Conceptual Process Design in Fermentation-Based Biomanufacturing

Vollmer, Nikolaus

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Conceptual Process Design in Fermentation-Based Biomanufacturing

Nikolaus I. Vollmer

PhD Thesis
January 2022

Department of Chemical and Biochemical Engineering
Technical University of Denmark
Kgs. Lyngby
Denmark

"What we know is a drop, what we don't know is an ocean."

Sir Isaac Newton

Dedicated to my grandfather Helmut,
who was noticeably proud of me pursuing a PhD,
but also slightly concerned
and thus insisting on me wearing a lab coat at all times.

I might have spend most of my time behind a screen,
and not with a pipet in my hand,
hence the lab coat was not very necessary,
nonetheless, I think you would be very proud of the result.

Preface

This thesis is submitted as partial fulfillment of the requirements for obtaining the degree of Doctor of Philosophy at the Technical University of Denmark. The work has been carried out at the Process and Systems Engineering Center (PROSYS) at the Department of Chemical and Biochemical Engineering from November 2018 until January 2022.

The thesis was supervised by Prof. Gürkan Sin, professor at the Department of Chemical and Biochemical Engineering of the Technical University of Denmark, and co-supervised by Prof. Solange I. Mussatto, professor at the Department of Biotechnology and Biomedicine of the Technical University of Denmark, and Prof. Krist V. Gernaey, professor at the Department of Chemical and Biochemical Engineering of the Technical University of Denmark.

This project is part of the Fermentation-Based Biomanufacturing Initiative at the Technical University of Denmark and received funding from the Novo Nordisk Foundation (Grant no: NNF17SA0031362).

Kgs. Lyngby, 31 January 2022

Nikolaus I. Vollmer
M.Sc. (TUM)

Acknowledgments

First of all, I would like to express my fullest gratitude to my principal supervisor Gürkan Sin and thank him for all the guidance and inputs, and also the support and confidence during the three years, without which this entire PhD would certainly not have been possible to master.

Also, I would like to thank my co-supervisor Krist Gernaey for the valuable inputs regarding all aspects of biotechnology, but also more general advice within and particularly outside the boundaries of science.

Furthermore, I'd like to acknowledge my co-supervisor Solange Mussatto for her scientific input regarding biomass pretreatment.

I would also like to especially thank all my collaborators during this project, in particular, Resul Al for our fruitful collaboration and all the advice and support on the framework paper, the entire Biomass Conversion and Bioprocess Technology group, namely Celina Yamakawa, for the collaboration on the biomass pretreatment paper, my host department DTU Biosustain and all colleagues there, and Stig Olsen for the valuable insights and highly constructive feedback into the topic of Life Cycle Assessment in the scope of our joint paper.

Moreover, I'd also like to acknowledge my master students Philip, Josep, Sejr, and Binoy, for whom I had the pleasure to be their supervisor during their projects and who gave me the opportunity to sharpen my skills as an advisor.

An integral part of a PhD thesis is the right surrounding: hence, a big thanks to past and present colleagues at PROSYS, for common coffee breaks and conversation, for Friday bars, and all other events that made this time unique and unforgettable. Having said this, I also would like to thank the core of PROSYS, Anja and Gitte, a lot for their help with all administrative issues and their open door in all situations.

Apart from my direct colleagues, I would also like to express my thanks to the entire FBM community, involving all PhD students with whom we formed a community and substantially shaped parts of the program, involving discussion days and the outreach mission. This would not have been possible with the support of Friederike, John, Andreas, Dina, Bjarke, and lastly also the Novo Nordisk Foundation. A particular thanks for this.

According to the main character of a viral German TV series, "friends are the salt in the soup of life." My heartfelt gratitude and thanks to all my friends from back home and over the world, in particular, all those friends I found here in Denmark during this time, which made the PhD and, which is undoubtedly even more important, the life apart from it as enjoyable and simply great as it is. Thank you, Sophia, Vivi, Michi, Ralf, Nils, Michi, Anja, Maggie, and all the other Germans in my life, which I will most definitely see more often in the near future than during the past three years. Thank you in particular, Jonas, for having stayed within four hours' reach and everything fun that we have lived through between Hamburg and Buenos Aires – I honestly don't know how to beat this, but we will for sure find a way. Thank

you, Kate, for Russian delights – sweet and salty, creativity and craziness, amazing holidays and more things than I could possibly count – mi sofa es su sofa. Thank you, Jasper, for the Hertog Jan that resides on the shelf next to me since the time we started this PhD, for all the good times on the way, and for the day in the very near future when we finally open that bottle. Thanks to my two roommates, Daniele and Bram, who made Denmark home, the pandemic somewhat worth living, our kitchen into an international hotspot of Carbonara, AVG and Cardamom buns, and, most importantly, for making Strandboulevarden great again – the beans are coming. Thanks to the Quarantine Club for hygge Saturday evenings in lockdown and great holidays in Gran Canaria, to Gisela for shared coffees, shared worries, shared paella, and much more, Vicente for the good talks, very refreshing swims in the frozen harbour, and very late summer raves, Simoneta for never giving up on convincing us to take the way to Amager, for a fun course in Delft and Amsterdam, Jess for being one of the most hospitable Dane I had the pleasure to get to know and thanks to Anne-Lise and Pau. Thank you, Arrate, Viviënne, Max, Chris, and all the others who lived up to the Beach Business.

In free adaption to Desmond Tutu, "you don't choose your family, you are a gift to them as they are to you." Hence, my heartfelt thanks to my family, and in particular to parents, Regine and Friedrich, for all your unconditional support, even for the idea of leaving Germany for an indefinite amount of time to do a PhD, which I was not even entirely sure about back then. This perception changed, your encouragement did not. Thank you very much for this!

And last, but surely not least, to my girlfriend Amalie, who I most certainly was "annoying" at times during the last months of this PhD, and still you made these Danish winter days brighter. Thank you!

English Abstract

The transition towards more sustainable production patterns as postulated in the 2030 Sustainable Development Agenda of the United Nations is one of the major challenges of our generation. In the wake of this transition, implementing fermentation-based biomanufacturing processes to replace the current chemical manufacturing processes is a key element. Nonetheless, major obstacles for the conceptual design of fermentation-based biomanufacturing processes remain. An example of these biomanufacturing processes are second-generation biorefineries. They utilize lignocellulosic biomass to produce different chemicals, fuels, energy, or materials. Despite their vast potential and dedicated research toward them over the past decades, the number of commercially operating plants is low and falls short of expectations. This is mainly due to their deficient economic potential.

In the light of the elaborated status quo and reviews on biomanufacturing, biorefinery concepts, and process design methodologies in the scope of this thesis, four research questions are formulated and investigated. For this, a case study for xylitol production in a biomanufacturing process is selected. The objective is to design the process for such a second-generation biorefinery conceptually and to investigate both the economic potential and the sustainability impact of the process in order to draw specific conclusions for the case study. Based on the results, conclusions and recommendations for future work regarding the conceptual process design of fermentation-based biomanufacturing processes from a more general perspective are formulated.

The first research question addresses the biomass pretreatment, which is an essential unit operation in the upstream process to fractionate and depolymerize the lignocellulosic biomass to use the monomers for the subsequent fermentation. After defining criteria for the fractionation and depolymerization, the two most promising out of all potential candidates of pretreatment technologies are selected, and experiments are designed and performed. Based on the results, the criteria for both investigated pretreatment methodologies are evaluated.

The second research question addresses the need for a suitable computational framework for the conceptual process design. Based on the evaluation of the shortcomings of the three most widespread design methodologies, namely heuristics and expert knowledge-based design, superstructure optimization, and simulation-based optimization, a synergistic optimization-based framework is proposed. An essential part of the framework is the use of surrogate models to enhance the computational tractability of the optimization. Different surrogate models are validated, and their metrics are compared to assess the framework. Based on this, the framework is applied to a case study to benchmark the different surrogate models and ultimately validate the framework regarding its suitability for the conceptual process design of biomanufacturing processes.

The third research question addresses the economic feasibility of a potential biomanufacturing process for xylitol production in a second-generation biorefinery setup. To begin with the end in mind, the production of value-added co-products apart from xylitol is considered, as well as the potential of process integration through the internal production

of steam and electricity to increase the economic feasibility. Mechanistic models for all potentially employed unit operations are developed and assessed. With the previously introduced framework, the biorefinery process is then designed. Based on the results, a risk-based economic analysis is performed to assess the impact of uncertainties in the capital expenditures, the operational expenses, and the market prices for the products, allowing for a conclusion regarding the economic feasibility of the biorefinery process.

Lastly, the fourth research question addresses the sustainability impact of the designed biorefinery process. A standardized four-step life cycle assessment for the plant is performed. The first step involves defining the assessment's goal, scope, and boundaries. In the second step, the assessment inventory subsumes all relevant flows going in and out of the system. The third step characterizes the impacts of each flow in respective impact categories that are defined through the choice of a specific impact assessment methodology. After normalizing the characterization results to a reference point, in the fourth step, the assessment results are interpreted. Furthermore, based on available life cycle assessment data from the current chemical production processes of xylitol, the process is compared to two different commercial alternatives, allowing for a conclusion regarding the sustainability impact of the biorefinery process.

Lastly, by answering the four research questions, the case study is evaluated as a whole, and potential strategies for improvement are indicated. By taking into account the results from this work, the developments in synthetic biology, and the fourth industrial revolution, a blueprint of a future design strategy for fermentation-based biomanufacturing processes is given to expedite the transition towards more sustainable production patterns.

Dansk Resumé

Overgangen til mere bæredygtige produktionsmønstre som postuleret i FN's 2030-dagsorden for bæredygtig udvikling er en af vores generations store udfordringer. Dette involverer forskellige tilgange, fx cirkulær økonomi og øget forbrug af vedvarende ressourcer. I den forbindelse er implementeringen af fermenteringsbaserede biofremstillingsprocesser til erstatning for de nuværende kemiske fremstillingsprocesser et centralt element. Ikke desto mindre er der stadig store hindringer for det konceptuelle design af fermenteringsbaserede biofremstillingsprocesser. Et eksempel på disse biofremstilling er anden generations bioraffinaderier. De bruger lignocelluloseholdig biomasse til at producere forskellige kemikalier, brændstoffer, energi eller materialer. På trods af deres enorme potentiale og dedikeret forskning inden for dette felt gennem de sidste årtier er antallet af kommercielt fungerende anlæg lavt og lever ikke op til forventningerne. Dette skyldes primært deres mangelfulde økonomiske potentiale.

I lyset af den udarbejdede status quo og anmeldelser af biofremstilling, bioraffinaderikoncepter og procesdesignmetoder, blev fire forskningsspørgsmål formuleret og undersøgt inden for rammerne af denne afhandling. Til dette blev et casestudie for produktion af xylitol via en biofremstillingsproces valgt. Målet er at konceptuelt designe processen for et sådan anden generations bioraffinaderi og at undersøge både det økonomiske potentiale og bæredygtighedspåvirkningen af processen for at drage konkrete konklusioner for casestudiet. På baggrund af resultaterne formuleres konklusioner og anbefalinger til det fremtidige arbejde vedrørende det konceptuelle procesdesign af fermenteringsbaserede biofremstillingsprocesser ud fra et mere generelt perspektiv.

Det første forskningsspørgsmål omhandler biomasseforbehandlingen, som er en nødvendig enhedsoperation i opstrømsprocessen for at fraktionere og depolymerisere den lignocelluloseholdige biomasse, så monomererne kan bruges til den efterfølgende fermentering. Efter at have defineret kriterier for fraktionering og depolymerisering, udvælges de to mest lovende af alle potentielle kandidater til forbehandlingsteknologier, og eksperimenter designes og udføres. På baggrund af resultaterne evalueres kriterierne for begge undersøgte forbehandlingsmetoder.

Det andet forskningsspørgsmål omhandler behovet for en passende beregningsramme for det konceptuelle procesdesign. På baggrund af evalueringen af manglerne ved de tre mest udbredte designmetoder, nemlig heuristik og ekspertviden-baseret design, superstructure-optimering og simuleringsbaseret optimering, foreslås en synergistisk optimeringsbaseret ramme. En væsentlig del af rammen er brugen af surrogatmodeller for at øge optimeringens beregningsmæssige håndterbarhed. Til vurdering af beregningsrammen valideres forskellige typer af surrogatmodeller, og deres metrikker vurderes. Baseret på dette anvendes rammen på et casestudie med henblik på at sammenligne de forskellige surrogatmodeller og i sidste ende validere rammen med hensyn til dens egnethed til det konceptuelle procesdesign af biofremstillingsprocesser.

Det tredje forskningsspørgsmål omhandler den økonomiske gennemførlighed af en potentiel biofremstillingsproces til produktion af xylitol i en anden generations bioraffinaderi. For at begynde med slutningen i tankerne overvejes produktionen af værdiøgende biprodukter bortset fra xylitol, samt potentialet for procesintegration med intern produktion af damp og elektricitet for at øge den økonomiske gennemførlighed. Mekanistiske modeller for alle potentielt anvendte enhedsoperationer udvikles og vurderes. Med den tidligere indførte ramme designes bioraffinaderiprocessen derefter. På baggrund af resultaterne udføres en risikobaseret økonomisk analyse for at inkludere usikkerheder i anlægsinvesteringerne, driftsomkostningerne og markedspriserne for produkterne, hvilket giver mulighed for en konklusion om bioraffinaderiprocessens økonomiske gennemførlighed.

Det fjerde forskningsspørgsmål omhandler endelig bæredygtighedspåvirkningen af den designede bioraffinaderiproces. Der udføres en standardiseret fire-trins livscyklusvurdering for anlægget. Det første trin involverer en definition af mål, omfang og grænser for vurderingen og definitionen af en funktionel enhed og et referencestrøm. I andet trin er opgørelsen til vurderingen sammensat af alle relevante strømme, der går ind og ud af systemet. Det tredje trin karakteriserer virkningerne af hver strøm i respektive påvirkningskategorier, der er defineret gennem valget af en specifik konsekvensanalysemetode. Efter normalisering af karakteriseringsresultaterne til et referencepunkt, i det fjerde trin fortolkes vurderingsresultaterne. Baseret på tilgængelige livscyklusvurderingsdata fra de nuværende kemiske produktionsprocesser af xylitol, sammenlignes processen med to forskellige kommercielle alternativer, hvilket giver mulighed for en konklusion vedrørende bæredygtighedspåvirkningen af bioraffinaderiprocessen.

Til sidst, ved at besvare de fire forskningsspørgsmål, evalueres casestudiet som en helhed, og potentielle strategier til forbedring er angivet. Ved at tage hensyn til resultaterne fra dette arbejde, udviklingen inden for syntetisk biologi og den fjerde industrielle revolution, gives der en blueprint for en fremtidig designstrategi for fermenteringsbaserede biofremstillingsprocesser for at fremskynde overgangen til mere bæredygtige produktionsmønstre.

Structure of the Thesis

This PhD thesis is structured into two parts:

The first part is a synopsis of the status quo in research and contextualizes the PhD research.

The second part comprises the main publications of the PhD, namely the four journal articles, submitted and partly published during the PhD in thematic order:

A. Model Development for the Optimization of Operational Conditions of the Pretreatment of Wheat Straw

B. Synergistic optimization framework for the process synthesis and design of biorefineries

C. Conceptual Process Design of an Integrated Xylitol Biorefinery with Value-Added Co-Products

D. Life Cycle Assessment of an Integrated Xylitol Biorefinery with Value-Added Co-Products

In this thesis, the journal articles are referred to by the indicated letter (A, B, C, and D). An exhaustive list of all dissemination activities connected to the PhD, including other published work, is given in the following section.

List of Dissemination Activities

The following section presents a detailed overview of all dissemination activities performed in connection with the PhD project between November 2018 and January 2022 in chronological order. All dissemination activities connected to the PhD scheduled for February 2022 and onwards are italicized.

List of Journal Publications

- **N.I. Vollmer**, R. Al, K.V. Gernaey & G. Sin, 2021. Synergistic Optimization Framework for the Process Synthesis and Design of Biorefineries, *Frontiers in Chemical Science and Engineering*, Vol. 16 (2), pp. 251-273.

- **N.I. Vollmer**, J.L.S.P. Driessen, C.K. Yamakawa, K.V. Gernaey, S.I. Mussatto & G. Sin, 2022. Model Development for the Optimization of Operational Conditions of the Pretreatment of Wheat Straw, *Chemical Engineering Journal*, Vol. 430, Part 4, 133106.

- **N.I. Vollmer**, K.V. Gernaey & G. Sin. Conceptual Process Design of an Integrated Xylitol Biorefinery with Value-Added Co-Products, *Frontiers in Chemical Engineering*, accepted/in press.

- **N.I. Vollmer**, C.L.C.L. Gargalo, K.V. Gernaey, S.I. Olsen & G. Sin. Life Cycle Assessment of an Integrated Xylitol Biorefinery with Value-Added Co-Products, *Journal of Cleaner Production*, submitted/under review.

List of Conference Proceedings

- **N.I. Vollmer**, K.V. Gernaey, S.I. Mussatto & G. Sin, 2020. Surrogate Modelling Based Uncertainty and Sensitivity Analysis for the Downstream Process Design of a Xylitol Biorefinery, *Computer Aided Chemical Engineering*, Vol. 48, pp. 1663-1668.

- **N.I. Vollmer**, R. Al & G. Sin, 2021. Benchmarking of Surrogate models for the Conceptual Process Design of Biorefineries, *Computer Aided Chemical Engineering*, Vol. 50, pp. 475-480.

- **N.I. Vollmer**, K.V. Gernaey & G. Sin, 2022. Sensitivity Analysis and Risk Assessment for the In-Silico Design and Use of Optimized Cell Factories in a Xylitol Biorefinery, *Computer Aided Chemical Engineering*, accepted.

- **N.I. Vollmer**, K.V. Gernaey & G. Sin, 2022. Value Chain Optimization of a Xylitol Biorefinery with Delaunay Triangulation Regression Models, *Computer Aided Chemical Engineering*, accepted.

List of Co-Authored Contributions

- H.Yu, **N.I. Vollmer**, C.L.C.L. Gargalo, K.V. Gernaey, Ahmad Arabkoohsar & G. Sin. A process systems engineering view of environmental impact assessment in renewable and sustainable energy production: status and perspectives, *Computers & Chemical Engineering*, submitted/under review.

List of Conference Participations & Presentations

- **N.I. Vollmer**, K.V. Gernaey, S.I. Mussatto & G. Sin, 2019. Sustainable Value Chain Design for Biorefineries, *ModLife Project Conference*, 23.5. 2019 – 24.5.2019, Kgs. Lyngby, Denmark, poster presentation.

- **N.I. Vollmer**, K.V. Gernaey, S.I. Mussatto & G. Sin, 2019. Sustainable Value Chain Design for Biorefineries, *KT Consortium Annual Meeting 2019*, 12.6. 2019 – 14.6.2019, Helsingør, Denmark, poster presentation.

- **N.I. Vollmer**, K.V. Gernaey, S.I. Mussatto & G. Sin, 2019. Sustainable Value Chain Design for Biorefineries, *Nordic Process Control Workshop 22*, 21.8.2019 – 23.8.2019, Kgs. Lyngby, Denmark, poster presentation.

- **N.I. Vollmer**, C.K. Yamakawa, K.V. Gernaey, S.I. Mussatto & G. Sin, 2019. Experiment Design, Modeling and Comparative Design Optimization of the Pretreatment of Wheat Straw for the Sustainable Production of Xylitol, *European Conference of Chemical Engineering 12 (ECCE 12)*, 15.9.2019 – 19.9.2019, Florence, Italy, poster presentation.

- **N.I. Vollmer**, K.V. Gernaey, S.I. Mussatto & G. Sin, 2019. Process Design, Optimization and Supply Chain Analysis of the Biotechnological Production of Xylitol, *PROSYS Annual Research Seminar 2019*, 25.10.2019, Kgs. Lyngby, Denmark, oral presentation.

- **N.I. Vollmer**, C.K. Yamakawa, K.V. Gernaey, S.I. Mussatto & G. Sin, 2019. Modification of the Upstream Process Design in a Biorefinery by Modeling and combined Uncertainty and Sensitivity Analysis, *AIChE Annual Meeting 2019*, 10.11.2019 – 15.11.2019, Orlando, United States of America, oral presentation.

- **N.I. Vollmer**, C.K. Yamakawa, K.V. Gernaey, S.I. Mussatto & G. Sin, 2019. Uncertainty and Sensitivity Analysis assessing the Robustness of Models employed in Superstructure Optimization for the Downstream Process Design of a Biorefinery, *AIChE Annual Meeting 2019*, 10.11.2019 – 15.11.2019, Orlando, United States of America, oral presentation.

- **N.I. Vollmer**, K.V. Gernaey & G. Sin, 2020. Surrogate Modelling Based Uncertainty and Sensitivity Analysis for the Downstream Process Design of a Xylitol Biorefinery, *30th European Symposium on Computer Aided Process Engineering (ESCAPE 30)*, 31.8.2020 – 2.9.2020, virtual event, oral presentation.

- **N.I. Vollmer**, R. Al, K.V. Gernaey & G. Sin, 2020. Process Design of a Xylitol Biorefinery with a Hybrid Optimization Approach, *KT Consortium Annual Meeting 2020*, 2.10.2020, virtual event, oral presentation.

- **N.I. Vollmer**, R. Al, K.V. Gernaey & G. Sin, 2020. Process Design of a Xylitol Biorefinery with a Hybrid Optimization Approach, *Computer Aided Process Engineering Forum 2020 (CAPEFORUM 2020)*, 8.10.2020 – 9.10.2020, Kgs. Lyngby, Denmark, oral presentation.

- **N.I. Vollmer**, R. Al, K.V. Gernaey & G. Sin, 2020. Process Design of a Xylitol Biorefinery with a Hybrid Optimization Approach, *PROSYS Annual Research Seminar 2020*, 23.10.2020, Copenhagen, Denmark, poster presentation.

- **N.I. Vollmer**, R. Al, K.V. Gernaey & G. Sin, 2020. Synergistic Process Synthesis and Design Framework for Integrated Biorefineries, *AIChE Annual Meeting 2020*, 16.11.2020 – 20.11.2020, virtual event, oral presentation.

- **N.I. Vollmer**, R. Al & G. Sin, 2021. Benchmarking of Surrogate Models for the Conceptual Process Design of Biorefineries, *31st European Symposium on Computer Aided Process Engineering (ESCAPE 31)*, 6.6.2021 – 9.6.2021, virtual event, oral presentation.

- **N.I. Vollmer**, K.V. Gernaey & G. Sin, 2021. Process Design for the Biotechnological Production of Xylitol and Value-Added Coproducts, *KT Consortium Annual Meeting 2021*, 15.6.2021 – 17.6.2021, virtual event, oral presentation.

- **N.I. Vollmer**, K.V. Gernaey & G. Sin, 2021. Process and Value Chain Design of the Biotechnological Production of Xylitol and Value-Added Co-Products, *PROSYS Annual Research Seminary 2021*, 15.10.2021, Kgs. Lyngby, Denmark, oral presentation.

- **N.I. Vollmer**, K.V. Gernaey & G. Sin, 2021. Process Design for the Biotechnological Production of Xylitol by a Synergistic Optimization, *AIChE Annual Meeting 2021*, 7.11.2020 – 11.11.2020, Boston, United States of America, oral presentation.

- ***N.I. Vollmer*** *& G. Sin, 2022. Which surrogate models to use? – Benchmarking of Superstructure Optimization Approaches for Process Design*, Nordic Process Control Workshop 23*, 17.3.2022 – 18.3.2022, Luleå, Sweden, poster presentation.*

- ***N.I. Vollmer****, K.V. Gernaey & G. Sin, 2022. Sensitivity Analysis and Risk Assessment for the In-Silico Design and Use of Optimized Cell Factories in a Xylitol Biorefinery,* 32nd European Symposium on Computer Aided Process Engineering (ESCAPE 32)*, 12.6.2022 – 15.6.2022, Toulouse, France.*

- ***N.I. Vollmer**, K.V. Gernaey & G. Sin, 2022. Value Chain Optimization of a Xylitol Biorefinery with Delaunay Triangulation Regression Models,* 14th International Symposium on Process Systems Engineering (PSE 2021+), *19.6.2022 – 23.6.2022, Kyoto, Japan.*

List of Newspaper Articles

- **N.I. Vollmer**, K.V. Gernaey, S.I. Mussatto & G. Sin, 2019. Konceptuelt design af biobaserede værdikæder – for en bæredygtig fremtid, *Dansk Kemi*, Vol. 100 (7), pp. 20-21.

List of other Dissemination Activities

- **N.I. Vollmer**, 2019. The Magical Sweet (Popular Science Video), *https://vimeo.com/390933089*

Content

List of Figures

List of Tables

Part

I

Synopsis

1. Introduction

1.1 The pathway for and challenges in the transition towards sustainable production patterns

At the beginning of the year 2022, the global scene is dominated by several major crises. First and foremost, the COVID-19 pandemic heavily influences the political agenda. It induced severe economic destabilization and recession around the globe due to necessary countermeasures. This involves social distancing, quarantines, and lockdowns, which consequently led to disrupted supply chains and shortages for a significant number of different products worldwide, as well as an escalated volatility of prices on the global energy market [1,2].

While the COVID-19 pandemic unfolded only recently, it shares the same root cause with an even bigger current global crisis: anthropogenic climate change [3]. The International Panel on Climate Change (IPCC) summarized these root causes in the latest reports of working groups I as part of the sixth assessment report: Due to the increasing emissions of greenhouse gasses since the first industrial revolution, the global average temperature started to increase. The caused global warming leads to melting polar and glacial ice shields and permafrost soil. As a consequence, this causes rising sea levels, the loss of marine and terrestrial habitats, a dramatically accelerated mass extinction in various domains, a high increase in probability for extreme weather events, ultimately leading to a magnification of uninhabitable zones that surge mass migration, and lastly an overwhelming potential for vast social and political conflicts [4,5].

A significant part of these greenhouse gas emissions can be attributed to industrial processes for the production of materials and energy, which up to this day mostly use fossil resources as feedstock, as assessed by the International Energy Agency in 2018 [6]. The contribution of process industries to the energy-related CO_2 emissions is around 12% of the overall energy-related emissions in 2020 [7]. According to BASF SE, the global chemical production volume in 2021 is anticipated to grow by around 4.4 %, even higher than pre-pandemic annual growth rates [8]. Equally, the global demand for coal, natural gas, and fossil oil is anticipated to increase by 4.5%, 3.2%, and 6%, respectively, which leads to an overall growth of CO_2 emissions of 4.8% in 2021, surpassing even pre-pandemic levels [9]. Historically speaking, this corresponds to an increase of almost 50% of the concentration of CO_2 in the atmosphere compared to levels before the first industrial revolution [9].

Essentially, this crisis is driven by how humanity currently produces and consumes. Having stated this, the solution lies at hand: Humanity needs to find sustainable alternatives for ensuring livelihood for all humankind, for how it provides energy, for how it produces and consumes, for how the global economy is supposed to be sustained responsibly, and for how to take action against climate change.

As a consequence, nations globally agreed multiple times on common frameworks on how to combat climate change, as the consequences for humanity started to come to the attention of a broader public audience by the publication of the report “The Limits to

Growth" by the Club of Rome in 1972 [10]. Exactly half a century later, global leaders have ratified the United Nations Framework Convention on Climate Change in Rio de Janeiro in 1992 and subsequently signed the Kyoto Protocol and the Paris Agreement in the eponymous cities in 1997 and 2016. They reconvene annually for a climate change conference – the last one being hosted in Glasgow in October 2021 – to reinforce commitments and expedite efforts towards the postulated goals of combatting climate change [11–13].
In this context, the sustainability aspect has been taken up by postulating 17 sustainable development goals (SDGs) in the 2030 Agenda for Sustainable Development of the United Nations to target the specific issues listed in the preceding paragraph [14]. In particular, the twelfth goal addresses responsible consumption and production by efficiently managing resources and achieving economic growth without exploiting environmental resources [14].

This sustainability paradigm has also widely been developed and discussed in industry, politics, and science: One observes trends towards a circular economy, describing an economic system where a maximally possible value is extracted from feedstocks while producing minimal waste, and thus being beneficial for reaching the SDGs [15–17]. Furthermore, concepts such as efficient resource recovery, paraphrased by the bon mot "one man's waste is another man's treasure," and the use of renewable instead of fossil resources are commonly seen as essential building blocks for circular economy concepts [18–23]. Moreover, the augmented utilization of biomanufacturing processes instead of chemical manufacturing processes in connection with the concepts mentioned in the preceding paragraph harbors significant potential in this regard [24–26]. What makes the use of biomanufacturing systems fascinating is the vast palette of products that can be produced, reaching far beyond those of chemical manufacturing systems. Particularly in the last decade, the progress made in cell factory engineering allows for novel strategies to perform targeted genetic, metabolic, or other modifications in cell factories to design them as hosts for – until now – inaccessible process routes [27–29]. Furthermore, their versatile and flexible applicability and their sustainability potential excel compared to their chemical counterparts due to the commonly milder process conditions [25,26].

In this context, inspired by long-established fossil refineries, biorefineries are an emerging concept: similar to their fossil counterparts, they utilize a feedstock and convert it to different products, like chemicals, fuels, or energy [30]. Whereas the conversion processes within the biorefinery are not exclusively biochemical, the distinction to fossil refineries comes with the utilization of renewable, biobased feedstocks [31]. As a way to classify them, biorefineries are grouped into generations. While the first generation of biorefineries utilizes sugar or starch from dedicated crops, second-generation biorefineries utilize lignocellulosic biomass, commonly in the form of by-products stemming from agriculture and forestry or dedicated non-crop plants [32,33]. Third and higher-generation biorefineries utilize algae, biogas, or other feedstocks [33].

Within the hitherto depicted concepts and trends, biorefineries combine several of the mentioned aspects that enable an industrial transformation and are thus considered beneficial for successfully managing a transition towards the sustainable development goals. Notwithstanding, they face demanding challenges that are difficult to overcome: On

the one hand, third and higher-generation biorefineries do not yet possess a sufficiently high technology readiness level to be successfully implemented at larger scales [32,34,35]. On the other hand, the major difficulty with first-generation biorefineries, also comprising commercially operated biomanufacturing processes in general, is the food-versus-fuel dilemma due to their utilization of crops that otherwise can be processed into food [36,37]. These conditions currently leave second-generation biorefineries as the only viable and technologically promising alternative for a fast implementation to achieve the SDGs [30,32,38,39]. However, to date, there are merely around a couple of dozen commercially operating biorefineries, mainly producing bioethanol [32,38,40]. The principal reason for this is a deficient economic potential for biorefineries at a commercial scale to compete against well-established chemical synthesis and production routes [40,41].

Technologically speaking, the main difference between a first-generation and a second-generation biorefinery employing biochemical conversion processes are biomass pretreatment and enzymatic hydrolysis units in the upstream process. They serve to decompose the lignocellulosic biomass into monomers, which then are used equally as in the first generation setup as feedstock for the conversion processes [42,43]. While enzymatic hydrolysis is a complex but relatively straightforward process, this does not account for the pretreatment in the same manner. The most common and most investigated pretreatment processes are physical and chemical pretreatment processes that operate at rather harsh operational conditions to achieve significant monomer yields [42,43]. Due to this fact, besides the monomers, a vast palette of undesired by-products – commonly referred to as inhibitors – are formed, which have a heavy impact on the operation of all following unit operations [42,43]. This leads to the conclusion that the transition from first-generation to second-generation biorefineries is strongly impacted by the pretreatment unit and the implied mediate consequences of its use. Besides the deficient economic feasibility, these considerations also imply consequences for the sustainability potential of these biorefineries, which is hardly investigated [44,45].

Another paradigm shift that influences the transition towards more sustainable production patterns is driven by industry itself. Historically speaking, what started during the first industrial revolution with the utilization of steam power and subsequently continued throughout the second and third industrial revolution with the utilization of electricity and computational systems, is now culminating in the fourth industrial revolution and the extended use of data-driven technologies, commonly denominated Industry 4.0 [46,47]. These include cloud computing, artificial intelligence, machine learning, big data, and other highly potent technologies [48,49]. In particular, big data and artificial intelligence/machine learning opened unprecedented possibilities. In the context of the transition towards more sustainable production patterns and biotechnology, for the aforementioned cell factory design and optimization, machine learning tools enhance the use of data analysis of existing cell factories to facilitate the mentioned targeted design of novel cell factories [50,51]. Moreover, they can be used to predict new molecules of interest or even protein folding by the sheer power of large quantities of data [52].

On a process level, the collection of data facilitates the development of so-called digital twins of whole processes and even companies, which paved the way for optimized operation, predictive maintenance, or the real-time connection of different plants within a company

through the internet of things [53–57]. However, as significant as the introduction of these technologies in their respective domains are, they are not a magic bullet: Certain tasks still require the input of expert knowledge and heavily rely on several iterations and improvement cycles before being implemented in the industry [58].

In particular, the conceptual design of novel chemical or biomanufacturing processes is complex and, up to this date, by far too complex to be designed by the use of artificial intelligence or machine learning [59,60]. In the particular case of biomanufacturing processes, several points have to be considered: Firstly, the behavior of engineered or wild-type cell factories in a full-scale reactor environment and associated issues can be hardly predicted and also the scale-up of such processes from laboratory scale into commercial operation is a unique venture for each novel process [61]. In addition, also hurdles with the transition of novel technologies to commercial operation introduce additional uncertainties [40]. Lastly, market conditions as feedstock and product supply and demand, the concurrent production of the same molecule of interest by alternative, existing processes, and other constraints are uncertain [40].

Within the research discipline of process systems engineering, which deals with applying computational methods to process engineering tasks, various research projects target issues associated with Industry 4.0 in general and process design in particular. In return, tools such as machine learning also experience increased use for solving or speeding up the solution of computational issues within the discipline [60,62]. While, traditionally, knowledge-driven approaches were used, they more and more get complemented or replaced by data-driven approaches [63]: For example, mechanistic models contain a lot of information and knowledge regarding the underlying physical, chemical, or biological phenomena of the respective system, expressed by mathematical equations. Once they are entirely designed, this creates excellent insights and possibilities, but they are notoriously intricate and laborious to build [64]. With this transition, they find their antagonists in machine learning models that can be fitted to any data set without prior knowledge of the underlying system and without being explicitly programmed for it [59,60]. This makes them an easy-to-apply solution, but they are strictly and only valid for the applied system and within the boundaries of the original data [59,60,63]. Although both approaches serve as concurrent schools of modeling, two approaches aim at leveraging synergies by using them conjointly, which is also applied in process design: Firstly, hybrid models, that contain both elements of mechanistic models and machine learning models, can model parts of a system that are known with greater detail, and unknown parts with a generic machine learning approach [65]. Secondly, surrogate models, which are fitted to data created from a mechanistic model, are used as an identical replica, which can alleviate certain computational burdens associated with mechanistic models [66].

As opposed to the earlier mentioned process design by expert knowledge, conceptual process design is defined as finding an optimal process out of several candidates for a given objective and imposed boundary constraints [67]. Mathematically speaking, this is the definition of an optimization problem [67]. Regarding the setup and the solution of this optimization problem, however, there exist two different approaches within process systems engineering. On the one hand, superstructure optimization (SSO) describes the simultaneous evaluation of a superstructure of all possible process configurations for a

globally optimal solution. Unit operations are represented by models, although mechanistic models in this approach heavily compromise the computational tractability of solvers for the optimization problem, especially when including uncertainty in the calculations [67].
On the other hand, simulation-based optimization (SBO) refers to the application of enumerative evaluation of mechanistic models for different process design candidates, which easily also allows the inclusion of uncertainty but is restricted by the available computational capacity due to the laborious process of performing a high number of simulations for many process alternatives [68,69]. Which approach is more suitable for a specific design task depends on the level of detail of the existing mechanistic models, the number of potential process alternatives, and many more, hence a standard consideration towards which of the approaches should be followed, particularly for the design of biomanufacturing processes, does not exist to date. Nonetheless, for both approaches, surrogate models can be used to enhance computational tractability and reduce the computational burden [66–68]. The selection of a specific surrogate model, suitable for the specific task and underlying mechanistic model out of the tremendous number of available surrogate models, is hardly investigated, although it can heavily influence the results, making conceptual process design challenging [66,70,71].

In conclusion, the transition towards more sustainable production patterns and both aspects of biomanufacturing in biorefinery contexts, as well as the conceptual design of such processes, face several challenges regarding various aspects. Increasing efforts from research, industry, politics, and society are required to master this transition successfully. In the scope of this PhD thesis, several of the mentioned challenges were investigated, corresponding research questions were defined, solutions for the respective challenges were developed and proposed, and ultimately published as articles in peer-reviewed journals. For that matter, the following section focuses on particularly those research questions and depicts the remainder of this PhD thesis.

1.2 Research questions in the scope of this PhD thesis

All listed research questions are answered based on a case study, an integrated second-generation biorefinery for the production of xylitol, which serves as an example, as a general investigation on all biomanufacturing processes cannot be done exhaustively within the scope of this work. However, all aspects that account for biomanufacturing processes in general are also highlighted and discussed.

As pointed out in section 1.1 regarding the commercialization of second-generation biorefineries, the major challenges are linked to biomass pretreatment as an additional unit operation within the biorefinery process. The conceptual design of the biomass pretreatment unit has to meet several criteria, namely 1) achieving a high monomer yield for the respective fractions in lignocellulosic biomass, 2) achieving a good fractionation of the different fractions present in lignocellulosic biomass, and 3) achieving a low formation of inhibitors in the pretreatment process.

Based on these criteria, the first research question is:

- **Research Question 1:**
  Which suitable biomass pretreatment technologies best fulfill these criteria, and how should they be selected?

Regarding the conceptual process design as a subdiscipline in process systems engineering, different and partly concurring approaches can potentially be used. Furthermore, they employ surrogate models in various ways to facilitate computational issues. Hence, the second research question is:

- **Research Question 2:**
  Which process design approach is most suitable for biomanufacturing processes, and how should a particular surrogate model or several surrogate models be selected?

Based on these two more fundamental questions, the overshadowing issue related to second-generation biorefineries is their deficient economic feasibility. Hence, given that satisfactory answers for research questions 1 and 2 are found, the imminently affiliated third research question is:

- **Research Question 3:**
  Can a conceptionally designed second-generation biorefinery be a cost-competitive alternative to existing chemical processes, and what products should be produced in such a setup?

Lastly, given the indications in section 1.1 regarding the sustainability potential of such biorefineries and their potential environmental impacts, apart from potentially lower greenhouse gas emissions, the fourth and last research question is:

- **Research Question 4:**
  Is a second-generation biorefinery more sustainable than the alternative chemical process, and can it play a key role in the transition towards more sustainable production patterns?

For a detailed analysis of all these questions and the detailed elaboration on the answers found in the scope of this PhD thesis, the remainder of part I of this thesis is structured as follows:

In chapter 2, the mentioned aspects regarding biomanufacturing and biorefinery concepts are reviewed and explained in more detail in section 2.1 and the specific topic of biomass pretreatment in section 2.2. Both serve as the basis for the actual explanation and thematic classification of the research in chapter 4 and conclusions in chapter 5. The chapter is concluded by a short analysis of current trends within biomanufacturing in section 2.3, which will be taken up again in connection with the outlook in chapter 5.

In chapter 3, the mentioned aspects regarding conceptual process design and approaches are reviewed and explained in more detail in section 3.1, as well as the aspects regarding mechanistic modeling and surrogate modeling for the conceptual process design in section 3.2. Aspects of chapter 2 are taken into account, both as a basis for the actual explanation and thematic classification of the research in chapter 4 and the conclusions from it in chapter 5. The chapter is concluded by a short analysis of current trends within process systems engineering in section 3.3, relating to the trends outlined in section 2.3, which will be taken up again in connection with the outlook in chapter 5.

Chapter 4 contains the explanation of the used case study, on which the findings of the research are applied, followed by the explanation of the research performed in the scope of this PhD thesis and the thematic classification, based on the four main journal publications in part II of the thesis in order to answer the postulated questions in this section.

