2). In contrast, in SST all the parameters can potentially vary, within the clusters. Therefore, enabling higher flexibility for model adaptation. This is in agreement with what is reported in Kim and Lee (2019), Nakama et al. (2020) and Chen et al. (2022), stating that transformation of the constraints provides more flexibility to the regularization. Additionally, the results of *Case 2* indicate that tighter bounds on the model parameters yield better accuracy of the glucose estimates, both for SST (RMSE 0.94) and OM (RMSE 0.91), reducing the effect of the bias. Therefore presenting a good compromise for the glucose estimates (Fig. 6b and Fig. 10b). Lastly, *Case 3* supports these results, since by relaxing the bounds the estimates exhibited deterioration in both states and parameters. Therefore showing that both methods necessitate additional bounds (i.e. inequality constraints). This necessity emerges mainly for two reasons: 1) The discontinuities in the derivatives, due to sudden changes in the process dynamics (e.g. 8, 20 and 23 hours, Fig. 7), which cause the parameters to hit the bounds, and 2) The absence of knowledge about the dynamics of the process (i.e. structural plant-model mismatch), which causes limitations on the model improvement. However, the results presented in *Case 3* show that SST (Fig. 8), by maintaining a fixed relationship between the parameters, presents less dependence on the bounds in comparison to OM (Fig. 12), where the bounds are the only limitation on the parameters. Since, as also reported in Kim et al. (2019), the choice of the bounds reflects the level of uncertainty in the parameters, their selection can be based on prior knowledge or historical data (i.e. $\theta_0 \pm \alpha \cdot \sigma_\theta$) depending on the level of conservativeness desired (i.e. *Case 2*) or by certain physical considerations (i.e. *Case 1* and *Case 3*).

At this point, it is possible to discuss the differences between OM and SST. In fact, the different selection of constraints on the parameters will change the search region of the optimization problem (i.e. differently regularise the problem, Fig. 1). For instance, SST will use a different orthonormal base for the constraint selection, in comparison to the OM, based on the eigenvectors of the approximated Hessian, creating clusters that maintain the same relation between the parameters as in the nominal case (θ_0). Conversely, OM will yield solutions that maintain singular parameters at nominal values (θ_0). Therefore, while SST allows more parameters to vary (i.e. clusters) but with fixed relationships among them, OM allows fewer parameters to vary freely (Fig. 1), and is therefore a less flexible approach. In other words, while OM reduces the number of decision variables, SST reduces their search region.

The optimization problem finds local optimal estimates (of states x and parameters θ) given the available information (i.e. model, measurements and constraints). This strongly depends on the available measurements and the regularization method selected, and therefore on the given constraints that intrinsically carry information (Psichogios

and Ungar, 1992). Indeed, due to the absence of multi-rate full state feedback, also the unmeasured states are affected, potentially presenting a large bias in the estimates (Dochain, 2003). Under this scenario, the available information over a finite horizon is limited. As a consequence, the model is likely to overfit the data, therefore resulting in poor model predictions (Bae et al., 2021). Additionally, as reported in Bonné and Jørgensen (2001), the use of regularization results in biased estimates, becoming a trade-off between reducing variance and bias. As shown in the results, *Case 1* and *Case 3* present an higher bias (Fig. 5b and 7b) but a lower variance (Figures 5c and 7c) compared to *Case 2* (Fig. 6b and c). It is therefore important to consider that, given the total absence of information on the sugar measurements, limitations on the parameter estimation, and consequently on the prediction capability of the model will arise. In fact, as visible in the comparison of the three cases presented (Sec. 6.2.2), a high variation in the parameters related to the sugar consumption (i.e. μ_{max}, K_S and Y_{XS}), although could improve model prediction, would decrease the accuracy of the sugar estimates. Because these estimates are dependent on the tuning of the weights in the optimization (i.e. MHE), it is reasonable to assume that one possible solution is an iterative correction of the weight Q_k for the model error cost (Dochain, 2003), to account for improvement in the model prediction. This, which is also in agreement with the work of Kim et al. (2019), where the authors state that, to avoid biased estimates when unanticipated model mismatch occurs, significant a priori knowledge about the structural model mismatch is necessary, will be further investigated in future works.

Finally, it is important mentioning that the results obtained by SST are satisfactory, given the simplicity of the model and its limitations. For instance, either the estimates or the model predictions can be improved, without the guarantee of an accurate tracking of the drifting model parameters (Bae et al., 2021). It can therefore be considered that the improvements in the estimates are due to the reduction of the effect of the uncertainty caused by plant-model mismatch and unmodeled disturbances (Kim et al., 2019), rather than the accurate estimation of model parameters (Bae et al., 2021). Thus, the parameters can only improve the model accuracy but they can not compensate for the lack of knowledge about its dynamics. In fact, it is clear from Figures 5a, 6a and 7a that the model is unable to properly describe the transition between growth phase and steady state for the biomass, especially after feeding. This applies also for the transient phase of the glucose (25-30 hours, Figures 5b, 6b and 7b). Indeed, despite the complexity of biological processes (Becker and Wittmann, 2012), bioprocesses are usually represented by oversimplified, unstructured Monod models. However, they are based on very simplistic representations of the cellular metabolism, by involving lumped parameters for the description of the intracellular phenomena (Jabarivelisdeh et al., 2020). To further illustrate that, we refer the reader to Section S3, which shows the difference between batch and fed-batch phases by the same model.

A possible solution to overcome this limitation is the use of hybrid (grey-box) models (Zendehboudi et al., 2018; Narayanan et al., 2019; Boiroux et al., 2019; Bradford et al., 2021). Some pioeneering works on the use of hybrid models to compensate for the lack of knowledge on the process dynamics can be found in process system engineering applications (Johansen and Foss, 1992), and more specifically for fed-batch bioreactors in Psichogios and Ungar (1992). As stated in Boiroux et al. (2019), hybrid models attempt to combine the advantages of first principle (i.e. white-box) and black-box (Narayanan et al., 2019) or Gaussian Process models (Bradford et al., 2021), by using the synergy between them (Narayanan et al., 2019). Indeed, in the case of absence of complete knowledge of the cell metabolism (i.e. lack of knowledge on the dynamics of the process), the hybrid model can correctly follow the physics of the process and therefore exhibit improved interpolation and extrapolation capabilities (Narayanan et al., 2019). These results therefore provide the opportunity to exploit the advantages of SST in the case of implementation of hybrid models.

8. Conclusions

This work presents the simultaneous estimation of states and parameters by an MHE in a microbial experimental fed-batch process. We here presented the comparison of two different regularization methods implemented in an MHE, for the selection of the additional decision variables (i.e. parameters or clusters). In addition, we proposed a stopping criterion based on structural identifiability to avoid the selection of cut-off values for the constraints selection. The results also present how, under presence of high model-mismatch, the necessity of additional inequality constraints (i.e. bounds on the parameters) is necessary when no full state feedback is available. Although under certain limitations, the results present accurate estimates and the possibility to adapt the model *on-line*.

CRediT authorship and contribution statement

Andrea Tuveri: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing and Visualization. **Caroline S.M. Nakama**: Conceptualization, Resources, Writing - Review & Editing, Visualization and Supervision. **José Matias**: Conceptualization, Writing - Review & Editing, Visualization and Supervision. **Haakon Eng Holck**: Software and Writing - Review & Editing. **Johannes Jäschke**: Writing - Review & Editing, Visualization and Funding acquisition. **Lars Imsland**: Conceptualization, Resources, Writing - Review & Editing, Visualization and Supervision. **Nadav Bar**:

Conceptualization, Writing - Review & Editing, Visualization, Supervision, Project administration and Funding acquisition.

Declaration of competing interest

The authors declare to not have known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

A. Tuveri is greatful to the colleagues Allyne M. dos Santos, Joakim R. Andersen, Rafael D. de Oliveira and Simen Bjorvand for useful discussion. The project was funded by the Bio Based Industries Joint Undertaking (JU) under the European Union's Horizon 2020 research and innovation programme under grant agreement N^o 790507. J. Matias acknowledges financial support from the Norwegian Research Council, SUBPRO, grant number: 237893. C. S. M. Nakama acknowledges financial support from the Norwegian Research Council through FRIPRO Project SensPATH.

References

Anane, E., López C, D.C., Barz, T., Sin, G., Gernaey, K.V., Neubauer, P., Cruz Bournazou, M.N., 2019. Output uncertainty of dynamic growth models: Effect of uncertain parameter estimates on model reliability. Biochemical Engineering Journal 150, 107247. doi:10.1016/j.bej.2019.107247.

Andersson, J.A.E., Gillis, J., Horn, G., Rawlings, J.B., Diehl, M., 2019. CasADi: a software framework for nonlinear optimization and optimal control. Mathematical Programming Computation 11, 1–36. doi:10.1007/s12532-018-0139-4.

Andersson, L.E., Scibilia, F., Imsland, L., 2016. An estimation-forecast set-up for iceberg drift prediction. Cold Regions Science and Technology 131, 88–107. doi:10.1016/j.coldregions.2016.08.001.

Bae, J., Kim, Y., Lee, J.M., 2021. Multirate moving horizon estimation combined with parameter subset selection. Computers & Chemical Engineering 147, 107253. doi:10.1016/j.compchemeng.2021.107253.

Balsa-Canto, E., Rodriguez-Fernandez, M., Banga, J.R., 2007. Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation. Journal of Food Engineering 82, 178–188. doi:10.1016/j.jfoodeng.2007.02.006.

Becker, J., Wittmann, C., 2012. Bio-based production of chemicals, materials and fuels - Corynebacterium glutamicum as versatile cell factory. Current Opinion in Biotechnology 23, 631–640. doi:10.1016/j.copbio.2011.11.012.

Boiroux, D., Mahmoudi, Z., Jorgensen, J.B., 2019. Parameter Estimation in Type 1 Diabetes Models for Model-Based Control Applications, in: 2019 American Control Conference (ACC), IEEE. pp. 4112–4117. doi:10.23919/ACC.2019.8814933.

Bonné, D., Jørgensen, S.B., 2001. Batch to Batch Improving Control of Yeast Fermentation, in: Computer Aided Chemical Engineering. Elsevier. volume 9, pp. 621–626. doi:10.1016/S1570-7946(01)80098-5.

Bonvin, D., Srinivasan, B., Ruppen, D., 2001. Dynamic optimization in the batch chemical industry. Technical Report.

Bradford, E., Imsland, L., Reble, M., del Rio-Chanona, E.A., 2021. Hybrid Gaussian Process Modeling Applied to Economic Stochastic Model Predictive Control of Batch Processes, in: Faulwasser, T., Müller, M.A., Worthmann, K. (Eds.), Recent Advances in Model Predictive Control. Springer International Publishing, Cham, pp. 191–218. doi:10.1007/978-3-030-63281-6{_}8.

Chen, W., Wang, B., Biegler, L.T., 2022. Parameter estimation with improved model prediction for over-parametrized nonlinear systems. Computers & Chemical Engineering 157, 107601. doi:10.1016/j.compchemeng.2021.107601.

Chis, O.T., Villaverde, A.F., Banga, J.R., Balsa-Canto, E., 2016. On the relationship between sloppiness and identifiability. Mathematical Biosciences 282, 147–161. doi:10.1016/j.mbs.2016.10.009.

Chu, Y., Huang, Z., Hahn, J., 2009. Improving prediction capabilities of complex dynamic models via parameter selection and estimation. Chemical Engineering Science 64, 4178–4185. doi:10.1016/j.ces.2009.06.057.

Dochain, D., 2003. State and parameter estimation in chemical and biochemical processes: a tutorial. Journal of Process Control 13, 801–818. doi:10.1016/S0959-1524(03)00026-X.

Doyle III, F.J., 1998. Nonlinear inferential control for process applications. Journal of Process Control 8, 339–353. doi:10.1016/S0959-1524(98)00015-8.

Elsheikh, M., Hille, R., Tatulea-Codrean, A., Krämer, S., 2021. A comparative review of multi-rate moving horizon estimation schemes for bioprocess applications. Computers & Chemical Engineering 146, 107219. doi:10.1016/j.compchemeng.2020.107219.

Gatzke, E.P., Doyle III, F.J., 2002. Use of multiple models and qualitative knowledge for on-line moving horizon disturbance estimation and fault diagnosis. Journal of Process Control 12, 339–352. doi:10.1016/S0959-1524(01)00037-3.

Goffaux, G., Wouwer, A.V., 2008. Design of a robust nonlinear receding-horizon observer-Application to a biological system. IFAC Proceedings Volumes 41, 15553–15558.

Grossmann, I.E., Sargent, R.W.H., 1978. Optimum design of chemical plants with uncertain parameters. AIChE Journal 24, 1021–1028. doi:10.1002/aic.690240612.

Hermann, R., Krener, A., 1977. Nonlinear controllability and observability. IEEE Transactions on Automatic Control 22, 728–740. doi:10.1109/TAC.1977.1101601.

Hernández Rodríguez, T., Posch, C., Pörtner, R., Frahm, B., 2021. Dynamic parameter estimation and prediction over consecutive scales, based on moving horizon estimation: applied to an industrial cell culture seed train. Bioprocess and Biosystems Engineering 44, 793–808. doi:10.1007/s00449-020-02488-1.

Isidori, A., 1985. Nonlinear control systems: an introduction. Springer.

Jabarivelisdeh, B., Carius, L., Findeisen, R., Waldherr, S., 2020. Adaptive predictive control of bioprocesses with constraint-based modeling and estimation. Computers & Chemical Engineering 135, 106744. doi:10.1016/j.compchemeng.2020.106744.

Johansen, T.A., Foss, B.A., 1992. Representing and Learning Unmodeled Dynamics with Neural Network Memories, in: 1992 American Control Conference, IEEE. pp. 3037–3043. doi:10.23919/ACC.1992.4792705.

Kim, B., Huusom, J.K., Lee, J.H., 2019. Robust Batch-to-Batch Optimization with Scenario Adaptation. Industrial & Engineering Chemistry Research 58, 13664–13674. doi:10.1021/acs.iecr.8b06233.

Kim, B., Lee, J.H., 2019. Parameter subset selection and biased estimation for a class of ill-conditioned estimation problems. Journal of Process Control 81, 65–75. doi:10.1016/j.jprocont.2019.05.015.