Lastly, chapter 5 contains a conclusion of the finding in section 5.1. In sections 5.2 and 5.3, an overview of potential future directions in research is provided, based on the particular findings in this thesis and in connection with other trends as depicted in sections 2.3 and 3.3 for a perspective outlook.

2. Review on Biomanufacturing

2.1 The biorefinery concept

2.1.1 Classification and promising concepts

Humans have used biotechnology in general for several millennia, initially to process or preserve food, e.g., ethanol or lactic acid fermentation to produce wine, yogurt, or sauerkraut. While the term biotechnology refers to the entire scientific domain, biotechnological production is henceforth denominated as biomanufacturing. This refers to production processes based on wild-type or engineered cell factories, also referred to as microorganisms, or enzymes [26]. The latter is, however, excluded from the scope of this thesis. The cell factory-based conversion in the biomanufacturing process is also called fermentation, following the early roots of this technology [26,72].

Biorefineries are a conceptual alternative to existing fossil refineries that produce a vast palette of products while utilizing renewable feedstocks instead. Biorefineries can be differentiated according to the products they produce, the employed conversion processes or the utilized feedstock, or the platform intermediates, which serve as a substrate for the conversion processes [33]. The conversion processes within the biorefinery can be either chemical, thermochemical, physical, or biochemical processes, or combinations of those, depending on the desired product [33]. These are commonly biofuels, like bioethanol, biodiesel, or synthetic fuels, gaseous products, biochemicals, like polymers, chemical building blocks, fine chemicals, and specialty chemicals, like fertilizers or food products for human or animal nutrition [33].

In the scope of this thesis, biorefineries are classified according to the utilized feedstock. As enumerated in chapter 1, biorefineries that utilize glucose or starch from dedicated food crops are considered first-generation biorefineries [30]. In opposition, biorefineries that utilize lignocellulosic biomass in the form of agricultural or forestal residues or dedicated non-food crops are considered second-generation biorefineries [30]. This involves either using a pretreatment unit to break down the polymeric structure of the lignocellulose or using a gasification unit to gasify the biomass and provide bio-syngas as a substrate for fermentation [32]. Third-generation biorefineries use micro- or macroalgae, and fourth-generation biorefineries integrate biogas, other fossil gas sources, or even CO_2 as carbon source and electricity from potentially renewable production systems as electron source [32].

Given the definition by feedstock, all current industrial biomanufacturing processes represent first-generation biorefinery processes. However, while existing biomanufacturing processes only require a small fraction of the globally produced sugar and starch, a more widespread implementation of first-generation biorefineries, especially for the production of fuels, would intensify the concurrence for these crops, the already mentioned fuel-vs-food dilemma [36,37,39]. Ultimately, this development would counteract the achievement of the SDGs as it stands diametrically to, e.g., the goal of ending hunger on a global scale and achieving food security [14,36,37]. Hence, producing a globally sufficient amount of biofuels, particularly bioethanol, is seen critically and is generally not considered

viable. Nonetheless, more than 1000 first-generation biorefineries exist, particularly in the US and Brazil, producing bioethanol from corn or sugar cane [32].

For third or fourth-generation biorefineries, major problems lie within the current technology readiness level of the associated technologies. While algae biorefineries are viable processes, globally relevant production volumes implicate large cultivation areas for algae. Their growth is based on photosynthesis and thus only happens in proximity to a light source and thus scales by the area and not by the volume. This requires extensive cultivation areas for higher production capacities [32,73]. For 4G biorefineries, the problems lie much more within actual technological barriers as the actual challenge of providing electrons to the cell factory is not entirely solved yet but rather an active research topic [74]. Hence, as for now, these fourth-generation biorefineries represent potentially interesting technologies, but the potential is not utilizable at an industrial scale at the moment [32,74,75].

As described in chapter 1, this leaves second-generation biorefineries as the only viable alternative for the large scale production of chemicals and fuels through biomanufacturing processes on a long-term perspective: on the one side, the technology readiness level of second-generation biorefineries is advanced enough, and on the other hand, the feedstock – lignocellulosic biomass – is abundantly available, in fact being the largest amount of organic matter on earth [43]. Due to this fact, a second-generation biorefinery is potentially versatile in its use: The biorefinery can process different kinds of feedstock that are available in a specific location, reducing the efforts associated with the supply chain, i.e., costs for transportation [76]. This opens manifold possibilities for the use in developed and developing countries, where biorefineries with smaller scales could be successfully implemented in biobased value chains to provide sustainable economic income and livelihood for people, according to the sustainable development goals [77–81].

Most first-generation and the few existing second-generation biorefineries have in common that they mainly produce ethanol [32,82]. Despite a tremendous number of studies and investigations on a laboratory scale regarding various products for biorefinery setups, almost none of them managed the transfer into a commercial context. This raises critical questions about the underlying problems of their implementation and the sense of purpose of the various studies regarding biorefinery setups and the myriads of possible products [40,83,84]. In addition, the global market price for ethanol is comparatively low due to low fossil oil and gas prices, which decreases the operative margins for second-generation biorefineries [83,84]. Besides this, the current global industrial and political focus shifts towards alternative forms of terrestrial transportation, namely battery electric vehicles or fuel-cell electric vehicles, and prioritized investment in rail-bound transportation [85]. This development will significantly reduce the market size for biofuels, putting another question mark behind ethanol as the sole product for future biorefineries [85–87]. From a conceptional point of view, this raises the question about which second-generation biorefinery setup shows most potential from both a technological as well as an economic perspective. Hence, the following paragraphs will depict several of these aspects.

A first approach is the production of products with more promising features than bioethanol: the US Department of Energy worked on this issue before and subsequently

released a list of the Top 12 most promising chemicals produced in a biorefinery setup in 2004, e.g., levulinic acid, itaconic acid, 3-hydroxybutyrolacetone or xylitol [88]. This list has been revised and appended several times, e.g., with building blocks as succinic acid, whereas others have been removed again, as their potential turned out to be minor compared to the initial expectations [89,90]. Lignocellulosic biomass contains different fractions, which will be further explained in section 2.2, allowing the production of different products but inherently limiting the yield of a product that is dependent on one particular fraction. One strategy for cell factory engineering is to genetically modify the cell factory to metabolize different substrates, according to different fractions in the biorefinery, to increase the overall product yield in the fermentation [91]. Alternatively, the strategy to produce value-added co-products, apart from the main product, can be followed, increasing the economic resilience of the biorefinery significantly [88,91].

Both these concepts incorporate the idea of a circular economic pattern by using as much of the available feedstock as possible. While both concepts aim to utilize the feedstock in a conversion process for production, process integration is an alternative way of increasing feedstock utilization and other utilities to reduce operational expenses (OPEX). Besides classic forms of process integration, such as the synthesis of a heat exchanging network or recycling streams, another possible way of integration in second-generation biorefineries is using the lignin in the lignocellulosic biomass for combustion instead of production. Here, the released energy is used for steam and power generation, which can be subsequently integrated with the downstream processing of the biorefinery [92,93]. Besides process integration, process intensification is a proposed way to increase the economic feasibility by decreasing the capital expenditures (CAPEX) through a reduced number or reduced size of unit operations. Examples are simultaneous saccharification and fermentation units, where the enzymatic hydrolysis and the fermentation are performed in the same unit operation, or the intensification of the fermentation unit through in-situ product removal [26,94].

2.1.2 The status quo of their implementation and associated challenges

For a more detailed overview of the current implementation status of biorefineries, the results from different databases are compared. The Joint Research Council of the European Union lists a total of 2186 commercially-operating biorefineries of all types and generations within the EU for the year 2020, of which 376 are second-generation biorefineries that produce chemicals from lignocellulosic biomass [95]. Specifications regarding the operating companies or the types of products are not indicated. Hence, this number provides a very optimistic guess. The European Technology and Bioinnovation (ETB) platform database only lists 312 biorefineries worldwide for the year 2019, of which 30 are commercially operating second-generation biorefineries that process lignocellulosic biomass [96]. All those 30 biorefineries produce ethanol, pyrolysis oil, or other fuel products. A similar picture is drawn based on the database of the Bio-Based Industries Consortium in 2017: they list a total of 224 biorefineries, of which only five biorefineries utilize lignocellulosic biomass for the production of fuels [97]. Lastly, a list of the Biorefineries blog regarding commercially operating biorefineries within Europe in the year 2018 shows 35 commercially operating biorefineries in total, of which one single plant by GFBiochemicals

produces levulinic acid from lignocellulosic biomass [98]. The categorized results from the database of the ETP are additionally displayed in Figure 1.

Figure 1: Overview of the numbers of biorefineries worldwide according to different levels of categorization.

The obtained picture is rather pessimistic. Firstly it is visible that around 75% of the commercial-scale biorefineries are not operating but either still in construction or not in operation anymore due to diverse associated issues. Furthermore, the total number of biorefineries that use lignocellulosic biomass as a feedstock only amounts to a third of all of these, the other two thirds utilize either gasification processes or first or other-generation feedstocks. Lastly, none of those lignocellulosic biorefineries produce chemicals other than the mentioned fuels.

A report by biofuelswatch, a UK-based non-profit organization, draws an even more dystopic picture: out of eleven commercial-scale biorefineries producing ethanol, the product which is most investigated and seen as the one with the slightest technological hurdles, all of them either failed and are not operating anymore, or report technical issues, e.g., lower production capacities as projected or others, hence declaring biorefineries a dead-end road [99]. This listing could be easily expanded with other examples, e.g., biomanufacturing processes for succinic acid or others [100,101]. Notwithstanding, in the near future, according to the EU biorefinery outlook for 2030, the conservative scenario calculates with three to five new second-generation biorefineries with value-added

products from different lignocellulosic fractions while assuming potential failure rates during their construction and implementation[102].

This underlines the apparent divergence between the perceived benefits of this technology and the actual number of these plants on a commercial level due to their deficient economic feasibility. As the fragile economic potential of biorefineries is also commonly known, the increasing number of publications in this area and the high number of announced construction projects indicate a huge confirmative bias in academia and other institutional domains towards their economic feasibility [31,38,43,94,99,103,104]. In point of fact, an optimistically estimated maximum of a couple fully implemented and operating plants today do not reflect the high number of research projects from a realistic point of view [32,38,40,99,105].

A first contributing factor to the deficient economic potential is the current fossil oil price, which is generally low compared to other potential feedstocks and, largely speaking, did on average only increase slightly over the past 40 years [106]. While from a sustainability perspective, the relegation of fossil resources is rather evident, from an economic perspective, the transition towards biomanufacturing is rigorously dictated by competitive product prices through alternative production routes. To this day, the overwhelming majority of research on biorefineries is based on an expected increase of the fossil oil price, which increases operative margins for biomanufacturing processes [40].

A second main reason is the additional requirement for a pretreatment and an enzymatic hydrolysis unit to process lignocellulosic biomass. While the technical difficulties regarding this aspect will be explained and evaluated in section 2.2, from an economic perspective, these additional unit operations imply increased CAPEX and, indirectly, increased OPEX for the entire biorefinery. Equally, additional necessary equipment for implementing a multi-product strategy or increasing the process integration leads to an increased CAPEX, causing the same effect on the overall economics. Lastly, an inherent problem for second-generation biorefineries are lower substrate concentration levels. This induces greater efforts for the downstream processing, which increases the OPEX of the plant significantly [107]. Overall, several of these contributing factors in combination affect the economic potential to an extent where most of these projects are rendered uncompetitive [40].

Notwithstanding, this issue is known. In order to elaborate more closely on this, the following paragraphs will summarize the topic from a political, industrial, and societal perspective: When taking a closer look at global, national, and even regional policies, the promotion of technologies that can be summarized as biorefinery concepts is considered important, especially with respect to the mitigation of the effects of climate change [83]. However, despite funding general research initiatives, state institutions are rarely initiating the actual construction of such biorefineries themselves: Instead, they are holding responsibility for creating reliable frameworks for industrial stakeholders to decide on investments in biorefineries [40]. Potential options to facilitate the promotion of biorefineries, therefore, target the competitiveness of biorefineries compared to fossil fuel-based process alternatives: This either can happen through direct subsidies or tax reliefs for these plants or otherwise indirect support through, e.g., the taxation of fossil CO_2 emissions, which would counterbalance the low margins due to the low fossil oil price

[13,83]. Nonetheless, this is generally not advanced in most policy-making frameworks [40,83].

From an industrial perspective, the decision in favor or against installing a biorefinery depends on the calculated profitability of the future project. For a few projects, this is doomed profitable, for example, in the case of a new second-generation bioethanol plant of Clariant or Energochemia [108,109]. However, these cases are rare exemptions. In general, these investments are still considered high-risk ventures [40]. One potential and often-considered way of mitigating these risks are joint ventures of larger industrial companies together with smaller pioneering companies to leverage synergies between existing production facilities and specialized knowledge [40]. Nonetheless, this is not always successful, as the case of bio-based production of succinic acid shows, but might play a more important role in the future [40,101].

Thirdly, the public debate regarding sustainability gains more interest and importance from a societal perspective: The potential of biotechnology as a whole is commonly seen with interest, not only in developed but also in developing countries [40,72]. On the other hand, society's awareness regarding its consumption behavior, the origin of the consumed products, and the effects on the environment also increases [72]. Advances, e.g., in genetic engineering, also face skepticism [72,110]. A very current example are mRNA vaccines: despite their immense potential in the fight against the COVID-19 pandemic, the communication regarding potential benefits and risks from a scientific perspective for a more general audience is challenging and can either decrease this skepticism or increase it even more [111,112]. Hence, to successfully promote these technologies and gain societal acceptance, appropriate communication about the safe and responsible use of these technologies is of crucial importance [113,114].

2.2 Lignocellulosic biomass as feedstock and its pretreatment

Lignocellulose is the constituting structural polymer of plant biomass and is hence one of the most abundantly available biomasses on earth. In general, lignocellulosic biomass consists of three main fractions, namely hemicellulose (HF), cellulose (CF), and lignin (LF). All three possess a polymer structure and are composed of typical monomers: Hemicellulose consists mainly of pentose sugars, namely xylose, and smaller amounts of arabinose. Furthermore, hexoses as glucose, mannose, and galactose are found in smaller quantities. Weak α-1,4-glycosidic bonds link them to longer chains, from which side chains with acetyl groups, galacturonic acid, or glucuronic acid branch off [115]. Cellulose consists mainly of glucose monomers connected by stronger β-1,4-glycosidic bonds, forming longer chains [115]. Lignin is a mostly heterogeneous macromolecule consisting of the three monolignols p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. Their fractions depend on the type of feedstock [115].

In general lignocellulosic biomass as feedstock for conversion processes has its provenance in agriculture and forestry, of which the latter can be divided into softwood and hardwood crops [43]. As an example for the agricultural sector, most crop residues are lignocellulosic biomass [43]. There are also dedicated crop plants solemnly used as feedstock for lignocellulosic biorefineries and stand aside from the food-energy nexus of other crops [43].

However, residues from other industries constitute a significant amount of unused resources at low costs, thus fulfilling the demand for sustainability much better [43]. Furthermore, waste streams from municipal waste treatment plants or paper mills also can be used as feedstock for a biorefinery. Table 1 shows the different classes of lignocellulosic biomass and their average percentage of hemicellulose, cellulose, and lignin.

Table 1: Overview of different biomass types and their typical composition [43]**.**

Biomass Type	Examples	Cellulosic Fraction [%]	Hemicellulosic Fraction [%]	Lignin Fraction [%]
Forestal residues				
- hardwood	Eucalyptus, willow, oak	37-50	12-23	18-29
- softwood	Pine, spruce	42-50	22-27	20-27
Agricultural residues	Sugarcane bagasse, brewers spent grains, rice straw, wheat straw, corn cob, sorghum	11-40	16-36	6-27
Herbaceous crops	Switchgrass, miscanthus	22-39	15-29	10-32
Waste	Cellulose sludge, waste paper	30-70	10-20	5-15

There exist several different pretreatment methods, which serve different purposes. Despite the efforts, there is no standard pretreatment procedure as this highly depends on the feedstock beyond other factors [42]. Furthermore, there is still a significant knowledge gap regarding most pretreatment methods' underlying mechanisms, which is also due to the high number of substances present and differences in the three-dimensional structure of the lignocellulosic biomass [116,117]. However, all the methods have in common that they break up the recalcitrant and polymeric structure of the lignocellulosic biomass by either biological, chemical, or physical means or even a combination of these [42,43].

All listed methods have in common that similar fragmentation and depolymerization reactions occur depending on the pretreatment reaction conditions. However, their conversion rates, selectivities, and yields vary accordingly. The released monomers of the hemicellulosic and cellulosic fraction participate in subsequent degradation reactions, leading to the formation of smaller molecules [116]. Those mostly have inhibitory effects on catalytic and microbial processes, depending on the catalyst's or cell factory's tolerance to each of these substances [118,119]. When several inhibitors are present, which is usually the case, different inhibitors can have reciprocally potentiating effects [120]. The main inhibitors are furfural, 5-hydroxymethylfurfural (5-HMF), acetic acid, formic acid, and – to a certain extent – also phenolic compounds, e.g., vanillin, depending on the particular pretreatment [116].

In Table 2, an overview of the most used pretreatment methods, their mechanism, and purpose are indicated according to Galbe and Wallberg [42]:

Table 2: Overview over different pretreatment methods, their mechanism, and the purpose [41].

Pretreatment method	Mechanism	Purpose
Acid (standard)	Catalysis with protons (added acid)	Breakdown of CF & HF
Acid (dilute)	Catalysis with protons (added dilute acid)	Breakdown of HF
Alkali	Catalysis with hydroxyl ions	Breakdown of LF
Hydrothermal	Catalysis with protons (autoprotolysis)	Breakdown of HF
Steam explosion	Pressure drop + Expansion of water	Breakdown of HF
Ionic Liquids	Large organic cation and small inorganic anion	Extraction of LF
Deep-Eutectic Solvents	Mixture of Lewis and Brønsted acids and bases	Extraction of LF
Organosolv	Organic solvents (+catalyst)	Extraction of LF
Biological	Enzyme catalysis with enzymes from xylophagous fungi	Fractionation and breakdown of either HF, CF, or LF

Several criteria should be fulfilled optimally by the chosen pretreatment method. These are namely:

- a sharp split between fractions,
- high respective yield(s), and
- low inhibitory compound formation [119].

It lies in the nature of any pretreatment that not all the criteria can be fulfilled to the same extent. On the other hand, this opens possibilities for manipulating the respective pretreatment method in the context of the whole process and should be investigated conceptually.

2.3 Advances in biomanufacturing for industrial processes

Several quantum leaps as the discovery of DNA by Rosalind Franklin, James Watson, and Francis Crick, the development of the polymerase chain reaction to amplify DNA, developed by Kary Mullis, and most recently, the development of CRISPR/Cas as a method for targeted editing of DNA, developed by Emmanuelle Charpentier and Jennifer Doudna, catapulted biotechnology to today's level and beyond [121–123]. DNA as information storage for the genome of cell factories thus presents one of several primary targets for engineering them to induce beneficial properties for biomanufacturing processes.

While early biomanufacturing processes and even up to this date many processes use wild type cell factories, engineered cell factories have several advantages – but also limitations – that will be briefly summarized in the following section.

Wild-type cell factories, as they can be found in nature, experience the influence of evolutionary pressure through the conditions in their habitat and are commonly perfectly adapted to their environment through long periods of natural evolution. The underlying process is called mutagenesis, referring to non-targeted mutations in the genome of organisms, which ultimately translates into altered properties that can be either favorable or unfavorable for the organism's survival in its environment [124]. As more beneficial properties provide fitness benefits to the mutated organisms, those commonly become predominant over time, while those with less beneficial properties become extinct within a population. Naturally, this process can take days, weeks, years, or even millennia, depending on the organism and the external conditions [124]. As they are perfectly adapted to their natural environment, artificial environments as a biomanufacturing process pose challenges to their performance. Consequently, this instigates research about their optimization through biotechnological tools [32].

The three most prominent key performance indicators for cell factories in biomanufacturing processes are the product yield, the volumetric productivity, and the product titer: While the product yield indicates how much of the used substrate was converted into the desired project, the volumetric productivity indicates how much of the product is produced per reactor volume unit and time unit [26]. The product titer indicates how much of the product is present per reactor volume at the end of the fermentation [26]. In general, while being interdependent to a certain level, all three should be evaluated to assess the suitability of a cell factory for a biomanufacturing process.

Principal engineering strategies for optimizing cell factories aim at one or multiple of those three key performance indicators. Besides these indicators, the overall performance of the cell factory in the reactor environment with respect to its stability and reaction to rapidly changing conditions due to gradients in the reactor can be other targets [26,125]. Common strategies for this are the overexpression, downregulation, or even knockout of genes to alternate the metabolism, the shift of co-factors, or the exchange or addition of genes [126]. While the prior strategies aim for the optimization of existing metabolic pathways in cell factories, in the case of the latter strategy, this ultimately allows the *de novo* introduction of heterologous metabolic pathways into a cell factory, opening the door to adaptable platform cell factories [29,127]

The employed techniques are scattered throughout various research disciplines within biotechnology, e.g., genetic engineering or metabolic engineering, and can be commonly described as rational or irrational approaches [126]. Genetic or metabolic engineering approaches are usually considered rational approaches and were only enabled recently through the manifold developments in biotechnology, as described in the first paragraph of this section. In addition, irrational approaches, e.g., adaptive laboratory evolution, mimicking extreme environmental conditions to promote mutation and expedite the random mutagenesis, are valuable tools in cell factory optimization [126,128].

Several examples of these approaches can be given for biorefinery applications, as these vary in particular due to the specific requirements concerning the process conditions in the biorefinery. For example, to increase the tolerance of cell factories to lower pH values of the substrate due to acid pretreatment, adaptive laboratory evolution as an irrational design approach can be successfully used to adapt cell factories to these environments [128]. Also, the co-utilization of hemicellulose and cellulose sugars from the pretreatment in one cell factory is a strategy to increase the overall product yield. This can be achieved by either overexpressing transporters for the hemicellulosic sugars to reduce the effect of catabolic repression of glucose or by substituting co-factors to allow the cell factory to process both substrates simultaneously [129]. Lastly, the titer for certain products can be increased through rational or irrational design approaches to reach industrially relevant concentrations and reduce downstream processing costs for products that are supposed to be produced in alternative biomanufacturing processes beyond others [130].

It needs to be emphasized that cell factory engineering usually has to be performed in several cycles to achieve a satisfactory level of improvement of the cell factory's performance: this is commonly referred to as the design-build-test-learn cycle in biotechnology [131]. The modification of target genes or metabolic pathways can increase the metabolic burden or decrease the performance of the cell factory in undesired ways, e.g., increasing the product yield but significantly decreasing the productivity of the cell factory due to the limited availability of metabolites and co-factors [131,132].

3. Review on Conceptual Process Design

3.1 Process design methodologies

3.1.1 Classification of existing methodologies

Within process systems engineering, process design is one of the major subdisciplines [67]. As mentioned in the introduction in chapter 1, process design is a highly complex and somewhat difficult task, as it has to deal with knowledge input from various scales and levels of details, which all are inherently uncertain to a respective degree. The general dilemma in process design resulting from this nexus is a balance between design resource efficiency, optimality, and exploration of all possible realizations of a future process.

The traditional way of designing a by expert knowledge commonly involves experts from different disciplines with prior knowledge on already existing processes from former design tasks [67,133]. Based on this prior knowledge, discussions about the design task commonly include this knowledge and other heuristics, estimations, and even guesses, depending on the influence of the decision on the entire design task [67,133,134]. Errors that have been committed in prior design tasks or targets from earlier process optimization steps ideally are also integrated at this point in order to find a good process design. This kind of procedure is also referred to as heuristic-based design [133,134]. While the design resource efficiency and the exploration of potential alternatives largely depend on the experts performing the process design, the found process design is not necessarily an optimal solution to the problem [67,133,134]. Nonetheless, this design strategy is still widely followed throughout all relevant branches of industry [67].

The first computational strategy for the conceptional design of processes is superstructure optimization (SSO) and is based on the idea of treating the design process as an optimization problem [135]. All potential process realizations are initially combined into a superstructure, which figuratively speaking is a process flowsheet with decision elements representing all possible topologies [67,135]. The superstructure can theoretically contain any possible design decision as a binary variable, e.g., different process trains, alternative unit operations, or potential heat exchanging networks [67,135]. The entire superstructure has to be described mathematically and set up as a mathematical program, which allows for the formulation of an optimization problem [135]. Subsequently, the optimization problem can be solved by a suitable optimization solver to global optimality, which yields the optimal process design out of all investigated options [135].

Some of the more common superstructure formulation strategies are state-task networks, or, complementarily, p-graph representations, where states describe the thermodynamic state of a mass or energy flow, and tasks describe a conversion, separation, or another process [135]. Alternatively, state-equipment networks, or, complementarily, r-graph representations, can be used. Here, states describe the thermodynamic state of a mass or energy flow, and equipments describe a piece of respective equipment to perform the desired task [135]. Furthermore, other more general or specific approaches, such as the state-space representation, generalized modular framework representation, group contribution-based methodology, phenomena-based building blocks, unit-port-

conditioning-stream representation, or building block representation, were all developed to fulfill specific requirements for the individual design cases [135].

Different types of models are commonly used in each type of superstructure formulation. They can be differentiated based on their included level of detail: Low-fidelity models commonly focus on predicting one variable, governed by the most dominant effects within the underlying system, e.g., the yield of a reaction, calculated through simple equations [67]. While the model is relatively simple, the prediction quality can be somewhat inaccurate. Medium-fidelity models commonly also predict a smaller number of variables based on several dominating effects and hence represent a compromise between the level of detail and accuracy in prediction [67]. On the other end, high-fidelity models include all levels of details of the underlying system and aim to predict several variables commonly based on a higher number of equations and with great accuracy, depending on the quality of the model parameters [67].

In general, due to the occurring phenomena in unit operations being of nonlinear nature and the involved decision variables to indicate selections between different unit operations, translating a superstructure formulation into a mathematical program yields a mixed-integer nonlinear program (MINLP) [67]. Even with small superstructure sizes and low-fidelity models, the resulting MINLP can become hard to solve or even intractable from a computational point of view [67,135]. As this is the major problem with performing SSO, several solution strategies have been developed in order to circumvent this issue: The first one is called general disjunctive programming, which, especially when using p-graph or r-graph superstructure formulations, allows for the introduction of certain logic statements and useful reformulations of the superstructure formulation, which facilitate the computational tractability significantly [135]. Secondly, as mentioned in chapter 1 and further explained in section 3.2, surrogate models offer a way to replace high-fidelity models with medium or low-fidelity models that still possess similar prediction capacities. This maintains the prediction quality in the optimization problem while facilitating the computational tractability through simpler model formulations [135]. Lastly, decomposition algorithms, which decompose the MINLP in, e.g., an outer and an inner optimization problem, facilitate the computational tractability by reducing the program size and structure [135]. Different implementation platforms and computational solvers can be chosen depending on the specific superstructure formulation, employed models, type of optimization problem, and solution strategy. For a more extensive overview regarding these points, the reader is referred to the excellent reviews of Chen & Grossmann and Mencarelli et al. [67,135].

A general difficulty for SSO approaches is the inclusion of uncertainty into the optimization problem. Per se, the resulting mathematical program has a deterministic solution, which does not account for any kind of uncertainty. It is possible to transform the deterministic optimization problem into an optimization problem including uncertainty, which can be solved with different approaches, e.g., robust optimization or stochastic optimization; however, this introduces even more complexity which goes to the detriment of the computational tractability [67].

As opposed to superstructure optimization, simulation-based optimization presents another major computational strategy for process design as the solution to an optimization problem. In comparison to superstructure optimization, the scope of simulation-based optimization is rather broad and comprises the use of models for stochastic or deterministic simulations to solve optimization problems [68,69]. Hence, fully elaborated flowsheet models can be utilized for the simulations. Nonetheless, the underlying simulation model in SBO is treated as a black-box simulation model, which allows for simplification, e.g., the use of surrogate models instead of the original flowsheet model.

As the applicability for simulation-based optimization is rather broad, so are the optimization algorithms, which are used in combination with the original models, depending on the type of optimization problem, the type of variables, the type of uncertainty, among others [68,69]. These reach from simple enumerative approaches, or other comparison procedures, over gradient-based methods and model-based methods, involving a surrogate model or response surface that is evaluated in order to find the global optimum, over to derivative-free methods, random search algorithms, e.g., particle swarm optimization or genetic algorithms and direct search algorithms [68,69].

For conceptual process design, where several alternatives are supposed to be evaluated, this implicates several runs of simulations. In general, convergence towards a global optimum in SBO settings can be challenging, so the computational burden for performing an entire process design task can be immense [68,69]. Despite advances in computational power and the applicability of parallel computing strategies, this is still a major limiting factor to the applicability of simulation-based optimization. For a more extensive overview regarding simulation-based optimization, the reader is referred to the excellent reviews of Amaran et al. and Bhosekar & Ierapetritou [68,69]. Compared to the other two approaches, simulation-based optimization strategies are somewhat novel to the field. However, through the increased computational power and the increased availability of flowsheet simulation models through particular software or proprietary models, it gains attraction [69].

3.1.2 Challenges associated with their application

Given the three criteria design resource efficiency, optimality, and exploration of all potential realizations of a future process, all three proposed methods are impaired in one of the mentioned categories. To this end, most process designs for biorefineries are conducted by expert-based approaches, followed by SSO. This fact is also reflected by the number of publications on this topic, as shown in Figure 2.

Therefore, the following paragraphs will give a short overview of formerly published manuscripts that deal with the conceptional process design of biorefineries, referring to frameworks or case studies to determine general trends and draw conclusions for the work in this PhD thesis.

Figure 2 shows the number of search results in Web of Science for the queries “Biorefinery,” “Biorefinery Process Design,” “Biorefinery Process Systems Engineering,” “Biorefinery

Superstructure Optimization," and "Biorefinery Simulation-Based Optimization" in the title, abstract, and keywords of the respective manuscript.

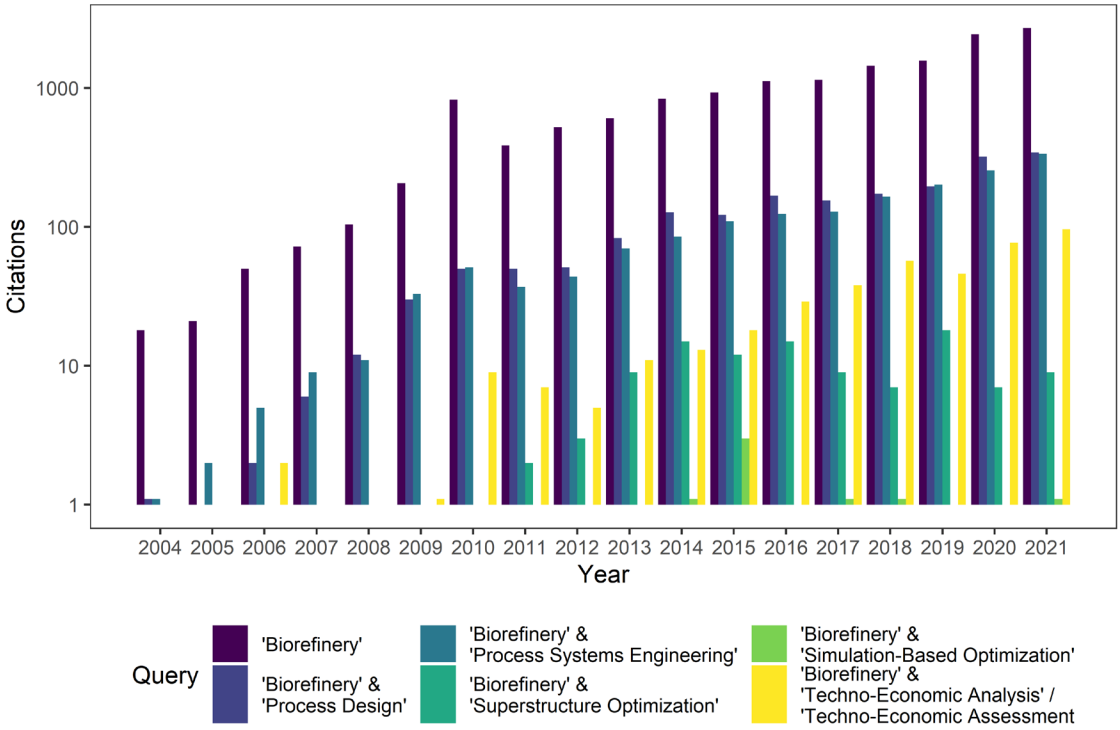

Figure 2: Overview of Web of Science publications resulting from the indicated queries.

In general, all the data sets show an increasing trend since 2004, with a particular peak in 2010 for the keywords "Biorefinery" and "Biorefinery Process Design." Secondly, out of all biorefinery publications, roughly 10% relate to either process design or process systems engineering, and of these, again roughly 10% relate to superstructure optimization and simulation-based optimization, although the vast majority is accounted for the prior one.

Assuming that the publications dealing with process design and process systems engineering have significant overlap, around 90% of all biorefinery publications deal with other topics than their process design, e.g., experimental investigations about different feedstocks, conversion processes, products or unit operations, and others. Secondly, given an average of hundred publications regarding the process design of biorefineries and an average of ten publications regarding the superstructure optimization for biorefinery processes, being a method suitable for the conceptual process design, the overall number of around a dozen operating biorefineries at commercial scale stands diametrical to these developments. In addition, this image solidifies when considering the increasing number of publications regarding biorefineries that include some sort of a techno-economic assessment. This partly confirms the conclusions regarding the confirmative bias regarding this technology, as mentioned in section 2.1.2, also within the discipline of process systems engineering.

The following concluding remarks can be made for the conceptual process design: (1) There is a need for incorporating and evaluating as many potentially beneficial elements

mentioned in chapter 2, (2) an intelligent combination of different process design strategies to overcome computational issues regarding the desired level of detail incorporated in the process, the desired degree of optimality of the solution and the desired minimal use of resources, and finally, (3) a rigorous techno-economic analysis, to realistically evaluate and judge the economic feasibility considering relevant sources of uncertainty.

3.2 Mechanistic modeling of biomanufacturing processes and the use of surrogate models

For the incorporation of expert knowledge into mathematical models for process design, as mentioned in section 1.1, the favorable approach and best practice is knowledge-based modeling, or mechanistic or high-fidelity models, as they are referred to in section 3.1. Mechanistic models of unit operations are essentially based on mass and energy balances, describing mass flows and thermodynamics in the unit operation, as well as reaction, conversion, or transport equations, describing the kinetics in the unit operation [64]. Potentially, also other balances can be employed, e.g., population balances or impulse balances, to describe other dynamics in the unit operation. This represents one of the most accurate ways of representing physical, chemical, or biological phenomena in a mathematical model. With sufficient data for estimating the model parameters and a comprehensive uncertainty and sensitivity analysis on the model robustness, it becomes a potent tool, not only able to interpolate but especially to extrapolate results for initially unseen input data with great accuracy [136].

The set of balances and kinetic equations yield a set of ordinary differential equations, which are commonly nonlinear. It is evident that such a model quickly renders any superstructure optimization problem intractable and significantly increases the burden of any simulation-based optimization problem. Therefore, for both optimization methods, surrogate models represent a viable instrument [66].

The available and investigated types of surrogate models are manifold and depend highly on the specific application and the required level of accuracy. Their complexity spans from simple linear regression models, over piecewise linear regression models, algebraic models, or machine learning models, with the most complex models being deep neural networks [66,68]. They are built based on a sampled input data set and the corresponding model outputs and subsequently fitted to them and ultimately validated. Sampling techniques and sizes can vary: the most commonly used ones are random sampling, Latin Hypercube Sampling, or Sobol Sampling, beyond others [66,70]. Also, adaptive sampling strategies gain importance, as this strategy allows for including more data points in regions where the predictive quality is desired to be high, while other regions can be predicted with lower numbers of sampling points [137].

Overall, surrogate modeling is a straightforward technique with broad applicability. Nonetheless, with the plethora of surrogate models available, there is no clear consensus on which surrogate model, especially in superstructure optimization, is the most suitable, as they all have different strengths and weaknesses. Besides general guidelines on suitable model types for certain dimensionalities and sizes of data sets, a couple of benchmarks have been performed but have not yielded any clear favorite candidates [66,70,71]. However, as

the performance of the surrogate models is crucial for a successful superstructure or simulation-based optimization, this topic is of significant importance in the context of this thesis.

3.3 Closing the gap

Commonly, while the thermodynamics of the fermentation process are expressed by mass and energy balances, the kinetics are usually expressed by so-called black-box kinetics [138]. These lumped kinetics describe the entire metabolism of the cell factory with three equations: (1) the uptake of the substrate into the cell, (2) a distribution relation to describe how much of the substrate is directed towards the growth of the cell factory, how much is directed towards the production of the respective product and how much is used for other maintenance tasks, and (3) a production equation to describe the production as a function of growth [138]. Based on these three equations, the entire cell factory can be dynamically described as a function of the substrate concentration, other possible co-substrates, and potentially present inhibitors [139]. The overall process reaction is calculated based on the stoichiometry of the process[138]. Furthermore, through additional equations, the behavior of the cell factory as a function of temperature or pH can be displayed [140].

However, from a biological point of view, the cell factory is segregated into different biologically active centers and composed of an extensive metabolic network dictated by the activities of the involved enzymes. The enzymes are limited in their activity by the availability of co-factors and other transcriptional and regulatory constraints [124]. Hence, while the replication through three mathematical expressions can give a very good approximation to the actual behavior, most of the introduced expressions are empirical and usually have to be fitted to a particular system to which the parameters of the model were fitted, hence the accuracy of the model when being used outside these bounds is not given per se and has to be evaluated [136].

From a process systems engineering perspective, the applicability of mechanistic fermentation models with black-box kinetics is, broadly speaking, sufficient for the desired applications, covering various scales of simulations, such as process design and control and value chain design at higher, and also reactor design at lower scales. The former task is usually formulated as an optimization problem similar to superstructure optimization problems with additional constraints; hence the same difficulties and solution strategies apply [141]. However, particularly the latter task is of major importance for fermentation-based biomanufacturing processes, as the behavior of cell factories largely depends on their immediate reactor environment, which is dictated by the fluid dynamics in the fermenter but does not scale linearly with the volume. This causes immense challenges when attempting to scale up results that are achieved on a laboratory scale to commercial scales and thus commonly involves several iterations with small increases in reactor volume and further detailed studies [142].

The gold standard for modeling the behavior of fluids is computational fluid dynamics (CFD). With such reactor models, the fluid behavior can be modeled with extremely high accuracy. Furthermore, CFD models allow the coupling with kinetic models, e.g., a black-box kinetic model of a cell factory, allowing a precise prediction of the behavior of an entire cell

culture over time of a fermentation, based on the black-box kinetics [143]. While these coupled models allow for a targeted scale-up of fermentation processes, their major drawback is the immense computational burden and extremely long simulation times [143]. A solution for this is the development of compartmental models, which, based on the CFD simulations, define compartments within the reactor with similar reaction conditions [144]. This reduces the number of necessary calculations by several magnitudes and accelerates them significantly while maintaining similar accuracy [144]. Despite this approach still being dependent on original CFD simulations, it is seen as a major approach to process design and systems engineering over multiple scales in the future [143–145]. Overall, in all these approaches, the lower end of the scale is commonly the cell membrane.

In biotechnology, on the other end, the upper end of the scale is also the cell membrane. Based on the developments summarized in section 2.3, a lot of data can be created covering the entire central dogma of molecular biology [51,146]. Through advanced analytical methods and a significant price drop for sequencing DNA, data regarding the genome, the transcriptome, the metabolome, and even the cell's proteome, describing the entity of molecules belonging to the respective class at a certain time is readily available [146]. Catalyzed through data-driven modeling approaches, a new branch of science focused on the comparative data analysis of this information emerged. Like process systems engineering, systems biology focuses on applying computational tools in biology relating to the data analysis of the genome, transcriptome, and other domains commonly referred to as omics. All this knowledge, gained through data-driven models, can be transferred into knowledge-driven models: genome-scale metabolic models (GSM) [147]. Instead of assuming three dominating metabolic reactions as in the black-box kinetic models, advanced genome-scale metabolic models can resemble the entire metabolism to the single molecule in the different elements of a cell. Within the cell, the metabolic reactions, including their catalyzing enzyme, are linked to the respective passage in the genome, which is ultimately transcribed for the production of the enzyme. Depending on whether transcriptome and proteome information is available, even the rate of expression and regulation of metabolic pathways can be modeled [147,148].