Kolås, S., Foss, B.A., Schei, T.S., 2009. Noise modeling concepts in nonlinear state estimation. Journal of Process Control 19, 1111–1125. doi:10.1016/j.jprocont.2009.03.002.

Kravaris, C., Hahn, J., Chu, Y., 2013. Advances and selected recent developments in state and parameter estimation. Computers & Chemical Engineering 51, 111–123. doi:10.1016/j.compchemeng.2012.06.001.

Kühl, P., Diehl, M., Kraus, T., Schlöder, J.P., Bock, H.G., 2011. A real-time algorithm for moving horizon state and parameter

estimation. Computers & Chemical Engineering 35, 71–83. doi:10.1016/j.compchemeng.2010.07.012.

Küpper, A., Diehl, M., Schlöder, J.P., Bock, H.G., Engell, S., 2009. Efficient moving horizon state and parameter estimation for SMB processes. Journal of Process Control 19, 785–802. doi:10.1016/j.jprocont.2008.10.004.

Küpper, A., Wirsching, L., Diehl, M., Schlöder, J.P., Bock, H.G., Engell, S., 2010. Online identification of adsorption isotherms in SMB processes via efficient moving horizon state and parameter estimation. Computers & Chemical Engineering 34, 1969–1983. doi:10.1016/j.compchemeng.2010.07.005.

Leu, G., Baratti, R., 2000. An extended Kalman filtering approach with a criterion to set its tuning parameters; application to a catalytic reactor. Computers & Chemical Engineering 23, 1839–1849. doi:10.1016/S0098-1354(00)00298-2.

Liu, J., Gnanasekar, A., Zhang, Y., Bo, S., Liu, J., Hu, J., Zou, T., 2021. Simultaneous State and Parameter Estimation: The Role of Sensitivity Analysis. Industrial & Engineering Chemistry Research 60, 2971–2982. doi:10.1021/acs.iecr.0c03793.

López C., D.C., Barz, T., Körkel, S., Wozny, G., 2015. Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. Computers & Chemical Engineering 77, 24–42. doi:10.1016/j.compchemeng.2015.03.002.

Lund, B.F., Foss, B.A., 2008. Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models. Automatica 44, 278–281. doi:10.1016/j.automatica.2007.04.006.

Marchetti, A., Chachuat, B., Bonvin, D., 2008. Real-time optimization via adaptation and control of the constraints, in: Computer Aided Chemical Engineering. Elsevier. volume 25, pp. 393–398. doi:10.1016/S1570-7946(08)80070-3.

McLean, K.A.P., McAuley, K.B., 2012. Mathematical modelling of chemical processes-obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. The Canadian Journal of Chemical Engineering 90, 351–366. doi:10.1002/cjce.20660.

Mohd Ali, J., Ha Hoang, N., Hussain, M., Dochain, D., 2015. Review and classification of recent observers applied in chemical process systems. Computers & Chemical Engineering 76, 27–41. doi:10.1016/j.compchemeng.2015.01.019.

Monod, J., 1949. The growth of bacterial cultures. Annual Review of Microbiology 3, 371–394. doi:10.1146/annurev.mi.03.100149.002103.

Nakama, C.S., Le Roux, G.A., Zavala, V.M., 2020. Optimal constraint-based regularization for parameter estimation problems. Computers & Chemical Engineering 139, 106873. doi:10.1016/j.compchemeng.2020.106873.

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

Nijmeijer, H., van der Schaft, A., 1990. Nonlinear Dynamical Control Systems. volume 175. Springer New York, New York, NY. doi:10.1007/978-1-4757-2101-0.

Powel, N.D., Morgansen, K.A., 2015. Empirical observability Gramian rank condition for weak observability of nonlinear systems with control, in: 2015 54th IEEE Conference on Decision and Control (CDC), IEEE. pp. 6342–6348. doi:10.1109/CDC.2015.7403218.

Psichogios, D.C., Ungar, L.H., 1992. A hybrid neural network-first principles approach to process modeling. AIChE Journal 38, 1499–1511. doi:10.1002/aic.690381003.

Raïssi, T., Ramdani, N., Candau, Y., 2005. Bounded error moving horizon state estimator for non-linear continuous-time systems: application to a bioprocess system. Journal of Process Control 15, 537–545. doi:10.1016/j.jprocont.2004.10.002.

Rao, C.V., 2000. Moving horizon strategies for the constrained monitoring and control of nonlinear discrete-time systems. The University of Wisconsin-Madison.

Rao, C.V., Rawlings, J.B., Mayne, D.Q., 2003. Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations. IEEE Transactions on Automatic Control 48, 246–258. doi:10.1109/TAC.2002.808470.

Robertson, D.G., Lee, J.H., Rawlings, J.B., 1996. A moving horizon-based approach for least-squares estimation. AIChE Journal 42, 2209–2224. doi:10.1002/aic.690420811.

Schei, T.S., 2008. On-line estimation for process control and optimization applications. Journal of Process Control 18, 821–828. doi:10.1016/j.jprocont.2008.06.014.

Schneider, R., Georgakis, C., 2013. How to NOT make the extended kalman filter fail. Industrial and Engineering Chemistry Research 52, 3354–3362. doi:10.1021/ie300415d.

Strang, G., 2016. Introduction to Linear Algebra. (5th edn). Wellesley-Cambridge Press, revised , 206–240.

Thompson, D.E., McAuley, K.B., McLellan, P.J., 2009. Parameter Estimation in a Simplified MWD Model for HDPE Produced by a Ziegler-Natta Catalyst. Macromolecular Reaction Engineering 3, 160–177. doi:10.1002/mren.200800052.

Tuveri, A., Holck, H.E., Nakama, C.S., Matias, J., Jäschke, J., Imsland, L., Bar, N., 2022. Bioprocess Monitoring: A Moving Horizon Estimation Experimental Application. IFAC-PapersOnLine 55, 222–227. doi:10.1016/j.ifacol.2022.07.448.

Tuveri, A., Pérez-García, F., Lira-Parada, P.A., Imsland, L., Bar, N., 2021. Sensor fusion based on Extended and Unscented Kalman Filter for bioprocess monitoring. Journal of Process Control 106, 195–207. doi:10.1016/j.jprocont.2021.09.005.

Vajda, S., Rabitz, H., Walter, E., LeCourtier, Y., 1989. Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. Chemical Engineering Communications 83, 191–219. doi:10.1080/00986448908940662.

Vajda, S., Valko, P., Turányi, T., 1985. Principal component analysis of kinetic models. International Journal of Chemical Kinetics 17, 55–81. doi:10.1002/kin.550170107.

Valipour, M., Ricardez-Sandoval, L.A., 2021. Assessing the Impact of EKF as the Arrival Cost in the Moving Horizon Estimation under Nonlinear Model Predictive Control. Industrial & Engineering Chemistry Research 60, 2994–3012. doi:10.1021/acs.iecr.0c06095.

Valipour, M., Ricardez-Sandoval, L.A., 2022. A robust moving horizon estimation under unknown distributions of process or measurement noises. Computers & Chemical Engineering 157, 107620. doi:10.1016/j.compchemeng.2021.107620.

Valipour, M., Ricardez-Sandoval, L.A., 2022. Extended moving horizon estimation for chemical processes under non-Gaussian noises. AIChE Journal 68, e17545. doi:10.1002/aic.17545.

Vercammen, D., Logist, F., Impe, J.V., 2016. Online moving horizon estimation of fluxes in metabolic reaction networks. Journal of Process Control 37, 1–20. doi:10.1016/j.jprocont.2015.08.014.

Villaverde, A.F., 2019. Observability and Structural Identifiability of Nonlinear Biological Systems. Complexity 2019, 1–12. doi:10.1155/2019/8497093.

Wächter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming 106, 25–57. doi:10.1007/s10107-004-0559-y.

Wu, S., McLean, K.A., Harris, T.J., McAuley, K.B., 2011. Selection of optimal parameter set using estimability analysis and MSE-based model-selection criterion. International Journal of Advanced Mechatronic Systems 3, 188. doi:10.1504/IJAMECHS.2011.042615.

Yao, K.Z., Shaw, B.M., Kou, B., McAuley, K.B., Bacon, D.W., 2003. Modeling Ethylene/Butene Copolymerization with Multi-site Catalysts: Parameter Estimability and Experimental Design. Polymer Reaction Engineering 11, 563–588. doi:10.1081/PRE-120024426.

Zendehboudi, S., Rezaei, N., Lohi, A., 2018. Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. Applied Energy 228, 2539–2566. doi:10.1016/j.apenergy.2018.06.051.

Supplementary material

This article contains a supplemetary document.

Supplementary Document for: A Regularized Moving Horizon Estimator for Combined State and Parameter Estimation in a Bioprocess Experimental Application

Andrea Tuveri[a], Caroline S.M. Nakama[a], José Matias[a], Haakon Eng Holck[a], Johannes Jäschke[a], Lars Imsland[b], Nadav Bar[a,*]

[a]*Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway*
[b]*Department of Engineering Cybernetichs, Norwegian University of Science and Technology (NTNU), Trondheim, Norway*

Contents

S1. Arrival Cost Update

The arrival cost (Γ), which is the first term in the cost function (Eq. 37), takes into account the effect of measurements previous to the estimation horizon (up to t_L). As presented in Elsheikh et al. (2021), Γ can be calculated using different approaches and this can affect the stability of the MHE itself. In this work we update it by using the single QR-factorization approach as described in Kühl et al. (2011). The QR-factorization is a linearised technique as well as the Extented Kalman Filter, however, it also holds the numerical properties of the square-root Kalman Filter. Moreover, the past information can influence the optimization only within the limits given by the process noise covariance $\bar{Q}_k$ (Kühl et al., 2011). When we shift the horizon to a new start point at t_{L+1}, the arrival cost would ideally be defined as:

$$\begin{aligned}\Gamma(x_{L+1},\theta_{L+1}) = \min_{x_L,\theta_L} \quad & \left(\left\|\begin{matrix}\hat{x}_L - x_L\\ \hat{\theta}_L - \theta_L\end{matrix}\right\|^2_{P_L} + \|y_L - h(x_L)\|^2_V + \left\|\begin{matrix}w_L\\ w_L^\theta\end{matrix}\right\|^2_{\bar{W}_L}\right)\\ \text{s.t.} \quad & w_L = x_{L+1} - F(x_L,\theta_L,u_L)\\ & w_L^\theta = \theta_{L+1} - \theta_L\end{aligned} \tag{A.1}$$

However, due to the nonlinearity of the model (i.e. x_{L+1} is described by a nonlinear function), the arrival cost Γ in Eq. A.1 does not have an analytical expression. Hence, we define the term $x(t_{L+1}|x_L)$ as the solution of the DAEs (Eq. 36) in the interval from $t \in [t_L, t_{L+1}]$ with x_L as initial value. To have an analytically solvable least-squares problem, $x(t_{L+1}|x_L)$ (Eq. A.2) is linearized around the best available estimate x^* and θ^*, to obtain an explicit solution of Eq. A.1:

$$\begin{aligned}x(t_{L+1}|x_L) &\approx x(t_{L+1}|x^*,\theta^*) + A\cdot(x_L - x^*) + A_\theta(\theta_L - \theta^*)\\ &\approx \tilde{x} + Ax_L + A_\theta\theta_L\end{aligned} \tag{A.2}$$

here $\tilde{x} := x(t_{L+1}|x^*) - Ax^* - A_\theta\theta^*$ and matrices A and A_θ are the derivatives of $x(t_{L+1}|x_L)$ with respect to x_L and θ_L:

$$A = \left.\frac{\partial F(x_L,\theta_L,u_L)}{\partial x_L}\right|_{x^*,\theta^*}, \qquad A_\theta = \left.\frac{\partial F(x_L,\theta_L,u_L)}{\partial \theta_L}\right|_{x^*,\theta^*}$$

Besides, similarly should be done for $h(x_L)$. However, since $h(x_L)$ is here linear, we can represent it, by using a selector matrix H, as $h(x_L) = Hx_L$. This way it becomes possible to solve, analytically, Eq. A.1 by rewriting it as:

$$\min_{x_L} \left\| \begin{matrix} P_L\begin{pmatrix}\hat{x}_L - x_L\\ \hat{\theta}_L - \theta_L\end{pmatrix}\\ V(y_L - Hx_L)\\ W_L\begin{pmatrix}x_{L+1} - \tilde{x} - Ax_L - A_\theta\theta_L\\ \theta_{L+1} - \theta_L\end{pmatrix}\end{matrix} \right\|^2_2 \tag{A.3}$$

by transforming the problem in Eq. A.3 with free variables x_Γ (i.e. x_L, x_{L+1}, θ_L and θ_{L+1}) by QR-factorization ($A_\Gamma = \mathcal{QR}$), we obtain:

$$A_\Gamma = \begin{pmatrix} P_L & 0\\ -VH & 0\\ -\bar{W}_L\begin{pmatrix}A & A_\theta\\ 0 & \mathbb{I}\end{pmatrix} & \bar{W}_L\end{pmatrix} = \mathcal{Q}\begin{pmatrix}\mathcal{R}_1 & \mathcal{R}_{12}\\ 0 & \mathcal{R}_2\\ 0 & 0\end{pmatrix} \tag{A.4}$$

The QR-factorization decomposes the matrix in the objective function (Eq. A.3) into the product of an orthogonal

*Corresponding author
Email address: nadi.bar@ntnu.no (Nadav Bar)

matrix $\mathcal{Q}$ and an upper triangular matrix $\mathcal{R}$ (Elsheikh et al., 2021). From Eq. A.4, we then obtain an equivalent problem of the form:

$$\min_{x_L,\theta_L} \left\| \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} + \begin{pmatrix} \mathcal{R}_1 & \mathcal{R}_{12} \\ 0 & \mathcal{R}_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_L \\ \theta_L \\ x_{L+1} \\ \theta_{L+1} \end{pmatrix} \right\|_2^2 \tag{A.5}$$

where, given that Eq. A.5 is solving $-\mathcal{Q}^{-1}A_\Gamma x_\Gamma + \mathcal{R}x_\Gamma = 0$, and the first term is γ_L, we can express it as:

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} = -\mathcal{Q}^{-1} \begin{pmatrix} P_L & 0 \\ -VH & 0 \\ -\bar{W}_L \begin{pmatrix} A & A_\theta \\ 0 & \mathbb{I} \end{pmatrix} & \bar{W}_L \end{pmatrix} \begin{pmatrix} x_L \\ \theta_L \\ x_{L+1} \\ \theta_{L+1} \end{pmatrix} \tag{A.6}$$

Finally, the approximated quadratic expression for the arrival cost results from the analytical solution of the problem in Eq. A.5:

$$\Gamma'(x_{L+1}) = \|\gamma_3\|_2^2 + \left\| \gamma_2 + \mathcal{R}_2 \begin{pmatrix} x_{L+1} \\ \theta_{L+1} \end{pmatrix} \right\|_2^2 \tag{A.7}$$

Since the first term of Eq. A.7 is given (Eq. A.6), the arrival cost updates are obtained as:

$$\begin{pmatrix} \hat{x}_{L+1} \\ \hat{\theta}_{L+1} \end{pmatrix} = -\mathcal{R}_2^{-1}\gamma_2, \qquad P_{L+1} = \mathcal{R}_2 \tag{A.8}$$

S2. Parameter variation and inequality constraints effect

In order to show to the reader the importance of the utilization of bounds (i.e. inequality constraints), to avoid that the plant-model mismatch and the discontinuities in the derivatives (due to sudden changes in the process dynamics) can influence the estimation performance, we report here the estimated parameters for *Case 1* (Fig.s S1 and S2), *Case 2* (Fig.s S3 and S4) and *3* (Fig.s S5 and S6).