Furthermore, this amount of information allows subsequently for manipulation of this information, and hence the in-silico engineering of cell factories with the potentially highest level of detail becomes possible[147,148]. Current research in this field aims to include more information of different scales into these models to increase their predictive capabilities and move forward from a mere genotypic description of the cell factory to a detailed phenotypic description [149,150]. Also, most of these predictions do not involve kinetic information about the rate of enzymatic conversions. However, if this information is available and implemented in a GSM and coupled with overall mass and energy balances, a GSM can replace the simple black-box kinetics, as it is used currently in process systems engineering [151]. Given a successful implementation, a reactor model with a GSM core can both predict the course of fermentation and the fluid dynamics within the reactor if coupled with a CFD model or compartmental model, and at the same time, the metabolic fluxes within the cell factory [144]. As both dynamics happen on very different time scales, this would be the key to successfully modeling fermentations at different volume scales and overcoming issues with scale-up in process design [145]. However, this largely depends on the amount and even more on the quality of the provided data [152]. As this field is still

under development, only very few GSMs are curated at a level where they can potentially replace the black-box kinetic core for a full-scale fermentation model [143,150].

In conclusion, both disciplines, process systems engineering and systems biology, provide models that solve engineering problems on multiple scales. Nonetheless, the level of integration is almost non-existent. The gap between the two disciplines literally follows the cell membrane, despite the massive potential for integration. Few research studies on the topic of reactor design tried to address this in the last years by coupling CFD simulations with a GSM of an existing reactor and an existing cell factory with promising results [143]. The rationale is that a successful scale-down enables a successful scale-up when beginning with the end in mind. While this approach is promising for the design of reactors, process design at an initial stage rarely includes detailed CFD simulations of not yet existing reactors.

A potential strategy of how to begin with the end in mind and integrating systems biology and process systems engineering in the context of process design for biomanufacturing processes could expedite conceptional process design to the next level will be illustrated in chapter 5.3, together with the results of this thesis.

4. Contribution of this PhD

The same case study concept is used throughout the four main journal publications for this thesis. The molecule of interest is xylitol, a sugar substitute with manifold beneficial health properties [153]. It is also one of the top 12 chemicals to be produced in a biorefinery, as declared by the US Department of Energy [88]. Up to this date, it is only produced in a chemical manufacturing process and is thus considered an ideal candidate for investigation in this thesis [118,119]. Furthermore, an average product price of around 4.80 US$ in 2021 potentially opens a higher operative margin than a comparative bioethanol production [154].

- **Research Question 1:**
  Which suitable biomass pretreatment technologies best fulfill these criteria, and how should they be selected?

As reported in paper A, the overarching aim is to find a suitable pretreatment method for a subsequent xylitol production. In lignocellulosic biomass, the fraction of interest is hence the hemicellulosic fraction containing xylose monomers, which can then be converted to xylitol in cell factories. Therefore, as a feedstock, wheat straw is chosen as a widely available agricultural waste in the northern hemisphere. Furthermore, according to the compositions listed in Table 1, the hemicellulosic fraction is comparatively large compared to the cellulosic and the lignin fraction, potentially increasing the product yield. Given the summary of pretreatment methods in Table 2, the pretreatment methods that are supposedly most suitable for the desired task – to fractionate the hemicellulose from the lignocellulosic biomass and break down the polymers – are autothermal, dilute acid, and steam explosion pretreatment. Given the third criterion of producing low amounts of inhibitors, the third pretreatment method is discarded from the selection, as the formation of inhibitors is comparatively much higher.

Subsequently, for both the autothermal and the dilute acid pretreatment, designs of experiments were created and performed. The evaluation of the experiments shows that the autothermal pretreatment yields a xylose monomer concentration of less than 10% of the theoretical maximum, while the dilute acid pretreatment yields a xylose monomer concentration of more than 90% of the theoretical maximum. Also, the amount of inhibitors formed in the dilute acid pretreatment is naturally higher than in the autothermal pretreatment, as the dilute acid acts as a catalyst for the degradation reactions. Nonetheless, the concentration of inhibitors is still tolerable for cell factories, as reported in the literature. With a potential upconcentration step after the pretreatment to increase the sugar concentrations, naturally, also the inhibitor concentrations increase. Hence, cell factories potentially need to be engineered, e.g., through adaptive laboratory evolution to tolerate higher inhibitor concentrations.

Further investigations in the scope of the manuscript are a comparison of three different models, which are fitted to the experimental data, and the subsequent optimization of the operational condition of the wheat straw with respectively each model. The investigated models are two data-driven models, a simple response surface model and a Gaussian

process regression (GPR) model as representative of machine learning models, and one knowledge-driven model, namely a mechanistic model as described in section 3.2. While all models are fitted successfully to the experimental data, the mechanistic model and especially the GPR model show significantly better validation metrics than the response surface model. Regarding the optimization of the operational conditions with all three models and experimental validation of the prediction, the response surface model again shows the weakest performance. Both the GPR and the mechanistic model predict the global optimum reliably. However, it is visible that the GPR model predicts the better, the closer to the experimentally found points, while its interpolative capacities are not entirely reliable [155].

- **Research Question 2:**
  Which process design approach is most suitable for biomanufacturing processes, and how should a particular surrogate model or several surrogate models be selected?

As reported in paper B, based on the desired use of mechanistic models for the process design, and given the consideration regarding the advantages and disadvantages of the different process design methods, the overarching aim is to develop a framework that synergistically incorporates all three methodologies to utilize their advantages as much as possible. For the xylitol production case study, other mechanistic models than the introduced pretreatment model were developed. These are a fermentation model based on a wild-type cell factory, an evaporation model for the upconcentration of the hydrolysate of the pretreatment and the downstream process, and a crystallization model for the downstream process.

In the first step of the S3O framework, product(s), feedstock, and unit operations are selected that are deemed suitable by available expert knowledge to constitute a potential process. Based on this selection, a superstructure is formulated in a bottom-up manner, based on feasibility considerations, to keep the size of the superstructure as small as possible. In the second step, in order to utilize the mechanistic models in the superstructure formulation, surrogate models are employed. Dependent on the surrogate model, an optimization problem is formulated and solved with a suitable solver. In this particular case, not only the global optimum is considered, but rather several candidate process topologies that returned feasible and potentially employable solutions for the process design. Lastly, in the third step, all the candidate process topologies are subjected to simulation-based optimization under uncertainty to find a truly feasible solution among the initial candidates. The overall idea behind the framework is hence to employ as much valuable expert knowledge as possible, harness the power of SSO while maintaining the computational tractability, and lastly to utilize the strength of SBO to incorporate uncertainties in a straightforward manner while decreasing the computational burden through a reduced number of necessary simulations.

The main focus in the manuscript lies in evaluating different surrogate models and validating the framework based on the case study for producing xylitol. In the scope of the manuscript, four different surrogate models were evaluated regarding their validation metrics and their performance in the optimization problem. The evaluated surrogate

models are ALAMO (automated learning of algebraic models for optimization) surrogate models, based on algebraic terms, Delaunay Triangulation Regression (DTR) surrogate models, based on piecewise-linear expressions, a GPR surrogate model, and an Artificial Neural Network (ANN) model, two examples of popular machine learning models. With the ALAMO surrogate model, the optimization problem can still be treated as MINLP and is solved with the BARON solver. The DTR surrogate model is piecewise linear, and hence the MINLP transfers into a mixed-integer linear problem (MILP), commonly easier to solve. The employed solver for it is the GUROBI solver. Both the GPR and the ANN do not consider integer variables. Hence, the MINLP converts into a set of nonlinear problems (NLP), which are solved in an enumerative manner.

Throughout the manuscript, the entire framework is applied to the xylitol case study: In step 1, as product and feedstock are considered to be xylitol and wheat straw, the superstructure is considered to have a compulsory pretreatment unit, a facultative upconcentration unit, a compulsory fermentation unit, a facultative evaporation unit, a compulsory first crystallization unit, and a second facultative crystallization unit, yielding overall eight potential process topologies. Subsequently, in step 2, flowsheet simulations with all operational variables for all topologies are performed. They are used for a global sensitivity analysis to determine the most sensitive operational variables. Based on the five most sensitive variables, flowsheet simulations with a reduced number of samples and variables are performed to fit the surrogate models. The models are subsequently validated. The validation of the coefficient of determination and the root mean squared error of the surrogate models shows that the ALAMO surrogate models overall perform worst. Furthermore, the DTR surrogate models also show inferior validation metrics. On the other hand, both the ANN and especially the GPR show excellent validation metrics.

However, in the SSO, the best results are achieved with the DTR surrogate models when compared to the validation simulations. The GPR and ANN-assisted SSO yield solutions are predicted too optimistically or prove to infringe the boundary constraints when compared to the validation simulations. The SSO with ALAMO surrogate models mainly yields infeasible solutions or far from the global optimum, compared to the other three. These results can be explained with the properties of the surrogate models. While the machine learning surrogates' predictive capacities around the sampling data are good, the prediction quality decreases for predictions that are further away from known data points. On the other hand, despite the bad validation metrics of the DTR surrogate models, the interpolation is relatively reliable despite being based on piecewise linear functions. In conclusion, it is of crucial importance to not only select surrogate models based on their validation metrics but to perform a case-dependent benchmark to evaluate the best option for the specific application.

In the third step of the framework, three candidate process topologies are subjected to SBO, and for all of them, the global optimum is improved compared to the SSO solution even under uncertainty, underlining the potential of SBO. In conclusion, the proposed framework is highly functional for the conceptual process design of biomanufacturing processes and synergistically combines the advantages of all three introduced process design methods [156].

- **Research Question 3:**
  Can a conceptionally designed second-generation biorefinery be a cost-competitive alternative to existing chemical processes, and what products should be produced in such a setup?

With the successfully validated framework, to answer research question 3, paper C deals with the detailed process design of an integrated biorefinery, henceforth denominated as xylitol biorefinery. Furthermore, a detailed techno-economic analysis is conducted to evaluate the biorefinery's economic feasibility. Here, for the first step of the framework, besides the main product xylitol and the feedstock wheat straw, the question regarding potential biorefinery concepts and setups is answered. As discussed in section 2.1, either the use of engineered cell factories to increase the product yield by utilizing several of the fractions in lignocellulosic biomass or the production of multiple products are two promising options. As xylitol is produced from xylose in a one-step conversion, the production of xylitol from glucose, the main monomer in the cellulosic fraction, would require significant engineering efforts. The second option is hence considered to be easier to implement. In this case, succinic acid is chosen, as another of the top 12 biochemicals, with a market price of around 2.2 US$ and multiple applications as platform chemical [101]. For the lignin fraction, the monomers commonly are not utilizable as a substrate for fermentation. However, chemical or physical conversions to polymers or other chemicals or fuels are potential alternatives. In this case, the production of sustainable aviation fuel is considered. The high aromatic content in lignin potentially yields a product that can be used as blend-in fuel for fossil kerosene. Also, in comparison to terrestrial transport, aerial transport will still depend on fossil fuels for several decades. Lastly, the option to use the lignin entirely for combustion and the generation of steam and electricity was also considered. Relevant unit operations were investigated for all three processes.

For the sizing and costing of equipment, an NREL report with detailed quotations for the entire equipment of a bioethanol refinery was taken to obtain as realistic results as possible. The costs for the xylitol biorefinery process were calculated through the plant capacity ratio method. Furthermore, auxiliary equipment, namely feedstock handling, product, and intermediate storage, a wastewater treatment plant, and a combustion, steam, and power generation unit, were considered. For the techno-economic analysis, different key performance indicators (KPIs), namely the net present value (NPV) and the minimum selling price of products (MSEP), were calculated.

With all considered options, the initial superstructure comprises 128 process topologies. As a preliminary check for feasibility of each process topology, coarse sampling with the flowsheets is performed, referring to a small number of Latin Hypercube samples, while checking whether the process topology does produce both xylitol and succinic acid. This analysis shows that most process topologies are not viable and can be omitted in the subsequent steps. All operational variables are again taken as input in the subsequent global sensitivity analysis to determine the most sensitive ones. Based on the global sensitivity analysis results, flowsheet samples with the five most sensitive operational variables are performed to fit DTR, GPR, and ANN surrogate models for the SSO. Again, the validation results of the surrogate models are comparable to prior findings. Also, the performance of each surrogate model in the SSO shows similar results compared to prior findings in paper

B. Regarding the actual process, the process topologies that show the highest NPV are those with all unit operations or without the second crystallization unit.

Furthermore, the production of sustainable aviation fuel instead of heat does not improve the NPV due to the low yield of the process. Hence, the option of producing sustainable aviation fuel is neglected, and the lignin is supposed to be used for steam and power generation entirely. Lastly, in the third step of the framework, four candidates are subjected to SBO, with uncertainty in the CAPEX, the OPEX, and the product price. Under the given uncertainties, the SBO solver cannot find a feasible solution with a positive NPV under the given iteration limit.

A detailed look at the results of the techno-economic analysis shows the highest CAPEX contribution stemming from the wastewater treatment unit and the highest OPEX contribution stemming from the evaporation units, which is expected in that form. Furthermore, the overall product yield for both fermentation processes could be significantly improved. Lastly, a risk-based analysis shows that only the flowsheet with all unit operations reaches a realistic MSEP for 50% of the realizations of uncertainty, conversely indicating a failure rate of 50% [157].

- **Research Questio**n 4:
  Is a second-generation biorefinery more sustainable than the alternative chemical process, and can it play a key role in the transition towards more sustainable production patterns?

Lastly, after having evaluated the economic feasibility of the xylitol biorefinery, in order to answer research question 4, the sustainability of the xylitol biorefinery is also assessed. As reported in paper D, the entire analysis is based on the consolidated process design obtained with the S3O framework in paper C. The life-cycle assessment is performed following a standardized procedure. First, the goals, scope, and boundaries are defined: The goal is to determine the environmental impacts of the xylitol biorefinery and assess it compared to the existing chemical manufacturing processes for xylitol, analyzing potential benefits of the biomanufacturing process. Regarding the scope, the xylitol biorefinery is considered from cradle to gate, and the reference unit is set to one ton of feedstock – wheat straw – per year. In the second step, all mass and energy flows into and out of the system boundaries are subsumed. In the third step, for each flow, the impacts are calculated. The impacts are assessed with the ReCiPe 2016 Midpoint H V1.05 method. The characterization factors in each impact category are retrieved from the Ecoinvent 3 database. Subsequently, the characterization factors for all flows in all impact categories are summed up and normalized. Lastly, these impacts are analyzed.

The global warming potential in CO_2 equivalents is comparatively low, as the xylitol biorefinery only emits biogenic CO_2. On the other hand, freshwater and marine ecotoxicity impacts are comparatively high, as well as the human carcinogenic potential. The prior two can be related to the electrical energy requirement of the biorefinery, as it is assumed that the energy is retrieved from wind power plants. The windmills require copper for their generators, which is usually harvested using toxic chemicals that can damage freshwater and marine life if not disposed of properly.

Furthermore, the sustainability impact is compared to the existing chemical manufacturing process for producing xylitol, firstly as a classical chemical process, using corn as feedstock, and secondly as a wood biorefinery process with a chemical conversion, run by Dupont, with improved sustainability potential. The data for the chemical manufacturing processes is retrieved from a white paper of Dupont, for which the impacts are calculated with the IMPACT2002+ method. The impacts for the xylitol biorefinery are hence also calculated through this method to enable a direct comparison. While the xylitol biorefinery shows lower impact metrics for all categories than the classic process, the metrics are higher than those of the wood biorefinery process. Hence, overall, while the process clearly has a low impact on sustainability regarding the global warming potential, it is not possible to conclude that the biorefinery process per se is better than already existing chemical processes [158].

5. Conclusions & Outlook

5.1 Conclusions in this thesis

After evaluating all results from the four main journal publications in the scope of this PhD thesis, an answer to the posed research questions in section 1.2 can be given, taking the xylitol biorefinery process as the main example and deducting more general conclusions for fermentation-based biomanufacturing processes in general.

Regarding research question one, for the particular case of xylitol production, the most suitable pretreatment method seems to be dilute acid pretreatment, as it best fulfills the three given criteria on a good fractionation, a high monomer yield, and low inhibitor yields. However, there is no straightforward answer to this question, as this evaluation still requires a lot of expert knowledge and different considerations, as the research within pretreatment technologies is instead targeted on exploring different feedstocks and products than to gain a deeper understanding of the detailed mechanism for the depolymerization and degradation reactions. Nonetheless, with this expert knowledge at hand, a very precise and well-validated mechanistic model for dilute acid pretreatment could be built for the use in further process design steps, which also confirms the conclusions of the findings in section 3.2 [155].

Regarding research question two, the proposed S3O framework manages to leverage synergies from different process design concepts, namely expert-based process design, superstructure optimization, and simulation-based optimization in a hybrid manner. By this, the framework allows for the use of mechanistic models, incorporating expert knowledge of the underlying unit operations using surrogate models in the design process. Evaluating four different surrogate models by analyzing their validation metrics and their performance in the superstructure optimization, it becomes evident that there are no preferred candidates for the use in superstructure optimization but that the solution is case-dependent. This indicates the requirement of a benchmark of different candidates in the specific optimization problem, as the validation metrics do not necessarily correspond to the performance in the optimization problem.

Regarding research question three, when applying the S3O framework to design the production of xylitol in an integrated second-generation biorefinery with value-added co-products, the framework yields a technically feasible process. However, the techno-economic analysis shows, besides known issues for CAPEX and OPEX, that the product yield remains relatively low while using wild-type cell factories in the process. A risk-based economic analysis shows that the process will only be feasible with a probability of around 50%. From an industrial perspective, this is a high-risk venture, confirming the predominant estimates about second-generation biorefineries not being feasible at an industrial scale.

Regarding research question 4, the life cycle analysis of the xylitol biorefinery process clearly shows a reduced impact on greenhouse gas emissions. However, the impacts on freshwater and marine toxicity and the human carcinogenic potential associated with the

xylitol biorefinery process are significant. The impact can be associated with the use of renewable electric energy from windmills. The energy source could be potentially replaced by another source of electricity, which is, however, dependent on the availability at the potential plant location. Nonetheless, the results confirm the general potential in replacing fossil process alternatives while keeping in mind that impacts on the ecosphere cannot be avoided entirely. This illustrates how complex the impact on ecological categories through primarily more sustainable production patterns can be and indicates a clear need for quantitative analysis of the sustainability impact of such biorefineries.

Given the potential and importance of second-generation biorefineries in terms of the sustainable transition, these conclusions are rather pessimistic. Despite a conceptionally designed biorefinery with a good selection for the biomass pretreatment and comprising many considerations based on current research, both the economic feasibility and the sustainability impact do not reflect the potential. Hence, in the following sections, both considerations regarding possible points of improvement and general considerations regarding the conceptual process design of fermentation-based biomanufacturing processes are discussed.

5.2 Potential improvements for an economically feasible biorefinery

For the given biorefinery in specific, but also for other conceptually designed biorefinery setups, the taken considerations yielded a technically very well feasible process design, which is not reflected in its economic feasibility. In short, there is a technical design that can produce xylitol using a biorefinery concept, yet it is highly costly. To this end, one consideration targeted the technology readiness level of the used unit operations. Since the biorefinery process is aqueous, a primary difficulty for the fermentation and the downstream process is the separation of water. One way of optimizing this process step is the possibility of using advanced evaporation units, e.g., through mechanical vapor compression or by multiple-effect evaporation, the required amount of heat can be reduced, which leads to a lower OPEX [107,159]. Alternatively, different unit operations, e.g., membrane separation or filtration units, can be used instead. Those potentially separate chemicals from aqueous streams in a more targeted manner and improve the efficiency; however, other issues are associated with these unit operations. Their technology readiness level is not comparable to evaporation units, which would require further research regarding their applicability [160]. Moreover, further steps of process integration and especially process intensification could help to lower both CAPEX and OPEX of the process, as mentioned in section 2.1.

A second significant aspect that can heavily impact the economic feasibility are considerations regarding the biorefinery's value chain. While the presented techno-economic analysis only takes factors into account, which directly are relevant on a process level, also higher-scale effects influence it directly. In Vollmer et al. (2022), a simple case study based on the presented process design was investigated by taking into account further constraints regarding feedstock availability, logistic constraints, and product market size constraints. These are directly implemented in the S3O framework, as explained in section 2.3. The results of the value chain optimization clearly show an effect of the economies of scale on the KPIs of the biorefinery, as well as an actual limitation by the

market size of the products. Recalling that the deciding factor for most biorefinery setups in the past was the assumed increasing price for fossil oil, which remained relatively constant over the past decades, thus causing rather constant product prices, these findings gain even more importance. Hence, it seems logical to consider a conceptual process design and integrate considerations regarding a conceptually designed value chain in a multi-scale approach already at initial stages of the design process to find an economically feasible biomanufacturing process solution [161].

Lastly, the designed process was based on two wild-type cell factories for the xylitol fermentation and succinic acid fermentation. With the considerations in section 2.3 and the low product yields for the final process design, it is rather conclusive to consider using such engineered cell factories instead of wild-type cell factories. In Vollmer et al. (2022), a simple case study on the in-silico design of such engineered cell factories and their impact on the KPIs of the biorefinery have been investigated. The proposed design of the cell factories was validated with findings in the literature, and subsequently, the techno-economic analysis of the process was performed based on the engineered cell factories. The results show that using these engineered cell factories improves the KPIs significantly, and the risk-based evaluation shows that the process becomes feasible in around 99% of the cases. As engineered cell factories seem to be a major lever to improve the KPIs, it seems logical to consider a conceptual process design and integrate considerations regarding a conceptually designed cell factory in a multi-scale approach at initial stages of the design process to find an economically feasible biomanufacturing process solution [162].

It is essential to point out that in the scope of this thesis, the question regarding the conceptual process design of biomanufacturing processes was discussed based on a case study of producing xylitol and value-added co-products. The selected two products are good candidates regarding a second-generation biorefinery setup, with sufficiently high product prices to enable an economically feasible biorefinery, particularly in light of the results discussed in this section. While this also might be the fact for other candidate combinations, it might not be for others. Hence, while the question of how such second-generation biorefinery processes should be designed can be answered, the question, if the potential for second-generation biorefineries, after all, is significant enough to lead to an increased number of conceptual designs that find their way into reality and a commercial-scale production will only be answered retrospectively in a couple of years or decades. Nonetheless, second-generation biorefineries are only one particular type of biomanufacturing processes. Therefore, the conclusions from the answers to the four research questions and the particular highlights in this section regarding the integration of value chain design, cell factory design, and process design for biomanufacturing processes in general are taken into account in the following last section of this thesis.

5.3 A glimpse into the future

The potential of fermentation-based biomanufacturing processes is indisputable. Hence, this last section is dedicated to drafting a blueprint for the future design of fermentation-based biomanufacturing processes and to shed light on the last remaining aspect, namely, how more sustainable production patterns can be implemented to ultimately achieve sustainable development goals.

While the proposed S3O framework manages to yield conceptually designed biomanufacturing processes, considering the elaboration in section 5.2, and based on all described developments both in biotechnology and systems engineering in sections 2.3 and 3.3, the scope of this framework should be extended: GSMs, if adequately supplied with data, can facilitate the in-silico cell factory design and are constructed as an optimization problem [163]. This fact can be utilized to implement them in a superstructure optimization problem instead of the black-box kinetics for the fermentation models. The superstructure then would contain an inner optimization problem – the GSM – and an outer optimization problem – the SSO. Both optimization problems can be solved simultaneously and under the same constraints if they are the same type of optimization problem. This allows for the inclusion of further considerations, e.g., the effects of increased production of another metabolite for the downstream processing. If the optimization problems are not of the same type, a surrogate model for the GSM can facilitate the implementation, while another benchmark of different surrogate models might be required.

Furthermore, the entire SSO can also be expanded with further constraints into a value chain optimization problem, which equally can be solved simultaneously, either directly or with the process flowsheet as another surrogate model. This nested optimization problem allows the simultaneous design of cell factory, process, and value chain. From an industrial perspective, considerations do start from a business perspective when analysing current markets and potential new products. By applying this extended framework in such cases, the entire process can be conceptionally designed and modeled. With a combined techno-economic analysis and a life cycle assessment, this would ultimately be a significant advance compared to current approaches and mean to truly "begin with the end in mind" and yield economically feasible and sustainable process solutions.

In conclusion, with the proposed blueprint in this section, the entirety of future biomanufacturing processes will benefit from such conceptual approaches. As a matter of fact, first developments in this direction are already happening at this very moment: Entire companies, such as Ginkgo Bioworks or Zymergen, are establishing so-called biofoundries where high-throughput systems apply the entire design-build-test-learn cycle in an automated way while continuously creating omics and other data for the creation of a digital replica of the biomanufacturing process [164,165]. At the same time, major biotechnological companies also put efforts into creating digital twins of their manufacturing processes. Combining these two efforts and using the data for building models, they can be integrated as suggested for the extended framework. In this way, the knowledge about these processes is extended, which can expedite the entire development process of novel biomanufacturing process significantly. In more general terms, the mentioned joint ventures between different players can provide fruitful outcomes for the whole industrial sector. Furthermore, more supportive policymaking frameworks can also contribute to supporting such initiatives. With the conduction of both a techno-economic analysis and a life-cycle assessment, the obtained design can be evaluated from an economic and an environmental allowing for thorough conclusions regarding the benefits of these novel processes, which will guide industry and research towards beneficial solutions. Lastly, appropriate communication to society will also contribute to anchoring the benefits of these ideas within the public discussion while not playing down potential risks to ensure widespread support on a higher level.

Hence, having the end in mind, the goal is to enable more sustainable production patterns. Biomanufacturing processes of the first generation for various products apart from fuels are already taking a great part in it, and their number is rapidly growing. When conceptually designed and with all the considerations given in the scope of this PhD thesis, second-generation biorefineries can play an important role in this transition. However, also third and higher-generation biorefineries will be further investigated and might soon become a concurrent concept. Furthermore, alternative developments beyond biotechnology, e.g., carbon capture and storage and power-to-x concepts, using traditional chemical process routes, are also emerging alternatives for the sustainable production of chemicals. Most importantly, however: whether a future sustainable production pattern contains a second-generation biorefinery, a power-to-x plant, a combination of the two, or another entirely novel process will be determined by their corresponding economics as well as their sustainability metrics. In the end, humankind has now one more suitable set of technologies at hand to reach the sustainable development goals and subsequently other climate change targets that we urgently have to reach to sustain ourselves as a whole on this planet in the future.

References

[1] G. Gopinath, The Great Lockdown: Worst Economic Downturn Since the Great Depression – IMF Blog, Int. Monet. Fund Blog. (2020). https://blogs.imf.org/2020/04/14/the-great-lockdown-worst-economic-downturn-since-the-great-depression/ (accessed December 8, 2021).

[2] A.G. Christopoulos, P. Kalantonis, I. Katsampoxakis, K. Vergos, COVID-19 and the Energy Price Volatility, Energies. 14 (2021) 6496. https://doi.org/10.3390/en14206496.

[3] R. Fuentes, M. Galeotti, A. Lanza, B. Manzano, COVID-19 and Climate Change: A Tale of Two Global Problems, Sustain. 2020, Vol. 12, Page 8560. 12 (2020) 8560. https://doi.org/10.3390/SU12208560.

[4] International Panel on Climate Change, Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, Cambridge Univ. Press. (2021) 3949. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report.pdf.

[5] W. Leal Filho, A.M. Azul, L. Brandli, P.G. Özuyar, T. Wall, eds., Climate Action, Springer International Publishing, Cham, 2020. https://doi.org/10.1007/978-3-319-95885-9.

[6] International Energy Agency, The Future of Petrochemicals – Analysis, Int. Energy Agency. (2018) 11–25. https://www.iea.org/reports/the-future-of-petrochemicals.

[7] IRENA, IRENA (2019), Global Energy Transformation: A Roadmap to 2050, 2019. https://www.irena.org/publications/2019/Apr/Global-energy-transformation-A-roadmap-to-2050-2019Edition.

[8] BASF, Outlook for the Chemical Industry, 2020. https://report.basf.com/2020/en/managements-report/forecast/economic-environment/chemical-industry.html.

[9] International Energy Agency, Global Energy Review 2021 - Assessing the effects of economic recoveries on global energy demand and CO 2 emissions in 2021 Global Energy, 2021. www.iea.org/t&c/.

[10] D.H. Meadows, D.L. Meadows, J. Randers, W.W.I. Behrens, The limits to growth, 1972. https://doi.org/10.5860/choice.42-1517.

[11] U. Nations, United Nations Framework Convention on Climate Change, 1992.

[12] United Nations, Kyoto Protocol to the United Nations Framework Convention On Climate Change, 1998. https://doi.org/10.51663/pnz.58.2.07.

[13] United Nations, The paris agreement, 2019. https://doi.org/10.4324/9789276082569-2.

[14] United Nations, Transforming our world: The 2030 agenda for sustainable

development, 2015. https://sdgs.un.org/sites/default/files/publications/21252030 Agenda for Sustainable Development web.pdf (accessed March 1, 2021).

[15] P. Schroeder, K. Anggraeni, U. Weber, The Relevance of Circular Economy Practices to the Sustainable Development Goals, J. Ind. Ecol. 23 (2019) 77–95. https://doi.org/10.1111/jiec.12732.

[16] M. Lieder, A. Rashid, Towards circular economy implementation: A comprehensive review in context of manufacturing industry, J. Clean. Prod. 115 (2016) 36–51. https://doi.org/10.1016/j.jclepro.2015.12.042.

[17] M.M. Bjørnbet, C. Skaar, A.M. Fet, K.Ø. Schulte, Circular economy in manufacturing companies: A review of case study literature, J. Clean. Prod. 294 (2021) 126268. https://doi.org/10.1016/j.jclepro.2021.126268.

[18] J. Singh, I. Ordoñez, Resource recovery from post-consumer waste: important lessons for the upcoming circular economy, J. Clean. Prod. 134 (2016) 342–353. https://doi.org/10.1016/j.jclepro.2015.12.020.

[19] J. Singh, R. Laurenti, R. Sinha, B. Frostell, Progress and challenges to the global waste management system, Waste Manag. Res. 32 (2014) 800–812. https://doi.org/10.1177/0734242X14537868.

[20] N. Gregson, M. Crang, S. Fuller, H. Holmes, Interrogating the circular economy: the moral economy of resource recovery in the EU, Econ. Soc. 44 (2015) 218–243. https://doi.org/10.1080/03085147.2015.1013353.

[21] J.F. Jenck, F. Agterberg, M.J. Droescher, Products and processes for a sustainable chemical industry: A review of achievements and prospects, Green Chem. 6 (2004) 544–556. https://doi.org/10.1039/b406854h.

[22] R. Ahorsu, F. Medina, M. Constantí, Significance and challenges of biomass as a suitable feedstock for bioenergy and biochemical production: A review, Energies. 11 (2018) 3366. https://doi.org/10.3390/en11123366.

[23] P.A. Østergaard, N. Duic, Y. Noorollahi, H. Mikulcic, S. Kalogirou, Sustainable development using renewable energy technology, Renew. Energy. 146 (2020) 2430–2437. https://doi.org/10.1016/j.renene.2019.08.094.

[24] M. Gavrilescu, Y. Chisti, Biotechnology - A sustainable alternative for chemical industry, Biotechnol. Adv. 23 (2005) 471–499. https://doi.org/10.1016/j.biotechadv.2005.03.004.

[25] Y. Lokko, M. Heijde, K. Schebesta, P. Scholtès, M. Van Montagu, M. Giacca, Biotechnology and the bioeconomy—Towards inclusive and sustainable industrial development, N. Biotechnol. 40 (2018) 5–10. https://doi.org/10.1016/j.nbt.2017.06.005.

[26] J.M. Woodley, Towards the sustainable production of bulk-chemicals using biotechnology, N. Biotechnol. 59 (2020) 59–64. https://doi.org/10.1016/j.nbt.2020.07.002.

[27] K.K. Hong, J. Nielsen, Metabolic engineering of Saccharomyces cerevisiae: A key cell factory platform for future biorefineries, Cell. Mol. Life Sci. 69 (2012) 2671–2690. https://doi.org/10.1007/s00018-012-0945-1.

[28] J. Nielsen, C. Larsson, A. van Maris, J. Pronk, Metabolic engineering of yeast for production of fuels and chemicals, Curr. Opin. Biotechnol. 24 (2013) 398–404. https://doi.org/10.1016/j.copbio.2013.03.023.

[29] A.M. Davy, H.F. Kildegaard, M.R. Andersen, Cell Factory Engineering, Cell Syst. 4 (2017) 262–275. https://doi.org/10.1016/j.cels.2017.02.010.

[30] F. Cherubini, The biorefinery concept: Using biomass instead of oil for producing energy and chemicals, Energy Convers. Manag. 51 (2010) 1412–1421. https://doi.org/10.1016/j.enconman.2010.01.015.

[31] International Energy Agency, Biorefineries: adding value to the sustainable utilisation of biomas, 2009. https://doi.org/10.1007/978-1-349-06888-3_42.

[32] A.J.J. Straathof, S.A. Wahl, K.R. Benjamin, R. Takors, N. Wierckx, H.J. Noorman, Grand Research Challenges for Sustainable Industrial Biotechnology, Trends Biotechnol. 37 (2019) 1042–1050. https://doi.org/10.1016/j.tibtech.2019.04.002.

[33] F. Cherubini, G. Jungmeier, M. Wellisch, T. Willke, I. Skiadas, R. van Ree, E. de Jong, Toward a common classification approach for biorefinery systems, Biofuels, Bioprod. Biorefining. 3 (2009) 534–546. https://doi.org/10.1002/bbb.172.

[34] M.H.M. Eppink, G. Olivieri, H. Reith, C. van den Berg, M.J. Barbosa, R.H. Wijffels, From current algae products to future biorefinery practices: A review, in: Adv. Biochem. Eng. Biotechnol., Springer Science and Business Media Deutschland GmbH, 2019: pp. 99–123. https://doi.org/10.1007/10_2016_64.

[35] A. Schievano, T. Pepé Sciarria, K. Vanbroekhoven, H. De Wever, S. Puig, S.J. Andersen, K. Rabaey, D. Pant, Electro-Fermentation – Merging Electrochemistry with Fermentation in Industrial Applications, Trends Biotechnol. 34 (2016) 866–878. https://doi.org/10.1016/j.tibtech.2016.04.007.

[36] S.N. Naik, V. V. Goud, P.K. Rout, A.K. Dalai, Production of first and second generation biofuels: A comprehensive review, Renew. Sustain. Energy Rev. 14 (2010) 578–597. https://doi.org/10.1016/j.rser.2009.10.003.

[37] M.W. Rosegrant, S. Msangi, T. Sulser, R. Valmonte-santos, Bioenergy and Agriculture: Promises and Challenges. Biofuels and the Global Food Balance, 2020 Vis. Briefs. (2006) 2005–2006. https://ideas.repec.org/p/fpr/2020br/14(3).html (accessed January 4, 2022).

[38] S.S. Hassan, G.A. Williams, A.K. Jaiswal, Lignocellulosic Biorefineries in Europe: Current State and Prospects, Trends Biotechnol. 37 (2019) 231–234. https://doi.org/10.1016/j.tibtech.2018.07.002.

[39] S.S. Hassan, G.A. Williams, A.K. Jaiswal, Moving towards the second generation of lignocellulosic biorefineries in the EU: Drivers, challenges, and opportunities, Renew. Sustain. Energy Rev. 101 (2019) 590–599. https://doi.org/10.1016/j.rser.2018.11.041.

[40] S. Alfano, F. Berruti, N. Denis, A. Santagostino, The future of second-generation biomass, (2016) 1–5. https://www.mckinsey.com/business-functions/sustainability/our-insights/the-future-of-second-generation-biomass.

[41] J. Zetterholm, E. Bryngemark, J. Ahlström, P. Söderholm, S. Harvey, E. Wetterlund, Economic evaluation of large-scale biorefinery deployment: A framework

integrating dynamic biomass market and techno-economic models, Sustain. 12 (2020). https://doi.org/10.3390/su12177126.

[42] M. Galbe, O. Wallberg, Pretreatment for biorefineries: A review of common methods for efficient utilisation of lignocellulosic materials, Biotechnol. Biofuels. 12 (2019) 1–26. https://doi.org/10.1186/s13068-019-1634-1.

[43] S.I. Mussatto, G.M. Dragone, Biomass Pretreatment, Biorefineries, and Potential Products for a Bioeconomy Development, in: Biomass Fractionation Technol. a Lignocellul. Feed. Based Biorefinery, Elsevier Inc., 2016: pp. 1–22. https://doi.org/10.1016/B978-0-12-802323-5.00001-3.

[44] S. Vieira, M.V. Barros, A.C.N. Sydney, C.M. Piekarski, A.C. de Francisco, L.P. de S. Vandenberghe, E.B. Sydney, Sustainability of sugarcane lignocellulosic biomass pretreatment for the production of bioethanol, Bioresour. Technol. 299 (2020) 122635. https://doi.org/10.1016/j.biortech.2019.122635.

[45] M. Bilal, H.M.N. Iqbal, Recent Advancements in the Life Cycle Analysis of Lignocellulosic Biomass, Curr. Sustain. Energy Reports. 7 (2020) 100–107. https://doi.org/10.1007/s40518-020-00153-5.

[46] H. Kagermann, W. Wahlster, J. Helbig, Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0 – Securing the Future of German Manufacturing Industry, 2013. https://www.bibsonomy.org/bibtex/25c352acf1857c1c1839c1a11fe9b7e6c/flint63%0Ahttp://forschungsunion.de/pdf/industrie_4_0_final_report.pdf.

[47] H. Lasi, P. Fettke, H.G. Kemper, T. Feld, M. Hoffmann, Industry 4.0, Bus. Inf. Syst. Eng. 6 (2014) 239–242. https://doi.org/10.1007/s12599-014-0334-4.

[48] C. Yang, Q. Huang, Z. Li, K. Liu, F. Hu, Big Data and cloud computing: innovation opportunities and challenges, Int. J. Digit. Earth. 10 (2017) 13–53. https://doi.org/10.1080/17538947.2016.1239771.

[49] J. Lee, H. Davari, J. Singh, V. Pandhare, Industrial Artificial Intelligence for industry 4.0-based manufacturing systems, Manuf. Lett. 18 (2018) 20–23. https://doi.org/10.1016/j.mfglet.2018.09.002.

[50] D.M. Camacho, K.M. Collins, R.K. Powers, J.C. Costello, J.J. Collins, Next-Generation Machine Learning for Biological Networks, Cell. 173 (2018) 1581–1592. https://doi.org/10.1016/j.cell.2018.05.015.

[51] C.E. Lawson, J.M. Martí, T. Radivojevic, S.V.R. Jonnalagadda, R. Gentz, N.J. Hillson, S. Peisert, J. Kim, B.A. Simmons, C.J. Petzold, S.W. Singer, A. Mukhopadhyay, D. Tanjore, J.G. Dunn, H. Garcia Martin, Machine learning for metabolic engineering: A review, Metab. Eng. 63 (2021) 34–60. https://doi.org/10.1016/j.ymben.2020.10.005.

[52] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature. 596 (2021) 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[53] A. Fuller, Z. Fan, C. Day, C. Barlow, Digital Twin: Enabling Technologies, Challenges and Open Research, IEEE Access. 8 (2020) 108952–108971. https://doi.org/10.1109/ACCESS.2020.2998358.

[54] P. Aivaliotis, K. Georgoulias, G. Chryssolouris, The use of Digital Twin for predictive maintenance in manufacturing, Int. J. Comput. Integr. Manuf. 32 (2019) 1067–1080. https://doi.org/10.1080/0951192X.2019.1686173.