S2.1. Case 1

Orthogonalization Method. The estimated parameters obtained are reported in Fig. S1, together with the number of degrees of freedom available (Eq. 19). The DOF_θ (Fig. S1f) present a minimum of 1 and a maximum value of 3. The number of constrained parameters is therefore obtained by Eq. 20. Moreover, for the OM, the degrees of freedom are reduced when the determinant of $Y_k Y_k^T$ approaches zero ($\leq 10^{-8}$, implemented in the algorithm Fig. 3), causing the DOF_θ to chatter (Fig. S1f).

Several relevant observations can be found. Firstly, the discrepancy between model and *on-line* measurements (Fig. 9a and c, 4-7 hours) is compensated by increase in μ_{max}, leading to a faster growth rate (Fig. S1a) whereas Y_{CO_2} is decreased for the same reason (Fig. S1e).

Secondly, both μ_{max} and Y_{XS} (Fig. S1a and d, 7-9 hours) are decreased while the substrate is being consumed. The parameter K_s (Fig. S1b, 7-9 hours) however is increased together with Y_{XCO_2}, to adapt (in Fig. 9b) the glucose consumption (Eq. 1) to the estimated values and compensate the mismatch between measured CO_2 and the model (Fig. 9c). Moreover, it is worth mentioning that abrupt changes in the process dynamics led to discontinuities in the derivatives and therefore sharp spikes in the estimated parameters (8, 20 and 23 hours, Fig. S1).

Thirdly, between 20 and 22 hours biomass increases due to the addition of glucose (Fig. 9a), requiring the need to increase μ_{max} (Fig. S1a). At the same time, sugar (Fig. 9b) is simultaneously consumed, diluted and added to the system (u, Eq. 1). However, only one of these phenomena (consumption) is dependent on the parameters. It becomes therefore reasonable that the optimization problem will therefore bring K_s (Fig. S1b) and Y_{XS} (Fig. S1d) up to the boundaries to compensate for the plant-model mismatch (Fig. 9b). Similarly, Y_{CO_2} (Fig. S1e) is also increased to reduce the discrepancy between model and *on-line* outputs (Fig. 9c).

Lastly, when $DOF_\theta = 1$ (Fig. S1f), only k_d is optimized (Fig. S1c). Between 10 and 20 hours its value reach the upper bound since there is no growth and death is dominant in the biomass dynamics (Fig. 9a). In contrast, we see that k_d chatters between lower and upper bounds after 30 hours (Fig. S1c). This is reasonable since we can see that the uncertainty in the outputs increased (Fig. 9c) and therefore the quality of the data is reflected in it.

Subset Selection by Transformation. The parameters reported in Fig. S2, although sometimes limited at the boundaries, present a different dynamic behaviour (compared to OM, Fig. S1) due to the differences in the two methods. It is important to note that the main difference between the OM and SST is the fact that, the OM constrains individual parameters whereas SST constrains linear combinations of parameters (Fig. 1). This implies that more than one parameter can vary within each cluster (Fig. S2). In order to understand this difference, we compare the results obtained in Fig. S1 (OM) and Fig. S2 (SST). While in Fig. S1 (1-9 hours) $\text{DOF}_\theta = 3$ and it is possible to see that these degrees of freedom are used only to estimate μ_{max}, Y_{XS} and Y_{XCO_2} (keeping the other two constrained at their nominal values), in the SST method all the parameters are estimated (Fig. S2).

We observed several differences in the SST performance, in comparison to OM. Firstly, similarly to what we have seen in OM (Fig. S1), the discrepancy between model and outputs (Fig. 5a and c, 4-8 hours), is compensated by an increase in μ_{max}, for a faster growth rate (Fig. S2a) and a decrease in Y_{CO_2} (Fig. S2e). In addition, since more parameters can vary simultaneously, also K_s and k_d compensate for the plant-model mismatch (Fig. S2b and c).

Secondly, during substrate consumption μ_{max} (Fig. S2a, 8-9 hours) is decreased and Y_{XCO_2} is increased. In contrast, Y_{XS} (Fig. S2d) is increased together with k_d (Fig. S2c) while K_s (Fig. S2b, 8-9 hours) is first decreased and than

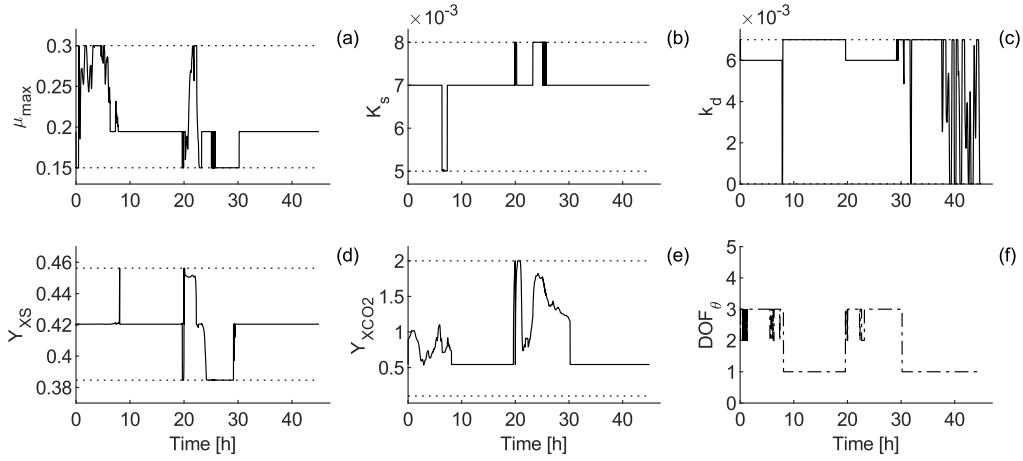


Figure S1: Orthogonalization Method, Case 1 - Estimated parameters (continuous black lines) μ_{max} (a), K_s (b), k_d (c), Y_{XS} (d) and Y_{CO_2} (e) with their bounds (inequality constraints, dotted lines). In addition, the available degrees of freedom (DOF_θ) are reported (f).

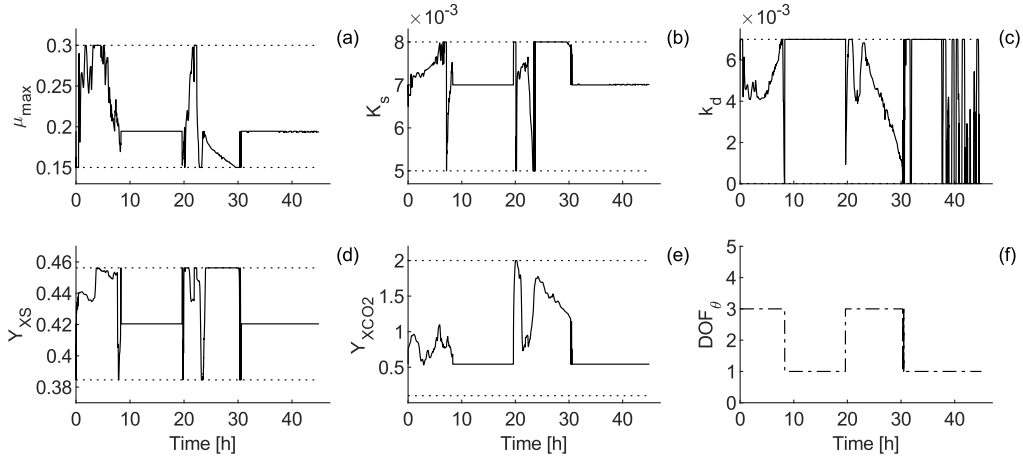


Figure S2: Subset Selection by Transformation, Case 1 - Estimated parameters (continuous black lines) μ_{max} (a), K_s (b), k_d (c), Y_{XS} (d) and Y_{CO_2} (e) with their bounds (inequality constraints, dotted lines). In addition, the available degrees of freedom (DOF_θ) are reported (f). In contrast to OM, here all parameters are updated.

increased, although presenting a spike due to discontinuities in the derivative (caused by sudden changes in the process dynamics), to adapt (in Fig. 5) the outputs or the estimated value to the model.

Thirdly, we can see that between 20 and 22 hours the addition of glucose (Fig. 5b) led to an increase in μ_{max} (Fig. 5a and S2a). This sugar will be consumed by the organism, however, dynamically the addition of glucose prevails over consumption and dilution (Eq. 1). However, all the parameters are unrelated to the input (u) and the increase in K_s and Y_{XS} will take that into account. In addition, k_d (Fig. S2c) and Y_{CO_2} (Fig. S2e) increase to compensate for the model discrepancy as well (Fig. 5a and c). Moreover, differently from OM, here we can see that during sugar consumption (22-30 hours, Fig. 5b) all the parameters except Y_{CO_2} are differently estimated. Indeed, k_d first increases and than decreases (Fig. 5c) together with K_s and μ_{max}, while Y_{XS} saturates up to the boundaries due to lack of information of glucose measurements in order to compensate for the plant-model mismatch in Fig. 5b. Lastly, when $DOF_\theta = 1$ (Fig. S2f), only k_d is optimized (Fig. S2c) and, similarly to OM, chatters after 30 hours (Fig. S1c).

S2.2. Case 2 and 3

While in *Case 2* the boundaries are tight and give a better estimation in terms of RMSE, they limit the possibility

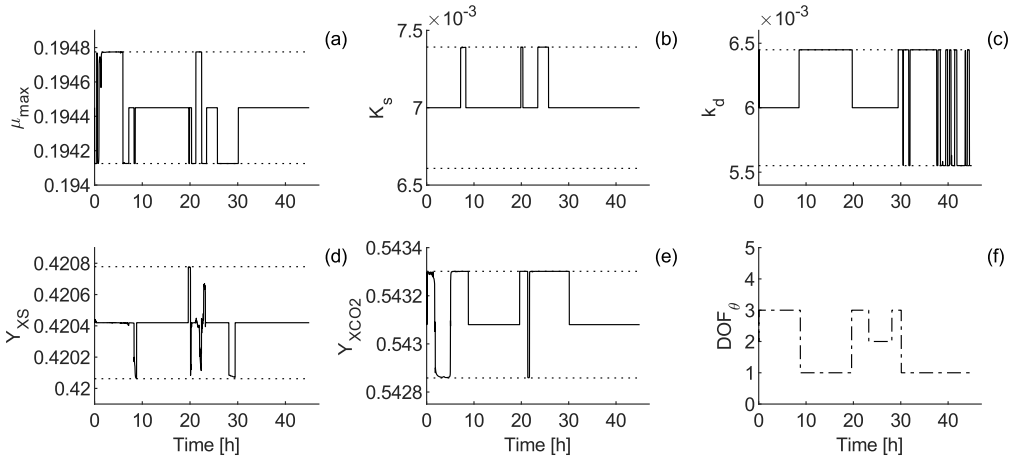


Figure S3: Orthogonalization Method, Case 2 - Estimated parameters (continuous black lines) μ_{max} (a), K_s (b), k_d (c), Y_{XS} (d) and Y_{CO_2} (e) together with their bounds (dotted line). In addition, the available degrees of freedom (DOF_θ) are reported (f). Decrease in the DOF_θ is due to the precaution taken in the OM to avoid utilizing singular matrices.

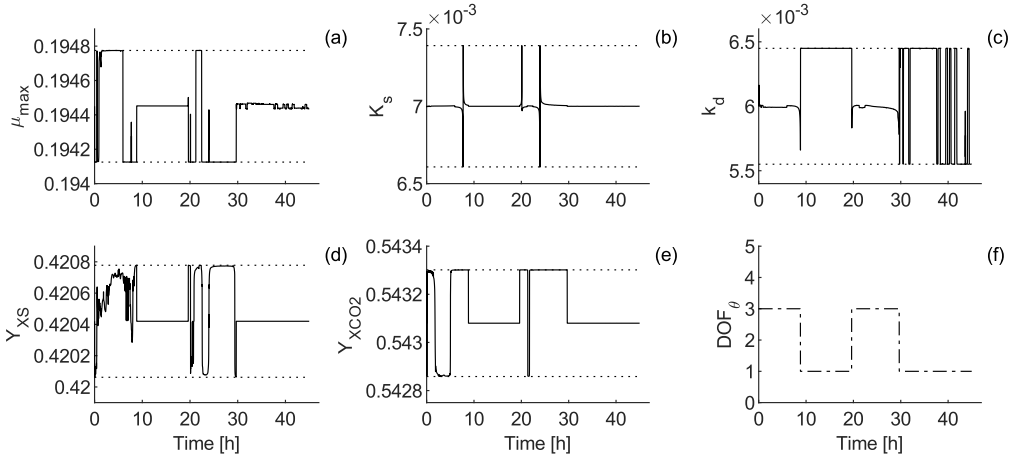


Figure S4: Subset Selection by Transformation, Case 2 - Estimated parameters (continuous black lines) μ_{max} (a), K_s (b), k_d (c), Y_{XS} (d) and Y_{CO_2} (e) together with their bounds (dotted line). In addition, the available degrees of freedom (DOF_θ) are reported (f).

of the model to adapt to the dynamics (Fig.s 10 and 6). However, when looking at the results for *Case 3*, we see that, although the model gains flexibility (Fig.s 11 and 7), the limitations due to the plant-model mismatch and the discontinuities in the derivatives, still require the use of inequality constraints to limit the deterioration of the estimates (Fig.s S5 and S6).