[55] R. He, G. Chen, C. Dong, S. Sun, X. Shen, Data-driven digital twin technology for optimized control in process systems, ISA Trans. 95 (2019) 221–234. https://doi.org/10.1016/j.isatra.2019.05.011.

[56] I.A. Udugama, P.C. Lopez, C.L. Gargalo, X. Li, C. Bayer, K. V. Gernaey, Digital Twin in biomanufacturing: challenges and opportunities towards its implementation, Syst. Microbiol. Biomanufacturing. 1 (2021) 257–274. https://doi.org/10.1007/s43393-021-00024-0.

[57] J. Smiatek, A. Jung, E. Bluhmki, Towards a Digital Bioprocess Replica: Computational Approaches in Biopharmaceutical Development and Manufacturing, Trends Biotechnol. 38 (2020) 1141–1153. https://doi.org/10.1016/j.tibtech.2020.05.008.

[58] I.E. Grossmann, I. Harjunkoski, Process systems Engineering: Academic and industrial perspectives, Comput. Chem. Eng. 126 (2019) 474–484. https://doi.org/10.1016/j.compchemeng.2019.04.028.

[59] A.M. Schweidtmann, E. Esche, A. Fischer, M. Kloft, J. Repke, S. Sager, A. Mitsos, Machine Learning in Chemical Engineering: A Perspective, Chemie Ing. Tech. 93 (2021) 2029–2039. https://doi.org/10.1002/cite.202100083.

[60] J.H. Lee, J. Shin, M.J. Realff, Machine learning: Overview of the recent progresses and implications for the process systems engineering field, Comput. Chem. Eng. 114 (2018) 111–121. https://doi.org/10.1016/j.compchemeng.2017.10.008.

[61] F. Delvigne, R. Takors, R. Mudde, W. van Gulik, H. Noorman, Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond, Microb. Biotechnol. 10 (2017) 1267–1274. https://doi.org/10.1111/1751-7915.12803.

[62] A.M. Schweidtmann, A. Mitsos, Deterministic Global Optimization with Artificial Neural Networks Embedded, J. Optim. Theory Appl. 180 (2019) 925–948. https://doi.org/10.1007/s10957-018-1396-0.

[63] V. Venkatasubramanian, The promise of artificial intelligence in chemical engineering: Is it here, finally?, AIChE J. 65 (2019) 466–478. https://doi.org/10.1002/aic.16489.

[64] L. Mears, S.M. Stocks, M.O. Albaek, G. Sin, K. V. Gernaey, Mechanistic Fermentation Models for Process Design, Monitoring, and Control, Trends Biotechnol. 35 (2017) 914–924. https://doi.org/10.1016/j.tibtech.2017.07.002.

[65] M. von Stosch, R. Oliveira, J. Peres, S. Feyo de Azevedo, Hybrid semi-parametric modeling in process systems engineering: Past, present and future, Comput. Chem. Eng. 60 (2014) 86–101. https://doi.org/10.1016/j.compchemeng.2013.08.008.

[66] K. McBride, K. Sundmacher, Overview of Surrogate Modeling in Chemical Process Engineering, Chemie-Ingenieur-Technik. 91 (2019) 228–239.

https://doi.org/10.1002/cite.201800091.

[67] Q. Chen, I.E. Grossmann, Recent developments and challenges in optimization-based process synthesis, Annu. Rev. Chem. Biomol. Eng. 8 (2017) 249–283. https://doi.org/10.1146/annurev-chembioeng-080615-033546.

[68] A. Bhosekar, M. Ierapetritou, Advances in surrogate based modeling, feasibility analysis, and optimization: A review, Comput. Chem. Eng. 108 (2018) 250–267. https://doi.org/10.1016/j.compchemeng.2017.09.017.

[69] S. Amaran, N. V. Sahinidis, B. Sharda, S.J. Bury, Simulation optimization: a review of algorithms and applications, 4OR. 12 (2014) 301–333. https://doi.org/10.1007/s10288-014-0275-2.

[70] B.A. Williams, S. Cremaschi, Surrogate Model Selection for Design Space Approximation And Surrogatebased Optimization | Elsevier Enhanced Reader, Comput. Aided Chem. Eng. 47 (2019) 353–358. https://reader.elsevier.com/reader/sd/pii/B9780128185971500564?token=3402D4F9F2DC6DB6D0B523692BCC782B775EA03BF948AD2B5815440726F7AA66D1D77566571AF3F9FEA418355439513D (accessed December 10, 2020).

[71] N.I. Vollmer, R. Al, G. Sin, Benchmarking of Surrogate Models for the Conceptual Process Design of Biorefineries, in: Comput. Aided Chem. Eng., Elsevier, 2021: pp. 475–480. https://doi.org/10.1016/B978-0-323-88506-5.50075-9.

[72] W. Soetaert, E. Vandamme, The impact of industrial biotechnology, Biotechnol. J. 1 (2006) 756–769. https://doi.org/10.1002/biot.200600066.

[73] L. Brennan, P. Owende, Biofuels from microalgae-A review of technologies for production, processing, and extractions of biofuels and co-products, Renew. Sustain. Energy Rev. 14 (2010) 557–577. https://doi.org/10.1016/j.rser.2009.10.009.

[74] R. Moscoviz, J. Toledo-Alarcón, E. Trably, N. Bernet, Electro-Fermentation: How To Drive Fermentation Using Electrochemical Systems, Trends Biotechnol. 34 (2016) 856–865. https://doi.org/10.1016/j.tibtech.2016.04.009.

[75] H.M. Woo, Solar-to-chemical and solar-to-fuel production from CO2 by metabolically engineered microorganisms, Curr. Opin. Biotechnol. 45 (2017) 1–7. https://doi.org/10.1016/j.copbio.2016.11.017.

[76] A. Sultana, A. Kumar, Optimal configuration and combination of multiple lignocellulosic biomass feedstocks delivery to a biorefinery, Bioresour. Technol. 102 (2011) 9947–9956. https://doi.org/10.1016/j.biortech.2011.07.119.

[77] M.E. Bruins, J.P.M. Sanders, Small-scale processing of biomass for biorefinery, Biofuels, Bioprod. Biorefining. 6 (2012) 135–145. https://doi.org/10.1002/bbb.1319.

[78] R.C. Kolfschoten, M.E. Bruins, J.P.M. Sanders, Opportunities for small-scale biorefinery for production of sugar and ethanol in the Netherlands, Biofuels, Bioprod. Biorefining. 8 (2014) 475–486. https://doi.org/10.1002/bbb.1487.

[79] E. de Jong, G. Jungmeier, Biorefinery Concepts in Comparison to Petrochemical Refineries, in: Ind. Biorefineries White Biotechnol., Elsevier, 2015: pp. 3–33. https://doi.org/10.1016/B978-0-444-63453-5.00001-X.

[80] S. Nanda, R. Azargohar, A.K. Dalai, J.A. Kozinski, An assessment on the sustainability of lignocellulosic biomass for biorefining, Renew. Sustain. Energy Rev. 50 (2015) 925–941. https://doi.org/10.1016/j.rser.2015.05.058.

[81] C. De Visser, R. van Ree, Small-scale biorefining, 2016. http://www.plasticseurope.org/documents/document/20111107113205-e_ghg_packaging_denkstatt_vers_1_1.pdf.

[82] L.R. Lynd, X. Liang, M.J. Biddy, A. Allee, H. Cai, T. Foust, M.E. Himmel, M.S. Laser, M. Wang, C.E. Wyman, Cellulosic ethanol: status and innovation, Curr. Opin. Biotechnol. 45 (2017) 202–211. https://doi.org/10.1016/j.copbio.2017.03.008.

[83] H. Eggert, M. Greaker, Promoting second generation biofuels: Does the first generation pave the road?, Energies. 7 (2014) 4430–4445. https://doi.org/10.3390/en7074430.

[84] A. Duque, C. Álvarez, P. Doménech, P. Manzanares, A.D. Moreno, Advanced bioethanol production: From novel raw materials to integrated biorefineries, Processes. 9 (2021) 1–30. https://doi.org/10.3390/pr9020206.

[85] IEA, Net Zero by 2050: A Roadmap for the Global Energy Sector, Int. Energy Agency. (2021) 224.

[86] R. Sathre, L. Gustavsson, A lifecycle comparison of natural resource use and climate impact of biofuel and electric cars, Energy. 237 (2021) 121546. https://doi.org/10.1016/j.energy.2021.121546.

[87] J.E. Campbell, D.B. Lobell, C.B. Field, Greater transportation energy and GHG offsets from bioelectricity than ethanol, Science (80-.). 324 (2009) 1055–1057. https://doi.org/10.1126/science.1168885.

[88] T. Werpy, G. Petersen, Top Value Added Chemicals from Biomass Volume I, 2004. https://doi.org/10.2172/15008859.

[89] J.J. Bozell, G.R. Petersen, Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "top 10" revisited, Green Chem. 12 (2010) 539–55. https://doi.org/10.1039/b922014c.

[90] E. De Jong, H. Stichnothe, G. Bell, H. Jorgensen, Bio-Based Chemicals: A 2020 Update, 2020. https://task42.ieabioenergy.com/wp-content/uploads/sites/10/2020/02/Bio-based-chemicals-a-2020-update-final-200213.pdf.

[91] L.R. Jarboe, Progress and challenges for microbial fermentation processes within the biorefinery context, in: A-Z Biorefinery, Elsevier, 2022: pp. 447–471. https://doi.org/10.1016/b978-0-12-819248-1.00019-1.

[92] M. Martín, I.E. Grossmann, On the systematic synthesis of sustainable biorefineries, Ind. Eng. Chem. Res. 52 (2013) 3044–3064. https://doi.org/10.1021/ie2030213.

[93] A.J. Ragauskas, G.T. Beckham, M.J. Biddy, R. Chandra, F. Chen, M.F. Davis, B.H. Davison, R.A. Dixon, P. Gilna, M. Keller, P. Langan, A.K. Naskar, J.N. Saddler, T.J. Tschaplinski, G.A. Tuskan, C.E. Wyman, Lignin valorization: Improving lignin processing in the biorefinery, Science (80-.). 344 (2014). https://doi.org/10.1126/science.1246843.

[94] N. Thongchul, P. Charoensuppanimit, A. Anantpinijwatna, R. Gani, S.

Assabumrungrat, Perspectives, challenges and future directions, in: A-Z Biorefinery, Elsevier, 2022: pp. 739–750. https://doi.org/10.1016/b978-0-12-819248-1.00018-x.

[95] Bio-based industry, (n.d.). https://datam.jrc.ec.europa.eu/datam/mashup/BIOBASED_INDUSTRY/ (accessed January 30, 2022).

[96] Production Facilities, (n.d.). https://www.etipbioenergy.eu/databases/production-facilities (accessed January 30, 2022).

[97] Mapping European Biorefineries | Bio-Based Industries Consortium, (n.d.). https://biconsortium.eu/news/mapping-european-biorefineries (accessed January 30, 2022).

[98] EUROPEAN ADVANCED BIOREFINERIES AT COMMERCIAL SCALE, (n.d.). https://biorrefineria.blogspot.com/p/listado-de-biorrefiern.html?m=1 (accessed January 30, 2022).

[99] A. Ernsting, R. Smolker, Dead End Road: The false promises of cellulosic biofuels, 2018. http://www.biofuelwatch.org.uk/wp-content/uploads/Cellulosic-biofuels-report-low-resolution.pdf.

[100] E. Mancini, S.S. Mansouri, K. V. Gernaey, J. Luo, M. Pinelo, From second generation feed-stocks to innovative fermentation and downstream techniques for succinic acid production, Crit. Rev. Environ. Sci. Technol. 50 (2020) 1829–1873. https://doi.org/10.1080/10643389.2019.1670530.

[101] Orion Market Research, Global Succinic Acid Market Forecast, 2016-2026, (2020).

[102] European Commission, EU Biorefinery Outlook to 2030, 2021. https://doi.org/10.2777/103465.

[103] J. Wenger, T. Stern, Reflection on the research on and implementation of biorefinery systems – a systematic literature review with a focus on feedstock, Biofuels, Bioprod. Biorefining. 13 (2019) 1347–1364. https://doi.org/10.1002/BBB.2021.

[104] F. Bauer, L. Coenen, T. Hansen, K. McCormick, Y.V. Palgan, Technological innovation systems for biorefineries: a review of the literature, Biofuels, Bioprod. Biorefining. 11 (2017) 534–548. https://doi.org/10.1002/BBB.1767.

[105] D. Biello, The false promise of biofuels., Sci. Am. 305 (2011) 58–65. https://doi.org/10.1038/SCIENTIFICAMERICAN0811-58.

[106] Our World in Data, Crude Oil Price in USD/barrel, (n.d.).

[107] A.A. Kiss, J.P. Lange, B. Schuur, D.W.F. Brilman, A.G.J. van der Ham, S.R.A. Kersten, Separation technology–Making a difference in biorefineries, Biomass and Bioenergy. 95 (2016) 296–309. https://doi.org/10.1016/j.biombioe.2016.05.021.

[108] Clariant SE, sunliquid, n.d. https://www.clariant.com/de/Business-Units/New-Businesses/Biotech-and-Biobased-Chemicals/Sunliquid.

[109] biofuelsdigest, Beta Renewables, Biochemtex ink deal for commercial-scale cellulosic biofuels project in Slovakia, (n.d.). https://www.biofuelsdigest.com/bdigest/2014/10/06/beta-renewables-biochemtex-ink-deal-for-commercial-scale-cellulosic-biofuels-project-in-slovakia/.

[110] L. Sjöberg, Principles of risk perception applied to gene technology., EMBO Rep. 5 Spec No (2004) S47–S51. https://doi.org/10.1038/sj.embor.7400258.

[111] P. Rzymski, L. Borkowski, M. Drąg, R. Flisiak, J. Jemielity, J. Krajewski, A. Mastalerz-Migas, A. Matyja, K. Pyrć, K. Simon, M. Sutkowski, J. Wysocki, J. Zajkowska, A. Fal, The strategies to support the COVID-19 vaccination with evidence-based communication and tackling misinformation, Vaccines. 9 (2021) 1–9. https://doi.org/10.3390/vaccines9020109.

[112] B. Hyland-Wood, J. Gardner, J. Leask, U.K.H. Ecker, Toward effective government communication strategies in the era of COVID-19, Humanit. Soc. Sci. Commun. 8 (2021) 1–11. https://doi.org/10.1057/s41599-020-00701-w.

[113] S.J. Cooke, A.J. Gallagher, N.M. Sopinka, V.M. Nguyen, R.A. Skubel, N. Hammerschlag, S. Boon, N. Young, A.J. Danylchuk, Considerations for effective science communication, FACETS. 2 (2017) 233–248. https://doi.org/10.1139/facets-2016-0055.

[114] T. Gascoigne, J. Leach, Science communication is more important than ever. Here are 3 lessons from around the world on what makes it work, Conversat. (2020). https://theconversation.com/science-communication-is-more-important-than-ever-here-are-3-lessons-from-around-the-world-on-what-makes-it-work-147670 (accessed January 6, 2022).

[115] T. Chaturvedi, A.I. Torres, G. Stephanopoulos, M.H. Thomsen, J.E. Schmidt, Developing process designs for biorefineries-definitions, categories, and unit operations, Energies. 13 (2020) 1493. https://doi.org/10.3390/en13061493.

[116] H. Rasmussen, H.R. Sørensen, A.S. Meyer, Formation of degradation compounds from lignocellulosic biomass in the biorefinery: Sugar reaction mechanisms, Carbohydr. Res. 385 (2014) 45–57. https://doi.org/10.1016/j.carres.2013.08.029.

[117] Y. Zheng, J. Shi, M. Tu, Y.-S. Cheng, Principles and Development of Lignocellulosic Biomass Pretreatment for Biofuels, 2 (2017) 1–68. https://doi.org/10.1016/bs.aibe.2017.03.001.

[118] Y. Delgado Arcaño, O.D. Valmaña García, D. Mandelli, W.A. Carvalho, L.A. Magalhães Pontes, Xylitol: A review on the progress and challenges of its production by chemical route, Catal. Today. 344 (2020) 2–14. https://doi.org/10.1016/j.cattod.2018.07.060.

[119] A.F. Hernández-Pérez, P.V. de Arruda, L. Sene, S.S. da Silva, A. Kumar Chandel, M. das G. de Almeida Felipe, Xylitol bioproduction: state-of-the-art, industrial paradigm shift, and opportunities for integrated biorefineries, Crit. Rev. Biotechnol. 39 (2019) 924–943. https://doi.org/10.1080/07388551.2019.1640658.

[120] E. Palmqvist, B. Hahn-Hägerdal, Fermentation of lignocellulosic hydrolysates. II: Inhibitors and mechanisms of inhibition, Bioresour. Technol. 74 (2000) 25–33. https://doi.org/10.1016/S0960-8524(99)00161-3.

[121] J.D. Watson, F.H.C. Crick, Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid, Nature. 171 (1953) 737–738. https://doi.org/10.1038/171737a0.

[122] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, H. Erlich, Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction, Cold Spring Harb.

Symp. Quant. Biol. 51 (1986) 263–273. https://doi.org/10.1101/sqb.1986.051.01.032.

[123] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, Science (80-.). 337 (2012) 816–821. https://doi.org/10.1126/science.1225829.

[124] M. Madigan, K. Bender, D. Buckley, W. Sattley, D. Stahl, Brock Biology of Microorganisms, 16th ed., Pearson Education Limited, 2021.

[125] G. Larsson, M. Törnkvist, E. Ståhl Wernersson, C. Trägårdh, H. Noorman, S.O. Enfors, Substrate gradients in bioreactors: Origin and consequences, Bioprocess Eng. 14 (1996) 281–289. https://doi.org/10.1007/BF00369471.

[126] G. Guirimand, N. Kulagina, N. Papon, T. Hasunuma, V. Courdavault, Innovative Tools and Strategies for Optimizing Yeast Cell Factories, Trends Biotechnol. 39 (2021) 488–504. https://doi.org/10.1016/j.tibtech.2020.08.010.

[127] N.M. Markina, A.A. Kotlobay, A.S. Tsarkova, Heterologous Metabolic Pathways: Strategies for Optimal Expression in Eukaryotic Hosts, Acta Naturae. 12 (2020) 28–39. https://doi.org/10.32607/actanaturae.10966.

[128] M. Dragosits, D. Mattanovich, Adaptive laboratory evolution - principles and applications for biotechnology, Microb. Cell Fact. 12 (2013) 64. https://doi.org/10.1186/1475-2859-12-64.

[129] Y. Hua, J. Wang, Y. Zhu, B. Zhang, X. Kong, W. Li, D. Wang, J. Hong, Release of glucose repression on xylose utilization in Kluyveromyces marxianus to enhance glucose-xylose co-utilization and xylitol production from corncob hydrolysate, Microb. Cell Fact. 18 (2019) 24. https://doi.org/10.1186/s12934-019-1068-2.

[130] S. Choi, H. Song, S.W. Lim, T.Y. Kim, J.H. Ahn, J.W. Lee, M.H. Lee, S.Y. Lee, Highly selective production of succinic acid by metabolically engineered Mannheimia succiniciproducens and its efficient purification, Biotechnol. Bioeng. 113 (2016) 2168–2177. https://doi.org/10.1002/bit.25988.

[131] P. Carbonell, A.J. Jervis, C.J. Robinson, C. Yan, M. Dunstan, N. Swainston, M. Vinaixa, K.A. Hollywood, A. Currin, N.J.W. Rattray, S. Taylor, R. Spiess, R. Sung, A.R. Williams, D. Fellows, N.J. Stanford, P. Mulherin, R. Le Feuvre, P. Barran, R. Goodacre, N.J. Turner, C. Goble, G.G. Chen, D.B. Kell, J. Micklefield, R. Breitling, E. Takano, J.L. Faulon, N.S. Scrutton, An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals, Commun. Biol. 1 (2018) 1–10. https://doi.org/10.1038/s42003-018-0076-9.

[132] G. Wu, Q. Yan, J.A. Jones, Y.J. Tang, S.S. Fong, M.A.G. Koffas, Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications, Trends Biotechnol. 34 (2016) 652–664. https://doi.org/10.1016/j.tibtech.2016.02.010.

[133] X. Li, A. Kraslawski, Conceptual process synthesis: Past and current trends, Chem. Eng. Process. Process Intensif. 43 (2004) 583–594. https://doi.org/10.1016/j.cep.2003.05.002.

[134] S.D. Barnicki, J.J. Siirola, Process synthesis prospective, Comput. Chem. Eng. 28 (2004) 441–446. https://doi.org/10.1016/j.compchemeng.2003.09.030.

[135] L. Mencarelli, Q. Chen, A. Pagot, I.E. Grossmann, A review on superstructure

optimization approaches in process system engineering, Comput. Chem. Eng. 136 (2020) 106808. https://doi.org/10.1016/j.compchemeng.2020.106808.

[136] G. Sin, K. V. Gernaey, A.E. Lantz, Good modeling practice for PAT applications: Propagation of input uncertainty and sensitivity analysis, Biotechnol. Prog. 25 (2009) 1043–1053. https://doi.org/10.1002/btpr.166.

[137] J. Eason, S. Cremaschi, Adaptive sequential sampling for surrogate model generation with artificial neural networks, Comput. Chem. Eng. 68 (2014) 220–232. https://doi.org/10.1016/j.compchemeng.2014.05.021.

[138] J.J. Heijnen, W.M. van Gulik, Section II - Balances and Reaction Models, in: C.D. Smolke (Ed.), Metab. Pathw. Eng. Handb. Fundam., CRC Press, Boca Raton, 2009: pp. II-1-11–20.

[139] 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. Biorefining. 14 (2020) 1046–1060. https://doi.org/10.1002/bbb.2108.

[140] G. Sin, P. Ödman, N. Petersen, A.E. Lantz, K. V. Gernaey, Matrix notation for efficient development of first-principles models within PAT applications: Integrated modeling of antibiotic production with Streptomyces coelicolor, Biotechnol. Bioeng. 101 (2008) 153–171. https://doi.org/10.1002/bit.21869.

[141] C.L. Gargalo, A. Carvalho, K. V. Gernaey, G. Sin, Optimal Design and Planning of Glycerol-Based Biorefinery Supply Chains under Uncertainty, Ind. Eng. Chem. Res. 56 (2017) 11870–11893. https://doi.org/10.1021/acs.iecr.7b02882.

[142] B.H. Junker, Scale-up methodologies for Escherichia coli and yeast fermentation processes, J. Biosci. Bioeng. 97 (2004) 347–364. https://doi.org/10.1016/S1389-1723(04)70218-2.

[143] C. Haringa, W. Tang, G. Wang, A.T. Deshmukh, W.A. van Winden, J. Chu, W.M. van Gulik, J.J. Heijnen, R.F. Mudde, H.J. Noorman, Computational fluid dynamics simulation of an industrial P. chrysogenum fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization, Chem. Eng. Sci. 175 (2018) 12–24. https://doi.org/10.1016/j.ces.2017.09.020.

[144] G. Nadal-Rey, D.D. McClure, J.M. Kavanagh, B. Cassells, S. Cornelissen, D.F. Fletcher, K. V. Gernaey, Development of dynamic compartment models for industrial aerobic fed-batch fermentation processes, Chem. Eng. J. 420 (2021) 130402. https://doi.org/10.1016/j.cej.2021.130402.

[145] G. Wang, C. Haringa, H. Noorman, J. Chu, Y. Zhuang, Developing a Computational Framework To Advance Bioprocess Scale-Up, Trends Biotechnol. 38 (2020) 846–856. https://doi.org/10.1016/j.tibtech.2020.01.009.

[146] B.B. Misra, C. Langefeld, M. Olivier, L.A. Cox, Integrated omics: Tools, advances and future approaches, J. Mol. Endocrinol. 62 (2019) R21–R45. https://doi.org/10.1530/JME-18-0055.

[147] B.J. Sánchez, J. Nielsen, Genome scale models of yeast: towards standardized evaluation and consistent omic integration, Integr. Biol. 7 (2015) 846–858. https://doi.org/10.1039/C5IB00083A.

[148] A.M. Feist, M.J. Herrgård, I. Thiele, J.L. Reed, B. Palsson, Reconstruction of biochemical networks in microorganisms, Nat. Rev. Microbiol. 7 (2009) 129–143. https://doi.org/10.1038/nrmicro1949.

[149] B.J. Sánchez, C. Zhang, A. Nilsson, P. Lahtvee, E.J. Kerkhoven, J. Nielsen, Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints, Mol. Syst. Biol. 13 (2017) 935. https://doi.org/10.15252/msb.20167411.

[150] H. Lu, F. Li, B.J. Sánchez, Z. Zhu, G. Li, I. Domenzain, S. Marcišauskas, P.M. Anton, D. Lappa, C. Lieven, M.E. Beber, N. Sonnenschein, E.J. Kerkhoven, J. Nielsen, A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism, Nat. Commun. 10 (2019) 1–13. https://doi.org/10.1038/s41467-019-11581-3.

[151] F.A. Vargas, F. Pizarro, J.R. Pérez-Correa, E. Agosin, Expanding a dynamic flux balance model of yeast fermentation to genome-scale, BMC Syst. Biol. 5 (2011) 75. https://doi.org/10.1186/1752-0509-5-75.

[152] I. Thiele, B. Palsson, A protocol for generating a high-quality genome-scale metabolic reconstruction, Nat. Protoc. 5 (2010) 93–121. https://doi.org/10.1038/nprot.2009.203.

[153] S.S. Da Silva, A.K. Chandel, D-Xylitol: Fermentative production, application and commercialization, Springer-Verlag Berlin Heidelberg, 2012. https://doi.org/10.1007/978-3-642-31887-0.

[154] IMARC, Xylitol Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2021-2026, 2021.

[155] N.I. Vollmer, J.L.S.P. Driessen, C.K. Yamakawa, K. V. Gernaey, S.I. Mussatto, G. Sin, Model development for the optimization of operational conditions of the pretreatment of wheat straw, Chem. Eng. J. 430 (2022) 133106. https://doi.org/10.1016/j.cej.2021.133106.

[156] N.I. Vollmer, R. Al, K. V. Gernaey, G. Sin, Synergistic optimization framework for the process synthesis and design of biorefineries, Front. Chem. Sci. Eng. 16 (2022) 251–273. https://doi.org/10.1007/s11705-021-2071-9.

[157] N.I. Vollmer, K. V. Gernaey, G. Sin, Conceptual Process Design of an Integrated Xylitol Biorefinery with Value-Added Co-Products, Front. Chem. Eng. (2022), accepted/in press.

[158] N.I. Vollmer, C.L. Gargalo, K. V. Gernaey, S.I. Olsen, G. Sin, Life Cycle Assessment of an Integrated Xylitol Biorefinery with Value-Added Co-Products, J. Clean. Prod. (n.d.), submitted/under review.

[159] B. Sarup, Overview of downstream product recovery technology in biorefineries, A-Z Biorefinery. (2022) 509–521. https://doi.org/10.1016/B978-0-12-819248-1.00006-3.

[160] C. Abels, F. Carstensen, M. Wessling, Membrane processes in biorefinery applications, J. Memb. Sci. 444 (2013) 285–317. https://doi.org/10.1016/J.MEMSCI.2013.05.030.

[161] N.I. Vollmer, K. V. Gernaey, G. Sin, Value Chain Optimization of a Xylitol Biorefinery with Delaunay Triangulation Regression Models, Comput. Aided Chem. Eng. (2022),

accepted.

[162] N.I. Vollmer, K. V. Gernaey, G. Sin, Sensitivity Analysis and Risk Assessment for the In-Silico Design and Use of Optimized Cell Factories in a Xylitol Biorefinery, Comput. Aided Chem. Eng. (2022), accepted.

[163] N.D. Price, J.L. Reed, B. Palsson, Genome-scale models of microbial cells: Evaluating the consequences of constraints, Nat. Rev. Microbiol. 2 (2004) 886–897. https://doi.org/10.1038/nrmicro1023.

[164] N. Hillson, M. Caddick, Y. Cai, J.A. Carrasco, M.W. Chang, N.C. Curach, D.J. Bell, R. Le Feuvre, D.C. Friedman, X. Fu, N.D. Gold, M.J. Herrgård, M.B. Holowko, J.R. Johnson, R.A. Johnson, J.D. Keasling, R.I. Kitney, A. Kondo, C. Liu, V.J.J. Martin, F. Menolascina, C. Ogino, N.J. Patron, M. Pavan, C.L. Poh, I.S. Pretorius, S.J. Rosser, N.S. Scrutton, M. Storch, H. Tekotte, E. Travnik, C.E. Vickers, W.S. Yew, Y. Yuan, H. Zhao, P.S. Freemont, Building a global alliance of biofoundries, Nat. Commun. 10 (2019) 1038–1041. https://doi.org/10.1038/s41467-019-10079-2.

[165] M. Chui, M. Evers, J. Manyika, A. Zheng, T. Nisbet, The Bio Revolution, 2020. https://www.mckinsey.com/mgi/overview.

Part

II

Publications

Paper

A

Model development for the optimization of operational conditions of the pretreatment of wheat straw

Nikolaus I. Vollmer, Jasper L.S.P. Driessen, Celina K. Yamakawa, Krist V. Gernaey, Solange I. Mussatto, Gürkan Sin,

Chemical Engineering Journal,

Vol. 430, Part 4, 133106,

https://doi.org/10.1016/j.cej.2021.133106

Abstract

The underlying study presents models for the optimization of operational conditions of the pretreatment of what straw. Experiments for hydrothermal and dilute acid pretreatment are performed and analyzed. The highest xylose monomer yield obtained for dilute acid pretreatment is $Y_{Xyl} = 98$ % at a temperature of $T = 195$℃, a reaction time of $t = 18\ min$ and a dilute acid concentration of $C_{acid} = 1.25\ wt$ %. The data is used to fit a response surface model (RSM), a Gaussian process regression model (GPR), and a mechanistic model based on thermodynamic principles and first-order rate equations. Each model is used in an optimization problem to predict the optimal operational conditions that maximize the xylose yield. The conditions found by the mechanistic model ($T = 191.6$ ℃, $t = 18\ min, C_{acid} = 1.13\ wt.\%$) with $C_{Xyl,mech} = 23.47\ wt.\%$ and the GPR ($T = 195$ ℃, $t = 18\ min, C_{acid} = 1.25\ wt.\%$) with $C_{Xyl,GPR} = 23.23\ wt.\%$ are in agreement and stand out compared to the RSM metamodeling approach ($T = 182.4$ ℃, $t = 26.2\ min, C_{acid} = 1.25\ wt.\%$), which yields $C_{Xyl,RSM} = 25.72\ wt.\%$. Considering the scenario of uncertainty in the feedstock composition, the optimization under this uncertainty with the mechanistic model yields slightly different conditions ($T = 182.6$℃, $t = 18\ min, C_{acid} = 0.84\ wt.\%$) and $C_{Xyl,mech,uc} = 20.88\ wt.\%$. Given the underlying phenomena in the biomass pretreatment, all models have shortcomings; however, the mechanistic model is validated best overall and is thus recommended for further engineering purposes as, e.g., the conceptual process design of biorefineries.

A.1 Introduction

Biorefineries as a replacement for traditional chemical process routes play an essential role in developing more sustainable production patterns demanded by the 2030 Sustainable Development Agenda of the United Nations [1,2]. Nevertheless, despite harboring this immense potential, there are still significant knowledge gaps in industrial biotechnology regarding the use of optimal cell factories in fermentation processes and along the value chain. The latter mainly refers to questions about the optimal use of feedstocks for a set of various products, as well as questions about process intensification and integration, which remain unanswered [3–6].

There is a multitude of explanations for this fact, but foremost the fragile economic potential due to different cost-drivers has impaired a breakthrough of this concept for the time being [7,8]. The fundamental problem is a lack of facilitation in practical applications of the conceptual design of a biorefinery. Such conceptual approaches can describe the inherent complexity of bioprocesses under the use of non-conventional feedstocks to find an optimal and cost-efficient setup. However, such conceptual approaches are seldom employed despite many approaches being published [3,9–12]. From a technical perspective, many problems remain unsolved: lignocellulosic biomass as a sugar source for fermentation processes requires an extensive pretreatment process of the biomass to reduce the recalcitrance of the feedstock and to release both the sugars from the hemicellulosic fraction and the cellulosic fraction respectively. There has been steady progress in developing new pretreatment methods or improving existing methods further, but despite the abundant number of methods, selecting a suitable pretreatment is far from straightforward [13,14]. Furthermore, all pretreatment methods are associated with high capital and operational expenditures due to the extreme process conditions, which remains a prominent issue regarding the economics of a biorefinery [15].

From a holistic perspective, the biomass pretreatment both has a critical influence on the overall process economics and also a crucial role in the actual design process for a biorefinery, as critical decisions on the recovery of the fractions in the lignocellulosic biomass and subsequently the possible product sets for the biorefinery are made [16]. Independent from the potentially viable products, the optimal pretreatment unit fulfills three criteria: 1) a precise split between the fractions in the biomass, 2) high respective yields for the monomers in the fractions, and 3) a low formation of inhibitors in the pretreatment process. Depending on which products are supposed to be produced, the applied method can vary, but the requirements stay the same [13].

For a biorefinery with a viable economic potential, the utilization of the feedstock is the key factor. Hence, for its conceptual design, computational alternatives provide the potential to yield conceptually feasible solutions. The crucial factor in this is the availability of models that accurately describe the underlying unit operations to design the process [9].

This work focuses on the development of a biomass pretreatment model for this precise purpose. As a case study, the biotechnological production of xylitol in a biorefinery is chosen: in 2004, the US Department of Energy declared xylitol as one of the top 12 chemicals to be produced in a biorefinery [17]. Xylitol can be produced from xylose via fermentation

in a suitable cell factory. Xylose is the main constituent of the hemicellulosic fraction in lignocellulosic biomass. These conditions make it an ideal product for the production in a biorefinery; however, research on xylitol production mainly focuses either on the pretreatment, or the fermentation, or the downstream processing, and a holistic perspective on the whole process is lacking [18–21].

In this study's scope, a suitable feedstock and pretreatment method for the given case are selected for the introduced criteria. Based on this, experiments for the selected feedstock and pretreatment methods are designed and performed. After analyzing the results of these experiments, different model candidates are calibrated to the data and subsequently validated, including an assessment of their robustness. All models are then taken to optimize the pretreatment conditions towards a selected objective for the case study. Lastly, these results are compared, and the most suitable model candidate for the conceptual process design of a biorefinery is selected.

The remainder of this study is structured as follows: in section 2, biomass pretreatment methods and feedstocks, in general, are introduced, and also the criteria applicable for selecting a method and a suitable feedstock in process design. Section 3 describes the experimental procedure of feedstock analysis, the design, performance, and analysis of the pretreatment experiments. Section 4 introduces both the data-driven models and the knowledge-driven models, which are employed in this study. For the knowledge-driven model, it is explained in detail how the parameters of the model are identified and estimated, as well as the employed procedure for the robustness assessment via uncertainty and sensitivity analysis. The procedure of validating all models is explained, as well as the optimization with all models. Section 5 includes the results of the feedstock analysis, both pretreatment experiments, the calibration and validation of data-driven models, the calibration, validation, and the robustness assessment of the knowledge-driven model, as well as the optimization study. Lastly, in section 6, the study's primary outcomes are summarized, and instigations for further work are given.

A.2 Materials and Methods

A.2.1 Feedstock Composition

For the performed experiments, wheat straw of the variety Triticum aestivum from a field in Freerslev sogn (Faxe, Denmark) of the harvest of fall 2018 was taken as feedstock. The wheat was harvested and dried. The chaff was separated from the wheat; the straw was dried again and then milled and ground. The resulting particle size after milling is in the range between 18 and 40 mm.

First, the respective feedstock is analyzed regarding its composition, particularly the hemicellulosic, the cellulosic, and the lignin fraction. This is performed with extractive-free biomass as described by the National Renewable Energy Laboratory (NREL) [26]. The extractives have been removed as described by the NREL [27]. Glucose, xylose, arabinose, acetic acid, formic acid, furfural, and 5-HMF were quantified by high-performance liquid chromatography (HPLC), using a Dionex Ultimate 3000 UHPLC+ Focused System (Dionex Softron GmbH, Germany) with a Bio-Rad Aminex column HPX-87H (300mm × 7.8 mm) at 60 °C, and 5.0 mM sulfuric acid as mobile phase at a flow rate of 0.6 mL/min. Sugars and

acids were detected using a Shodex RI-101 refractive index detector, whereas 5-HMF and furfural were detected using ultraviolet measurements at 254 nm.

A.2.2 Pretreatment Experiments

For each point in the design of experiments, a batch experiment is performed. The used batch vessel is a non-stirred pressure vessel (Parr Series 4760, 600 mL, Parr Instrument Company, Moline, IL). In preparation for the experiments, the milled biomass's moisture content and its dry mass are determined with an automated moisture analyzer (MB 163-M, VWR International) as a first step. For the hydrothermal pretreatment, the corresponding amount of water is weighed to meet the set solid-to-liquid ratio. For the dilute acid pretreatment, the sulfuric acid and its water content are determined according to the experiment's acid concentration. The amount of water is weighed respectively. Then, the acid is mixed with the water. Subsequently, the biomass and the liquid are added alternatingly in small portions into the reaction vessel until it is filled. Lastly, the mixture is stirred thoroughly by hand to achieve an equal distribution of the liquid.
As soon as this is achieved, the reactor is closed and sealed, and put into a heating bath that is set to the desired temperature for the experiment. After a transition period of t=5 min when the reaction vessel's temperature rises to the silicone bath's temperature level, the time measurement is started. After the desired residence time, the reaction vessel is taken out of the heating bath and cooled rapidly to inhibit further degradation reactions. After approximately 30 min, the reactor vessel is opened, and the liquid phase is separated from the solid phase through a simple sieve. The solid residue is put in a mechanical press to extract the maximum amount of free liquid phase, and hereafter its moisture content is determined again. The volume of the total amount of liquid hydrolysate is determined as well as its pH value. Lastly, the hydrolysate is filtered once with a vacuum filter in order to remove solid residues. The hydrolysate samples are also analyzed by HPLC, as depicted in section 2.1.

A.2.3 Post-hydrolysis Experiments

The hydrothermal pretreatment with its neutral pH range proves to yield a significant amount of oligomeric sugars, indicating slower depolymerization reactions [13]. Therefore, the hydrolysate is commonly subjected to a post-hydrolysis step to increase the degradation of the oligomers into sugar monomers. This post-hydrolysis is performed as described by the NREL [28]. Hereafter, the samples are again analyzed by HPLC to determine the concentrations of the created monomers.

A.2.4 Response Surface Methodology

RSM as statistical analysis is described by Box and Wilson as follows: The actual response surface is a second-order polynomial fitted to experimental data to predict optimal conditions for the given set of factors D of the design of experiments. The general form of the polynomial is the following:

$$y = \alpha + \beta \cdot b + \sum_{i \in D} \gamma_i \cdot x_i + \sum_{i \in D} \sum_{j \geq i \in D} \delta_{ji} \cdot x_i \cdot x_j, \tag{A-1}$$

In which x_i denote the factors, y the response and α, β, γ and δ the coefficients of the polynomial [30]. The RSM model with a second-order polynomial simplifies the underlying system significantly, but its use for statistical analysis of experiments remains high [31]. In this study, the RSM model in the *rsm* library in R is used [32]. After calibrating the model by

fitting its parameters to experimental data, the validation of the model is crucial to confirm both the goodness of fit of the estimation and its predictive capability for the system it was calibrated for [45]. In the case of the RSM model, the model is validated by an analysis of variance (ANOVA). Furthermore, common validation metrics which quantitatively express both the goodness of fit and the predictive capacities are the coefficient of determination R^2 and the root mean square error $RMSE$, which are both used in this study. A detailed description is provided with the supplementary material.

For the hydrothermal pretreatment, the factors for the design of experiments are chosen to be the reaction temperature T in the interval $T = [160,195]$ °C and the reaction time in the interval $t = [20,60]\ min$. The central composite design (CCD) is chosen with $\alpha = 0.714$, two cube center points, and one star center point, which yields eleven experimental points. For the dilute acid pretreatment, the factors for the design of experiments are chosen to be the reaction temperature T in the interval $T = [160,195]$ °C, the reaction time t in the interval $t = [20,60]\ min$ and the acid concentration C_{ac} in the interval $C_{ac} = [0.8,1.3]\ wt\ \%$. The CCD is chosen to be circumscribed with $\alpha = 1.69$, two cube center points, and one star center point, which yields seventeen experimental points. The response variable y for all is the xylose concentration. A detailed description of the setup of the design of experiments can be found in the supplementary material. The scripts for the design of experiments are provided through a GitHub repository [33].

A.2.5 Gaussian Process Regression

A GPR model's predictive capacities derive from the eponymous stochastic process: the prediction of interpolated values is governed by prior covariances of training data points and described by specific kernel functions, whose parameters are fitted to these points [29]. This is expressed as follows:

$$y = \mu(x) + \sigma^2 \cdot \mathcal{Z}(x, \omega), \qquad \mu(x) = \rho \cdot \beta(x). \tag{A-2}$$

Here, the predicted output y is described by $\mu(x)$ as the mean value of the stochastic process with the input x and σ^2 as its variance; $\mathcal{Z}(x, \omega)$ denotes a zero mean unit variance stochastic process with the mentioned kernel function ω. For the mean value, ρ relates to parameters which are fitted based on the training data and $\beta(x)$ describes a set of basis functions. Regarding the employable basis and kernel functions, the reader is referred to the book by Rasmussen [34]. In this study, the *fitrgp* function of the Statistics and Machine Learning Toolbox in MATLAB is used for fitting the GPR model. Analogously to the RSM model, the model input x consists of the three design factors T, t and C_{ac}, and the model output y respectively corresponds to the measured concentrations. The model validation is performed by cross-validation equally by calculating the coefficient of determination, and the root mean square error for the model, as described in section A.3.5. Further explanation on the validation is given in the supplementary material.

A.2.6 Mechanistic Model

The backbone of most published models for the pretreatment of lignocellulosic biomass are mass and energy balances; however, different types of reaction equations are employed. Mostly, these are pseudo-first or second-order reaction kinetics [35–37]. Generally, it is rather difficult to develop a wholesome model, as some reaction mechanisms and some components are unknown and vary highly between different feedstocks and pretreatment methods [38]. Furthermore, measuring these components in experimental setups is not straightforward, as subsequent reactions occur relatively fast at a particular stage. Hence,

mechanistic models for biomass pretreatment commonly only capture one specific pretreatment method in a certain range of operational conditions for a selected number of components, depending on the provided experimental data and the respectively estimated parameters.

For the model in this study, we choose an equivalent approach based on mass balances and reaction kinetics. The considered set of components J in this model is listed in Table 3:

Table 3: List of considered components in the pretreatment model.

Number	Component	Shorthand symbol in model
1.	Xylan	(Xyn)
2.	Xylose	(Xyl)
3.	Cellulose	(Cel)
4.	Glucose	(Glu)
5.	Arabinan	(Arn)
6.	Arabinose	(Ara)
7.	Acetyl Groups	(Act)
8.	Acetic Acid	(Aac)
9.	Furfural	(Fur)
10.	5-Hydroxymethylfurfural (5-HMF)	(Hmf)
11.	Formic acid	(Fac)
12.	Further Degradation Products	(Deg)

The set of occurring reactions I which are considered in the model are illustrated in the following Figure 3:

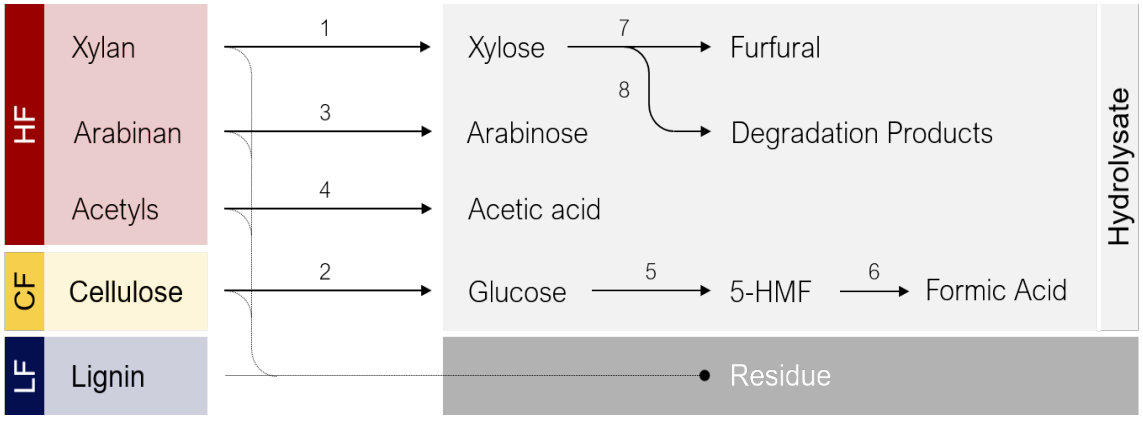

Figure 3: Illustration of the considered reactions occurring during the pretreatment, describing the transfer of components of the HF, CF, and LF into hydrolysate and residue.

Hence, the number of reactions is $|I| = 8$. The reaction equations, including their stoichiometry, are provided in the supplementary material. Each reaction $i \in I$ occurs at a reaction rate r_i which can be calculated with the rate constant k_i of the reaction and the concentrations of the participating reactants C_j in $wt\%$ or $g/100g$ of dry biomass for $j \in J_i \subset J$ with J_i as a subset of the set of all reactants J.

$$r_i = k_i \cdot \prod_{j \epsilon J_i} C_j \quad \forall\, i \in I. \tag{A-3}$$

The rate constant k_i for every reaction i can be determined by the Arrhenius law as follows:

$$k_i = A_i \cdot exp\left(-\frac{E_{A,i}}{\tilde{R} \cdot T}\right) \cdot C_{acid}^{n_i} \quad \forall\, i \in I, \tag{A-4}$$

with A_i as the frequency factor in min^{-1}, $E_{A,i}$ as the activation energy in $kJ \cdot mol^{-1}$, $\tilde{R}$ as the universal gas constant in $kJ \cdot mol^{-1} \cdot K^{-1}$, T as the temperature of the reaction in K, C_{acid} as the concentration of supplied acid in case of dilute acid pretreatment in $wt\%$ or $g/100g$ of dry biomass and n_i as reaction order exponent for the participation of acid in each respective reaction.
For each component C_j a component balance can be formulated:

$$\frac{dC_j}{dt} = \sum_{i \in I_J} s_i \cdot r_i \quad \forall\, j \in J, \tag{A-5}$$

with $I_J \subset I$ as the subset of reactions which involve component j and s_i as the stoichiometric factor for the reaction r_i. All stoichiometric factors can be summarized in the stoichiometric matrix S with the dimensions $|I| \times |J|$:

$$S = \begin{pmatrix} -1 & 1.136 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1.111 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1.136 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1.364 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{A-6}$$

The values for the reaction towards xylose, glucose arabinose, and acetic acid take into account the anhydrous form in which the individual molecule is present in the biomass's polymeric structure [28]. All equations as indicated are implemented in MATLAB and solved subsequently. The differential equations are solved with the *ode15s* solver. The model validation is performed by a train & test single-split validation. Also, the coefficient of determination and the root mean square error for the model are used as described in section A.3.5. Further explanation regarding the validation is provided through the supplementary material.

A.2.7 Parameter Estimation & Identifiability Analysis

A.2.7.1 Parameter Estimation

The approach in this section is based on maximum-likelihood estimation described by Sin and Gernaey (2016) [39]. A detailed summary can be found in the supplementary material. The values which are determined by the maximum-likelihood estimation are the mean estimate $\hat{\theta}_i$ of the parameter i, the standard deviation σ_i and the lower and upper bound of the 95 % confidence interval of the estimate l_i and u_i.

A.2.7.2 Identifiability Analysis

Depending on the modeled system, a mechanistic model can comprise a very high number of parameters that are supposed to be estimated. Depending on the model structure and the

amount of available experimental data, this can lead to high standard deviations for the estimated values. Hence, it is paramount to identify a subset of parameters $\theta_k \in \theta$ which is significant for the model output and uniquely estimable by parameter estimation from a distinct data set [40]. The described identifiability analysis methodology is based on a local sensitivity analysis known as one-factor-at-a-time (OAT) method as described by Brun et al. (2002) [41]. The analysis quantifies the significance of each parameter – expressed by the value δ^{msqr} – and describes the collinearity of each possible combination of datasets by a collinearity index γ. A detailed explanation regarding the calculation is given in the supplementary material.

A.2.8 Uncertainty and Sensitivity Analysis

A.2.8.1 Monte Carlo-Based Uncertainty Analysis

Due to several error sources introduced in the model, e.g., measurement errors in experimental data, there is a need to quantify the uncertainty in the model output [40]. Several ways of performing this assessment exist, of which a Monte Carlo method is employed here. According to Sin et al. (2009), Monte Carlo-based uncertainty analysis is practically performed in four steps: 1) definition of input uncertainty, in this case, the uncertainty from measurement errors in the parameter estimation, 2) sampling, in this study Latin Hypercube Sampling, 3) Monte Carlo simulations, and 4) result analysis. A detailed description of the procedure can be found in the supplementary material. The entire procedure is implemented in MATLAB.

A.2.8.2 Variance-Based Sensitivity Analysis

After performing an uncertainty analysis, which describes the model output uncertainty with uncertain input, the complementary analysis is a sensitivity analysis, aiming to apportion the output uncertainty on the different inputs [40]. For this, commonly, the first-order sensitivity index S_i and the total sensitivity index S_{Ti} are calculated for each variable. They describe respectively the sensitivity of the model output to the respective parameter solitarily, hence the first order, and the sensitivity of the model output to the parameter including all interactions with other parameter, hence the total sensitivity [42,43]. In particular, in this study, a global sensitivity analysis based on the work of Saltelli et al. (2010) is implemented: the numerical calculation of both sensitivity indices is performed with Sobol sequence sampling: As a first step, two sampling matrices A and B are generated by Sobol sampling. Subsequently, two mixed matrices A_B^i and B_A^i are generated, in which column i from the one matrix is replaced by the same column of the respective other matrix, and all other columns are maintained. Then, the model outputs are calculated for all four sampling matrices, and the respective sensitivity measures can be calculated. The first-order sensitivity index is calculated as:

$$S_i = V(y) - \frac{1}{2N}\sum_{j=1}^{N}\left(y_B(j) - y_{ABi}(j)\right)^2, \tag{A-7}$$

and the total sensitivity index as:

$$S_i = \frac{1}{2N}\sum_{j=1}^{N}\left(y_A(j) - y_{ABi}(j)\right)^2, \tag{A-8}$$

according to the referred methods [44]. A detailed description of the backgrounds of the method can be found in the supplementary material. The *easyGSA* toolbox in MATLAB is used for the analysis [42].

A.2.9 Optimization

A.2.9.1 Optimization Excluding Uncertainty

To maximize the monomer yield by optimizing the operative conditions of the biomass pretreatment, all presented data and knowledge-driven models can be employed. The optimization problem, which needs to be solved for this is independent of the model and can be formulated as follows:

$$\begin{aligned} &\max y = f(x) \\ &x \in X \subset \mathbb{R}^n \end{aligned} \tag{A-9}$$

The optimization problem is defined as a constrained nonlinear program (NLP). Dependent on each model, different optimization techniques are applicable; they are described in detail for each model in the supplementary material. For the optimization with the RSM model, the *rsm* package in R is chosen [32]. For the optimization with the GPR model, the *ga* solver in MATLAB is chosen. Finally, for the optimization with the mechanistic model, the *fmincon* solver with a sequential quadratic programming algorithm and a multi-start setup with the *MultiStart* function in MATLAB is chosen.

A.2.9.2 Optimization Including Uncertainty

As models predict results with inherent uncertainties, ideally, these should be considered when optimizing. Furthermore, the assumptions taken before building a model can significantly impact the prediction of the model output. Therefore, taking these into account while optimizing the operational conditions can potentially find a more robust optimum under more robust conditions. A more detailed introduction is given in the supplementary material. For this, the *MOSKopt* solver in MATLAB is used [46].

A.3 Results

A.3.1 Feedstock Analysis

The weight percentages per total dry weight of biomass of the composition of the wheat straw were determined according to the description in section A.2.1 and are listed in the following Table 4:

Table 4: Results of the composition analysis of the fiber fraction of wheat straw.

Component	**This study**	**Vassilev et al., 2012** [47]
Cellulose	40.7 wt. % ($\sigma = 0.0003$)	44.5 wt. %
Hemicellulose	33.6 wt. % ($\sigma = 0.0025$)	33.2 wt. %
Lignin (soluble & insoluble)	24.9 wt. % ($\sigma = 0.0041$)	22.3 wt. %

Vassilev et al. (2012) reviewed nine different research papers that indicate the composition of wheat straw and present the mean value [47], which is used for a comparison here. Thus, it becomes evident that the wheat straw composition in this study is in good agreement with the literature values; only the present amount of cellulose is slightly lower than the reference value.

Regarding the composition of the hemicellulosic fraction, also several values are reported in the literature. Hence, it is assumed that an average value of 80% is attributed to xylan, 10% to arabinan, and 10% to acetyl groups for straw [47,48]. This assumption will be revisited and also taken into account in the optimization under uncertainty in subsection A.2.9.2.

A.3.2 Hydrothermal Pretreatment

For the experiments, a solid-to-liquid ratio of $1:10$ (w/w) is chosen with a dry mass of feedstock of $30\ g$. The experiments are performed as described in section A.2.2 and the results are listed in Table 5.

Table 5: Results from the HPLC analysis of the autothermal pretreatment experiments.

Number	Temperature ℃	Time Min	Xyl wt%	Glu wt%	Ara wt%	Aac wt%	Hmf wt%	Fac wt%	Fur wt%
1	165	40	1.62	0.56	0.15	0.61	0.00	0.13	0.00
2	177.5	54	1.67	0.46	0.54	1.31	0.01	0.23	0.02
3	177.5	40	1.70	0.54	0.41	0.92	0.01	0.18	0.01
4	177.5	26	1.62	0.57	0.11	0.47	0.00	0.10	0.00
5	190	40	1.69	0.39	0.67	1.33	0.01	0.29	0.03
6	160	20	1.59	0.57	0.07	0.39	0.00	0.00	0.00
7	160	60	1.66	0.58	0.86	0.75	0.00	0.15	0.00
8	177.5	40	1.54	0.49	0.32	0.85	0.01	0.17	0.01
9	177.5	40	1.65	0.53	0.39	0.92	0.00	0.18	0.01
10	195	60	1.66	0.29	0.84	2.95	0.05	0.87	0.54
11	195	20	1.55	0.54	0.14	0.54	0.00	0.12	0.00

The yield of xylose Y_{Xyl} (w/w) in the hydrothermal pretreatment lies at around $Y_{Xyl} = 5 - 6\ \%$. It becomes evident that the selected pretreatment method combined with the feedstock and the experimental condition is not a suitable combination for the given criteria in section A.1. The most obvious reason for this is possibly the high recalcitrance of the feedstock. Due to the mild conditions of the hydrothermal pretreatment, the polymeric chains are not entirely broken down into monomers, but many oligomeric sugars are released from the biomass and are not depolymerized furtherly [49,50].

For a further analysis of the hydrolysate, a post-hydrolysis is performed as described in section A.2.3. The detailed results are listed in Table 6. As a result, the only notable increase in the concentration of xylose monomers occurred for the experiment with the highest temperature and the longest reaction time; the yield is $Y_{Xyl} = 65.7\ \%$. Other studies also report a maximum hemicellulosic sugar recovery for post-hydrolyzed fractions of hydrothermally pretreated wheat straw of around 60 % to 70 % [49,50]. Considering that such reaction conditions are unfavorable from an economic perspective, the additional post-hydrolysis step does not significantly improve the feasibility of the hydrothermal pretreatment in this setup.

Table 6: Results from the HPLC analysis of the posthydrolyzed autothermal pretreatment experiments.

Number	Temperature °C	Time Min	Xyl wt%	Glu wt%	Ara wt%	Aac wt%	Hmf wt%	Fac wt%	Fur wt%
1	165	40	1.82	1.12	0.34	0.43	0.14	0.18	0.04
2	177.5	54	3.95	1.42	1.45	1.63	0.10	0.27	0.17
3	177.5	40	1.60	1.50	1.12	0.96	0.11	0.26	0.12
4	177.5	26	2.18	1.31	0.31	0.46	0.18	0.22	0.04
5	190	40	5.84	1.59	1.76	1.87	0.09	0.31	0.36
6	160	20	2.13	1.23	0.24	0.35	0.16	0.18	0.03
7	160	60	1.97	1.09	0.50	0.53	0.11	0.20	0.05
8	177.5	40	1.47	1.37	1.24	1.66	0.12	1.70	0.11
9	177.5	40	1.53	1.39	1.13	0.91	0.11	0.23	0.12
10	195	60	15.29	1.63	1.70	3.62	0.12	0.66	1.16
11	195	20	2.34	1.36	0.43	0.54	0.14	0.25	0.06

A.3.3 Dilute Acid Pretreatment

For the experiments, a solid-to-liquid ratio of $1:10$ (w/w) is chosen with a dry mass of feedstock of $30\ g$. The chosen acid is sulfuric acid and the acid concentration C_{ac} concerning the solid mass is chosen to be in the interval $C_{ac} = [0.8,1.3]$ %. The experiments are performed as described in section A.2.2, and the results are listed in Table 7:

Table 7: Results from the HPLC analysis of the dilute acid pretreatment experiments.

No.	Temperature	Time	Acid	Xyl	Glu	Ara	Aac	Hmf	Fac	Fur	Xyl yield
	°C	Min	wt. %	wt. %	wt. %	wt. %	wt. %	wt. %	wt. %	wt. %	%
1	160	10	0.8	1.47	0.58	1.18	0.38	0.00	0.00	0.00	6.21
2	160	26	1.7	17.84	1.80	2.25	2.82	0.17	0.00	0.28	75.38
3	173	18	1.25	20.28	1.63	2.66	2.85	0.12	0.00	0.09	85.69
4	186	26	1.7	19.48	0.40	2.63	4.11	0.22	0.31	1.99	82.35
5	186	10	0.8	4.61	0.75	2.15	0.88	0.03	0.00	0.01	19.48
6	160	10	1.7	3.15	0.51	1.50	0.81	0.01	0.00	0.00	13.31
7	160	26	0.8	12.79	1.06	2.37	1.65	0.07	0.00	0.03	54.07
8	173	18	1.25	21.04	1.84	2.64	2.95	0.15	0.00	0.13	88.92
9	186	26	0.8	23.27	2.48	2.72	3.57	0.28	0.24	0.56	98.34
10	195	18	1.25	23.29	2.27	2.72	3.63	0.23	0.19	0.34	98.42
11	151	18	1.25	1.44	0.61	1.68	1.49	0.01	0.00	0.00	6.07
12	173	4.55	1.25	1.44	0.56	0.96	0.31	0.00	0.00	0.00	6.10
13	173	18	1.25	20.45	1.72	2.55	2.95	0.14	0.22	0.12	86.43
14	173	31.45	1.25	21.30	2.56	2.62	3.71	0.26	0.28	1.01	90.02
15	173	18	2.0	21.03	2.12	2.56	4.62	0.20	0.23	0.28	88.89
16	173	18	0.5	13.80	1.24	1.70	2.04	0.10	0.13	0.10	58.32
17	186	10	1.7	1.66	0.67	1.57	0.38	0.01	0.00	0.00	7.01

For the sugar monomer yield, it becomes apparent that the dilute acid pretreatment releases a significantly higher amount of xylose monomers compared to the results of the

hydrothermal pretreatment in section A.3.2. Especially points in the octants with a higher temperature than the center point, longer reaction times than the central point, and higher and lower acid concentrations than the central point – points 3, 4, 8, 9, 10, 13, 14, and 15 – show very high monomer yields. Taking into account the stoichiometric factor, an average hydrolysate volume of $V = 250\,mL$, as well as the amount of xylan in hemicellulose, the highest yield is obtained for point ten with $Y_{Xyl} = 98\,\%$. Similar high yields have been reported for a combined dilute-acid and steam explosion pretreatment of wheat straw, for the pretreatment of wheat straw with a subsequent enzymatic hydrolysis step, or for lower absolute monomer concentrations. [51–53]. The degradation of the cellulosic fraction is comparatively small. As the furfural production occurs by the degradation of xylose, the amount of furfural for these conditions is low. On the other hand, the amount of acetic acid for these conditions is high, indicating an equally high yield for this reaction. In conclusion, the dilute acid pretreatment for combining the given feedstock, pretreatment method, and experimental conditions seems advantageous and a good option for the study.

A.3.4 Model Calibration of Mechanistic Model

All model calibration and validation will only be performed based on the dilute acid pretreatment. Furthermore, published studies indicate that the prediction of xylose with a rate constant, as shown in equation (A-4), might impair predictions over the large design space considered in this study [35,54]. Preliminary analyses with the presented experimental data confirm this. Hence, the mechanistic model will only be fitted to a spherical sector of the design space described earlier in section A.2.7, as the resulting xylose concentrations were the highest over the whole design space. The reasoning behind this is the potential application in a biorefinery context, where a maximum amount of monomers, as one of the defined criteria, is required. Hence, the data points for estimation are points 3, 8, 10, 13, 14, 15, and 16, as listed in Table 7. Data points 4 and 9 are used for the model validation. The design space including these points is illustrated in Figure 4.

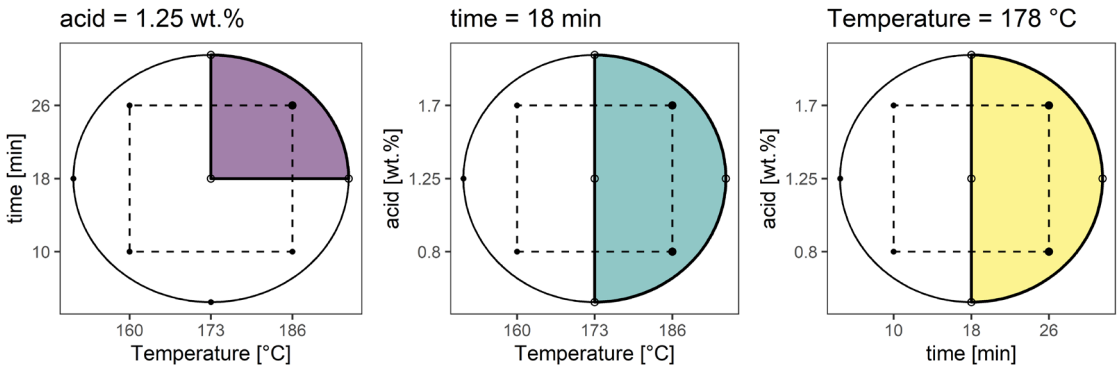

Figure 4: Design space of the experiments for all 17 operational conditions(• not estimated, ○ training data set, ● testing data set).

The scripts for the identifiability analysis and the parameter estimation are provided through a GitHub repository [33].

A.3.4.1 Identifiability Analysis

The Identifiability analysis for the kinetic model is performed as described in section A.2.7.2. The values which are determined by the maximum-likelihood estimation are the mean estimate $\hat{\theta}_i$ of the parameter i, the standard deviation σ_i and the upper and lower bound of the 95 % confidence interval of the estimate l_i and u_i.

The used values for the parameters are obtained by a preliminary run of the parameter estimation as described in subsection A.2.7 with arbitrary initial values: all frequency factors were set to $A_i = 1 \cdot 10^{10}\ s^{-1}$, all activation energies were set to $E_i = 100\ kJ \cdot mol^{-1}$ and all reaction order exponents were set to $n_i = 1$. The values for the initial parameter values are provided in the supplementary material. The resulting values for δ_{msqr} are illustrated in Figure 5.

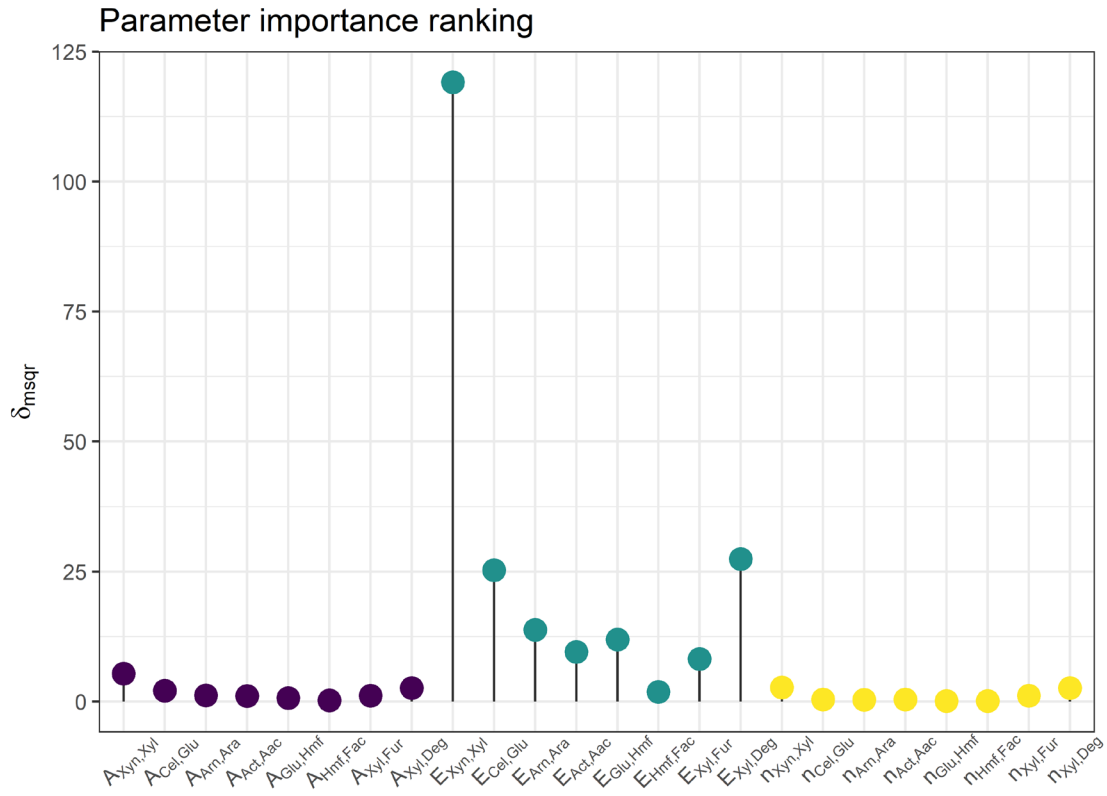

Figure 5: Values for the parameter significance calculated as δ_{msqr} for all parameter values.

The figure clearly illustrates that the most significant parameters are all activation energies, whereas both the frequency factors and the reaction order exponents are of minor significance. This is in agreement with similar investigated results, e.g., in Prunescu et al. (2015) [37]. In consequence, the activation energies are selected as the parameters to be estimated.

A.3.4.2 Parameter Estimation

The mechanistic model's parameter estimation is performed as described in section A.2.7.1 for the eight parameters selected in section A.3.4.1, namely the activation energies. The

estimated parameters' initial values are the results from the initial estimation, as performed in section A.3.4.2. They are listed in the supplementary material. With these set values, the estimated values for the activation energies, their standard deviations, and upper and lower bounds for the confidence intervals result are given in Table 8:

Table 8: Values of the estimated parameters of the pretreatment model.

Parameter	$\hat{\theta}_i$	σ_i	l_i	u_i
$E_{Xyn,Xyl}$	121.29	0.1048	121.08	121.5
$E_{Cel,Glu}$	39.246	0.2618	38.718	39.775
$E_{Arn,Ara}$	56.917	0.3761	56.158	57.677
$E_{Act,Aac}$	38.66	0.2469	38.161	39.159
$E_{Glu,HMF}$	66.425	1.2743	63.851	68.998
$E_{HMF,Fac}$	38.036	2.5472	32.891	43.18
$E_{Xyl,Fur}$	25.907	0.7316	24.429	27.384
$E_{Xyl,Deg}$	37.374	0.2053	36.959	37.789

Also, the following correlation matrix results from the parameter estimation:

$$CORR = \begin{pmatrix} 1 & 0 & 0.02 & 0 & 0 & 0 & 0.01 & 0.56 \\ 0 & 1 & 0 & 0 & 0.69 & 0.01 & 0 & 0 \\ 0.02 & 0 & 1 & 0 & 0 & 0 & 0.57 & 0.28 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.69 & 0 & 0 & 1 & 0.01 & 0 & 0 \\ 0 & 0.01 & 0 & 0 & 0.01 & 1 & 0 & 0 \\ 0.01 & 0 & 0.57 & 0 & 0 & 0 & 1 & 0.51 \\ 0.56 & 0 & 0.28 & 0 & 0 & 0 & 0.51 & 1 \end{pmatrix}.$$

The resulting standard deviations of the parameter estimation are low and thus indicate a very good fit. This is equally reflected in the correlation matrix, where only a few estimated parameters have a significant but not critical covariance. In comparison with other reported values, the found values for the activation energies in this study are in good agreement with findings in other studies [37,55,56]. Overall, it can be concluded that the parameter estimation in combination with the identifiability analysis leads to an excellent estimation result that can be further investigated.

A.3.5 Model Validation

With the given experimental data points for the dilute acid pretreatment, the RSM, the GPR, and the mechanistic model are fitted. For the GPR model. the used basis function β is constant, and the used kernel function ω is an exponential function. For the mechanistic model, the split between training and testing data set is described in section A.3.4. The scripts for all model validations are provided through a GitHub repository [33].

The metrics for each model are calculated as indicated in section A.2. This comprises the analysis of variance (ANOVA) for the RSM model, a k-fold cross-validation for the GPR with $k = 5$ and a single-split validation with training and testing data set for the mechanistic model. The validation metrics for all models for training and testing data sets (if applicable) are indicated in Table 9:

Table 9: Validation metrics of the RSM, the GPR, and the mechanistic model.

	RSM	GPR	Mechanistic Model
R^2_{train}	0.914	0.999	0.988
R^2_{test}	−	0.141	0.305
$RMSE_{train}$	2.537	0.969	0.842
$RMSE_{test}$	−	1.277	0.682

The ANOVA returns $F = 6.403$ on 10 and 6 degrees of freedom, in connection with $p = 0.0169 < 0.05$. Hence, the null hypothesis is rejected, and the model is considered validated. In conclusion, the RSM model combined with a CCD appears to be a viable combination to build a data-driven model from experimental data in this study's scope. The GPR training data set metrics indicate an excellent fit, whereas the testing data set $RMSE$ increases significantly, indicating a possibly inferior predictive capacity. For the mechanistic model, both metrics for both data sets show a very good fit with even lower $RMSE$ for the testing set than the GPR; hence, the model also validates well.

A.3.6 Uncertainty and Sensitivity Analysis

A.3.6.1 Uncertainty Analysis

The uncertainty analyses for all model outputs of the mechanistic model were performed as described in section A.2.8. The considered uncertainty here is the one deriving from the estimation of the model parameters. The errors are assumed to be normally distributed according to their standard deviation, as measurement errors from the experimental data as error source are assumed to have random character. With the correlation matrix and the standard deviation of the parameters, a multivariate normal random distribution is generated for $N = 1000$ Monte Carlo samples. The Monte Carlo procedure is performed N times for all 17 operational conditions. The scripts for the uncertainty analysis are provided through a GitHub repository [33]. The results for the uncertainty in the predictions of xylose, acetic acid, and furfural are illustrated in the following violin plot in Figure 6. The violin plots for the other components are provided with the supplementary material.

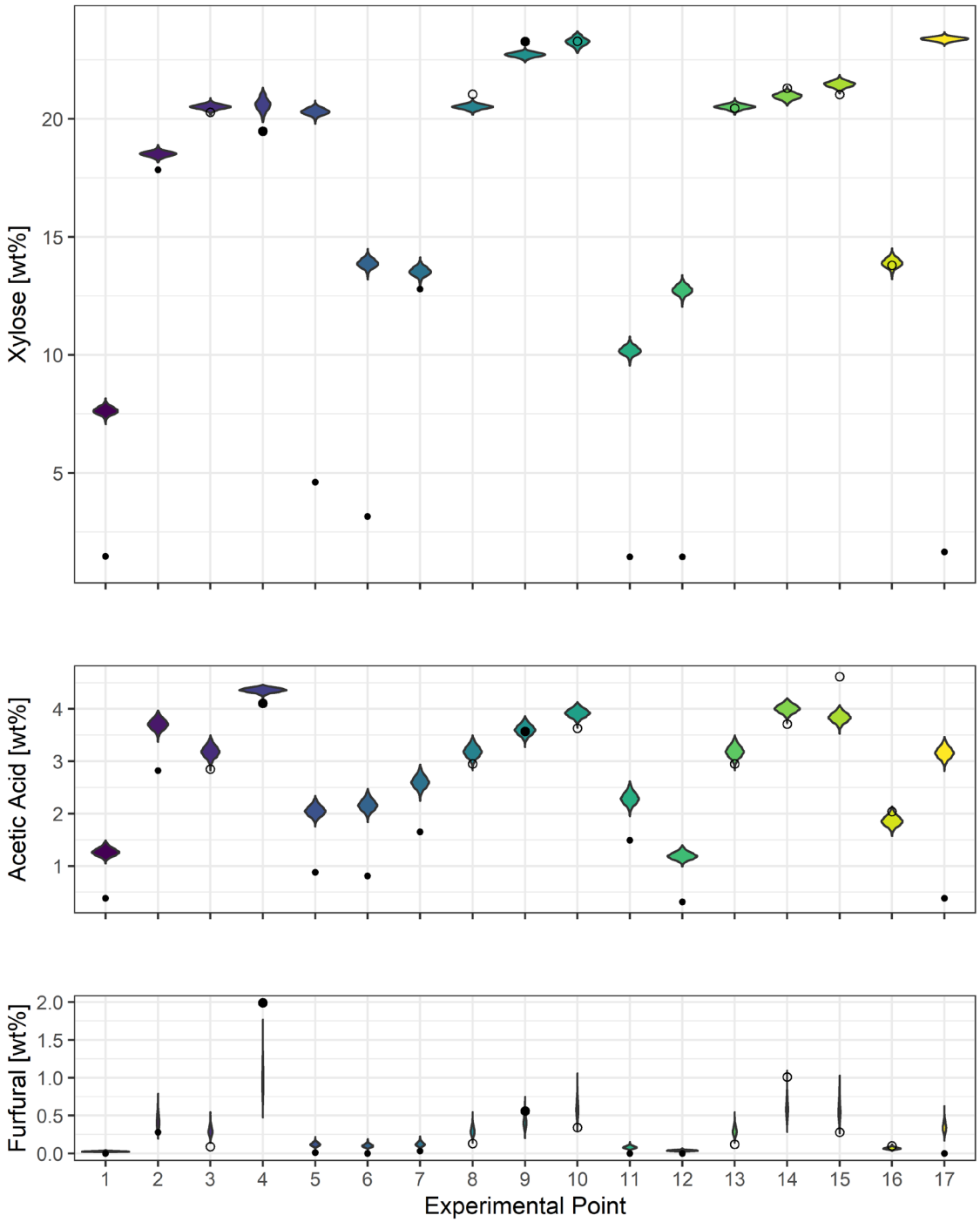

Figure 6: Violin plots for the results of the Monte Carlo-based uncertainty analysis for N=1000 samples for all 17 operational conditions(• not estimated, ○ training data set, ●testing data set) for the concentrations of xylose, acetic acid, and furfural.

For the uncertainty in the xylose prediction, it becomes apparent that the uncertainty in the prediction is generally very low and that the model predicts closely the experimental results from both the training and the testing data set. For data points that were excluded from the estimation, however, the model overpredicts these values significantly. As already mentioned in the introduction of section 5.5, this is expected and reflects the fact that based

on this expected behavior, only a subset of all available data points was chosen for the estimation as described in subsection A.3.4.

Similar behavior for the uncertainty in the acetic acid prediction is displayed, however at a smaller scale as the concentrations are generally lower: the data points which were not used for estimation are slightly overpredicted, and all but one data point in the training and testing data set are predicted within the range of the model uncertainty. For point 15, the model underpredicts the acetic acid concentration. The used acid concentration for this data point is $C_{ac} = 2.0\ wt\%$, which possibly explains a higher release of acetyls from the polymeric structure than predicted by the model as the reaction might be more than proportionally dependent on the acid concentration.

The uncertainty in the furfural prediction also happens at a generally low level; in contrast to the xylose and acetic acid predictions, the model does mostly not overpredict the values for data points, even for points that have not been estimated. Solely for point 15, the model underpredicts the furfural concentration. The operational conditions for this data point with $T = 186$ °C, $t = 26\ min$ and $C_{ac} = 2.0\ wt\%$ are all close to the upper bound of their range. Hence, these conditions are rather severe for the pretreatment, potentially leading to increased xylose degradation and a high furfural concentration. In conclusion, the mechanistic model shows low uncertainty in predicting all considered component concentrations.

A.3.6.2 Sensitivity Analysis

Similar to the uncertainty analyses, the sensitivity analyses were performed for all mechanistic model outputs for all three operational parameters. The considered ranges for the variables are $T = [173,195]$ °C, $t = [18,30]\ min$ and $C_{ac} = [0.5,2.0]\ wt$ %. All sensitivity analyses were performed with $N = 8192$ Sobol samples. The scripts for the sensitivity analysis are provided through a GitHub repository [33]. The resulting first-order and total sensitivity indices for all three operational variables for all output concentrations are presented in the heatmap in Figure 7

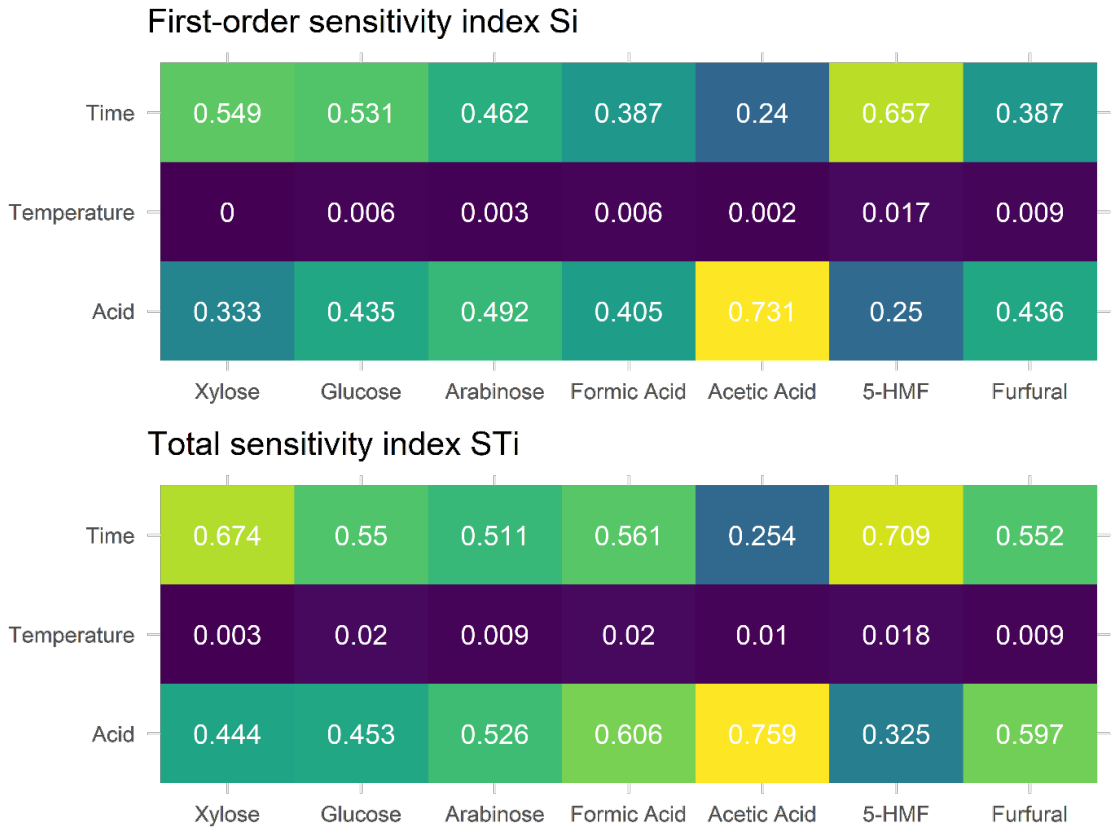

Figure 7: Heat map for the values of the sensitivity analysis of all operational parameters for all model outputs of the mechanistic model.