S3. Plant-model mismatch

To show the limitation of using a simple unstructured Monod model, we present the results obtained by fitting separately the batch (Fig. S7) and the fed-batch phase (Fig. S8) together with the respective parameters (Table S5).

Table S5: Model parameters $\theta \in \mathbb{R}^{n_\theta}$ values obtain for batch and fed-batch phases only, compared with the nominal values used in the estimation problem.

Parameter	Description	Nominal	Batch	Fed-batch
μ_{max}	Maximum growth rate	0.19445	0.2290	0.2015
K_S	Monod growth constant	0.007	0.0051	0.0096
k_d	Death rate constant	0.006	0.0061	0.0037
Y_{XS}	S from X yield	0.42042	0.4104	0.4407
Y_{XCO_2}	CO_2 from X yield	0.54308	0.5740	0.6284

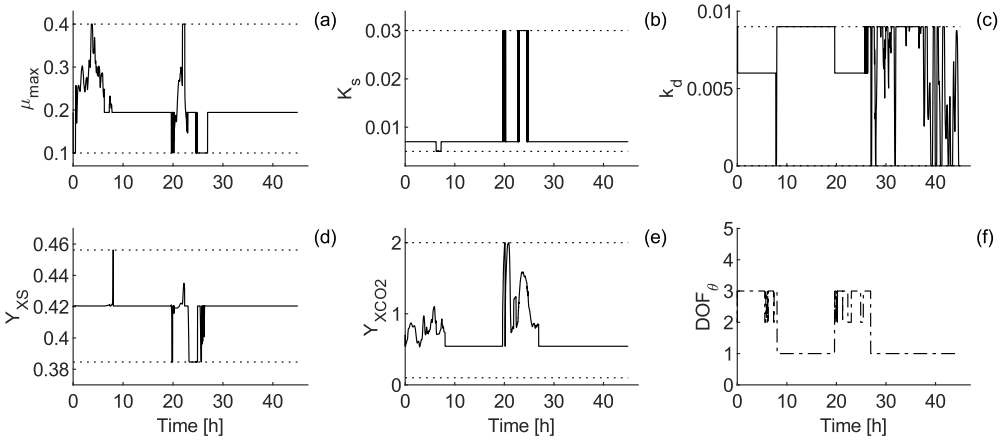


Figure S5: Orthogonalization Method, Case 3 - Estimated parameters (continuous black lines) μ_{max} (a), K_s (b), k_d (c), Y_{XS} (d) and Y_{CO_2} (e) together with their bounds (dotted line). In addition, the available degrees of freedom (DOF_θ) are reported (f). Decrease in the DOF_θ is due to the precaution taken in the OM to avoid utilizing singular matrices. This case presents the largest bounds (among all three cases). The parameters hit the bounds only in the case of discontinuities in the derivatives, and therefore bounds are necessary to not let this penalize the estimation performance.

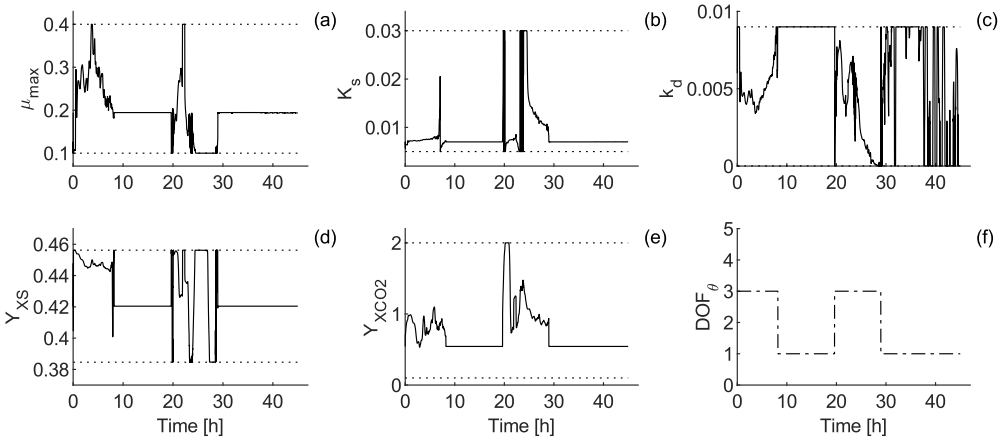


Figure S6: Subset Selection by Transformation, Case 3 - Estimated parameters (continuous black lines) μ_{max} (a), K_s (b), k_d (c), Y_{XS} (d) and Y_{CO_2} (e) together with their bounds (dotted line). In addition, the available degrees of freedom (DOF_θ) are reported (f). This case presents larger bounds in the inequality constraints. The parameters stay mostly within the bounds, which they only hit in the case of discontinuities in the derivatives. Moreover, the variation of Y_{XS} (d) is more limited than the other parameters in order to avoid that it can penalize the variation of the model without having enough information.

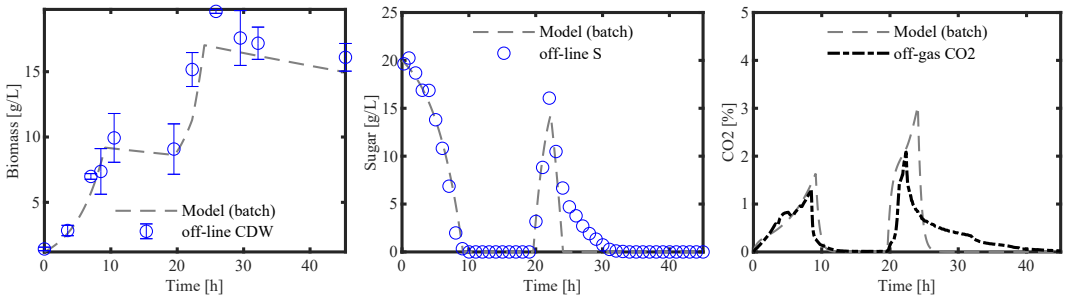


Figure S7: Model fit obtained using solely the Batch data-set.

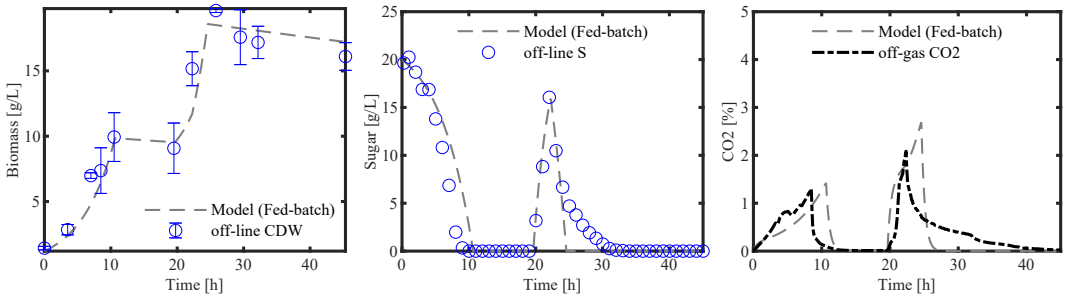


Figure S8: Model fit obtained using solely the Fed-batch data-set.

5 | A practical implementation of Moving Horizon Estimation with Delayed Measurements in a Bioprocesses Experimental Application

"The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge."

STEPHEN W. HAWKING (1942-2018)

This paper is awaiting publication and is not included in NTNU Open

6 | Discussion

"To teach details is to bring confusion; to establish the relationship between things is to bring knowledge."

MARIA MONTESSORI (1870-1952)

In this chapter, to enhance the contributions of this thesis in regard to monitoring of biological processes under the presence of uncertainty and plant-model mismatch, we will first summarize the results presented in Chapters 3, 4 and 5 (Chapters 6.1) and consecutively contextualize them with respect to the available literature (Chapters 6.2).

6.1 Research Questions

In this section we will take into consideration the research questions given in Chapter 1 and give a short summary on them:

Research Question 1. *What are the available state estimator to handle uncertainty in nonlinear biological processes? How can these estimators accurately estimate the unmeasured variables of interest under plant-model mismatch?*

In Chapter 3, to investigate the possibility of handling uncertainty in a nonlinear biological process, we implemented three different Bayesian estimators (i.e. EKF, UKF and MHE) using experimental data-sets from a *Corynebacterium glutamicum* fed-batch cultivation process. Because Bayesian estimators assume that variables have a stochastic nature and take probability distributions into account (Chapter 2), they resulted a valuable approach to handle uncertainty. Indeed, although the use of a simple Monod model with structural plant-model mismatch, all the estimators presented the possibility of accurately estimating the variables of interest, presenting an improvement compared with the model predictions.

To ensure an accurate estimation of the unmeasured variable (i.e. glucose), we firstly analyzed the observability of the system. The local observability analysis presented in Chapter 3 showed that to ensure the local observability of the system under consideration, measurements of volume, biomass and CO_2 were necessary. The analysis also showed that, under continuous feeding (i.e. CSTR) or in absence of sugar depletion, the sole use of biomass and CO_2 or volume and CO_2 were sufficient. However, it is important to note that the local observability of the system does not guarantee the accuracy of the estimates, because the local observability analysis performed through Lie derivatives is only a structural property of the system and does not use any information of the uncertainty of either the measurements or the model.

Secondly, to obtain an accurate estimation of the glucose concentration, we gave importance to the tuning of the process covariance matrix (Q, Sec. 2.3.1). This was done by selecting a state-dependent variable process noise covariance matrix (Q_k, Eq. 2.7), following the works of [14, 15]. Under dynamic conditions, the concentration of the variables of interest (i.e. metabolites) will change (i.e. increase or decrease) and therefore the only use of constant, additive process noise covariance

matrix Q can result in unsatisfactory estimates or divergence of the estimator, if the orders of magnitude of dynamic changes and model uncertainty do not remain consistent. Therefore, to obtain a variable, state dependant process noise covariance matrix Q_k, we considered that the model uncertainty mostly comes from the uncertain, lumped parameters used in the unstructured Monod model [39]. Additionally, to handle the possible changes in metabolism in the second batch phase (i.e. batch-to-batch variation), we increased the trust on the measurements by increasing the uncertainty in two of the model parameters (i.e. K_s and Y_{CO_2}), relying less on the model predictions to achieve more accurate estimates.

Thirdly, to avoid negative (i.e. infeasible) concentrations, we implemented state constraints through a QP-problem, in both EKF and UKF. Moreover, because of the necessity of implementing state constraints, we investigated the MHE as an alternative approach, due to its ability to easily handle state constraints, and discussed the importance of hyperparameter tuning (i.e. R, Q_k and horizon lenght T) to ensure that the data within the horizon can overcome the bias of the arrival cost approximation.

Lastly, we compared EKF and UKF which showed a comparable performance. This might be related to the fact that 1) the updates were faster (i.e. minutes) than the dynamics of the system (i.e. hours) and therefore the linear approximation did not hampered the performance of the EKF; 2) while the measurement function implemented throughout this work was linear, the presence of nonlinearities in the measurement functions could work in favour of the UKF. Additionally, we implemented the MHE separately. Although the EKF and UKF estimators are not directly comparable with the MHE because of different tuning of σ_V^2 and $\sigma_{CO_2}^2$, they presented similar performances. However, the reader must consider that, while the computational effort of EKF and UKF is lower, the MHE has the advantage of easily implementing state constraints.

Research Question 2. *How can we implement on-line model adaptation to further improve the accuracy of the estimated states under plant-model mismatch?*

To handle the structural plant-model mismatch, in Chapter 4 we investigated the possibility of capturing the changing or missing dynamics through parameter adaptation. This was done in an MHE, by adding the parameters as single degrees of freedom in the optimization problem.

Firstly, to include the parameters as degrees of freedom in the optimization problem, we augmented the model by considering them as additional states. However, the improvements in the estimates are due to the reduction of the effect of the uncertainty caused by plant-model mismatch and unmodeled disturbances [87] rather

than the accurate estimation of model parameters [21]. Because the estimation of the internal states (i.e. parameters) was not our main focus, instead of considering the parameters as additional states, it is possible to consider their changes as disturbances (i.e. noise) as reported in [20]. This needs some attention. Indeed, although considering them as disturbances would have the advantage of reducing the number of states and therefore the size of the optimization problem solved at every iteration, this approach would not ensure that only one degree of freedom is used for every single parameter. Alternatively, when considering the parameters as states, we make sure that one degree of freedom is used for every single parameter accordingly to its statistics (i.e. this is enforced by $\dot{\theta}_i = w_{\theta_i}$).

Secondly, to overcome the ill-conditioness of the problem that arises by adding parameters as degrees of freedom, we implemented and compared two regularization methods in an MHE, namely Orthogonalization Method (OM) [25, 88] and Subset Selection by Transformation (SST) [24, 87]. Although parameters can be estimated also using EKF or UKF by considering the parameters as additional states (i.e. model augmentation), the MHE was selected because it results in a more straightforward approach to include regularization methods. The results presented in Chapter 4 showed how the use of regularization methods leads to a trade-off between reduction of variance and bias. Indeed, the parameter adaptation penalizes the estimates of the unmeasured state, becoming a trade-off for the practitioner, for either improvement of estimates or model predictions, therefore resulting that significant previous knowledge on the structural mismatch is necessary to avoid biased estimates [87]. Moreover, comparing OM and SST, we showed how SST, by incorporating previous knowledge on the parameters, results in a more flexible and robust approach under higher parameter uncertainty. Although the results presented an improvement with comparison to the estimates obtained without the addition of parameter adaptation, an accurate tracking of the drifting model parameters was not guaranteed. Indeed, the structural plant-model mismatch posed a limitation on the improvement of the model adaptation and the improvements of either estimates or model predictions resulted from the correction of the plant-model mismatch and unmodelled disturbances.

Lastly, to determine the number of degrees of freedom for the optimization problem (and therefore the number of parameters to estimate at each iteration), we proposed a stopping criterion based on the structural identifiability of the system. This enabled us to have the possibility to update a variable number of parameters per iteration, avoiding the selection of a stopping criterion based on a threshold, which would be simply based on trial and error [89, 90], or based on the minimization of the Mean Squared Error (MSE) proposed by [89], which instead would not have been possible in our case because of absence of full-state feedback (i.e.

measurements of glucose were not available for the estimator).