Reviewing the results for the first-order sensitivity indices in detail, a tendency for moderate sensitivities in the time and acid concentration becomes apparent. Equally, a generally low sensitivity for the temperature is seen. Reciprocally, the results for the total sensitivity indices illustrate that only the influence of the temperature in the given interval would be negligible. Referring back to the uncertainty analysis results in subsection 5.6.1, the conclusions regarding the underprediction of acetic acid for high acid concentrations agree with a high first-order sensitivity index for the acid concentration on the output of acetic acid. Similarly, the furfural output is susceptible to the acid concentration due to a higher release of xylose monomers; hence the higher the acid concentration, the higher the furfural output. Combined with a high temperature and time, this can yield very high furfural outputs, whereas a maximal acid concentration with moderate temperature and time yields moderate furfural outputs. The importance of all three parameters in different absolute orders is also seen in other studies [35,54].

A.3.7 Optimization of Operational Conditions

As the last part of the presented study, the operational conditions of the dilute acid pretreatment are optimized with all developed models. The scripts for all optimization setups are provided through a GitHub repository [33]. The considered objective is the output concentration of xylose, which is to be maximized. The objective function, as stated in (A-9), is the respective model evaluation function. The considered bounds for the input variables as stated in (A-9) are $T = [173,195]$ °C, $t = [18,30]\ min$ and $C_{ac} = [0.5,2.0]\ wt.$ However a further constraint is included that the optimal point can only lie within the

design space of the CCD for which the design of experiments is valid. This is expressed with the following equation:

$$\left|\frac{T \cdot °C^{-1} - 173}{13}\right| + \left|\frac{t \cdot min^{-1} - 18}{8}\right| + \left|\frac{C_{ac} - 1.25}{0.45}\right| \leq \alpha, \qquad \text{(A-10)}$$

With the value of $\alpha = 1.69$ for the CCD of the dilute acid pretreatment experiments.

A.3.7.1 Results

The optimized variables, as well as the value for the objective function, are listed in Table 10:

Table 10: Results for the optimized operational variables and the value for the objective function of all models.

Variable	Unit	RSM	GPR	Mechanistic Model
T	°C	182.4	195	191.6
t	min	26.2	18	18
C_{ac}	$wt\%$	1.25	1.25	1.13
$\boldsymbol{C_{Xyl}}$	$\boldsymbol{wt\%}$	**25.72**	**23.23**	**23.47**

For the RSM, the point does not lie in the vicinity of already measured data points but rather within the design space. A single experiment for these conditions is performed to verify these conditions, as described in section A.2.4. The resulting concentration of xylose is $C_{Xyl} = 20.97\ wt\%$ and hence a deviation from the predicted value by 18.5 %. The RSM model hence overpredicted the maximum concentration significantly.
For the GPR, although the predicted concentration is notably lower than the one predicted with the RSM model, the experimental value for this point is $C_{Xyl} = 23.29\ wt\%$, which is an excellent prediction. It is to mention that the point is part of the training data set of the GPR, namely point 10.
The predicted xylose concentration is even higher for the mechanistic model than the one predicted by the GPR model, while the difference within both the selected conditions and the output concentrations is marginal. This point is not verified experimentally due to the close vicinity of this point to point 10 of the experimental design and the minimal difference in the output concentrations within the uncertainty range of the experimental values. Nonetheless, given the sensitivity analysis results in subsection A.3.6.2, a slightly reduced acid concentration compared to point 10 is conclusive as the most sensitive operational condition for the xylose concentration is the acid concentration.

A.3.7.2 Comparison

After calibrating all models to experimental data and calibrating them respectively, conclusions regarding their comparability and performance shall be drawn in this section. Figure 8 shows contour plots for all three introduced models, with the output being the xylose concentration for respectively two of the three operational conditions, while the third operational condition is fixed at its mean value.

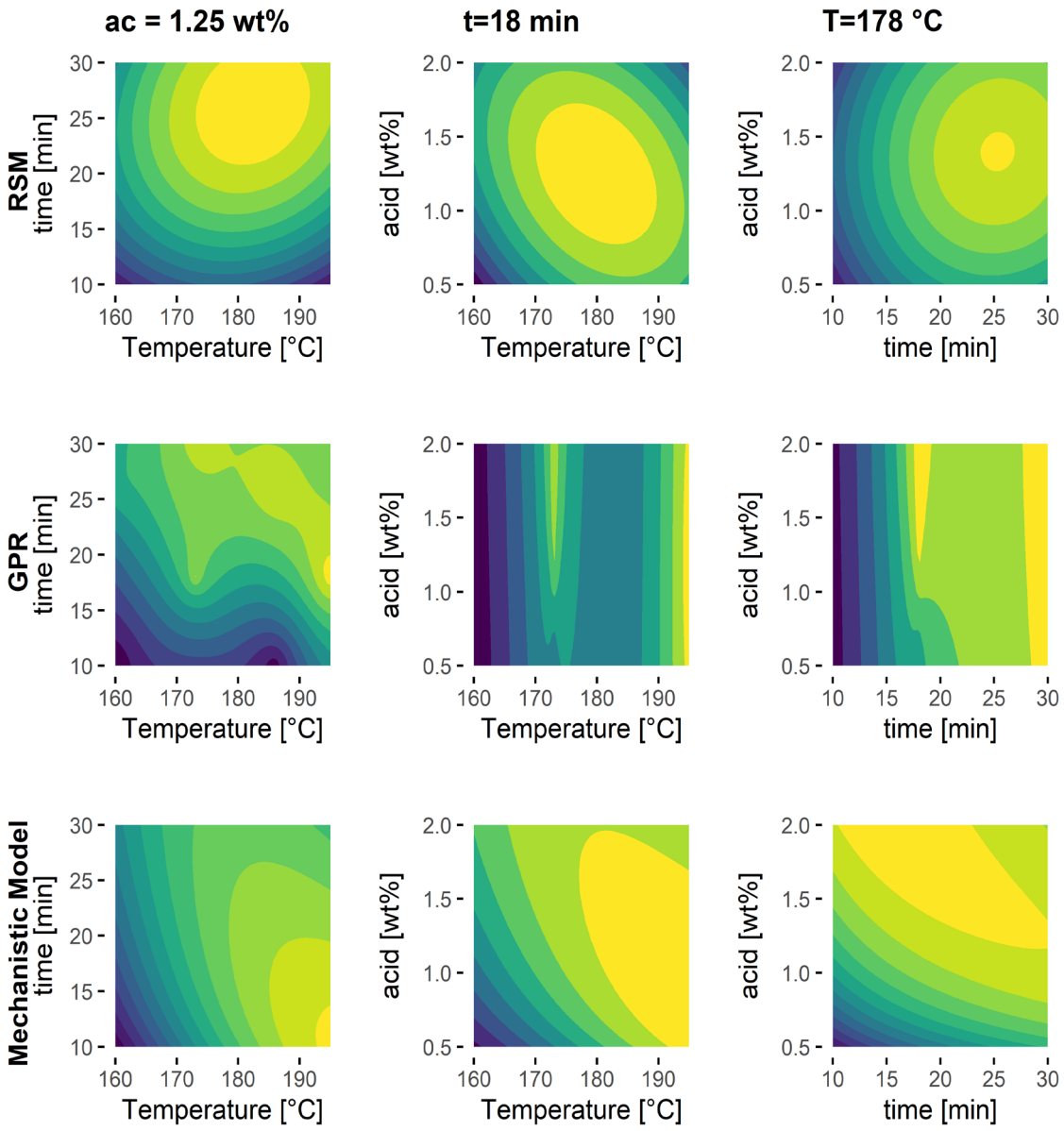

Figure 8: Contour plots of the xylose concentration for two out of three operational conditions, with the third fixed at its mean value.

For the RSM model, the contour of the xylose output is characterized by the underlying second-order polynomial, illustrating the global optimum for a maximum xylose output as a single point within the design space. For the GPR model, the contour plot illustrates the general behavior, which accounts for GPR models: The prediction in close vicinity of known points is excellent, whereas the uncertainty in prediction significantly increases for areas that do not lie in the vicinity to known data points. This also relates to the fact that the validation metrics – particularly the RMSE – for the testing data set were the highest of all models. This leads to the fact that several local maxima can be seen in the contours for the training data points. Lastly, the contour plots of the mechanistic model resemble the kinetic behavior of the first-order rate expressions, which predicts the global optimum towards higher temperatures, average times, and lower acid concentrations. Lower xylose concentrations are, however, overpredicted.

As a concluding remark, it should be stated that all models involve a certain level of heuristics and approximations since the exact mechanisms of all occurring reactions in biomass pretreatment with the inherently high number of components is infeasible to describe. Hence, using a particular model always indicates a trade-off between different characteristics, and there is no best one-fits-all candidate modeling approach. To this end, it is noted that data sets resulting from typical designs of experiments usually are analyzed by RSM, where our study highlights shortcomings of this methodology [31,57]. In point of fact, to determine optimal conditions, predicting unseen data with a small error is of vital importance. Therefore, a diligent approach is the benchmark of different modeling approaches against each other regarding the prediction of unseen or test data sets. Since the mechanistic model is validated best in this regard, it is used for the optimization under uncertainty and verifies the assumptions taken regarding the feedstock.

A.3.8 Optimization of Operational Conditions under Uncertainty

The purpose of the optimization step is to use the model to rigorously explore the design space for any conditions that can improve process yield. Therefore, the model is optimized under uncertainty to assess the effects of both the model uncertainties and the assumption regarding the feedstock. The uncertainty in the parameters is expressed through the covariance matrix and used for sampling. For the assumption in the feedstock composition, it is assumed that the original amount of xylan in hemicellulose of 80% is uniformly distributed in an interval of [70,80] %, and the arabinan and acetyl amounts are uniformly distributed in an interval of [10,15] % as a worst-case scenario. The solver is performing $N = 250$ Monte Carlo simulations for each solver iteration and initially performs 25 simulations and subsequently performs 75 iterations. The chosen infill criterion is *mcFEI,* and the corresponding infill solver is a particle swarm solver. The scripts for the optimization under uncertainty are provided through a GitHub repository [33].
Additionally, two further constraints for the concentration of acetic acid and furfural are introduced in order to assure that the predicted optimum does not yield higher concentrations of both components compared to the optimum, which was found without considering the uncertainty in the feedstock:

$$C_{Aac} \leq 3.705\ wt\%, \tag{A-11}$$

$$C_{Fur} \leq 0.5289\ wt\%. \tag{A-12}$$

When hedging against the uncertainty with the mean value of the predictions of the objective function and constraints, optimizing the conditions predicts a concentration of xylose of $C_{Xyl} = 22.68\ wt\%$ at a temperature of $T = 191.2$°C, $t = 18\ min$ and $C_{acid} = 1.12\ wt\%$. The resulting objective is only marginally smaller than the one predicted without considering uncertainty while maintaining the same operational conditions. Hence, it can be concluded that even with varying feedstock composition, the operational conditions remain optimal, and on average, the output yield is reduced by about 3.3 %.

When hedging against the uncertainty with the mean value plus one standard deviation of the predictions of objective function and constraints, optimizing the conditions predicts a concentration of xylose of $C_{Xyl} = 20.88\ wt\%$ at a temperature of $T = 182.6$°C, $t = 18\ min$ and $C_{acid} = 0.84\ wt\%$. The yield for the output concentration is now reduced by around 10.8 % compared to the deterministic optimum. The chosen operational conditions vary

significantly: the selected temperature is reduced, and the dilute acid concentration. This is because the formation of particularly acetic acid and partially furfural are influenced by the dilute acid concentration and the temperature in the case of acetic acid, as seen in the first-order sensitivity indices presented in Figure 7. In order to not infringe the constraints with the hedge against uncertainty, milder conditions are selected.

In conclusion, the optimization under uncertainty showed both that the assumption taken about the feedstock composition and the intrinsic uncertainty of the model due to the parameter estimation do not significantly influence the prediction of the objective to produce a maximum amount of xylose. A study with corn stover as feedstock confirms that feedstock variability – apart from trivial effects – does not significantly influence the monomer yields [58]. Moreover, the optimization under uncertainty with a variability of the feedstock composition only moderately influences the xylose concentration, as the chosen operational conditions are milder.

A.4 Conclusions

In the scope of this paper, a model for biomass pretreatment was developed based on a case study and validated for optimizing the operational conditions of the pretreatment to assist the conceptual process design of a biorefinery. Wheat straw was chosen as feedstock, and pretreatment experiments with both hydrothermal and dilute acid pretreatment were performed. The analysis of the experiments showed the clear favorability of the dilute acid pretreatment. A response surface model, a Gaussian process regression model, and a mechanistic model based on mass and energy balances and first-order reaction kinetics were fitted to the data. All models show considerably good validation metrics. However, the optimal conditions found with the three models differ, which is relatable to each model's properties. A comparative analysis shows that the predictions of the mechanistic model are most reliable for the underlying case study, which is thus recommended for use in conceptual process design. Furthermore, optimizing the operational conditions under different uncertainty scenarios shows that the model assumptions do not affect the predicted performance and consequently confirm the robustness of the model for use in process design applications under uncertainty.

Furthermore, more fundamental knowledge about the reaction mechanisms and the factors influencing them will significantly improve the predictive quality of knowledge-driven pretreatment models and broaden their validation range. Moreover, this allows for evaluating different pretreatment methods and feedstocks with the same knowledge-driven model, facilitating the conceptual process design even further and helping to assess viable biorefinery concepts for more sustainable production of chemicals.

Acknowledgments

The authors would like to acknowledge funding by the Novo Nordisk Foundation under grant no. NNF17SA0031362 and NNF20SA0066233. Furthermore, they would like to acknowledge Torbjørn Ølshøj Jensen for providing the wheat straw.

References

[1] United Nations, Transforming our world: The 2030 agenda for sustainable development, 2015. https://sdgs.un.org/sites/default/files/publications/21252030 Agenda for Sustainable Development web.pdf (accessed March 1, 2021).

[2] F. Cherubini, The biorefinery concept: Using biomass instead of oil for producing energy and chemicals, Energy Convers. Manag. 51 (2010) 1412–1421. https://doi.org/10.1016/j.enconman.2010.01.015.

[3] A.J.J. Straathof, S.A. Wahl, K.R. Benjamin, R. Takors, N. Wierckx, H.J. Noorman, Grand Research Challenges for Sustainable Industrial Biotechnology, Trends Biotechnol. 37 (2019) 1042–1050. https://doi.org/10.1016/j.tibtech.2019.04.002.

[4] T. Chaturvedi, A.I. Torres, G. Stephanopoulos, M.H. Thomsen, J.E. Schmidt, Developing process designs for biorefineries-definitions, categories, and unit operations, Energies. 13 (2020) 1493. https://doi.org/10.3390/en13061493.

[5] L.A.M. van der Wielen, S.I. Mussatto, J. van Breugel, Bioprocess intensification: Cases that (don't) work, N. Biotechnol. 61 (2021) 108–115. https://doi.org/10.1016/j.nbt.2020.11.007.

[6] H.J. Noorman, J.J. Heijnen, Biochemical engineering's grand adventure, Chem. Eng. Sci. 170 (2017) 677–693. https://doi.org/10.1016/j.ces.2016.12.065.

[7] S.S. Hassan, G.A. Williams, A.K. Jaiswal, Lignocellulosic Biorefineries in Europe: Current State and Prospects, Trends Biotechnol. 37 (2019) 231–234. https://doi.org/10.1016/j.tibtech.2018.07.002.

[8] S.S. Hassan, G.A. Williams, A.K. Jaiswal, Moving towards the second generation of lignocellulosic biorefineries in the EU: Drivers, challenges, and opportunities, Renew. Sustain. Energy Rev. 101 (2019) 590–599. https://doi.org/10.1016/j.rser.2018.11.041.

[9] M. Koutinas, A. Kiparissides, E.N. Pistikopoulos, A. Mantalaris, Bioprocess systems engineering: Transferring traditional process engineering principles to industrial biotechnology, Comput. Struct. Biotechnol. J. 3 (2012) e201210022. https://doi.org/10.5936/csbj.201210022.

[10] J. Moncada B, V. Aristizábal M, C.A. Cardona A, Design strategies for sustainable biorefineries, Biochem. Eng. J. 116 (2016) 122–134. https://doi.org/10.1016/j.bej.2016.06.009.

[11] K. Ulonska, A. König, M. Klatt, A. Mitsos, J. Viell, Optimization of Multiproduct Biorefinery Processes under Consideration of Biomass Supply Chain Management and Market Developments, Ind. Eng. Chem. Res. 57 (2018) 6980–6991. https://doi.org/10.1021/acs.iecr.8b00245.

[12] K. Darkwah, B.L. Knutson, J.R. Seay, A Perspective on Challenges and Prospects for Applying Process Systems Engineering Tools to Fermentation-Based Biorefineries, ACS Sustain. Chem. Eng. 6 (2018) 2829–2844. https://doi.org/10.1021/acssuschemeng.7b03762.

[13] M. Galbe, O. Wallberg, Pretreatment for biorefineries: A review of common methods for efficient utilisation of lignocellulosic materials, Biotechnol. Biofuels. 12 (2019) 1–26. https://doi.org/10.1186/s13068-019-1634-1.

[14] S.I. Mussatto, G.M. Dragone, Biomass Pretreatment, Biorefineries, and Potential Products for a Bioeconomy Development, in: Biomass Fractionation Technol. a Lignocellul. Feed. Based Biorefinery, Elsevier Inc., 2016: pp. 1–22. https://doi.org/10.1016/B978-0-12-802323-5.00001-3.

[15] R. Saini, C.S. Osorio-Gonzalez, K. Hegde, S.K. Brar, S. Magdouli, P. Vezina, A. Avalos-Ramirez, Lignocellulosic Biomass-Based Biorefinery: an Insight into Commercialization and Economic Standout, Curr. Sustain. Energy Reports. 7 (2020) 122–136. https://doi.org/10.1007/s40518-020-00157-1.

[16] J.K. Saini, R. Gupta, Hemansi, A. Verma, P. Gaur, R. Saini, R. Shukla, R.C. Kuhad, Integrated Lignocellulosic Biorefinery for Sustainable Bio-Based Economy, in: Sustain. Approaches Biofuels Prod. Technol., Springer, Cham, 2019: pp. 25–46. https://doi.org/10.1007/978-3-319-94797-6_2.

[17] S. Choi, C.W. Song, J.H. Shin, S.Y. Lee, Biorefineries for the production of top building block chemicals and their derivatives, Metab. Eng. 28 (2015) 223–239. https://doi.org/10.1016/j.ymben.2014.12.007.

[18] T.L. De Albuquerque, I.J. Da Silva, G.R. De MacEdo, M.V.P. Rocha, Biotechnological production of xylitol from lignocellulosic wastes: A review, Process Biochem. 49 (2014) 1779–1789. https://doi.org/10.1016/j.procbio.2014.07.010.

[19] L. Venkateswar Rao, J.K. Goli, J. Gentela, S. Koti, Bioconversion of lignocellulosic biomass to xylitol: An overview, Bioresour. Technol. 213 (2016) 299–310. https://doi.org/10.1016/j.biortech.2016.04.092.

[20] D. Dasgupta, S. Bandhu, D.K. Adhikari, D. Ghosh, Challenges and prospects of xylitol production with whole cell bio-catalysis: A review, Microbiol. Res. 197 (2017) 9–21. https://doi.org/10.1016/j.micres.2016.12.012.

[21] A. Felipe Hernández-Pérez, P.V. de Arruda, L. Sene, S.S. da Silva, A. Kumar Chandel, M. das G. de Almeida Felipe, Xylitol bioproduction: state-of-the-art, industrial paradigm shift, and opportunities for integrated biorefineries, Crit. Rev. Biotechnol. 39 (2019) 924–943. https://doi.org/10.1080/07388551.2019.1640658.

[22] H. Rasmussen, H.R. Sørensen, A.S. Meyer, Formation of degradation compounds from lignocellulosic biomass in the biorefinery: Sugar reaction mechanisms, Carbohydr. Res. 385 (2014) 45–57. https://doi.org/10.1016/j.carres.2013.08.029.

[23] Y. Zheng, J. Shi, M. Tu, Y.-S. Cheng, Principles and Development of Lignocellulosic Biomass Pretreatment for Biofuels, 2 (2017) 1–68. https://doi.org/10.1016/bs.aibe.2017.03.001.

[24] Y. Delgado Arcaño, O.D. Valmaña García, D. Mandelli, W.A. Carvalho, L.A. Magalhães Pontes, Xylitol: A review on the progress and challenges of its production by chemical route, Catal. Today. 344 (2020) 2–14. https://doi.org/10.1016/j.cattod.2018.07.060.

[25] E. Palmqvist, B. Hahn-Hägerdal, Fermentation of lignocellulosic hydrolysates. II: Inhibitors and mechanisms of inhibition, Bioresour. Technol. 74 (2000) 25–33. https://doi.org/10.1016/S0960-8524(99)00161-3.

[26] A. Sluiter, B. Hames, R.O. Ruiz, C. Scarlata, J. Sluiter, D. Templeton, D. of Energy, Determination of Structural Carbohydrates and Lignin in Biomass. Laboratory Analytical Procedure (LAP), 2004.

http://www.nrel.gov/biomass/analytical_procedures.html (accessed March 1, 2021).

[27] A. A. Sluiter, R. Ruiz, C. Scarlata, J. Sluiter, D. Templeton, Determination of Extractives in Biomass: Laboratory Analytical Procedure (LAP); Issue Date 7/17/2005 - 42619.pdf, Tech. Rep. NREL/TP-510-42619. (2008) 1–9. http://www.nrel.gov/biomass/analytical_procedures.html (accessed March 1, 2021).

[28] A. Sluiter, B. Hames, R. Ruiz, C. Scarlata, J. Sluiter, D. Templeton, Determination of Sugars , Byproducts , and Degradation Products in Liquid Fraction Process Samples Laboratory Analytical Procedure (LAP) Issue Date : 12 / 08 / 2006 Determination of Sugars , Byproducts , and Degradation Products in Liquid Fraction Proce, 2008. www.nrel.gov (accessed March 1, 2021).

[29] K. McBride, K. Sundmacher, Overview of Surrogate Modeling in Chemical Process Engineering, Chemie-Ingenieur-Technik. 91 (2019) 228–239. https://doi.org/10.1002/cite.201800091.

[30] G.E.P. Box, K.B. Wilson, On the Experimental Attainment of Optimum Conditions, J. R. Stat. Soc. Ser. B. 13 (1951) 1–38. https://doi.org/10.1111/j.2517-6161.1951.tb00067.x.

[31] D. Bas, I.H. Boyaci, Modeling and optimization i: Usability of response surface methodology, J. Food Eng. 78 (2007) 836–845. https://doi.org/10.1016/j.jfoodeng.2005.11.024.

[32] R. V. Lenth, Response-surface methods in R, using RSM, J. Stat. Softw. 32 (2009) 1–17. https://doi.org/10.18637/jss.v032.i07.

[33] N. Vollmer, Pretreatment Model, GitHub Repos. (2021). https://github.com/NikolausVollmer/Pretreatment-Model.

[34] C.E. Rasmussen, Gaussian Processes in machine learning, in: O. Bousquet, U. von Luxburg, G. Rätsch (Eds.), Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer Verlag, Berlin, Heidelberg, 2004: pp. 63–71. https://doi.org/10.1007/978-3-540-28650-9_4.

[35] J. Shen, C.E. Wyman, A novel mechanism and kinetic model to explain enhanced xylose yields from dilute sulfuric acid compared to hydrothermal pretreatment of corn stover, Bioresour. Technol. 102 (2011) 9111–9120. https://doi.org/10.1016/j.biortech.2011.04.001.

[36] S.E. Jacobsen, C.E. Wyman, Cellulose and Hemicellulose Hydrolysis Models for Application to Current and Novel Pretreatment Processes, in: Twenty-First Symp. Biotechnol. Fuels Chem., Humana Press, Totowa, NJ, 2000: pp. 81–96. https://doi.org/10.1007/978-1-4612-1392-5_6.

[37] R.M. Prunescu, M. Blanke, J.G. Jakobsen, G. Sin, Dynamic modeling and validation of a biomass hydrothermal pretreatment process-a demonstration scale study, AIChE J. 61 (2015) 4235–4250. https://doi.org/10.1002/aic.14954.

[38] N. Mosier, C. Wyman, B. Dale, R. Elander, Y.Y. Lee, M. Holtzapple, M. Ladisch, Features of promising technologies for pretreatment of lignocellulosic biomass, Bioresour. Technol. 96 (2005) 673–686. https://doi.org/10.1016/j.biortech.2004.06.025.

[39] S. Gürkan, G. Krist, Data Handling and parameter estimation, IWA Publishing, London, 2016.

[40] G. Sin, K. V. Gernaey, A.E. Lantz, Good modeling practice for PAT applications: Propagation of input uncertainty and sensitivity analysis, Biotechnol. Prog. 25 (2009) 1043–1053. https://doi.org/10.1002/btpr.166.

[41] R. Brun, M. Kühni, H. Siegrist, W. Gujer, P. Reichert, Practical identifiability of ASM2d parameters - Systematic selection and tuning of parameter subsets, Water Res. 36 (2002) 4113–4127. https://doi.org/10.1016/S0043-1354(02)00104-5.

[42] R. Al, C.R. Behera, A. Zubov, K. V. Gernaey, G. Sin, Meta-modeling based efficient global sensitivity analysis for wastewater treatment plants – An application to the BSM2 model, Comput. Chem. Eng. 127 (2019) 233–246. https://doi.org/10.1016/j.compchemeng.2019.05.015.

[43] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global Sensitivity Analysis. The Primer, John Wiley & Sons, Ltd, Chichester, UK, 2008. https://doi.org/10.1002/9780470725184.

[44] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, Comput. Phys. Commun. 181 (2010) 259–270. https://doi.org/10.1016/j.cpc.2009.09.018.

[45] N.R. Council, Assessing the reliability of complex models: Mathematical and statistical foundations of verification, validation, and uncertainty quantification, National Academies Press, 2012. https://doi.org/10.17226/13395.

[46] R. Al, C.R. Behera, K. V. Gernaey, G. Sin, Stochastic simulation-based superstructure optimization framework for process synthesis and design under uncertainty, Comput. Chem. Eng. 143 (2020) 107118. https://doi.org/10.1016/j.compchemeng.2020.107118.

[47] S. V. Vassilev, D. Baxter, L.K. Andersen, C.G. Vassileva, T.J. Morgan, An overview of the organic and inorganic phase composition of biomass, Fuel. 94 (2012) 1–33. https://doi.org/10.1016/j.fuel.2011.09.030.

[48] C. Wyman, Handbook on Bioethanol, Routledge, Boca Raton, 2018. https://doi.org/10.1201/9780203752456.

[49] J.A. Pérez, A. González, J.M. Oliva, I. Ballesteros, P. Manzanares, Effect of process variables on liquid hot water pretreatment of wheat straw for bioconversion to fuel-ethanol in a batch reactor, J. Chem. Technol. Biotechnol. 82 (2007) 929–938. https://doi.org/10.1002/jctb.1765.

[50] J.A. Pérez, I. Ballesteros, M. Ballesteros, F. Sáez, M.J. Negro, P. Manzanares, Optimizing Liquid Hot Water pretreatment conditions to enhance sugar recovery from wheat straw for fuel-ethanol production, Fuel. 87 (2008) 3640–3647. https://doi.org/10.1016/j.fuel.2008.06.009.

[51] M.A. Kärcher, Y. Iqbal, I. Lewandowski, T. Senn, Comparing the performance of Miscanthus x giganteus and wheat straw biomass in sulfuric acid based pretreatment, Bioresour. Technol. 180 (2015) 360–364. https://doi.org/10.1016/j.biortech.2014.12.107.

[52] A.M.J. Kootstra, H.H. Beeftink, E.L. Scott, J.P.M. Sanders, Comparison of dilute mineral and organic acid pretreatment for enzymatic hydrolysis of wheat straw, Biochem. Eng. J. 46 (2009) 126–131. https://doi.org/10.1016/j.bej.2009.04.020.

[53] E. Guerra-Rodríguez, O.M. Portilla-Rivera, L. Jarquín-Enríquez, J.A. Ramírez, M. Vázquez, Acid hydrolysis of wheat straw: A kinetic study, Biomass and Bioenergy. 36 (2012) 346–355. https://doi.org/10.1016/j.biombioe.2011.11.005.

[54] J.E. Morinelly, J.R. Jensen, M. Browne, T.B. Co, D.R. Shonnard, Kinetic characterization of xylose monomer and oligomer concentrations during dilute acid pretreatment of lignocellulosic biomass from forests and switchgrass, Ind. Eng. Chem. Res. 48 (2009) 9877–9884. https://doi.org/10.1021/ie900793p.

[55] X. Liu, M. Lu, N. Ai, F. Yu, J. Ji, Kinetic model analysis of dilute sulfuric acid-catalyzed hemicellulose hydrolysis in sweet sorghum bagasse for xylose production, Ind. Crops Prod. 38 (2012) 81–86. https://doi.org/10.1016/j.indcrop.2012.01.013.

[56] Q. Jin, H. Zhang, L. Yan, L. Qu, H. Huang, Kinetic characterization for hemicellulose hydrolysis of corn stover in a dilute acid cycle spray flow-through reactor at moderate conditions, Biomass and Bioenergy. 35 (2011) 4158–4164. https://doi.org/10.1016/j.biombioe.2011.06.050.

[57] D. Bas, I.H. Boyaci, Modeling and optimization II: Comparison of estimation capabilities of response surface methodology with artificial neural networks in a biochemical reaction, J. Food Eng. 78 (2007) 846–854. https://doi.org/10.1016/j.jfoodeng.2005.11.025.

[58] N.D. Weiss, J.D. Farmer, D.J. Schell, Impact of corn stover composition on hemicellulose conversion during dilute acid pretreatment and enzymatic cellulose digestibility of the pretreated solids, Bioresour. Technol. 101 (2010) 674–678. https://doi.org/10.1016/j.biortech.2009.08.082.

Paper

B

Synergistic optimization framework for the process synthesis and design of biorefineries

Nikolaus I. Vollmer, Resul Al, Krist V. Gernaey, Gürkan Sin,

Frontiers in Chemical Science and Engineering,

Vol. 16, Issue 2, pp. 251-273,

https://doi.org/10.1007/s11705-021-2071-9

Abstract

The conceptual process design of novel bioprocesses in biorefinery setups is an important task, which remains yet challenging due to several limitations. We propose a novel framework incorporating superstructure optimization and simulation-based optimization synergistically. In this context, several approaches for superstructure optimization based on different surrogate models can be deployed. By means of a case study, the framework is introduced and validated, and the different superstructure optimization approaches are benchmarked. The results indicate that even though surrogate-based optimization approaches alleviate the underlying computational issues, there remains a potential issue regarding their validation. The development of appropriate surrogate models, comprising the selection of surrogate type, sampling type, and size for training and cross-validation sets, are essential factors. Regarding this aspect, satisfactory validation metrics do not ensure a successful outcome from its embedded use in an optimization problem. Furthermore, the framework's synergistic effects by sequentially performing superstructure optimization to determine candidate process topologies and simulation-based optimization to consolidate the process design under uncertainty offer an alternative and promising approach. These findings invite for a critical assessment of surrogate-based optimization approaches and point out the necessity of benchmarking to ensure consistency and quality of optimized solutions.

B.1 Introduction

The global necessity of novel process solutions to meet the demands of a growing society and find adequate solutions, matching the increasing need for sustainable process solutions, is high [1]. In general, biotechnological process solutions are deemed to be conceptually more sustainable [2]. Recent developments in synthetic biology allow for producing a vast palette of biofuels, chemicals, foods, and even pharmaceutical ingredients in cell factories. The global biofoundry initiative aims at accelerating the acquisition of knowledge, the integration of data, and the development of new cell factories [3,4].

On the other side, the implementation of biorefinery concepts that incorporate these chemicals' sustainable production from sustainable feedstocks, as e.g. lignocellulosic biomass or other residues or dedicated crop plants is dramatically low. Up to the current date, a total number of less than 100 active lignocellulosic biorefineries are operated around the world [5]. This is mainly due to these biorefinery concepts' critical economic robustness, as the operative margins for most products are narrow and chemical processes are competitive [6]. Despite several ideas of designing these biorefineries and improving their economic robustness by mass and heat integration or producing several products simultaneously in a so-called multi-product biorefinery, a conceptual design strategy that significantly promotes the implementation of these biorefineries is still an active area of research [7–10]. Current research directions span from the integration and expansion of established approaches [11,12], over the use of novel approaches and models [13,14], up to the inclusion of further economic, environmental, and sustainability factors in particular [15–19], as well as the accommodation of specificities for fermentation-based processes [20]. The vast majority of these studies primarily follow three approaches for conceptual process design that are briefly introduced in the following paragraphs.

Classically, process design for biorefineries is performed by so-called hierarchical decomposition and involves domain knowledge from different expert areas to yield a consolidated solution [21]. However, this methodology has several shortcomings as it usually involves several iterations, which results in a long idea-to-process time, and does not necessarily yield an optimal solution for the process design, especially considering that biotechnological processes suffer from issues with the scale-up from laboratory scale to production scale [3,22].

A computational and more conceptual approach to process design is superstructure optimization (SSO) [23]. Out of all possible process options provided in a superstructure, the optimal configuration is chosen through mathematical optimization. In comparison to the previous alternative, the resulting process design is globally optimal. However, the methodology is initially limited by the number of alternatives that are included in the superstructure. Furthermore, the resulting optimization problem can become very large even though a considerable number of possible solutions can be infeasible a priori. Also, the use of high-fidelity models is limited in this approach as this can be computationally challenging to solve. Lastly, the incorporation of uncertainty increases the complexity of formulating the optimization problem through stochastic or robust optimization [24]. SSO has been successfully performed for several chemical production processes; however, the

design of biotechnological processes comprises several new challenges and uncertainties which are difficult to address [25].

A second, more recent approach to process design is simulation-based optimization (SBO). The methodology builds upon the evaluation of process simulations where high-fidelity models are used [26]. Furthermore, uncertainties can be easily included in a Monte Carlo-based approach [27]. Several proposed frameworks utilize machine learning surrogate models as e.g. stochastic kriging (SK), coupled with a Bayesian optimization approach to iteratively improve the objective function results, leading to an optimal process design [27,28]. Despite this approach being able to incorporate complex physiological models of cell factories and processes, it is heavily limited by available computational power. The number of simulations can get very high for several alternative process configurations, which constrains this approach severely to a small design space.

After elucidating the challenges within the process design of novel bioprocesses in biorefinery setups, it becomes evident that novel solutions are required, which incorporate knowledge from different fields and approach the problem in an interdisciplinary manner. Hence, we propose a novel synergistic optimization framework for the conceptual process design of biorefineries, based on a hybrid approach integrating surrogate-based SSO with SBO. It harnesses both the power of the SSO for process synthesis and the potential of SBO for detailed design optimization. The framework itself capitalizes biotechnological knowledge to guide decisions in a bottom-up strategy for both SSO and SBO, which finally yields a consolidated process design.

The remainder of the paper is structured as follows: The used surrogate models for the superstructure are introduced in section B.2.1 and the methods of performing SSO in section B.2.2, and the SBO in section B.2.3. The detailed components of the proposed framework and the applied workflow are introduced in section B.2.5. For the application of the framework, a case study is introduced in section B.3. The corresponding high-fidelity models for the specific case study are introduced in appendix B.5.1, including the corresponding methodology for uncertainty and sensitivity analysis in appendix B.5.2. The evaluation of the framework's superstructure optimization step is presented in section B.3.2 and the SBO step in section B.3.3. Conclusions and an outlook on future research work are given in section B.4.

B.2 Theoretical Background

B.2.1 Surrogate Models

Surrogate models describe various types of models, which mimic the behavior of original first-principle models. The primary reason for developing and using a surrogate model, described by McBride and Sundmacher (2019), is the reduction of computational costs for model evaluations, which is an inherent part of the SSO, as explained in section B.2.2. This comes at the cost of a certain error in the surrogate model's predictions due to imperfect resemblance. The applied technique for creating a surrogate model highly depends on the computational constraints regarding the original and the surrogate model, as well as the type of application of the surrogate model [29]. The simplest model types are linear, piecewise linear, or polynomial functions, which are fitted to the original data [30,31].

However, the majority of the used surrogate models are based on machine learning techniques, among others radial basis functions, Gaussian process regressors, and at the higher end artificial neural networks and even deep neural networks [29,32]. The following subsections describe first the development procedure for these model types and subsequently four different surrogate model types which are used in this framework, one based on algebraic equations, one piecewise linear model, and two machine learning models.

The development procedure for surrogate models – independent of the model type is similar and commonly starts with sampling the model input space by e.g. Latin Hypercube (LHS) or Sobol Sampling. To ensure a space-filling sampling for a sufficient interpolative quality of the model, the choice of an adequate sampling technique is crucial [33]. Subsequently, the surrogate model is fitted to a part of the data, dependent on the chosen validation strategy.

A standard method, which is also applied in this framework, is called k-fold cross-validation: the sampled dataset is divided into k equally sized subsets. In a routine, each subset is once used to validate the surrogate model, and the remaining $k-1$ subsets are used for the fitting of the model. The total validation error is calculated by the average error of the k validation folds [34].

B.2.1.1 Automated Learning of Algebraic Models

The Automated Learning for Algebraic Models (ALAMO) toolbox was developed to create algebraic models for applications like SSO [35,36]. As described in Wilson's and Sahinidis' (2017) work, the ALAMO toolbox first fits simple, algebraic models, consisting of several nonlinear terms and their linear combinations to the points in the input space. Then, by employing derivative-free optimization, the best suitable combination of terms is determined through error maximization sampling. Further constraints on the model outputs can be imposed, and the possibility to perform adaptive sampling to increase the number of points in the input space at specific locations [35]. Especially for process design purposes, the ALAMO toolbox has been applied widely as a surrogate modeling technique with promising results [22,37,38].

B.2.1.2 Delaunay Triangulation Regression

The second surrogate model is based on the concept of regression by a piecewise linear functional relationship. In the case of a one-dimensional input space, the piecewise linear functional relationship is a set of composed line segments. For a general definition of this piecewise linear functional relationship, be $P \subset \mathbb{R}^{n \times r}$ a set of points in the $n \times r$-dimensional space and a subset of real numbers. All elements $p \in P$ are considered vectors in the Euclidean vector space over $\mathbb{R}^{n \times r}$ and consist of two elements $p = \{p^x, p^z\}$. All elements p^x and p^z are respectively elements of the sets $P_x \subset P$ and $P_z \subset P$, both subsets of P. Furthermore, be $X \subset \mathbb{R}^n$ the input space and $Z \subset \mathbb{R}^r$ the output space of our functional relationship with all elements $x \in X$ and $z \in Z$ and $P \subset X \times Z$.