Research Question 3. *When infrequent and delayed at-line measurements are available, how can we best include them to enhance the quality of the real-time information?*

To enhance real-time information, in Chapter 5 we presented the different results obtained including infrequent and delayed measurement in the estimator. This was done, using experimental data, with a multi-rate MHE (MMHE), because of its ability, in comparison of EKF and UKF, to handle delayed measurements trough the use of a window of past measurements. However, as presented in [31], delayed measurements can also be included in recursive Bayesian state estimators as the EKF and the UKF.

Firstly, we implemented the MMHE always considering the fast measurements, while considering the slow measurements error only when those delayed measurements were available (i.e. variable structure) and putting them back within the horizon at the time the measurement was taken [81]. Additionally, because of the possibility to achieve smoother estimates, we also implemented the MMHE considering the slow measurements always available and keeping a constant slow measurement error between two consecutive sampling times (i.e. using a zero-order hold to extrapolate and spread the slow measurement error at the fast sampling rate) and therefore always have an error for the slow measurements at the same frequency as the fast ones (i.e. fixed structure) [28, 27, 29].

Secondly, to improve the estimation accuracy, we implemented both MMHE with variable and fixed structure using filter recalculation as proposed by [30] and also presented in [31] for EKFs. In this case the optimization problem (and therefore all the estimates) are recalculated *off-line* from the time at which the sample was taken until the current time, therefore using the information from the slow measurements throughout different horizons (Fig. 1.5). This improved the estimation accuracy for the variable structure, while it did not for the fixed structure. Most likely because of the stronger effect of the plant-model mismatch in the fixed structure.

Lastly, the results showed how the selected approach could differently influence the estimation accuracy. Indeed, while the implementation of a variable structure with filter recalculation resulted the most accurate in our case for monitoring purposes, we showed how a fixed structure MMHE would deliver smoother estimates, therefore being preferable in cases of more uncertain or erroneous *at-line* measurements and, most importantly, showed how the implementation of a fixed structure could be preferred for control purposes, because the sudden correction of the estimates could be reflected in the control actions, possibly resulting in un-

wanted output behaviours. Additionally, the results also presented the importance of tuning the uncertainty of the slow measurements, showing that their inclusion becomes a trade-off between rejecting the error due to plant-model mismatch and the newly acquired measurements [30]. We conclude that the choice of MMHE implementation therefore depends on the specific application and the desired output.

6.2 Development and Design of State Estimators

In this section, we will discuss the results presented, in the form of general guidelines for the application of state estimators in the bioprocess industry. This section is, by no means, exhaustive, however we aim to provide relevant comments and ideas on the design of the main components of the state estimators which emerged while conducting this work. This is therefore done, by complementing general guidelines present in literature, relevant for the development and implementation of state estimators in a bioprocess, in relation to the works presented in Chapters 3, 4 and 5.

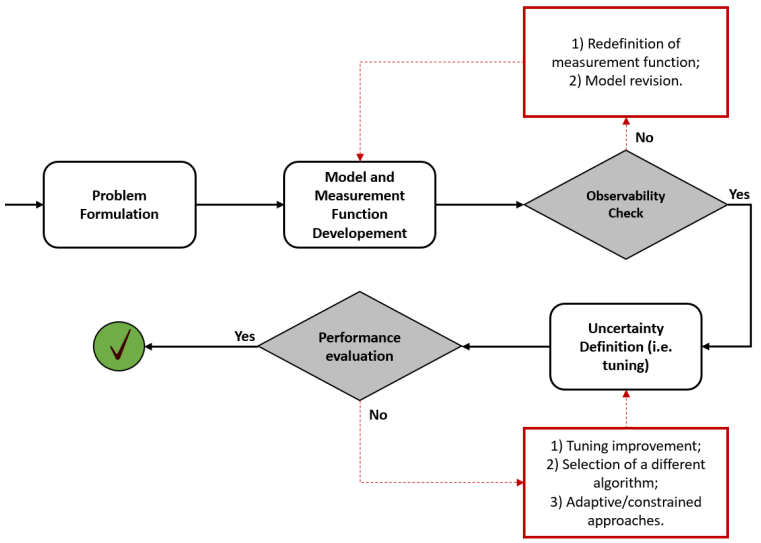


Figure 6.1: The figure, adapted from [9], reports a step-by-step workflow for the design of state estimators. In case the condition for observability is not satisfied, possibilities to ensure that are reported (red box). Similarly, in case the performance desired is not achieved, possibilities to improve that are reported (red box).

Most importantly, when developing a state estimator, as also stated in [9, 49, 50] and reported in Fig. 6.1, it is important to define:

1. the problem (i.e. what we want to estimate);
2. the available model and output measurements (y), together with their relationship, to evaluate if their combination allows to infer information on the variables of interest;
3. the assessment of the observability of the system (i.e. the possibility to infer information on the variables of interest, given the available information);
4. the understanding of the sources of uncertainty to properly tune the different parameters (i.e. hyperparameters);
5. the necessity to implement constrained or adaptive approaches;
6. the evaluation of the performance, coupled with the selection of the most suitable approach.

6.2.1 Problem Definition

The reactor set-up used in this work is reported in Fig. 6.2, with all the available *in-line* and *on-line* measurements. The interest, in this work, was to monitor the concentration of glucose, together with biomass and volume in the reactor, continuously. However, either direct or indirect measurement for the glucose were not available, neither *in-line* nor *on-line*. Therefore, it was necessary to infer this information using the available measurements and the physical insight (i.e. model) to develop a state estimator (i.e. soft sensor).

6.2.2 Model Development and Measurement Selection

In Sec. 2.2.1, we presented different possibilities to develop models, because of their central role in the development of state estimators, since they incorporate the physical or data-based knowledge we have on the process. Simply, we could say that they incorporate our experience, creating relations between the outputs (i.e. values that we measure) and the states (i.e. variables of interest).

As reported in Chapters 3, 4 and 5, we modelled the process using an unstructured first principle model, based on Monod-like kinetics (Sec. 2.3). Considering states (x) and measurements (y), the dynamics of the process (Fig. 6.2) can be formulated as a set of ordinary differential equations (ODEs) as:

$$\dot{x} = f(x, \theta, u) \tag{6.1a}$$

$$y = h(x) \tag{6.1b}$$

where x are the states, θ the model parameters, u the input and y the measurements. The selection of process dynamics and the measurement function in Eq. 6.1, was

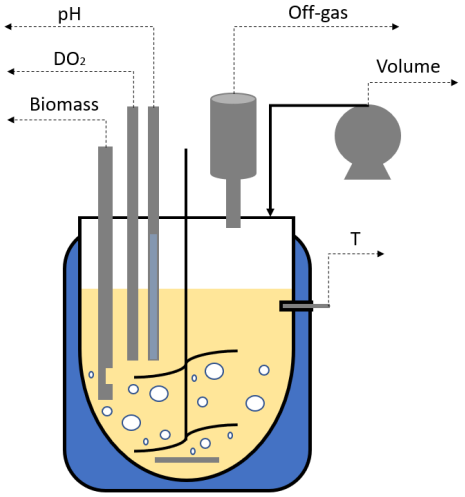


Figure 6.2: Bioreactor scheme with available *in-line* (i.e. absorbance for biomass, dissolved oxygen DO_2 and pH) and *on-line* measurements (i.e. off-gas composition for O_2 and CO_2, volume and temperature).

done to relate the latter to the variables that we intended to estimate (i.e. glucose), in absence of direct measurements.

Firstly, by considering the available measurements reported in Fig. 6.2, the biomass and the volume can be directly measured trough the absorbance probe and the input to the reactor respectively. Additionally, for the glucose, it seems reasonable to consider the dissolved O_2 and CO_2 in the broth (i.e. aqueous solution) and the off-gas measurements for O_2 and CO_2, because the substrate uptake (i.e. consumption) is related to respiration. Taking into account the phenomena occurring in both liquid and gas phase can be beneficial, as reported in [63]. However, in [63] the consideration of both O_2 and CO_2 was done to compensate for the lack of other measurement devices available, as the absorbance probe in our case, to infer information on all the variables of interest. Thus, it is important to note that complementing our system dynamics with those presented in [63], if not strictly necessary (e.g. the system is not observable or certain dynamics can not be neglected), would increase the number of model parameters and state variables, therefore increasing the size of the system and hampering the advantages of a simple dynamic model for the implementation of a state estimator. Indeed, by following the idea of [50] that a further improvement of first principle knowledge should be taken with care, and that the priority is to include the main process nonlinearities, we have adopted an approach similar to [18], by only taking the CO_2 composition in the off-gas into account. This is done considering that the amount of CO_2 in the

reactor (i.e. aqueous) is constant (i.e. quasi-steady state), due the rapid gas-phase dynamics, given constant head pressure and pH (i.e. tightly controlled) [63, 68].

By selecting as available measurements $y = [y_V, y_X, y_{CO_2}]$ and defining the state variables of the system in Eq. 6.1 given by $x = [V, X, S, CO_2]$, we can rewrite the discrete set of ODEs in Eq. 6.1:

$$x_k = F(x_k, \theta_k, u_k) \tag{6.2a}$$
$$y_k = Hx_k \tag{6.2b}$$

where H is a selector matrix that represents the linear relation between y and x.

The values of the nominal parameters $\theta = [\mu_{max}, K_S, k_d, Y_{XS}, Y_{CO_2}]$ used throughout the Chapters 3-5 were obtained *off-line*, by nonlinear least-squares data-fitting, on a single experimental data set. However, the practitioner is advised to use several data sets to better cover the *batch-to-batch* variation in the parameter estimates and therefore obtain a more robust measure of the parameters uncertainty.

Additionally, it is important to comment on the measurement selection. Indeed, in the presented case the composition of CO_2 in the off-gas was not of interest per se, but it was only added in the system dynamics to add information on the variables of interest (i.e. biomass and glucose). The use of additional measurements of secondary variables (i.e. CO_2 in our case) can be done by developing static or dynamic models, either mechanistic, data-based or hybrid. In the case of static models, as was recently presented in [12], those should be used as measurement functions ($h(x)$, Eq. 6.1b), while the dynamic ones should be considered as extra states (i.e. augmented system, Eq. 2.1d), therefore developing a model specifically for state estimation purposes.

6.2.3 Observability

After developing the model and selecting the measurements, an important step in the design of a state estimator is the evaluation of the observability of the system. This defines the possibility to infer information on the states (i.e. variables of interest) given the knowledge on the external outputs (i.e. measurements) [9, 49, 13, 60, 26]. The practitioner must note that, as reported in [18] and in Sec. 6.2.2, in the absence of full observability the system must be restructured for the purpose (Sec. 6.2.1). In this direction, the work of [60] presents an algorithm to systematically select the minimal set of required measurements to ensure full observability. Indeed, if measurements are not correctly selected to ensure observability or detectability [48, 49, 60], the estimator will be unable to infer information on the non-observable states, therefore leading to the prediction of those

through simulation results rather than estimation [18]. Alternatively, if infrequent and delayed measurements are available, these can be used to ensure observability [30, 31, 81]. Although infrequent measurement present additional information for the estimator, care must be taken on selecting the implementation for their inclusion, as mentioned in [31, 27, 28] and in Chapter 5. Indeed, the selected approach can strongly influence the estimation accuracy.

Observability is a property of the system, dependent on the given dynamics ($\dot{x} = f(x, \theta, u)$), inputs ($u$) and measurement function ($y = h(x)$), which for nonlinear sytems is usually evaluated using Lie derivatives, as reported in Chapters 3 and 4. However, besides the observability analysis through Lie derivatives, which is solely a qualitative property of the system (i.e. it only gives a yes or no answer), it exists the possibility to analyze the observability through the observability Grammian [49, 26], which instead gives a quantitative measure in addition to the qualitative one. A discussion on the two approaches discern from the interests of this work. Therefore, the most interested reader is referred to [49, 91, 92].

6.2.4 Uncertainty Quantification

As it was pointed out in the works of [14–16], most often practitioners give very little attention to the importance of uncertainty quantification, considering the process (Q), measurement (R) and initial measurement noise (P_0) covariance matrices as purely trial and error tuning parameters.

The tuning of the process noise covariance Q is usually considered the most challenging task. However, in the case of biological processes, models are usually developed by *off-line* parameter identification, through nonlinear least-squares data-fitting, using experimental data. Therefore, it must be considered that the noise does not always simply enters in an additive way, since quite often we might have some knowledge of it. Consequently, it is important to consider that the process of parameter identification not only gives the estimated value (i.e. mean), but also their uncertainty. Thus, it is with this mindset that the practitioner should approach the tuning of the process (or model) uncertainty Q, hence avoiding the tedious work of purely estimating uncertainty based on a trial and error approach. Indeed, this approach not only simplifies the selection of the values for Q, but also follows the idea that most often the noise does not only enters the model additively but also trough uncertainty in the model parameters or the inputs [14–16], therefore becoming state or input dependent. Indeed, it is our advice, when tuning the process noise covariance matrix Q, to first consider the parameter (Chapters 3 and 4) and input [15] uncertainties, while using the additive noise for fine tuning. The latter is also meant to take into account structural model mismatch and avoid numerical issues.

The tuning of the measurement noise covariance matrix R is, instead, usually considered a less challenging task, since most often the measurement noise covariance matrix can be considered time-invariant and readily available form the manufacturer of the device [16]. However, the possibility of having a time-varying R must be taken into account [93]. Indeed, 1) in the case of unstructured regression models to calibrate the sensors [12] or 2) in the case of structured state-dependant or input-dependant noise in the measurement devices, a similar approach to the one presented for the tuning of the process noise covariance Q (Sec. 2.3.1) must be considered. This is however feasible only in the presence of data-sets of a reasonable size, as it is often the case in industrial applications [12].

In contrast to the previous two, the initial measurement noise covariance matrix P_0 can not be evaluated experimentally and its tuning will mostly affect the speed of convergence of the estimator [16]. Following the work of [16], in the works presented throughout Chapters 3, 4 and 5, its calculation was based on the true initial state. However, in case this is not available, an average value based on the expected lower and upper bounds can be employed.