Defining the input space X as n-manifold with boundary, the triangulation $\mathcal{T}_X$ of X is a homogeneous simplicial k-complex Σ, being homeomorphic to $X$and $k = n$, referring to that the S consists only of n-simplices. An n-simplex σ is the special form of an n-polytope and

consists of exactly $n+1$ vertices v, so $\{v_1, \dots, v_{n+1}\} \subseteq \sigma \in \Sigma$. For illustrative purposes, for $n = 2$ the 2-simplex is a triangle. Hence the n-simplex is the equivalent to a triangle in any dimension n. At this point, we define the set of vertices V and every vertex $v \in V$ as the element p^x of the point p. Furthermore, an important property of an n-simplex σ is it being an affine and hence convex space, which allows any point x^* inside the simplex to be described as a linear combination of the vertices with the coefficients $\alpha_i < 1$:

$$x^* = \sum_{i=1}^{n+1} \alpha_i \cdot v_i = \sum_{i=1}^{n+1} \alpha_i \cdot p_i^x \quad \forall\, v_i, p_i^x \in \sigma. \tag{B.1}$$

Furthermore, the functional relationship of the surrogate model f is defined as:

$$f: \begin{cases} \mathbb{R}^n \to \mathbb{R}^r \\ x \mapsto z \end{cases}. \tag{B.2}$$

The set of real numbers is denoted as $\mathbb{R}$. Given this, the conclusion for the surrogate model is that any point z^* can also be described by a linear combination of elements p^z of the point p by equation (B.1), which is equivalent to the concept of regression. In conclusion, the triangulation of the input space $\mathcal{T}_X$ with the functional relationship to the output space $Z = f(\mathcal{T}_X)$ represents the surrogate model based on the described regression concept.
Reconsidering the one-dimensional case, a 1-simplex, also called edge, is nothing else than a line segment; the one 1-simplicial complex corresponds to the set of composed line segments, where each point on any line segment can be described as a linear combination of the two vertices or endpoints of the line segment it lies on. Lastly, the functional relationship describes the functional value of the point of the line segment by a linear combination of the vertices' functional values.

A prominent type of triangulation is the so-called Delaunay triangulation, which imposes the criterion that the circumcircle – or its higher-dimensional equivalent, its n-hypersphere – of the vertices of the simplex cannot contain another vertex [39]. There exist published algorithms to create a Delaunay triangulation regression (DTR) for a given set of points [40,41].

B.2.1.3 Gaussian Process Regression

The third surrogate model is based on the concept of regression by a stochastic process – the eponymous Gaussian process regression (GPR) – and particular kernel functions to determine parameters of the following functional relationship:

$$f: \begin{cases} \mathbb{R}^n \to \mathbb{R} \\ x \mapsto z \end{cases}, \tag{B.3}$$

with the expression, as explained by Al et al. (2019):

$$z = \mu_{GP}(x) + \sigma_{GP}^2 \cdot \mathcal{F}(x, \omega), \quad \mu_{GP}(x) = \rho \cdot \beta(x) \tag{B.4}$$

Here, $\mu_{GP}(x)$ denotes the mean value of the Gaussian process and σ_{GP}^2 its variance. For the mean value $\mu_{GP}(x)$, ρ are estimated parameters from the input data, as well as the variance σ_{GP}^2 and $\beta(x)$ relate to a group of basis functions. Furthermore, $\mathcal{F}(x, \omega)$ describes a zero

mean unit variance stochastic process. The correlation or kernel functions ω are used to correlate any point in the input space x^* with existing points x [42]. There are various available kernel functions; hence, the reader is referred to the book by Rasmussen [43]. A couple of remarks on GPR as a surrogate model shall be made: Firstly, it is very well capable of displaying highly nonfunctional relationships, and the amount of data necessary to obtain a good surrogate model is relatively low and favorably low-dimensional [29]. GPR surrogate models have seen a wide range of applications in machine learning, particularly also in different process design tasks [44–46].

B.2.1.4 Artificial Neural Network

The last type of surrogate model is the class of artificial neural networks (ANN). ANNs are used in a plethora of applications, especially complex machine learning tasks as image and voice recognition, natural language processing, and artificial intelligence. Al et al. (2019) give the following description: in general, neural networks consist of at least three layers, one input, one hidden, and one output layer. This case is considered as a shallow neural network and, in particular, a multi-layer perceptron. Each layer contains a certain number of nodes that relate their inputs to their outputs with a so-called transfer or activation function of a specific type. When several hidden layers are added to the network, it is referred to as a deep neural network [42]. Due to their flexible architecture and possibilities in terms of composing and learning a network, neural networks are an ideal candidate for the use as a surrogate model in process design tasks and have seen a widespread application in the area of process systems engineering [29,47–49].

B.2.2 Superstructure Optimization

SSO as a method for process synthesis and design is a computational method based on mathematical optimization. Following Chen & Grossmann's elaboration, SSO involves three steps, namely (1) the definition of a set of process alternatives in a superstructure representation, (2) the formulation of the corresponding optimization problem, and (3) solving the optimization problem with an adequate solver in order to obtain the optimal process design. The primary limitations remain firstly with the initial definition of the design space, implying that only solutions can be found that are initially considered. Secondly, these limitations are also fueled by the assumptions taken for the models and the capabilities of the employed solver for the optimization problem [23]. Moreover, a capital restriction is the inability to account for uncertainty in deterministic optimization approaches such as SSO [24].

The superstructure itself can be postulated in various ways. One of the most common superstructure formulations is a State-Task-Network, in which each unit operation model forms a task in the network, the flows are described as states, and both are connected via nodes that correspond to mixers or splitters. This composition highly resembles an actual process flowsheet, with the main difference being that the nodes represent binary decisions on whether a process path exists [50]. For the interested reader, a description of other existing superstructure formulations can be found in a recent review by Mencarelli et al. [51].

As described by Chen and Grossman (2017), the resulting optimization problem is a mixed-integer program (MIP) of the form:

$$\text{MIP:} \begin{cases} \min z = f(x,y) \\ s.t.\ \ h(x,y) = 0 \\ \quad g(x,y) \leq 0 \\ x \in X, y \in [0,1] \end{cases}, \tag{B.5}$$

with a defined objective function $z = f(x,y)$, referring to a certain metric, e.g. product purity, a key performance indicator or sustainability measures, and subject to equality and inequality constraints $h(x,y)$ and $g(x,y)$, describing physical constraints, system and equipment specifications and their limits, as well as other process constraints e.g. product purity. The continuous variables $x \in X$ denote process variables as states, mass and energy flows, and design parameters, all within a specific input space $X \subseteq \mathbb{R}$. The binary variables y denote the mentioned decisions on the existence of equipment or process paths. Depending on the underlying physical system, both the objective function f and the constraints can be nonlinear; hence the optimization problem is a mixed-integer nonlinear program (MINLP).

The objective function evaluation can be theoretically calculated based on the unit operation models and additional equations. However, the complexity of the formulation and the computational cost of evaluating high-fidelity models commonly makes the solution of the optimization problem by currently available solvers complicated or even intractable [49,51]. Consequently, there have been various research efforts in surmounting these hurdles in SSO, e.g. linearization of the objective function, surrogate model-assisted SSO, or decomposition algorithms [50,51]. Especially the capacity of surrogate model-assisted SSO alleviating the computational burden has been exploited in various studies [49,52,53]. This allows for an elegant solution to integrate complex high-fidelity models from different platforms indirectly via their surrogates into a simple superstructure formulation, compared to extensive equation-based approaches as, e.g., generalized disjunctive programming [49,51].

In the following section, four different superstructure formulations with respectively one of the introduced surrogate models in section B.2.1 are introduced. The benchmark of all four approaches is part of sections B.3.2.2 and B.3.2.3.

B.2.2.1 Surrogate-Assisted Mixed-Integer Nonlinear Program

The first option under investigation are ALAMO surrogate models in an MINLP and no reformulation of the actual optimization problem. The resulting algebraic model equations from the ALAMO surrogate models can be introduced as the function for the objective function $f(x,y)$ and the equality constraints $h(x,y)$ in the MINLP formulation described in equation (B.5), the problem can be solved with an adequate solver. As the ALAMO surrogates are fitted from flowsheet simulations, a binary variable y_m for all flowsheet options $m \in I_M$ is introduced, as well as a SOS1 constraint, indicating that the optimal solution can only lie in one flowsheet:

$$\sum_{m \in I_M} y_m = 1. \tag{B.6}$$

B.2.2.2 Surrogate-Assisted Mixed-Integer Linear Program

The second option for the SSO suggested here involves developing a superstructure based on DTR surrogate models and the reformulation of the underlying MINLP into a mixed-integer linear program (MILP).

For the set of possible process design configurations M with an index set $I_M = \{1,2,\dots,M\}$, a certain number of flowsheet simulations are performed for each configuration $m \in M$. A DTR surrogate model is then fitted for each configuration $m \in M$ from a set of sample points $P_{m,x}$ over the input space $X \subset \mathbb{R}^n$ and a set of simulation points $P_{m,z}$ over the output space $Z \subset \mathbb{R}^r$. Every sample point p_m^x and respectively p_m^z equally is a set of the type $p_m^x = \{p_{m,1}^x, p_{m,2}^x, \dots, p_{m,n}^x\}$ and $p^z = \{p_{m,1}^z, p_{m,2}^z, \dots, p_{m,r}^z\}$ with every $p_{m,d}^j$ being a scalar value. Besides, two index sets $I_X = \{1,2,\dots,n\} \subset \mathbb{N}$ and $I_Z = \{1,2,\dots,r\} \subset \mathbb{N}$ for the dimension of the input and output space are defined and an index set I_P for set P_x and P_z as $|P_x| = |P_z|$. The created triangulation $\mathcal{T}_X$ yields a set of simplices S, fulfilling the definitions stated in section B.2.1.2. Be $I_S = \{1,2,\dots,|S|\} \subset \mathbb{N}$ the index set of S, then each simplex $\sigma_{m,s}$ again is a specific set of points $p_{m,d,i}^x$ with $i \in I_{SP} = \{i \mid p_{m,d,i}^x \in \sigma_{m,s} \ \forall\, d \in I_X\} \subseteq I_P$ that are all part of the simplex. Lastly, a binary variable $y_{m,s}$ is introduced, which relates to whether a point x is part of a simplex $\sigma \in S$, in the flowsheet configuration $m \in M$.

The MILP is postulated as follows: The objective function is the minimum value of one element of the output variable z:

$$\min z_k, \tag{B.7}$$

with $z_k \in z$ and $k \leq r$. Furthermore, for any input variable x on the boundary ∂X of X and output variable z under the given functional relationship (B.1) for $\mathcal{T}_X$, we can rewrite equation (B.2) in the following way:

$$y_{m,s} \cdot x_d = \sum_{i \in I_{SP}} y_{m,s} \cdot \alpha_{m,s,i} \cdot p_{m,d,i}^x \quad \forall\, d \in I_X, s \in I_S, m \in I_M, \tag{B.8}$$

$$y_{m,s} \cdot z_d = \sum_{i \in I_{SP}} y_{m,s} \cdot \alpha_{m,s,i} \cdot p_{m,d,i}^z \quad \forall\, d \in I_Z, s \in I_S, m \in I_M. \tag{B.9}$$

Furthermore, due to the convex property of the simplex, the following equation considering the linear combination coefficients $\alpha_{m,s,i}$, which are treated as variables by the optimizer, must be fulfilled:

$$\sum_{i \in I_{SP}} \alpha_{m,s,i} = 1 \quad \forall\, s \in I_S, m \in I_M. \tag{B.10}$$

Since a MILP only allows the formulation of linear equations, a Big-M notation for equation (B.8) and (B.9) is introduced with the three variables ρ, φ and ψ replacing the three products $y_{m,s} \cdot x_d$, $y_{m,s} \cdot \alpha_{m,s,i}$ and $y_{m,s} \cdot z_d$. The following are the four equations for the product $y_{m,s,d} \cdot x_d$:

$$\left.\begin{aligned} \rho_{m,s,d} &\leq M_d^x \cdot y_{m,s} && \text{(B.11)}\\ \rho_{m,s,d} &\leq x_d && \text{(B.12)}\\ \rho_{m,s,d} &\geq x_d - M_d^x \cdot (1 - y_{m,s}) && \text{(B.13)}\\ \rho_{m,s,d} &\geq 0 && \text{(B.14)} \end{aligned}\right\} \forall s \in I_S, d \in I_X, m \in I_M.$$

In these equations, M denotes so-called Big-M parameters; their values are set to the upper bound of the respective variable they represent.

Similarly, also the Big-M notation for the second product $y_{m,s} \cdot \alpha_{m,s,i}$is written:

$$\left.\begin{aligned} \varphi_{m,s,i} &\leq y_{m,s} && \text{(B.15)}\\ \varphi_{m,s,i} &\leq \alpha_{m,s,i} && \text{(B.16)}\\ \varphi_{m,s,i} &\geq \alpha_{m,s,i} - \left(1 - y_{m,s}\right) && \text{(B.17)}\\ \varphi_{m,s,i} &\geq 0 && \text{(B.18)} \end{aligned}\right\} i \in I_{SP}, \forall s \in I_S, m \in I_M.$$

In this case, the Big-M parameter is equal to 1 and hence left out of the equations. Lastly, also the Big-M notation for the third product $y_{m,s} \cdot z_d$ is listed here:

$$\left.\begin{aligned} \psi_{m,s,d} &\leq M_d^y \cdot y_{m,s,d} && \text{(B.19)}\\ \psi_{m,s,d} &\leq z_d && \text{(B.20)}\\ \psi_{m,s,d} &\geq z_d - M_d^y \cdot (1 - y_{m,s,d}) && \text{(B.21)}\\ \psi_{m,s,d} &\geq 0 && \text{(B.22)} \end{aligned}\right\} \forall s \in I_S, d \in I_Z, m \in I_M.$$

Equations (B.11) - (B.22) now allow to express equations (B.8) and (B.9) without the multiplication of variables in the following way:

$$\rho_{m,s,d} = \sum_{i \in I_{SP}} \varphi_{m,s,d,i} \cdot p_{m,d,i}^x \quad \forall\, d \in I_X, s \in I_S, m \in I_M, \tag{B.23}$$

$$\psi_{m,s,d} = \sum_{i \in I_{SP}} \varphi_{m,s,d,i} \cdot p_{m,d,i}^z \quad \forall\, d \in I_Z, s \in I_S, m \in I_M. \tag{B.24}$$

To satisfy equations (B.23) and (B.24), there exists exactly one simplex in one flowsheet configuration in which the point x is located, and all other simplices in this flowsheet and all other flowsheets do not contain the point. Hence, a SOS1 constraint is added to the postulation in order to express this:

$$\sum_{m \in I_M} \sum_{s \in I_S} y_{m,s} = 1. \tag{B.25}$$

With equations (B.7) and (B.10), the MILP is now well defined and can be solved with a suitable optimization algorithm. Alternatively, further constraints on other elements of the output variables can be imposed, referring to other process metrics, e.g. economic indicators. For different types of triangulations, this MILP has been postulated in similar ways by Misener et al. [54,55].

B.2.2.3 Surrogate-Assisted Series of Nonlinear Programs

The third option for the SSO suggested by the authors involves developing a superstructure based on GPR or ANN surrogate models and the reformulation of the underlying MINLP into a series of nonlinear programs (NLP).

To remove the integer variables y from the problem as stated in (B.5), which are introduced for different flowsheet options, the series of NLP has to involve an NLP for each of these flowsheet options. For the set of possible process design configurations M with an index set $I_M = \{1,2,\dots,M\}$, a certain number of flowsheet simulations are performed for each configuration $m \in M$. For each flowsheet configuration $m \in M$ a GPR or ANN surrogate is then fitted from a set of sample points $P_{m,x}$ over the input space $X \subset \mathbb{R}^n$ and a set of simulation points $P_{m,z}^{OBJ}$ for the metric to be evaluated in the objective function over the output space $Z^{OBJ} \subset \mathbb{R}$. This procedure is performed for each of the metrics j to be considered as constraint with a set of simulation points $P_{m,z,j}^{CON}$ over the respective output space $Z_j^{CON} \subset \mathbb{R}$. The boundaries for the input variables are set to the bounds of the input space ∂X and a set of initial points in the input space $X_0 \in X$ are declared, in order to assure global optimality by performing a multi-start optimization with an amount of $|X_0| = s$. The objective function is equally formulated as for the MILP with equation (B.7). The series of NLP can now be solved sequentially for each NLP performing being solved s times with the different initial points $x_0 \in X_0$ by using a suitable optimization algorithm.

B.2.3 Simulation-Based Optimization

All the presented SSO approaches in section B.2.2 are deterministic approaches to optimization, relating to the found optimum, not incorporating any kind of uncertainty. In contrast to that, the process design of a novel chemical or biochemical process inherently represents a significant uncertainty by itself [56]. Due to the development of computational power and the ever-more growing use of simulation software, making the computational tractability of complex systems possible, the concept of SBO or, in particular, stochastic simulation optimization has seen increasing interest over the past years. This concept allows to incorporate stochastic considerations as uncertainty into an optimization formulation and solve them, given sufficient computational capacities [26–28]. In contrast to the prior presented approaches, the system to be optimized does not need to follow a particular mathematical structure, as the systems are commonly treated as black boxes [27].

Drawbacks of SBO are that the ability to find an optimum to a given optimization problem with SBO is heavily constrained by the computational tractability, which can be easily exceeded by excessively high computational costs for simulation evaluations, a high dimensionality of the problem, or the description of multiple objectives and constraints [26]. Furthermore, as information about derivatives in black-box systems is not readily available, the proof of global optimality for the obtained solution by SBO remains challenging [57].

Among the several approaches for performing SBO, a surrogate model-based method with stochastic kriging (SK) surrogate models will be elucidated in this section, as their use is favorable for computationally expensive simulations [44]. The interested reader is referred

to a comprehensive summary of the benefits and drawbacks of several other approaches by Amaran et al. [57]. Besides the mentioned benefits of SBO, a surrogate-based approach gives into the structure of the search space and location of a possible global optimum [58]. SK as surrogate model type is an extended variant to the presented GPR in section B.2.1.3. For the individual differences, the reader is referred to the original contributions [28,59].
Following the framework described by Al et al. (2020), the SK surrogate is described with the following functional relationship for any input point $x \in \mathbb{R}^n$, closely related to equation (B.4):

$$z = \mu + \mathcal{E}(x) + \mathcal{L}(x), \tag{B.26}$$

where μ is now a constant term referring to the mean value of the prediction and $\mathcal{E}(x)$ and $\mathcal{L}(x)$ representing extrinsic and intrinsic uncertainty in the prediction. Extrinsic uncertainty describes the uncertainty of the surrogate model with regards to the high-fidelity model due to an imperfect representation of the input space $X \subset \mathbb{R}^n$ by the set of sampling points P_X. Intrinsic uncertainty represents uncertainty in the original model with regards to the original physical system it represents. The model is fitted to the initial set of sampling points of the input space P_X and simulation points P_Z in the output space $Z \in \mathbb{R}$.
The performed optimization works as an evolutionary program, where an adaptive search with infill optimization under a given infill criterion is performed. With an initially small set of sample points, the infill optimization directs the search towards more promising areas in the search space by adding new sample points to the set P_X. In this iterative procedure, the SK model is updated and hence improved. The iterations are terminated upon reaching a specific criterion, and the simulation results from each step can be investigated, and the optimum can be determined from the set of sample points of the last iteration i^* [27]:

$$\min z = \min P_{i^*,x}. \tag{B.27}$$

Different infill criteria can be used for this task, and the reader is referred to the literature for an overview [27].

B.2.4 Optimization Under Uncertainty

Process design inherently involves various sources of uncertainty. These uncertainties can be accommodated in the optimization; however, this always implies a certain imparity in the calculated objective function, a so-called “price of robustness” [60]. Several factors influence this tradeoff: Grossmann et al. (2016) mention the availability of information on the type of uncertainty, as well as general data on the uncertainty, the way of hedging against the uncertainty, a tremendous computational burden, as well as difficult tractability of the results for this [24].

Established practices in performing optimization under uncertainty in connection with mathematical programming are robust optimization, chance-constraint programming, or stochastic programming, which are investigated for several decades and recently received attention by the use of data-driven modeling techniques as, e.g., deep and reinforcement learning in order to alleviate certain shortcomings of the classical mathematical programming approaches [61]. However, for increased tractability of the optimization problem, simulation-based approaches that use high-fidelity simulation models and a metamodeling approach seem to be a more viable solution [24,62].

Here, Monte Carlo methods, as introduced in section B.5.2, represent a straightforward approach to include the optimization under uncertainty in SBO and allow a simple statistical quantification of the objective and the constraint values. This has been performed successfully for similar process design tasks [27,63]. This procedure is integrated with the presented solver in sectionB.2.3 and will thus be used as such in the scope of this study.

B.2.5 Framework-S3O

As mentioned, the conceptual design of novel bioprocesses in a biorefinery setup is a highly complex challenge. They arise from the difficulty of utilizing sustainable feedstocks as lignocellulosic biomass, which requires additional unit operations in the upstream process, the problematic nature of microorganisms of showing scale-dependent dynamics due to the intricate interactions in the regulation within the transcriptional and metabolic network, and the arduous conception of the downstream process, which has to account for the specifications in the upstream process while being constrained by economic limitations. While all three introduced process design strategies can be utilized, the resulting process design can face several hurdles deriving from a lack of conceptuality, intrinsic limitations in their feasibility, or disproportionate computational burdens. To surmount these hurdles, the proposed framework aims to leverage synergies in applying all proposed strategies in a hybrid manner, where the benefits of each methodology are harnessed to expedite the global task of designing a process conceptually.

It comprises three sequential steps, namely (1) the selection of product sets, substrates, and operations, (2) SSO for determining candidate process topologies, and (3) simulation optimization for consolidating an optimal process design. It is illustrated in Figure 9.

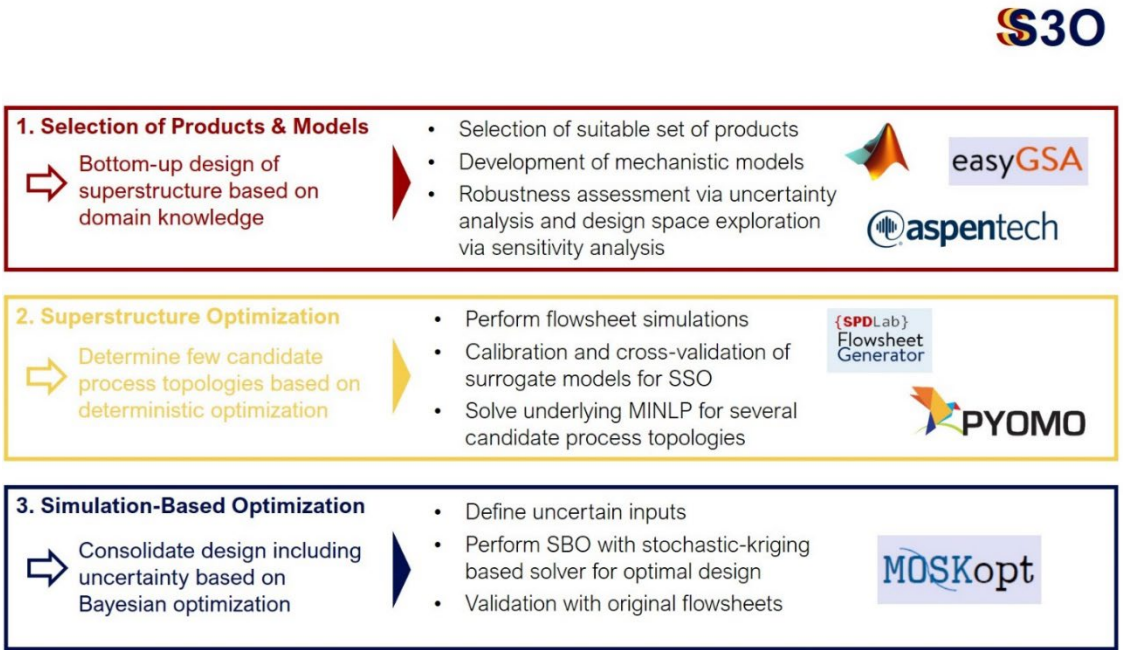

Figure 9: Illustration of the proposed framework S3O with its three stages: selection of products and models (1), SSO (2), and SBO (3), as well as the employed software and toolboxes.

- Selection of Product Sets, Substrates, and Operations

The overarching idea in the framework is to "begin with the end in mind" [64,65]. Applying this principle in a biorefinery context, the first thing to be defined is the set of products. Due to the critical economic viability of biorefinery concepts, it is of utmost importance to choose an appropriate portfolio of products, which exploits the available substrate to the maximally possible extent and potentially maximizes the economic key performance indicators of the biorefinery. Once the set of products is defined, the feedstock for the biorefinery has to be chosen accordingly, where a feedstock candidate should contain reasonable amounts of the respective substrate that is needed to produce the desired set of products.
Based on both a defined set of products and a feedstock, all potential process candidates' necessary unit operations can be defined. As an integral part of this framework, this step is heavily influenced by domain knowledge, which allows for a bottom-up assessment of the possible alternatives reducing the workload in this step immensely. With the defined number of alternatives for unit operations and process routes, a mechanistic model is developed for each unit operation, incorporating the domain knowledge in the form of mathematical equations to describe the underlying physical, chemical, and biological phenomena.

- Superstructure Optimization

With all developed models and the set of all process alternatives, a superstructure is composed. By the prior selection of possible process alternatives through expert knowledge and ultimately subjected to economic considerations, the size of the superstructure is a priori heavily reduced and hence faster to solve. Classically, by performing SSO, the result is a deterministically optimal process design. A second paramount benefit of SSO is the possibility to account for nontrivial design decisions, deriving from the nature of binary decisions in the process design. However, this does not allow directly to incorporate possible uncertainties, which is why the SSO in this framework serves as a selection tool for not only one deterministic optimal process design but rather several candidate process topologies. Furthermore, additional constraints regarding the operability of the process, product quality constraints, or others can be included in this stage.

- Simulation Optimization

By selecting several candidate process topologies through SSO, the number of candidates subjected to simulation optimization is again reduced and hence computationally facilitated. The candidate topologies, which are prone to be the best process design, are subjected to simulation optimization under uncertainty to ultimately yield a consolidated process design. Possible uncertainties that can be included span from uncertainties in the technical domain, e.g., scale-up issues for fermentation reactors or fluctuations in impurities that are separated in the downstream processing, over the operational domain, e.g., varying product and feedstock prices and supply, up to the computational domain, e.g., uncertainties in model parameters, error propagation properties of models and uncertainties in design parameters, as already mentioned in section B.1. Constraints from the prior stage can be added, as well as additional constraints. The entire workflow of the implemented framework is shown in Figure 10.

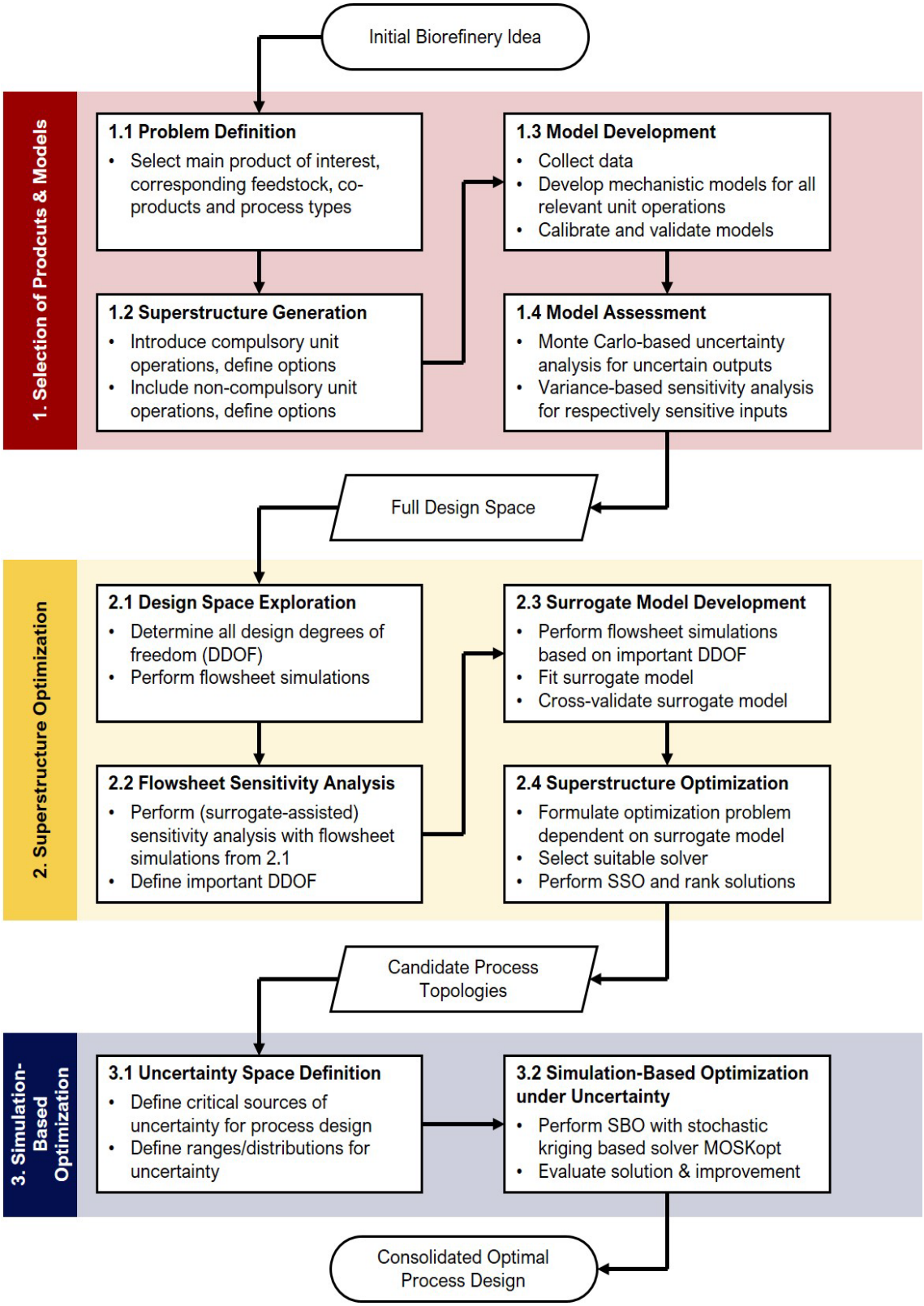

Figure 10: Workflow of the proposed framework S3O indicating the tasks in the three stages and its intermediate and the final results.

For the SSO, it is crucial to assess which option delivers the best results under the given objective of determining a small set of candidate process topologies. The constraints here are that 1) the methodology should be able to pick candidates that turn out to be optimal, 2) delivers results consistently with respect to differently shaped design spaces and flowsheet alternatives, 3) the SSO can be performed within a reasonable amount of time with reasonable computational resources, 4) the results are consistent with each other, e.g. considering different sample sizes for the input space and 5) should yield solutions which are close to the theoretical underlying global optimum of the original flowsheet model. In the following section B.3.2, all proposed options are evaluated and benchmarked regarding these criteria to define which methodology to choose ultimately.

B.3 Case Study & Results

To demonstrate the feasibility and capabilities of the introduced framework, a relevant industrial case study is selected. After a concise description of the case study and its relevance to industry, the application of the three steps of the framework and the results are described and analyzed.

B.3.1 Selection of Product Sets, Substrate & Operations

As the first and main produced chemical in the biorefinery concept in the case study, xylitol is selected. Xylitol is a sugar substitute with a similar taste to sucrose and manifold beneficial health properties, as around 40% fewer calories than sucrose, a very low glycemic index, which makes it very suitable for diabetic nutrition, and anti-cariogenic properties [66]. It has been determined as one of the top 12 chemicals to be produced in a biorefinery concept by the US Department of Energy already in 2004, maintained that status in 2010, and is still attracting researcher's interest in order to facilitate a biotechnological production in cell factories [67–71]. Nevertheless, and despite a market volume estimate of 1 Bio USD and more in 2022, there are only a few companies worldwide, e.g. DuPont Nutrition Biosciences, producing xylitol in a chemical process from wood biomass or corn in a chemical conversion process [70,72].

Proposed process designs in literature are scarce; however, those who approach the idea of a process design for the biotechnological production of xylitol either conclude that the chemical production process is economically more safe and promising or that xylitol can increase the added value of a biorefinery with other principal products, as for example ethanol [73–75]. Hence, the process design of a biotechnological production process for xylitol is a perfectly suitable case study for demonstrating the proposed framework.

Beginning with the end in mind, according to the framework, a suitable set of products is supposed to be chosen in the first step of the workflow. As xylitol is favorably produced from the hemicellulosic fraction of lignocellulosic biomass, the product set can involve a product for the cellulosic fraction of the biomass and one for the lignin fraction. For the cellulosic fraction, succinic acid is considered as a product. It is also one of the top 12 chemicals to be produced in a biorefinery setup and has potential as a platform chemical, which makes it attractive for several industrial branches, which benefits the economic resilience of the biorefinery [67,76]. The lignin fraction can be either used in a combustion process to provide heat for steam generation in order to integrate heat over the different

process trains in the biorefinery or can be further converted in a pyrolysis process into sustainable aviation fuels, for which there is a high demand with higher economic margins than for common biofuels as e.g. bioethanol [77–79]. The downstream process for both the cellulosic and the hemicellulosic process train can involve a classic setup with an evaporation unit and following crystallization units or involve alternative technologies as membrane separation; both approaches are applied commercially and have been investigated for their use in biorefinery downstream processes for both xylitol and succinic acid [53,68,76]. As possible feedstock, wheat straw with a high hemicellulosic content is chosen. A potential superstructure formulation for this base-case process design is illustrated in Figure 11.

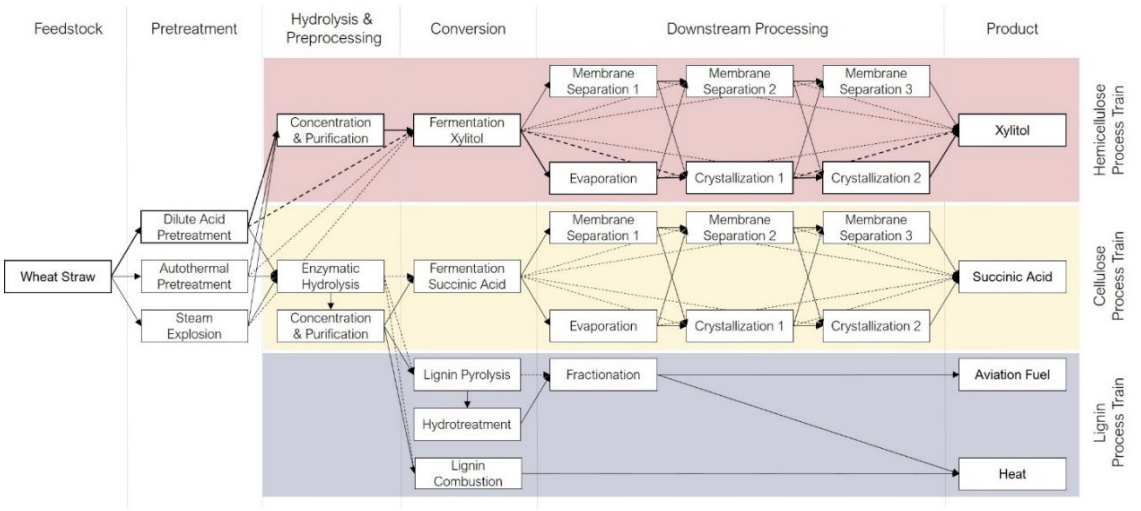

Figure 11: Illustration of the entire initial bottom-up composed superstructure for the base-case process design of the introduced case study with a hemicellulose, a cellulose, and a lignin process train; the reduced superstructure which will serve as the base case in this study is marked in bold.

It becomes evident that by following the concept of beginning with the end in mind and rigorously applying expert knowledge in a bottom-up composition approach for the superstructure, the initial search space is kept comparatively small, which expedites all the following steps in the workflow.

However, in order to analyze the results of the application of this framework more tractable and thus accessible to the reader, we reduce the superstructure of this base case design to a smaller subset of only the xylitol production train with a limited amount of unit operations. Once the framework is validated, the whole superstructure, as in Figure 11, can be processed nonetheless. The reduced superstructure involves six unit operations, namely a biomass pretreatment unit operated as dilute acid pretreatment (PT), an upconcentration unit (UCH), a fermentation unit operated as batch fermentation (FX), an evaporation unit (EX), a first crystallization unit operated as cooling crystallization (CX1) and a second crystallization unit operated as antisolvent crystallization (CX2). Out of these six unit operations, three are compulsory (PT, FX, CX1), and three are optional (UCH, EX, CX2), which results in eight binary decisions or eight different flowsheet alternatives. These are listed in Table 11 together with their configuration ID, which will be used as an identifier throughout the section.

Table 11: Overview of all flowsheet options with their respective configuration ID (cID) and the units composing the flowsheet.

Configuration ID (cID)	Flowsheet
1	PT-UCH-FX-EX-CX1-CX2
2	PT-UCH-FX-EX-CX1
3	PT-UCH-FX-CX1-CX2
4	PT-UCH-FX-CX1
5	PT-FX-EX-CX1-CX2
6	PT-FX-EX-CX1
7	PT-FX-CX1-CX2
8	PT-FX-CX1

With eight flowsheet options as listed in Table 11, the following step 2 of the framework could also be solved by a purely enumeration-based approach instead of formulating an SSO problem. However, to rigorously search through the entire design space for globally optimal solutions and to better account for nontrivial design decisions while accelerating calculation times, the SSO approach is the favorable option. The small problem size is chosen as mentioned to enhance the tractability in the scope of this manuscript.

For all the unit operations, mechanistic models are developed: the pretreatment model (PT) is set up as described in section B.5.1 while considering the same components as Prunescu et al. (2015) [80]. The data for the calibration of the model derives from proprietary experiments. The operational variables considered in the pretreatment model are the pretreatment temperature T_{PT} in °C, the pretreatment time t_{PT} in min and the acid concentration $acid$ in $wt\%$ to the original biomass. The fermentation model (FX) is set up as described in section B.5.1.1, while using data from batch experiments from Tochampa et al. (2005) [81]. The considered operational variables for the fermentation model are the fermentation time t_{FX} in h as well as the inoculum concentration $inoc$ in $g/g\ broth$. The evaporation model (UCH, EX) is set up as described in section B.5.1.2 in the ASPEN Plus process simulation software while using the DIPPR property database and the NRTL equation of state. The considered operational variable is the vapor fraction for either evaporation v_{EX} or upconcentration v_{UCH}. The crystallization model (CX1, CX2) is set up as described in section B.5.1.4 and subsequently validated and calibrated with proprietary experimental data. The considered operational variables are the crystallization time $t_{CX1,2}$ in h, the flowrate of coolant for CX1 $F_{C,CX1}$ in $kg \cdot s^{-1}$ and the flowrate of antisolvent for CX2 $F_{AS,CX2}$ in $kg \cdot s^{-1}$ as well as the cooling temperature $T_{C,CX1}$ in °C for CX1. The considered model outputs for this study are the mass of produced xylitol M_{Xyo} in kg, the concentration of the inhibitory compounds 5-hydroxymethylfurfural and acetic acid in the final process stage C_{5HMF}, C_{Aac} in $g \cdot L^{-1}$ as well as a CO_2 ratio φ in kg/kg, indicating how much CO_2 is produced per kilogram of xylitol by the generation of steam to provide heat in the pretreatment and the evaporation units. The considered uncertainty in the third step of the framework is the composition of the feedstock. All models are implemented in MATLAB. The evaporation model is interfaced with a COM interface to MATLAB. All models are assessed regarding their robustness by means of Monte Carlo-based uncertainty and sensitivity analysis as described in sections B.5.2.1 and B.5.2.2. Hence, the full design space for the case study is set up by factorial selection in SPDlab, while excluding infeasible

options a priori. All model implementations are available through the S3O GitHub repository [82].

B.3.2 Superstructure Optimization

B.3.2.1 Flowsheet Sensitivitiy Analysis

In order to perform the design space exploration, all operational variables for each flowsheet are considered as input variables. With the functionalities of SPDlab, $N = 2000$ samples with LHS sampling for each flowsheet (cID 1-8) are simulated [83]. The flowsheet samples are used as input for the flowsheet sensitivity analysis by using the easyGSA toolbox in MATLAB [42] to determine which of the operational variables are most important regarding the model output and should hence be considered as variables in the optimization problem. The design space exploration results are illustrated in Figure 12 in violin plots for each model output for each relevant flowsheet option. The results of the ANN-assisted flowsheet sensitivity analysis are illustrated in a heatmap in Figure 13. All scripts and implementations regarding the flowsheet sensitivity analysis are available through the S3O GitHub repository [82].