6.2.5 Constraints and Adaptive Approaches

When considering bioprocesses, since the majority of them are conducted in batch or fed-batch mode, it is often common that concentrations can become zero or close to zero. This can become problematic when performing state estimation, since the probability of an expected value close to zero can cause the estimate to find its realization in the negative domain. Thus, if such situation is encountered or expected, it is necessary to avoid it by implementing non-negativity constraints (e.g. $x \geq 0$). From a probability point of view, this consists in truncating the probability density function [62, 22]. In the case of MHEs, this can be done in a straightforward way, by simply adding inequality constraints in the optimization problem as presented in Chapters 3 and 4. Differently, when dealing with EKFs or UKFs, this can be done using different approaches and, the most commonly implemented one is clipping, which simply consists in projecting back into their bounds the estimates that violate the constraints [79, 62, 53]. However, different approaches have been proposed later in the works of [19, 94]. Indeed, in the application of EKF and UKF presented in Chapter 3, the utilization of a QP-problem was implemented, following the work of [19], only when constraints were violated.

Constraints however, not only can be useful to avoid infeasible values for the estimates, but they can also be implemented for adaptive estimation approaches, when interested in the joint estimation of states and parameters in the presence of plant-model mismatch. Thus incorporating robustness to modelling errors (Chapter 4) [13, 9, 21]. Indeed, when interested on the joint estimation of states and para-

meters, the practitioner must take into account that the estimation of all the model parameters might result in an ill-posed problem and that constraints might be used to avoid that [48], by limiting the number of the parameters being estimated or by limiting them through the selection of linear combinations of them, as reported in Chapter 4 [90]. This, as mentioned above, results in a more straightforward approach in MHEs, although also possible to implement in EKFs and UKFs.

6.2.6 Technique Selection

Given that the system is observable and that information on the uncertainties is available, we discourage the simple use of deterministic models, while encouraging for the use of state estimators, to also take the uncertainties into account. In this extent, it becomes therefore important to select the most convenient approach, given necessities and limitations of the specific application.

Firstly, in case the sole interest is to infer information on the unmeasured variables of interest, we believe EKFs and UKFs should be taken into account firstly, given their simplicity and their low computational requirements (Chapter 3).

Secondly, if the necessity of implementing constraints arises, constrained EKFs and UKFs [19, 53] can be considered, although MHEs might result in a more natural approach to implement constraints (Chapter 3). Moreover, also in the case of systems described by differential algebraic equations (DAE), algebraic equations can be simply added to the existing set of constraints in the formulation [13]. However, when choosing to implement an MHE, care must be taken in the selection of the time horizon, aiming to find a trade-off between computational requirement and performance (Chapter 3) [50].

Thirdly, under presence of plant-model mismatch, we believe that the use of MHEs presents the most flexible approach (Chapter 4) for *on-line* model adaptation. Indeed, *on-line* model adaptation would benefit from the use of past horizon information to deliver smoother estimates, simultaneously allowing to include different types of regularization approaches, in the form of penalties or algebraic equations implemented through constraints, to obtain the most accurate solution.

Lastly, when including infrequent and delayed samples, although EKF or UKF also present the possibility to handle them [31], MHE results to be the preferred approach because of its ability to easily handle delays, considering an horizon of past measurements [27–30, 81, 21]. However, as discussed in Chapter 5, for the inclusion of infrequent measurements the selection of the correct implementation is crucial. Indeed, the practitioner must select it depending on the specific application and the desired outcome.

7 | Concluding Remarks and Future Outlook

"We must have an industry where the machine takes over everything is able to take over, so that it can create us the profit we would like to use to go on a forest trip or watch the sunset." [95]

JENS GLAD BALCHEN (1926-2009)

7.1 Conclusion

In this work we presented the important role that state estimators have in the next-generation biotech industry (Chapters 1 and 2). Moreover, through the experimental application of three different Bayesian state estimators, we highlighted different possibilities to infer information on unmeasured variables under the presence of uncertainty and plant-model mismatch (Chapters 3, 4 and 5).

In Chapter 3:

- we implementated three different Bayesian state estimators, namely EKF, UKF and MHE;
- we focused on the importance of local observability analysis for the selection of the necessary measurements to ensure the observability of the system;
- we highlighted the necessity of careful hyperparameter tuning, most specifically the process noise covariance matrix Q_k and the horizon length T for the MHE;
- we considered the need to implement state constraints in both EKF and UKF and the alternative of using MHE to avoid negative concentrations during sugar depletion.

In Chapter 4:

- we evaluated the possibility to capture the changing or missing dynamics through parameter adaptation, improving the estimation accuracy under plant-model mismatch;
- we implemented two different regularization methods (i.e. Orthogonalization Method and Subset Selection by Transformation) in a MHE and gave their geometric interpretation for an easy comparison of the two;
- we proposed a stopping criterion based on the structural identifiability of the system.

In Chapter 5:

- we implemented a multi-rate MHE (MMHE) to include delayed and slow measurements to enhance the accuracy of the estimates;
- the MMHE was implemented with both variable and fixed structure;

- to further improve the estimation accuracy, we have implemented both fixed and variable structure with filter recalculation;
- we discussed as, in our case, a variable structure would be preferable for only monitoring purposes, while a fixed structure, delivering smoother estimates, is preferred for control purposes.

Additionally, through Chapter 6, we discussed our results contextualizing them within general guidelines for the development and design of Bayesian state estimators.

7.2 Future Outlook

Firstly, to extend the results obtained in this work, future works should investigate the use of hybrid models, to better characterise the system dynamics and therefore limit the hindering effect of structural plant-model mismatch.

Secondly, the use of redundant sensors is important, considering their industrial implementation, for either the detection of sensor failure and the implementation of algorithms to detect that or for the application to systems which present both space and time dynamics (i.e. described by partial differential equations), such as tubular reactors for the cultivation of microalgae.

Thirdly, focus should be placed on the implementation of optimal sample scheduling techniques, minimizing the uncertainty of the estimates. This is important to reduce costly and time consuming measurements when sensors are not available and infrequent measurements have to be adopted to ensure the observability of the system.

Lastly, as a continuation of this work, simultaneous state and parameter estimation should be tested in closed-loop, when controlling experimental cultivations in real time. During the past years, we have upgraded and improved the laboratory set-up at the Department of Chemical Engineering at NTNU and works are ongoing for the real-time control of the cultivation processes.

A Appendix

A.1 Basic Concepts of Probability and Statistics

The experimental practitioner is, most likely, used to take triplicates (or more generally multiple samples) to determine the concentration of a certain metabolite. Indeed, this is done to define the *mean* and the *sample variance*, which represent, respectively, the average and the distance from the mean [96]. We take, as example, five samples of the concentration of glucose $g = \begin{bmatrix} g_1, & g_2, & g_3, & g_4, & g_5 \end{bmatrix}$ in the cultivation broth at a certain time point (t = 4h) and analyze them obtaining:

$$g = \begin{bmatrix} 5.4 & 5.8 & 5.4 & 5.8 & 5.7 \end{bmatrix} \tag{A.1}$$

Given $N = 5$ as the sample size, we obtain:

$$m = \frac{\sum_{i=1}^{N} g_i}{N} = 5.62 \tag{A.2a}$$

$$S^2 = \frac{\sum_{i=1}^{N} (g_i - m)^2}{N - 1} = 0.042 \tag{A.2b}$$

which represent the *mean* (Eq. A.2a) and the sample variance (i.e. variation in our samples, Eq. A.2b) of our realizations (i.e. measured glucose concentrations).

If now we want to conduct a second experiment in the same conditions, we do not have a certainty of the values we will get, but we can expect that, by taking five samples at the same time point (t = 4h) we have an information on the probabilities from the previous experiment (Eq. A.1), which are 40% of obtaining 5.4, 40% of obtaining 5.8 and 20% of obtaining 5.7. Therefore, we can define the probability p for the expected concentration x:

$$p = \begin{bmatrix} p_1, & p_2, & p_3 \end{bmatrix} = \begin{bmatrix} 0.4, & 0.4, & 0.2 \end{bmatrix}$$
$$x = \begin{bmatrix} x_1, & x_2, & x_3 \end{bmatrix} = \begin{bmatrix} 5.4, & 5.8, & 5.7 \end{bmatrix}$$

This probability allow us to define the *expected value* ($E[x]$) and its *variance* (σ^2) as:

$$E[x] = \sum_{j=1}^{3} p_j x_j = 5.62 \tag{A.4a}$$

$$\sigma^2 = E\left[(x - E(x))^2\right] = \sum_{j=1}^{3} p_i (x_j - E(x))^2 = 0.042 \tag{A.4b}$$

However, most often, we are not only interested in measuring the glucose (g), but also the biomass (b) and the product (r), therefore expecting different concentrations. Given $M = 1, \ldots, 3$ the number of different concentrations, we will have $M = 3$ different random variables (RV). Each of them will have an *expected value* ($E[x_M]$). However, since the biomass grows and produces by consuming glucose, these RV will be correlated and they will have a *joint probability* (p_{gb}) which represents the probability that RV 1 produces g and RV 2 produces b. Therefore, we can define a *covariance* (i.e. measure of joint variability) as:

$$\sigma_{gb}^2 = \sum\sum p_{gb}(g - E[g])(b - E[b]) \tag{A.5}$$

where Eq. A.5, written in matrix form for $M = 1, 2$ (i.e. g and b), becomes the covariance matrix V:

$$V = \sum\sum p_{gb} \begin{bmatrix} (g-E[g])^2 & (g-E[g])(b-E[b]) \\ (g-E[g])(b-E[b]) & (b-E[b])^2 \end{bmatrix}$$

For a more in depth explanation the reader is referred to [96, 62], which we used as a reference here.

A.2 Eigenvalues, Eigenvectors and Singular Values Decomposition

Eigenvectors q are vectors which define the direction of Aq (i.e. directions of the invariant action), where $A \in \mathbb{R}^{n \times n}$. Eigenvalues λ instead, are scalars values which tell whether the eigenvector q is stretched, shrunk, reversed or unchanged (i.e. magnitude of the invariant action). Their relationship is defined as follows [96, 97]:

$$Aq = \lambda q$$

The eigenvalues λ are obtained by solving $det(A - \lambda I) = 0$ (i.e. the characteristic polinomial), where $I \in \mathbb{R}^{n \times n}$ is the identity matrix. For every λ, the correspondent eigenvectors are obtained by solving $(A - \lambda I)q = 0$. This allows us to decompose (or factorize) matrix $A \in \mathbb{R}^{n \times n}$ (i.e. eigendecomposition) as follows:

$$A = \sum_{i=1}^{n} \lambda_i q_i {q_i}^T = Q \Lambda Q^T \tag{A.6}$$

Eq. A.6 coincides with Singular Values Decomposition in the case the matrix A is symmetric and positive definite [97]. Indeed, more generally, given a matrix $B \in \mathbb{R}^{m \times n}$, it can be decomposed in three terms as follows:

$$B = \sum_{i=1}^{n} u_i \sigma_i {v_i}^T = U \Sigma V^T$$

The three terms geometrically correspond to rotation, stretching and rotation [96]. Where U and V are the matrices of the singular vectors, which form an orthonormal basis for the SVD and describe the directions of its maximum action. Matrix B can therefore be diagonalized as:

$$Av = \sigma u$$

where σ represents the singular values which instead describe the magnitude of the maximum action. For a further explanation the reader is referred to [96].

A.3 Cholesky Factorization

Given a matrix $A \in \mathbb{R}^{m \times m}$ symmetric and positive definite, then there exists a unique lower triangular matrix $C \in \mathbb{R}^{m \times m}$ such that [96, 98, 97]:

$$A = C^T C \tag{A.7}$$

Eq. A.7 represents the Cholesky factorization of the matrix A [96, 98, 97].

A.4 Gradient, Hessian, Jacobian and Sensitivity

Given a variable vector $x \in \mathbb{R}^{n_x}$ and a function $f : \mathbb{R}^{n_x} \to \mathbb{R}$ the vector of first order partial derivatives $\nabla f(x)$ (i.e. *gradient*) is defined as [97]:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_{n_x}} \end{bmatrix} \tag{A.8}$$

The matrix of the second order partial derivatives $\nabla^2 f(x)$ (i.e. *Hessian*, $\mathcal{H}$) is defined as [97]:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial {x_1}^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_{n_x}} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial {x_2}^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_{n_x}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_{n_x} \partial x_1} & \frac{\partial^2 f}{\partial x_{n_x} \partial x_2} & \cdots & \frac{\partial^2 f}{\partial {x_{n_x}}^2} \end{bmatrix}$$

If we now maintain the variable vector $x \in \mathbb{R}^{n_x}$ and define a new function $g(x)$ so that $g : \mathbb{R}^{n_x} \to \mathbb{R}^m$, we can define the *Jacobian* (J) as [97]:

$$J(x) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_{n_x}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_{n_x}} \end{bmatrix} \tag{A.9}$$

The *Jacobian* (Eq. A.9) is also called *Sensitivity* matrix $(S(x) = J(x))$ [99]. Indeed, also following [97], each of the elements in Eq. A.8 (i.e. partial derivatives) defines the sensitivity of the function f with respect to a component.

A.5 Linear Programming (LP)

Linear Programming (LP) refers to the solution of a problem where the objective function is linear as well as the constraints. Generally it can be written as [97]:

$$\min_x \quad \phi(x) = c^T x \tag{A.10a}$$

$$\text{s.t.} \quad Ax = a \tag{A.10b}$$

$$Bx \geq b \tag{A.10c}$$

where $x \in \mathbb{R}^{n_x}$ represents the vector of decision variables, while $c \in \mathbb{R}^{n_x}$ is the gradient of the objective $\phi(x)$ (Eq. A.10a). Equations A.10b and A.10c instead, are the equality and inequality constraints. The system in Eq. A.10 is written as a minimization problem, however, if the interest is to find the maximum, the objective function can be simply transformed by multiplying $\phi(x)$ by -1 (i.e. $\phi(x) = -c^T x$).

A.6 Quadratic Programming (QP)

Quadratic Programming (QP) refers to the solution of a QP-problem (i.e. an optimization problem where the objective function is quadratic and the constraints linear) and can generally be written as [97]:

$$\min_x \quad \phi(x) = \frac{1}{2} x^T \mathcal{Q} x + c^T x \tag{A.11a}$$

$$\text{s.t.} \quad Ax = a \tag{A.11b}$$

$$Bx \geq b \tag{A.11c}$$

where $x \in \mathbb{R}^{n_x}$ represents the vector of decision variables, while $\mathcal{Q} \in \mathbb{R}^{n_x \times n_x}$ and $c \in \mathbb{R}^{n_x}$ are respectively the Hessian and the gradient of the objective $\phi(x)$ (Eq. A.11a). Eq. A.11b and A.11c instead, represent respectively equality and inequality constraints. If the Hessian $\mathcal{Q}$ is positive semidefinite, the QP-problem is convex and its solution represents the global minimum. If not, the solution is a local minimum.