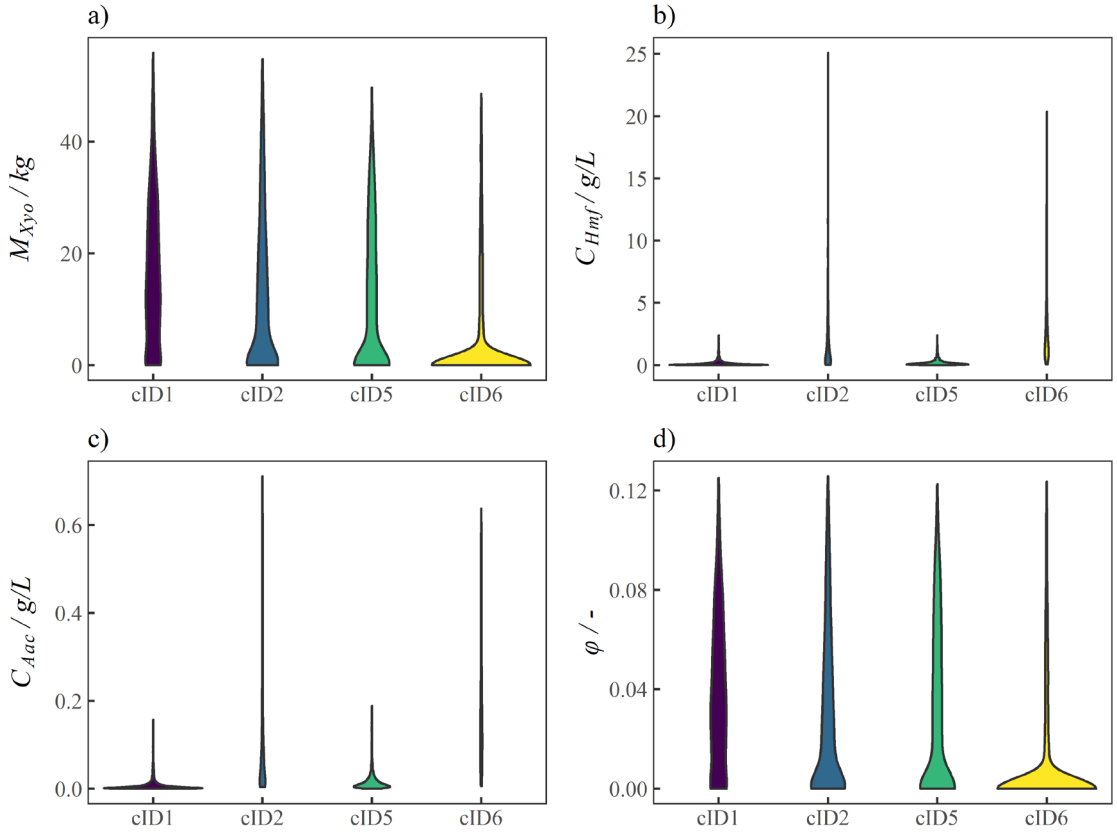

Figure 12: Violin plots of the results from the design space exploration of flowsheets a) cID 1, b) cID 2, c) cID 5, and d) cID 6 with the outputs: the mass of produced xylitol (upper left), the concentration of 5-HMF in the final stage (upper right), the concentration of acetic acid in the final stage (lower left) and the CO2 ratio (lower right).

Figure 13: Heatmap of the total sensitivity indices (S_{Ti}) calculated with the easyGSA toolbox by using ANN surrogates for all flowsheet options and all operational variables.

As a first major result, the flowsheets with the cID 3, 4, 7, and 8 turn out to be infeasible with the given design space, as no set of input variables results in produced xylitol. Furthermore, it becomes evident that there are major differences between the flowsheets' design spaces with cID 1,2,5, and 6. Where cID 6 seems to have an evenly distributed number of points regarding the mass of xylitol produced, for cID 6, most of the sets of input variables turn out to be infeasible with only a small feasible fraction. Regarding the inhibitory compounds, especially cID 2 and 6, as well as cID 5 to minor extents, seem to produce significant amounts of inhibitors. This allows for the conclusion that both inhibitory compounds should be constrained to a certain level in the optimization problem formulation to ensure product quality.

Regarding the flowsheet sensitivity analysis, a more uniform picture emerges. For all four flowsheets, similar operational variables seem vital despite a different order, and also similar variables are insensitive concerning the output. Hence, for cID 1, 5, and 6, we define $T_{PT}, acid, inoc, t_{FX}$ and v_{EX} and for cID 2, we define $acid, inoc, t_{FX}, v_{EX}$ and v_{UCH} as the variables to be considered in the optimization problem. All simulations consider an initial amount of lignocellulosic biomass of $M = 1000\ kg$.

In conclusion, the design space exploration helped reduce the initial search problem by 50%, and the flowsheet sensitivity analysis can be used as a tool for identifying crucial variables for SSO.

B.3.2.2 Surrogate Model Performance Assessment

For the surrogate model development and validation, $N = 500$ and $N = 1000$ flowsheet simulations with LHS sampling are performed to compare their performance with a different number of sampling points. The flowsheets simulations are again performed with the SPDlab functionalities. All four surrogate model types are fitted and cross-validated with these sampling points, as explained in section B.2.1. The ALAMO surrogates are fitted within the ALAMO software. For cross-validation, 20% of the input samples are split off, and

ALAMO is given this fraction as a validation dataset to calculate validation metrics internally. The DTR surrogates are all created with the Delaunay functionality in Python's scipy package, utilizing the QHull algorithm [41]. For their cross-validation, due to the model's inexistent extrapolative capabilities, the boundary points, which form the convex hull of the design space, are added to the sampling points. By random selection, a fraction of 20% of the sample points is split off to calculate the validation metrics for the training and testing dataset. This procedure is repeated ten times to calculate the range of the cross-validation metrics.

The GPR surrogates are fitted in MATLAB while utilizing the Statistics and Machine Learning Toolbox. For each dataset and each output, an internal routine to optimize hyperparameters, e.g. the kernel functions in MATLAB, is used with default settings to fit the GPR model to the data. Subsequently, a 5-fold cross-validation, as described in section B.2.1, is performed to obtain the validation metrics. The ANN surrogate is equally fitted in MATLAB while using the Deep Learning Toolbox. For each dataset and each output, a grid-search algorithm as implemented in the easyGSA toolbox is employed to optimize several parameters, e.g., the number of nodes in the hidden layer to fit the ANN model to the data. Afterward, also a 5-fold cross-validation is performed in order to obtain the validation metrics. All scripts and implementations regarding the surrogate performance analysis are available through the S3O GitHub repository [82].

The calculated validation metrics are the coefficient of determination R^2 and the root mean square error $RMSE$ for the whole dataset, as well as the testing and the training dataset, dependent on the applied validation method. Figure 14 shows the parity plots for both N=500 and N=1000 flowsheet samples for each surrogate model predicting the amount of produced xylitol. For the ALAMO surrogate models, the parity plots show a relatively high variance in the model and the tendency to underpredict higher values for the amount of xylitol. Secondly, the GPR surrogates show almost a perfect fit, apart from cID 6, where the accuracy is slightly lower than for the other three flowsheets. Lastly, the ANN surrogates show equally good fits to the data, however, with higher variances than the GPR models and a decrease in the prediction quality for cID 6. This can be attributed to the low number of feasible sampling points in the design space of cID 6, which is diametric to both surrogate model methodologies as they rely heavily on the amount of provided input data. The DTR surrogates are not shown as parity plots as the simulation points become an inherent part of the model, which makes the coefficient of determination $R^2 = 1$ by definition and a parity plot thus dispensable.

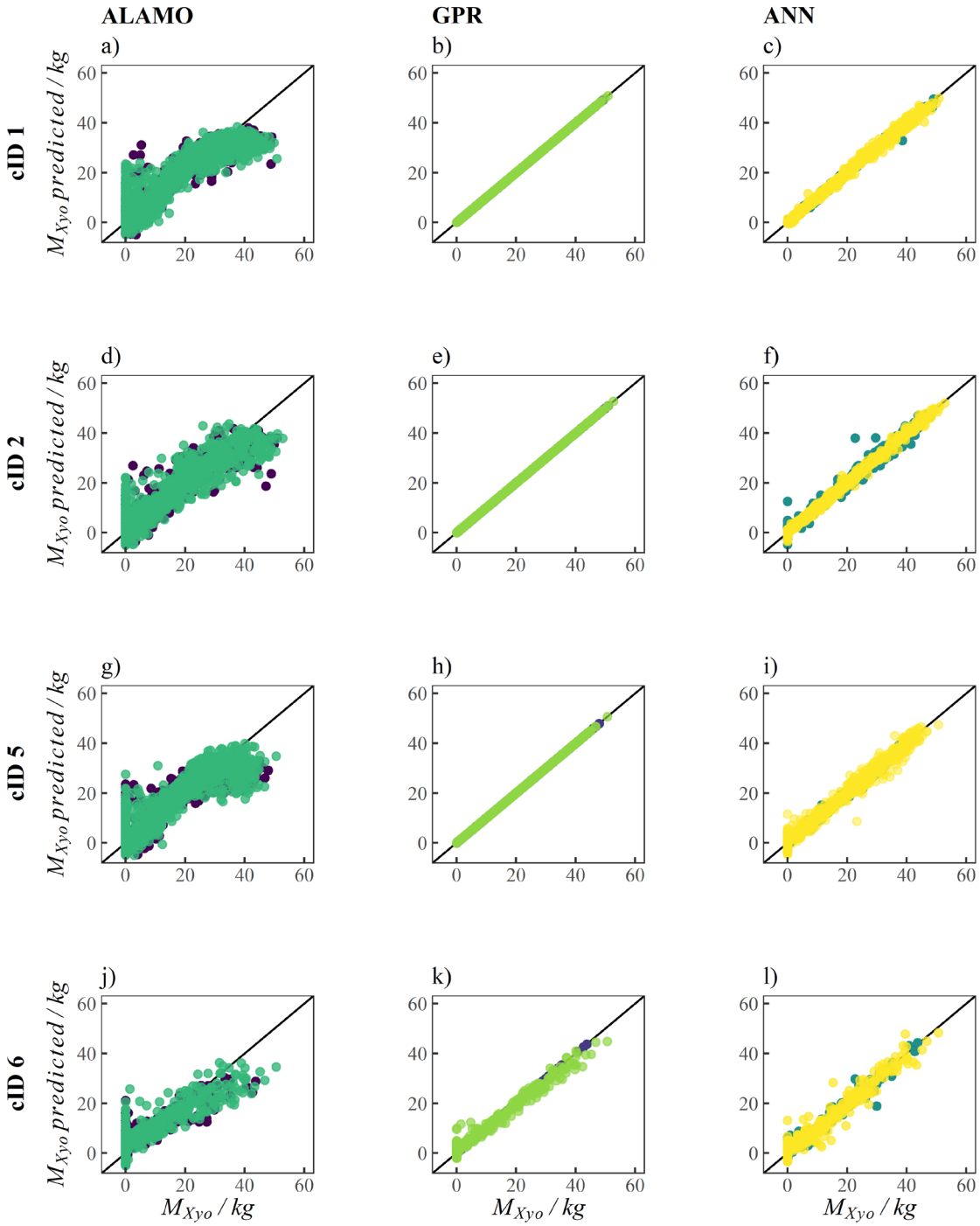

Figure 14: Parity plots of the ALAMO, the GPR, and the ANN surrogate models for all flowsheets (ALAMO: a) cID 1, d) cID2, g) cID5, j) cID6; GPR: b) cID1, e) cID2, h) cID5, k) cID6; ANN: c) cID 1, f) cID2, i) cID 5, l) cID6) indicating the predicted outputs over the simulated outputs for N=500 (dark blue, blue, turquoise) and N=1000 samples (green, bright green, yellow).

For all illustrated models in Figure 14, the cross-validation metrics for the full, the testing, and the training data set for flowsheet cID1 for the output variable being the amount of produced xylitol are listed in Table 12.

Table 12: Cross-validation metrics of all surrogate models for flowsheet option cID1 for both N=500 and N=1000 samples for the output variable being the amount of produced xylitol.

	ALAMO		DTR		GPR		ANN	
	N=500	N=1000	N=500	N=1000	N=500	N=1000	N=500	N=1000
R^2	0.822	0.765	1	1	1	1	0.997	0.994
$RMSE$	5.27	6.29	0	0	0.007	0.017	0.597	0.922
$R2_{train}$	0.817	0.762	1	1	0.997	1	0.997	0.994
$R2_{test}$	0.722	0.724	0.487	0.642	0.933	0.952	0.895	0.956
$RMSE_{train}$	5.35	6.31	0	0	0.423	0.121	0.674	0.944
$RMSE_{test}$	6.54	6.99	8.802	7.677	2.945	2.66	4.002	2.535

The cross-validation metrics for all other flowsheet cIDs and the other output variables are listed in the supporting material. For the ALAMO surrogate models, the impression from the parity plots is confirmed by average values for R^2 between 0.6 and 0.8 and $RMSE$ values, which are significantly high. As described earlier, the R^2 and $RMSE$ values for the DTR surrogate model are immanently 0 or 1 for the full and the training data set, but for the testing dataset, it becomes obvious that the quality of fit for unseen data is insufficient, expressed by R^2 values around 0.6 and $RMSE$ values up to almost 10. Equally, for the GPR and the ANN surrogate, the validation metrics confirm the parity plots' results; both model types show excellent validation scores even for the testing data sets. As a general trend, it is also to denote that for the DTR, the GPR, and the ANN surrogates, the model quality overall increases with $N = 1000$ instead of $N = 500$ samples, which is in agreement with the described properties in section B.2.1. The ALAMO surrogates do not show a consistent improvement of the validation metrics with increasing sample size, which is relatable to the difficulty of fitting a given number of algebraic terms to datasets of increasing size.

Overall, it is to state that both machine learning surrogate models show the best validation metrics and both ALAMO and especially the DTR surrogates reveal insufficient predictive abilities for unseen data points in the test dataset.

B.3.2.3 Superstructure Optimization Results

In order to define the underlying optimization problem to the SSO properly, the following objective function and constraints are introduced:

$$\text{MINLP:} \begin{cases} \max M_{Xyo} = f(x,y) \\ s.t.\ \ C_{Hmf} = g_1(x,y) \leq 0.5\ g \cdot L^{-1} \\ C_{Aac} = g_2(x,y) \leq 0.5\ g \cdot L^{-1} \\ \varphi = g_3(x,y) \geq 0{,}1 \\ x \in X, y \in [0{,}1] \end{cases}, \tag{B.28}$$

The operational variables for the solver to choose are $acid, inoc, t_{FX}, v_{EX}$ for all flowsheet cIDs, T_{PT} for flowsheet cIDs 1, 2, and 6 and v_{UCH} for flowsheet cID 2. In order to solve the underlying optimization to each SSO, suitable solvers have to be chosen. For the ALAMO surrogates and the resulting MINLP, the BARON solver is chosen [84]. All algebraic equations for each flowsheet option are instantiated in PYOMO. However, due to the reduced design space of four flowsheet options, the binary variable is removed, and the four flowsheets are solved separately to accelerate the calculation process. This converts the

MINLP given as in (B.28) into four NLPs; however, the results do not differ as the underlying optimization problem does not change. Hence this simplification is valid. All optimization problems are solved with the default solver settings. If a solution is found, the problem is solved to global optimality; otherwise, the optimization problem is found infeasible.
For the DTR surrogates and the resulting MILP, the Gurobi solver is chosen. The formulation of the problem is equally given with (B.28), with f, g_1, g_2 and g_3 being linear functional relationships as given with (B.2). Equations (B.6) through (B.25) are implemented in PYOMO. Equally for the MILP, the four flowsheets are solved separately, which reduces the dimensionality of the binary variable. However, the resulting simplified MILP yields the same solution; hence the simplification is valid. For $N = 500$ flowsheet samples the resulting optimization problem results in around 1,000,000 continuous and 50,000 integer variables and double the number for $N = 1000$ flowsheet samples. All problems are solved to global optimality.

For both the GPR and the ANN surrogates, the MATLAB solver fmincon with the sequential quadratic programming algorithm is chosen. The formulation of the problem is equally given with (B.28), with f, g_1, g_2 and g_3 being functional relationships as given with (B.3) while excluding the binary variable y. The surrogates are passed to the solver together with a multi-start option, indicating that the optimization problem should be solved for $q = 1000$ times in order to ensure global optimality. For flowsheets where the optimization problem yields a high fraction of all the multi-start solutions converging, the results are within the same magnitude as the other flowsheets' results. With cID 2 and $N = 500$ samples and also for cID 6 and the ANN, however, the observed solutions diverge, which is also reflected in the small number of multi-starts converging. All scripts and implementations regarding the SSO are available through the S3O GitHub repository [82].

All results from the SSO for all four flowsheets for N=500 samples are listed in Table 3 until Table 6: Results from the SSO of flowsheet cID6 with N=500 samples with all surrogate models and their respective solvers. All tables indicate the predicted objective function values and constraints (opt) and the results from the corresponding validation simulation with the original flowsheet for the same conditions (val) – according to optimal operational conditions and their lower (lb) and upper (ub) bounds.

Table 13: Results from the SSO of flowsheet cID1 with N=500 samples with all surrogate models and their respective solvers.

CID 1-500	ub	lb	ALAMO/BARON		DTR/Gurobi		GPR/fmincon		ANN/fmincon	
			opt	val	opt	val	opt	val	opt	val
T_PT	173	195	179.581		184.31		187.74		177.24	
acid	0.5	2	0.672		1.456		1.337		2.000	
inoc	0.5	3	3.000		1.523		1.497		1.191	
t_FX	8	16	47.938		43.207		42.656		47.727	
v_EX	0.99	0.998	0.995		0.996		0.998		0.998	
M_Xyo			59.852	0.000	49.094	48.964	54.083	43.410	85.240	45.682
C_Hmf		0.5	0.000	0.028	0.058	0.060	0.034	0.006	0.020	0.007
C_Aac		0.5	0.002	0.004	0.002	0.002	0.001	0.000	0.001	0.000
φ	0.1		0.140	0.000	0.118	0.117	0.116	0.100	0.100	0.114

Table 14: Results from the SSO of flowsheet cID2 with N=500 samples with all surrogate models and their respective solvers.

CID 2-500	ub	lb	ALAMO/BARON		DTR/Gurobi		GPR/fmincon		ANN/fmincon	
			opt	val	opt	val	opt	val	opt	val
Acid	0.5	2	0.685		0.762		0.874		2.000	
Inoc	0.5	3	2.998		1.493		0.500		2.736	
t_FX	12	48	47.999		46.427		40.493		12.490	
v_EX	0.99	0.998	0.996		0.997		0.990		0.998	
v_UCH	0.4	0.6	0.512		0.520		0.585		0.423	
M_Xyo			53.004	0.000	50.017	51.123	13.315	0.078	4.410	11.938
C_Hmf		0.5	0.500	3.935	0.500	0.451	1.830	0.779	1.467	0.136
C_Aac		0.5	0.000	0.289	0.022	0.021	0.082	0.048	0.057	0.008
φ	0.1		0.123	0.000	0.117	0.120	0.028	0.000	0.037	0.028

Table 15: Results from the SSO of flowsheet cID5 with N=500 samples with all surrogate models and their respective solvers.

CID 5-500	ub	lb	ALAMO/BARON		DTR/Gurobi		GPR/fmincon		ANN/fmincon	
			opt	val	opt	val	opt	val	opt	val
T_PT	173	195			193.69		185.50		186.6	
acid	0.5	2			0.776		1.188		1.127	
inoc	0.5	3			2.315		0.963		0.782	
t_FX	8	16			28.415		45.079		48.000	
v_EX	0.99	0.998			0.993		0.998		0.998	
M_Xyo			In-feasible		47.915	47.86	54.829	48.12	67.400	46.86
C_Hmf		0.5			0.152	0.152	0.057	0.038	0.044	0.022
C_Aac		0.5			0.012	0.126	0.003	0.002	0.002	0.001
φ	0.1				0.118	0.118	0.132	0.123	0.168	0.117

Table 16: Results from the SSO of flowsheet cID6 with N=500 samples with all surrogate models and their respective solvers.

CID 1-500	ub	lb	ALAMO/BARON		DTR/Gurobi		GPR/fmincon		ANN/fmincon	
			opt	val	opt	val	opt	val	opt	val
T_PT	173	195			184.00		184.00		191.04	
acid	0.5	2			0.960		1.265		1.609	
inoc	0.5	3			2.536		2.507		0.976	
t_FX	8	16			23.221		22.465		45.041	
v_EX	0.99	0.998			0.998		0.998		0.997	
M_Xyo			In-feasible		43.688	43.95	47.727	43.58	0.016	26.35
C_Hmf		0.5			0.373	0.347	0.500	0.322	11.321	4.627
C_Aac		0.5			0.022	0.021	0.014	0.018	0.243	0.134
φ	0.1				0.112	0.112	0.101	0.111	0.000	0.066

For the MINLP problems utilizing ALAMO surrogates and the BARON solver, it becomes clear that the low surrogate quality, expressed by the poor performance metrics, heavily affects the feasibility and the quality of the optimization results. For this option, none of the obtained solutions were satisfactory.

However, for the MILP problems utilizing the DTR surrogates, all solutions show a very close agreement between the predicted solutions and their validations, despite the insufficient validation metrics. Furthermore, the problem is solved for all flowsheets for all sample sizes. However, it becomes clear that for all flowsheets, the theoretical underlying global optimum is not reached. This is attributable to the number of sampling points provided as the DTR surrogates are strictly interpolative and can only predict according to the provided input samples.

For the GPR and the ANN surrogates, for $N = 500$ samples, the results are mostly not in agreement with the validation simulations despite the good validation metrics. For $N = 1000$ samples, however, the results from the optimization improve and converge with the validation results. This is explained by the general trend of machine learning models to show an improved prediction when provided with larger amounts of training data. Overall it shows that the GPR-assisted NLP predicts more consistency for different flowsheets and sampling sizes. In contrast to that, the ANN assisted NLP predicts more inconsistently, depending on the sampling size and flowsheet, but with successful predictions, the predicted values for the objective function are higher than the ones predicted by the GPR and thus closer to what would correspond to the global optimum for the rigorous flowsheet. The illustration Figure 15 indicates both the prediction of each surrogate model for each flowsheet and sample size as the center of the circle and the root mean square error of the testing data set of the surrogate model as the radius of the circle. Furthermore, the results from the validation simulations are added (cross).

Again, it becomes clear that the most consistent combination of the surrogate model, optimization problem, and solver is the choice of DTR surrogates despite its validation metrics, as overall, the consistency is highest. For both the GPR and the ANN, it is visible that the models predict higher objective function values, but the validation simulations are less in agreement than with the DTR surrogates. For the ALAMO surrogates, it becomes apparent that their performance in the given optimization problem is impaired. This can potentially be attributed to the underprediction surrogate models, as seen in Figure 14, which do not allow for an optimal solution under the given constraints.

Overall, it is to point out that after analyzing the quality of the surrogate models and the results from the SSO, the indication regarding the quality between the different surrogate models is ambiguous. However, regarding the underlying case study, it becomes apparent that flowsheet cID 1 shows the best objective function values for both sample sizes; hence it should be subjected to investigation in the third step of the framework. Both cID 2 and cID 5 show very similar objective function values and constraint values, which is why both should be equally subjected to investigation in the third step of the framework.

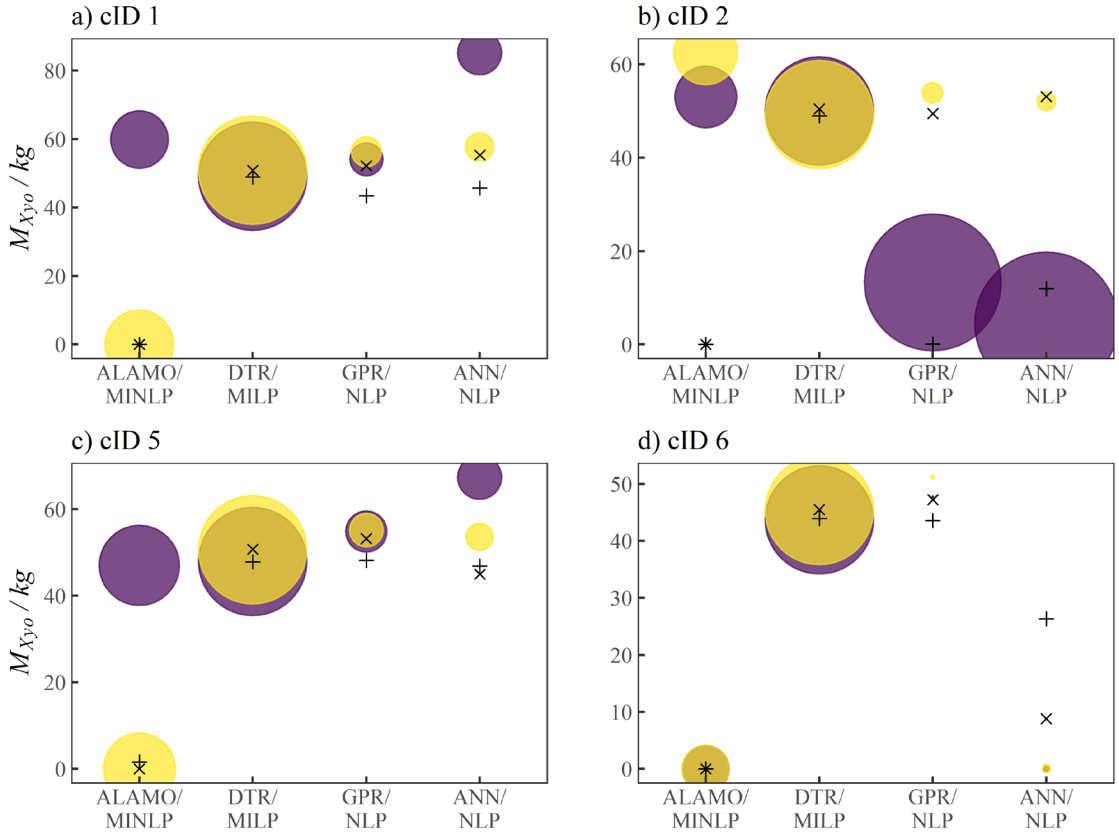

Figure 15: Bubble plot for the visualization of the consistency metrics of the different superstructure modeling approaches, the center of each sphere indicating the predicted value in the optimization problem, the radius of the sphere being the RMSE of the testing dataset in the cross-validation, and the cross/saltire indicating the respective validation simulation for a) cID1, b) cID 2, c) cID5 and d) cID 6 for respectively N=500 samples (blue, cross) and n=1000 samples (yellow, saltire).

B.3.3 Simulation-Based Optimization Results

All flowsheets considered candidate process topologies from step two of the framework are now subjected to SBO using the MOSKopt solver [27], utilizing stochastic kriging surrogate models. As uncertain input, the wheat straw composition is chosen to vary around 5% by the nominal value. The composition of the feedstock is highly dependent on climate effects as well as geological conditions of the fields, amongst others, which lead to varying compositions [85]. Logically, a varying feedstock composition influences the product yields, explaining the importance of critically assessing this uncertainty. The initially assumed composition is 31.3% hemicellulose, 42.7% cellulose, and the residual as lignin; the uncertainty is supposed to be uniformly distributed. For the SBO, 25 initial samples and each sample with 100 Monte Carlo samples are chosen. The optimization criterion for the underlying Bayesian Optimization is the multi-constraint FEI, as explained in Al et al.'s work (2020) [27]. The solver hedges against uncertainty with the simulation's mean values and performs 75 iterations, which results in 100 total calculation steps. The starting points are chosen to be the optimal results for each flowsheet from step two of the framework. All scripts and implementations regarding the SBO are available through the S3O GitHub

repository [82]. The results from the three SBO runs are listed in the following Table 17 (opt) together with the corresponding validation simulations (val).

Table 17: Results from the SBO for all candidate process topologies with the MOSKopt solver, using 25 initial sampling points, 75 iterations, the mean value as hedge against uncertainty, and the multi-constraint FEI criterion.

	ub	lb	cID 1		cID 2		cID 5	
			opt	val	opt	val	opt	val
T_PT	173	195	195.000				195	
acid	0.5	2	0.715		0.984		0.879	
inoc	0.5	3	1.611		3.000		3	
t_FX	8	16	44.994		30.367		24.271	
v_EX	0.99	0.998	0.997		0.998		0.998	
v_UCH	0.4	0.6			0.400			
M_Xyo			56.310	56.736	53.760	54.224	53.96	54.17
C_Hmf		0.5	0.045	0.032	0.492	0.471	0.062	0.061
C_Aac		0.5	0.001	0.001	0.019	0.020	0.002	0.002
φ	0.1		0.119	0.125	0.127	0.129	0.128	0.128

Firstly, the results from the SSO are confirmed, referring to that cID1 appears to be the optimal flowsheet in the given design space, including a pretreatment unit, an upconcentration unit, a fermentation unit, an evaporation unit, and two crystallization units. This is also the flowsheet option with the highest number of possible unit operations to maximize the product yield. Both cID2 and 5 seem to perform equally well, despite their difference being once performing an upconcentration before the fermentation and once performing two crystallization steps in the downstream processing. Overall, the yield for all flowsheet options is comparatively low. This indicates the possible use of alternative unit operations and operation modes in the initial superstructure to increase the amount of final product.

Secondly, it is noteworthy that for all the three flowsheets, the predicted optimum for each respectively, despite being subjected to uncertainty in the feedstock, is globally higher than by any of the introduced SSO alternatives. Figure 16 indicates for each cID how much the objective function value improved quantitatively over the 100 iterations when providing the operational conditions found to be optimal in step 2.

In conclusion, the performed SBO based on the SSO results from step 2 appears to be an excellent combinatorial solution leveraging synergistic effects between screening a multitude of alternatives and thoroughly designing few alternatives.

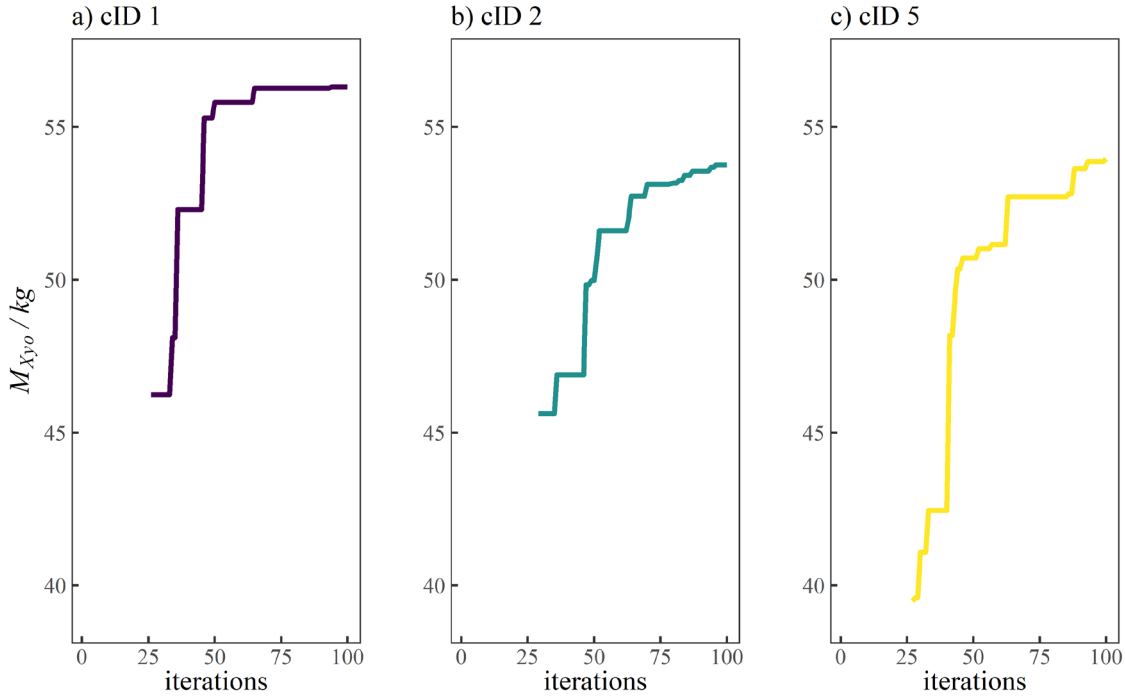

Figure 16: Visualization of the improvement of the objective function value over the iterations in the SBO with the MOSKopt solver for flowsheet a) cID 1, b) cID 2, and c) cID 5.

B.4 Conclusions & Future Research

B.4.1 Conclusions

In this paper, we presented a framework to expedite the conceptual process design of novel bioprocesses in biorefinery setups by leveraging synergistic effects from applying expert knowledge and combining SSO and SBO in a hybrid manner. In this process, four different SSO alternatives are investigated and benchmarked. The proposed bottom-up approach to compose the initial superstructure by beginning with the end in mind and first selecting a product set, a feedstock, and subsequently, the processing units reduce the initial maximum size of the superstructure heavily.

The second step of the proposed framework is based on using SSO to solve the process synthesis problem underlying the superstructure and determine several candidate process topologies. Four different alternatives of solving the optimization problem by surrogate model-assisted SSO are proposed: the use of ALAMO surrogate models in an MINLP, the use of Delaunay Triangulation Regression surrogate models in a MILP, and the use of either Gaussian Process Regression or Artificial Neural Network surrogate models in a series of NLP with a multi-start solution strategy. The latter three options serve the purpose of predicting several candidate process topologies. A thorough benchmark of all four alternatives reveals that all surrogate models have different shortcomings when applied in the framework. A general property of surrogate models is the shortcoming in terms of adhering to mass balance restrictions. Furthermore, they do not rely on physical constraints but rather on an alternative mathematical formulation and the quantity of provided data to fit the behavior of the underlying high-fidelity model accurately.

In consequence, the predictions of surrogate models inherently deviate to a certain extent. Hence, a combined validation of both surrogate performance and optimization results is essential to benchmark the proposed alternatives. In other words, the mere selection of a surrogate model based on convincing cross-validation metrics is not a guarantee to obtain a consistent and qualitatively good result from the resulting optimization problem in which the surrogate model is used. We have empirically presented in this study that cross-validation metrics alone, i.e. the coefficient of determination and the root mean square error, are solitarily not a good indicator of the surrogate model's quality in the subsequent optimization application. For example, the GPR and ANN surrogates suffer from several shortcomings in the optimization problem, whereas their validation metrics are superior. Opposed to this, the DTR surrogates' validation metrics are inferior compared to the GPR and the ANN, but the optimization results are consistent and qualitatively better. Hence, we recommend a critical assessment of the quality of a chosen surrogate model and its validation in the respective optimization problem. We also point out that the combination of DTR surrogates in a MILP has shown promising results in using the proposed framework in this particular case study.

Regarding the size of the resulting superstructure formulations for larger design cases, it is to state that also the computational burden will consequently increase but can be alleviated in several ways. For example, the necessary flowsheet simulations can be expedited by using both a parallelization approach to evaluate bigger numbers of flowsheet options and utilizing cloud or cluster computing solutions to accelerate the speed of a single simulation. Furthermore, by decreasing the initial size of the superstructure as in step 1 of the framework and solely focusing on several candidate process topologies in step 3, the overall computational burden is reduced optimally. It can also facilitate the solution of larger problems than the presented one.

Lastly, the third step of the framework involves the SBO of all candidate process topologies from step 2. The used solver MOSKopt can improve the objective function value consistently for all subjected candidates under a given uncertainty scenario and yields a consolidated optimal process design as the final result of the framework.

Overall, it remains to conclude that our proposed framework was validated successfully. It considerably facilitates and accelerates the conceptual process design of novel bioprocesses in biorefinery setups, as shown with an example of a xylitol production process in a biorefinery.

B.4.2 Future Research

Despite the impaired validation metrics, the proposed framework with the DTR surrogate models performed best in the optimization task and is hence the suggested alternative in this case. However, to improve the validation metrics and improve the predictions in terms of global optimality, a first option would be to base the framework on adaptive sampling strategies instead of a static sampling strategy as Latin Hypercube Sampling [86–88]. Alternatively, algorithms that reduce the mesh's original size by iteratively removing edges of the triangulation can be employed [89]. The improvement for the triangulation and subsequently the MILP is a reduced number of simplices by having fewer sampling points in regions, where a linear interpolation with a lower error is possible, and increasing the

number of sampling points in regions, where the linear interpolation causes a greater error and improving the prediction in the vicinity of the global optimum. This has been proposed in the literature by Chen et al. (2019) for the case of a triangulation model but also as assistance for improving the fit of other machine learning models as GPRs, which can be equally used for this optimization task, as shown in this work [90,91].

Overall, we acknowledge that the presented results here are specific to the selected case study. Therefore, the application in further case studies is necessary in order to consolidate these conclusions further. Indeed, we believe there are concrete empirical observations, which indicate that the validation issue of emerging surrogate-based/machine learning-based optimization approaches needs to be more critically analyzed and assessed to ensure their appropriate use in further process systems engineering applications.

From a more holistic perspective, the use of this framework is not only limited to process design but can equally be applied for the optimization of whole value chains, involving the choice of different feedstocks, plant locations, and logistic constraints. Lastly, especially for the used fermentation models, the black-box kinetics can be replaced by genome-scale metabolic models in order to perform also cell factory optimization, which can contribute to the further expedition of designing bioprocesses and promote the transition towards a bio-based and circular economy as instigated by the 2030 Sustainability Agenda of the United Nations.

B.5 Appendix

B.5.1 Biorefinery Unit Operations and High-Fidelity Models

As a holistic concept, a biorefinery setup can process a multitude of feedstocks into various products as fuels, chemicals, pharmaceuticals, foods, and energy; rather than the actual conversion processes, general biorefineries share the same objective of sustainable and efficient use of resources [92]. In particular, lignocellulosic biorefineries achieve this by utilizing agricultural and forestial residues or dedicated lignocellulosic crops as their raw material, making them a promising candidate for biorefinery setups due to the abundance and the low price of the feedstock [5,93]. Due to the properties of lignocellulosic biomass and the aim of developing biotechnological processes, there are several compulsory unit operations for biorefinery processes, namely a pretreatment unit, a fermentation unit, an evaporation unit, and a crystallization unit, for each of which a model is developed [8]. All mentioned units are widely used throughout industry and possess a high technology readiness level (TRL). There are further unit operations or alternatives to the four mentioned ones that have a sufficiently high TRL to be employable in a biorefinery setup; however, for the sake of the introduction of this framework, only the four mentioned units are described in detail and used in a case study for the validation of the framework. However, their use in the presented framework is not compulsory insofar as other models can equally be employed depending on the specific process design task, as the surrogate models for the framework are build independent of the underlying high-fidelity model. Furthermore, all model implementations are available through the S3O GitHub repository [82].

B.5.1.1 Pretreatment Model

Lignocellulosic biomass consists mainly of three fractions: the hemicellulosic fraction, the cellulosic fraction, and the lignin fraction. Classically, the conversion processes in lignocellulosic biorefineries use the monomers of the hemicellulosic and the cellulosic fraction as substrate; however, in the biomass, they are present in a polymeric form and cannot be converted directly [93]. Hence, lignocellulosic biorefineries employ a pretreatment process as the first unit operation to reduce the recalcitrance of the biomass and break down the polymeric structure of commonly one of the three fractions. It depends highly on the applied pretreatment technology and the operating conditions, which of the fractions is primarily targeted. However, all methodologies have in common that their objective is to be very selective on the specific fraction, have a high specific monomer yield, and have a low inhibitory compound yield [93]. The number of compounds occurring in the pretreatment is high; thus, a model commonly only describes essential components [94].
A general mass balance for a component i in an open thermodynamic system is given with the following