The QP-problem can be seen as a special case of an NLP-problem. Moreover, in case the problem is convex, its solution does not results much more difficult than an LP-problem [97].

A.7 Nonlinear Programming (NLP)

Nonlinear Programming (NLP) refers to the solution of a NLP-problem (i.e. an optimization problem where at least the objective function or the constraints are nonlinear). Generally it can be stated as [97]:

$$\min_{x} \quad \phi(x) \tag{A.12a}$$

$$\text{s.t.} \quad h_i(x) = a_i \qquad i = 1, \ldots, n \tag{A.12b}$$

$$g_j(x) \geq b_j \qquad j = 1, \ldots, m \tag{A.12c}$$

where $x \in \mathbb{R}^{n_x}$ represents the vector of decision variables. $\phi : \mathbb{R}^{n_x} \to \mathbb{R}$ instead represents a generic objective function (Eq. A.12a), while $h : \mathbb{R}^{n_x} \to \mathbb{R}^n$ and $g : \mathbb{R}^{n_x} \to \mathbb{R}^m$ represent two generic functions for equality and inequality constraints (Eq. A.12b and A.12c).

A.8 Least-Squares Problem

Least-squares problems are most often encountered for the identification of model parameters using experimental data. This is done through the minimization of a nonlinear least-squares function defined as follows [97]:

$$\phi(x) = \frac{1}{2} \sum_{i=1}^{n} {r_i}^2(x) \tag{A.13}$$

where $r_i(x)$ is the residual function and is given by:

$$r_i(x) = y_i - f(x, t_i) \tag{A.14}$$

with measurements y_i and model predictions (e.g. concentrations) given by $f(x, t_i)$. Therefore, the residual $r_i(x)$ (Eq. A.14) represents the distance, at each time point t_i, between the model and the measurement. This problem becomes an LP, QP or NLP whether it follows the definitions given in Sections A.5, A.6 or A.7. For a more extensive explanation the reader is referred to [97].

References

[1] Michalis Koutinas, Alexandros Kiparissides, Efstratios N Pistikopoulos, and Athanasios Mantalaris. Bioprocess systems engineering: transferring traditional process engineering principles to industrial biotechnology. *Computational and structural biotechnology journal*, 3(4):e201210022, 2012. ISSN 2001-0370.

[2] Yu Luo, Varghese Kurian, and Babatunde A Ogunnaike. Bioprocess systems analysis, modeling, estimation, and control. *Current Opinion in Chemical Engineering*, 33:100705, 2021. ISSN 2211-3398.

[3] Judith Becker and Christoph Wittmann. Bio-based production of chemicals, materials and fuels - Corynebacterium glutamicum as versatile cell factory. *Current Opinion in Biotechnology*, 23(4):631–640, 2012. ISSN 09581669. doi: 10.1016/j.copbio.2011.11.012.

[4] Jens Nielsen. Production of biopharmaceutical proteins by yeast: advances through metabolic engineering. *Bioengineered*, 4(4):207–211, 2013. ISSN 2165-5979.

[5] Gregory Stephanopoulos. Synthetic biology and metabolic engineering. *ACS synthetic biology*, 1(11):514–525, 2012. ISSN 2161-5063.

[6] Francesco Destro and Massimiliano Barolo. A review on the modernization of pharmaceutical development and manufacturing – Trends, perspectives, and the role of mathematical modeling. *International Journal of Pharmaceutics*, 620:121715, 5 2022. ISSN 03785173. doi: 10.1016/j.ijpharm.2022.121715.

[7] Harini Narayanan, Martin F Luna, Moritz von Stosch, Mariano Nicolas Cruz Bournazou, Gianmarco Polotti, Massimo Morbidelli, Alessandro Butté, and Michael Sokolov. Bioprocessing in the digital age: the role of process models. *Biotechnology journal*, 15(1):1900172, 2020. ISSN 1860-6768.

[8] Efstratios N Pistikopoulos, Ana Barbosa-Povoa, Jay H Lee, Ruth Misener, Alexander Mitsos, Gintaras V Reklaitis, Venkat Venkatasubramanian, Fengqi You, and Rafiqul Gani. Process systems engineering–the generation next? *Computers & Chemical Engineering*, 147:107252, 2021. ISSN 0098-1354.

[9] Ronald Alexander, Gilson Campani, San Dinh, and Fernando V Lima. Challenges and opportunities on nonlinear state estimation of chemical and biochemical processes. *Processes*, 8(11):1462, 2020. ISSN 2227-9717.

[10] Rimvydas Simutis and Andreas Lübbert. Bioreactor control improves bioprocess performance. *Biotechnology journal*, 10(8):1115–1130, 2015. ISSN 1860-6768.

[11] Volker F Wendisch, K Madhavan Nampoothiri, and Jin-Ho Lee. Metabolic Engineering for Valorization of Agri-and Aqua-Culture Sidestreams for Production of Nitrogenous Compounds by Corynebacterium glutamicum. *Frontiers in Microbiology*, 13, 2022.

[12] Tobias Wallocha and Oliver Popp. Off-Gas-Based Soft Sensor for Real-Time Monitoring of Biomass and Metabolism in Chinese Hamster Ovary Cell Continuous Processes in Single-Use Bioreactors. *Processes*, 9(11):2073, 2021. ISSN 2227-9717.

[13] Sachin C Patwardhan, Shankar Narasimhan, Prakash Jagadeesan, Bhushan Gopaluni, and Sirish L Shah. Nonlinear Bayesian state estimation: A review of recent developments. *Control Engineering Practice*, 20(10):933–953, 2012. ISSN 0967-0661.

[14] Giuseppe Leu and Roberto Baratti. An extended Kalman filtering approach with a criterion to set its tuning parameters; application to a catalytic reactor. *Computers & Chemical Engineering*, 23(11-12):1839–1849, 1 2000. ISSN 00981354. doi: 10.1016/S0098-1354(00)00298-2.

[15] S. Kolås, B. A. Foss, and T. S. Schei. Noise modeling concepts in nonlinear state estimation. *Journal of Process Control*, 19(7):1111–1125, 7 2009. ISSN 09591524. doi: 10.1016/j.jprocont.2009.03.002.

[16] René Schneider and Christos Georgakis. How to NOT make the extended kalman filter fail. *Industrial and Engineering Chemistry Research*, 52(9): 3354–3362, 2013. ISSN 08885885. doi: 10.1021/ie300415d.

[17] Georges Bastin and Denis Dochain. On-line estimation of microbial specific growth rates. *Automatica*, 22(6):705–709, 1986. ISSN 0005-1098.

[18] D. Krämer and R. King. A hybrid approach for bioprocess state estimation using NIR spectroscopy and a sigma-point Kalman filter. *Journal of Process Control*, 82:91–104, 2019. ISSN 09591524. doi: 10.1016/j.jprocont.2017.11.008.

[19] S Kolås, B.A. Foss, and T.S. Schei. Constrained nonlinear state estimation based on the UKF approach. *Computers & Chemical Engineering*, 33(8): 1386–1401, 8 2009. ISSN 00981354.

[20] Peter Kühl, Moritz Diehl, Tom Kraus, Johannes P Schlöder, and Hans Georg Bock. A real-time algorithm for moving horizon state and parameter estimation. *Computers & Chemical Engineering*, 35(1):71–83, 1 2011. ISSN 00981354. doi: 10.1016/j.compchemeng.2010.07.012.

[21] Jaehan Bae, Yeonsoo Kim, and Jong Min Lee. Multirate moving horizon estimation combined with parameter subset selection. *Computers & Chemical Engineering*, 147:107253, 4 2021. ISSN 00981354. doi: 10.1016/j.compchemeng.2021.107253.

[22] Christopher V Rao. *Moving horizon strategies for the constrained monitoring and control of nonlinear discrete-time systems*. The University of Wisconsin-Madison, 2000. ISBN 0599782234.

[23] Ravindra D Gudi, Sirish L Shah, and Murray R Gray. Adaptive multirate state and parameter estimation strategies with application to a bioreactor. *AIChE Journal*, 41(11):2451–2464, 11 1995. ISSN 00011541. doi: 10.1002/aic.690411111.

[24] Boeun Kim and Jay H Lee. Parameter subset selection and biased estimation for a class of ill-conditioned estimation problems. *Journal of Process Control*, 81:65–75, 9 2019. ISSN 09591524. doi: 10.1016/j.jprocont.2019.05.015.

[25] Berit Floor Lund and Bjarne A Foss. Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models. *Automatica*, 44(1):278–281, 1 2008. ISSN 00051098. doi: 10.1016/j.automatica.2007.04.006.

[26] Alejandro F Villaverde. Observability and Structural Identifiability of Nonlinear Biological Systems. *Complexity*, 2019:1–12, 1 2019. ISSN 1076-2787. doi: 10.1155/2019/8497093.

[27] Stefan Krämer and Ralf Gesthuisen. Multirate state estimation using moving horizon estimation. *IFAC Proceedings Volumes*, 38(1):1–6, 2005. ISSN 14746670. doi: 10.3182/20050703-6-CZ-1902.00654.

[28] S Kramer, Ralf Gesthuisen, and S Engell. Fixed structure multirate state estimation. In *Proceedings of the 2005, American Control Conference, 2005.*, pages 4613–4618. IEEE, 2005. ISBN 0-7803-9098-9. doi: 10.1109/ACC.2005.1470723.

[29] Francesco Scibilia and Morten Hovd. Multi-Rate Moving Horizon Estimation with Erroneous Infrequent Measurements Recovery. *IFAC Proceedings Volumes*, 42(8):1037–1042, 2009. ISSN 14746670. doi: 10.3182/20090630-4-ES-2003.00171.

[30] Luo Ji and James B Rawlings. Application of MHE to large-scale nonlinear processes with delayed lab measurements. *Computers & Chemical Engineering*, 80:63–72, 9 2015. ISSN 00981354. doi: 10.1016/j.compchemeng.2015.04.015.

[31] Ajit Gopalakrishnan, Niket S Kaisare, and Shankar Narasimhan. Incorporating delayed and infrequent measurements in Extended Kalman Filter based nonlinear state estimation. *Journal of Process Control*, 21(1):119–129, 1 2011. ISSN 09591524. doi: 10.1016/j.jprocont.2010.10.013.

[32] Jacques Monod. The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1):371–394, 10 1949. ISSN 0066-4227. doi: 10.1146/annurev.mi.03.100149.002103.

[33] Mohd Izuan Effendi Halmi, S A Ahmad, M A Syed, N A Shamaan, and M Y Shukor. Mathematical modelling of the molybdenum reduction kinetics in Bacillus pumilus strain Lbna. *Bulletin of Environmental Science and Sustainable Management (e-ISSN 2716-5353)*, 2(1):24–29, 2014. ISSN 2716-5353.

[34] Peter Sinner, Marlene Stiegler, Christoph Herwig, and Julian Kager. Noninvasive online monitoring of Corynebacterium glutamicum fed-batch bioprocesses subject to spent sulfite liquor raw material uncertainty. *Bioresource Technology*, page 124395, 11 2020. ISSN 09608524. doi: 10.1016/j.biortech.2020.124395.

[35] N I Marcos, Martin Guay, Denis Dochain, and Tao Zhang. Adaptive extremum-seeking control of a continuous stirred tank bioreactor with Haldane's Kinetics. *Journal of Process Control*, 14(3):317–328, 2004. ISSN 0959-1524.

[36] Benjamin J Hodgson, Christopher N Taylor, Misti Ushio, J R Leigh, Tatiana Kalganova, and Frank Baganz. Intelligent modelling of bioprocesses: a comparison of structured and unstructured approaches. *Bioprocess and Biosystems Engineering*, 26(6):353–359, 2004. ISSN 1615-7605.

[37] Krist V Gernaey, Anna Eliasson Lantz, Pär Tufvesson, John M Woodley, and Gürkan Sin. Application of mechanistic models to fermentation and biocatalysis for next-generation processes. *Trends in biotechnology*, 28(7): 346–354, 2010. ISSN 0167-7799.

[38] Lisa Mears, Stuart M Stocks, Mads O Albaek, Gürkan Sin, and Krist V Gernaey. Mechanistic fermentation models for process design, monitoring, and control. *Trends in biotechnology*, 35(10):914–924, 2017. ISSN 0167-7799.

[39] Banafsheh Jabarivelisdeh, Lisa Carius, Rolf Findeisen, and Steffen Waldherr. Adaptive predictive control of bioprocesses with constraint-based modeling and estimation. *Computers & Chemical Engineering*, 135:106744, 4 2020. ISSN 00981354. doi: 10.1016/j.compchemeng.2020.106744.

[40] Zoltan Szallasi, Jörg Stelling, and Vipul Periwal. System Modeling in Cellular Biology: From Concepts to Nuts and Bolts, 2007.

[41] Harini Narayanan, Michael Sokolov, Massimo Morbidelli, and Alessandro Butté. A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins. *Biotechnology and Bioengineering*, 116(10):2540–2549, 10 2019. ISSN 0006-3592. doi: 10.1002/bit.27097.

[42] Dimitris C Psichogios and Lyle H Ungar. A hybrid neural network-first principles approach to process modeling. *AIChE Journal*, 38(10):1499–1511, 10 1992. ISSN 0001-1541. doi: 10.1002/aic.690381003.

[43] Timur Bikmukhametov and Johannes Jäschke. Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. *Computers & Chemical Engineering*, 138:106834, 2020. ISSN 0098-1354.

[44] Yu Luo, Robert J Lovelett, J Vincent Price, Devesh Radhakrishnan, Kristopher Barnthouse, Ping Hu, Eugene Schaefer, John Cunningham, Kelvin H Lee, and Raghunath B Shivappa. Modeling the effect of amino acids and copper on monoclonal antibody productivity and glycosylation: a modular approach. *Biotechnology Journal*, 16(2):2000261, 2021. ISSN 1860-6768.

[45] Peter Harms, Yordan Kostov, and Govind Rao. Bioprocess monitoring. *Current Opinion in Biotechnology*, 13(2):124–127, 2002. ISSN 0958-1669.

[46] V Vojinović, J M S Cabral, and L P Fonseca. Real-time bioprocess monitoring: Part I: In situ sensors. *Sensors and Actuators B: Chemical*, 114(2): 1083–1091, 2006. ISSN 0925-4005.

[47] Reiner Luttmann, Daniel G Bracewell, Gesine Cornelissen, Krist V Gernaey, Jarka Glassey, Volker C Hass, Christian Kaiser, Christian Preusse, Gerald Striedner, and Carl-Fredrik Mandenius. Soft sensors in bioprocessing: a status report and recommendations, 2012.

[48] Denis Dochain. State and parameter estimation in chemical and biochemical processes: a tutorial. *Journal of Process Control*, 13(8):801–818, 12 2003. ISSN 09591524. doi: 10.1016/S0959-1524(03)00026-X.

[49] Jarinah Mohd Ali, N. Ha Hoang, M.A. Hussain, and Denis Dochain. Review and classification of recent observers applied in chemical process systems. *Computers & Chemical Engineering*, 76:27–41, 5 2015. ISSN 00981354. doi: 10.1016/j.compchemeng.2015.01.019.

[50] Tor Steinar Schei. On-line estimation for process control and optimization applications. *Journal of Process Control*, 18(9):821–828, 10 2008. ISSN 09591524. doi: 10.1016/j.jprocont.2008.06.014.

[51] Ali Mesbah, Adrie E M Huesman, Herman J M Kramer, and Paul M J Van den Hof. A comparison of nonlinear observers for output feedback model-based control of seeded batch crystallization processes. *Journal of Process Control*, 21(4):652–666, 2011. ISSN 0959-1524.

[52] Pau Cabaneros Lopez, Isuru A. Udugama, Sune T. Thomsen, Christian Roslander, Helena Junicke, Miguel Mauricio-Iglesias, and Krist V. Gernaey. Towards a digital twin: a hybrid data-driven and mechanistic digital shadow to forecast the evolution of lignocellulosic fermentation. *Biofuels, Bioproducts and Biorefining*, 14(5):1046–1060, 9 2020. ISSN 1932-104X. doi: 10.1002/bbb.2108.

[53] Rambabu Kandepu, Bjarne Foss, and Lars Imsland. Applying the unscented Kalman filter for nonlinear state estimation. *Journal of Process Control*, 18 (7-8):753–768, 8 2008. ISSN 09591524. doi: 10.1016/j.jprocont.2007.11.004.

[54] Balasubramaniam Srinivasan, Srinivas Palanki, and Dominique Bonvin. Dynamic optimization of batch processes: I. Characterization of the nominal solution. *Computers & Chemical Engineering*, 27(1):1–26, 2003. ISSN 0098-1354.

[55] Balasubramaniam Srinivasan, Dominique Bonvin, Erik Visser, and Srinivas Palanki. Dynamic optimization of batch processes: II. Role of measurements in handling uncertainty. *Computers & chemical engineering*, 27(1):27–44, 2003. ISSN 0098-1354.

[56] Dinesh Krishnamoorthy. Novel Approaches to Online Process Optimization Under Uncertainty: Addressing the limitations of current industrial practice. 2019. ISSN 8232641991.

[57] Benoît Chachuat, Balasubramaniam Srinivasan, and Dominique Bonvin. Adaptation strategies for real-time optimization. *Computers & Chemical Engineering*, 33(10):1557–1567, 10 2009. ISSN 00981354. doi: 10.1016/j.compchemeng.2009.04.014.

[58] Sigurd Skogestad. Control structure design for complete chemical plants. *Computers & Chemical Engineering*, 28(1-2):219–234, 2004. ISSN 0098-1354.

[59] Haeun Yoo, Ha Eun Byun, Dongho Han, and Jay H Lee. Reinforcement learning for batch process control: Review and perspectives. *Annual Reviews in Control*, 52:108–119, 2021. ISSN 1367-5788.

[60] Nina P G Salau, Jorge O Trierweiler, and Argimiro R Secchi. Observability analysis and model formulation for nonlinear state estimation. *Applied Mathematical Modelling*, 38(23):5407–5420, 2014. ISSN 0307-904X.

[61] Andrew H Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007. ISBN 0486462749.

[62] Dan Simon. *Optimal State Estimation*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 5 2006. ISBN 9780470045343. doi: 10.1002/0470045345.

[63] Gregory Stephanopoulos and Ka-Yiu-Y San. Studies on on-line bioreactor identification. I. Theory. *Biotechnology and Bioengineering*, 26(10):1176–1188, 1984. ISSN 10970290. doi: 10.1002/bit.260261006.

[64] Ka-Yiu San and Gregory Stephanopoulos. Studies on on-line bioreactor identification. II. Numerical and experimental results. *Biotechnology and bioengineering*, 26(10):1189–1197, 1984. ISSN 0006-3592.

[65] Ron Grosz, Gregory Stephanopoulos, and Ka-Yiu San. Studies on on-line bioreactor identification. III. Sensitivity problems with respiratory and heat evolution measurements. *Biotechnology and bioengineering*, 26(10):1198–1208, 1984. ISSN 0006-3592.

[66] Ka-Yiu San and Gregory Stephanopoulos. Studies on on-line bioreactor identification. IV. Utilization of pH measurements for product estimation. *Biotechnology and bioengineering*, 26(10):1209–1218, 1984. ISSN 0006-3592.

[67] R D Gudi and S L Shah. The role of adaptive multirate Kalman filter as a software sensor and its application to a bioreactor. *IFAC Proceedings Volumes*, 26(2):249–254, 1993. ISSN 1474-6670.

[68] R. D. Gudi, S. L. Shah, and M. R. Gray. Multirate state and parameter estimation in an antibiotic fermentation with delayed measurements. *Biotechnology and Bioengineering*, 44(11):1271–1278, 1994. ISSN 10970290. doi: 10.1002/bit.260441102.

[69] Simon J Julier, Jeffrey K Uhlmann, and Hugh F Durrant-Whyte. A New Approach for Filtering Nonlinear Systems. In *Proceedings of American Control Conference - ACC*, pages 1628–1632, 1995. doi: 10.1109/ACC.1995.529783.

[70] Guillaume Goffaux, A. Vande Wouwer, and Olivier Bernard. Improving continuous–discrete interval observers with application to microalgae-based bioprocesses. *Journal of Process Control*, 19(7):1182–1190, 7 2009. ISSN 09591524. doi: 10.1016/j.jprocont.2009.03.009.

[71] L. Dewasme, M. Sbarciog, E. Rocha-Cózatl, F. Haugen, and A. Vande Wouwer. State and unknown input estimation of an anaerobic digestion reactor with experimental validation. *Control Engineering Practice*, 85:280–289, 4 2019. ISSN 09670661. doi: 10.1016/j.conengprac.2019.02.003.

[72] Dominik Krämer and Rudibert King. On-line monitoring of substrates and biomass using near-infrared spectroscopy and model-based state estimation for enzyme production by S. cerevisiae. *IFAC-PapersOnLine*, 49(7):609–614, 2016. ISSN 24058963. doi: 10.1016/j.ifacol.2016.07.235.

[73] Aydin Golabgir and Christoph Herwig. Combining Mechanistic Modeling and Raman Spectroscopy for Real-Time Monitoring of Fed-Batch Penicillin Production. *Chemie Ingenieur Technik*, 88(6):764–776, 6 2016. ISSN 0009286X. doi: 10.1002/cite.201500101.

[74] Julian Kager, Christoph Herwig, and Ines Viktoria Stelzer. State estimation for a penicillin fed-batch process combining particle filtering methods with online and time delayed offline measurements. *Chemical Engineering Science*, 177:234–244, 2018. ISSN 00092509. doi: 10.1016/j.ces.2017.11.049.

[75] L. Dewasme, S. Fernandes, Z. Amribt, L. O. Santos, Ph Bogaerts, and A. Vande Wouwer. State estimation and predictive control of fed-batch cultures of hybridoma cells. *Journal of Process Control*, 30:50–57, 2015. ISSN 09591524. doi: 10.1016/j.jprocont.2014.12.006.

[76] Abdolrahimahim Yousefi-Darani, Olivier Paquet-Durand, Jörg Hinrichs, and Bernd Hitzmann. Parameter and state estimation of backers yeast cultivation with a gas sensor array and unscented Kalman filter. *Engineering in Life Sciences*, 21(3-4):170–180, 2021. ISSN 1618-0240.

[77] Guillaume Goffaux and A Vande Wouwer. Design of a robust nonlinear receding-horizon observer-Application to a biological system. *IFAC Proceedings Volumes*, 41(2):15553–15558, 2008. ISSN 1474-6670.

[78] Niels Krausch, Jong Woo Kim, Tilman Barz, Sergio Lucia, Sebastian Groß, Matthias C Huber, Stefan M Schiller, Peter Neubauer, and Mariano Nicolas Cruz Bournazou. High-throughput screening of optimal process conditions using model predictive control. *Biotechnology and Bioengineering*, 2022. ISSN 0006-3592.

[79] Eric L Haseltine and James B Rawlings. Critical Evaluation of Extended Kalman Filtering and Moving-Horizon Estimation. *Industrial & Engineering Chemistry Research*, 44(8):2451–2460, 4 2005. ISSN 0888-5885. doi: 10.1021/ie034308l.

[80] Rudolph Van Der Merwe, Eric Wan, and Simon Julier. Sigma-point Kalman filters for nonlinear estimation and sensor-fusion: Applications to integrated navigation. In *Aiaa guidance, navigation, and control conference and exhibit*, page 5120, 2004.

[81] Mohamed Elsheikh, Rubin Hille, Alexandru Tatulea-Codrean, and Stefan Krämer. A comparative review of multi-rate moving horizon estimation schemes for bioprocess applications. *Computers & Chemical Engineering*, 146:107219, 3 2021. ISSN 00981354. doi: 10.1016/j.compchemeng.2020.107219.

[82] Douglas G Robertson, Jay H Lee, and James B Rawlings. A moving horizon-based approach for least-squares estimation. *AIChE Journal*, 42(8):2209–2224, 8 1996. ISSN 0001-1541. doi: 10.1002/aic.690420811.

[83] James B Rawlings and Bhavik R Bakshi. Particle filtering and moving horizon estimation. *Computers & Chemical Engineering*, 30(10-12):1529–1541, 9 2006. ISSN 00981354. doi: 10.1016/j.compchemeng.2006.05.031.

[84] Vinay A Bavdekar, R. Bhushan Gopaluni, and Sirish L Shah. A Comparison of Moving Horizon and Bayesian State Estimators with an Application to a pH Process. *IFAC Proceedings Volumes*, 46(32):160–165, 12 2013. ISSN 14746670. doi: 10.3182/20131218-3-IN-2045.00152.

[85] Kenneth R Muske, James B Rawlings, and Jay H Lee. Receding horizon recursive state estimation. In *1993 American Control Conference*, pages 900–904. IEEE, 1993. ISBN 0780308603.

[86] Leif Erik Andersson, Francesco Scibilia, and Lars Imsland. An estimation-forecast set-up for iceberg drift prediction. *Cold Regions Science and Technology*, 131:88–107, 11 2016. ISSN 0165232X. doi: 10.1016/j.coldregions.2016.08.001.

[87] Boeun Kim, Jakob K Huusom, and Jay H Lee. Robust Batch-to-Batch Optimization with Scenario Adaptation. *Industrial & Engineering Chemistry Research*, 58(30):13664–13674, 7 2019. ISSN 0888-5885. doi: 10.1021/acs.iecr.8b06233.

[88] K Zhen Yao, Benjamin M Shaw, Bo Kou, Kim B McAuley, and D W Bacon. Modeling Ethylene/Butene Copolymerization with Multi-site Catalysts: Parameter Estimability and Experimental Design. *Polymer Reaction Engineering*, 11(3):563–588, 1 2003. ISSN 1054-3414. doi: 10.1081/PRE-120024426.

[89] Yunfei Chu, Zuyi Huang, and Juergen Hahn. Improving prediction capabilities of complex dynamic models via parameter selection and estimation. *Chemical Engineering Science*, 64(19):4178–4185, 10 2009. ISSN 00092509. doi: 10.1016/j.ces.2009.06.057.

[90] Caroline S.M. Nakama, Galo A.C. Le Roux, and Victor M Zavala. Optimal constraint-based regularization for parameter estimation problems. *Computers & Chemical Engineering*, 139:106873, 8 2020. ISSN 00981354. doi: 10.1016/j.compchemeng.2020.106873.

[91] Robert Hermann and Arthur Krener. Nonlinear controllability and observability. *IEEE Transactions on Automatic Control*, 22(5):728–740, 10 1977. ISSN 0018-9286. doi: 10.1109/TAC.1977.1101601.

[92] Nathan D Powel and Kristi A Morgansen. Empirical observability Gramian rank condition for weak observability of nonlinear systems with control. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 6342–6348. IEEE, 12 2015. ISBN 978-1-4799-7886-1. doi: 10.1109/CDC.2015.7403218.

[93] Zeinab Mahmoudi, Niels Kjølstad Poulsen, Henrik Madsen, and John Bagterp Jørgensen. Adaptive Unscented Kalman Filter using Maximum Likelihood Estimation. *IFAC-PapersOnLine*, 50(1):3859–3864, 7 2017. ISSN 24058963. doi: 10.1016/j.ifacol.2017.08.356.

[94] Dan Simon. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory & Applications*, 4(8):1303–1318, 8 2010. ISSN 1751-8644. doi: 10.1049/iet-cta.2009.0032.

[95] Gard Paulsen. *Alltid rabiat: Jens Glad Balchen og den kybernetiske tenkemåten*. Fagbokforlaget, 2019. ISBN 8245025461.

[96] Gilbert Strang. *Introduction to linear algebra*. Wellesley-Cambridge Press Wellesley, MA, 5th edition, 2016.

[97] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 1999.

[98] Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013. ISBN 1421408597.

[99] S Vajda, P Valko, and T. Turányi. Principal component analysis of kinetic models. *International Journal of Chemical Kinetics*, 17(1):55–81, 1 1985. ISSN 05388066. doi: 10.1002/kin.550170107.

ISBN 978-82-326-5395-9 (printed ver.)
ISBN 978-82-326-6334-7 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology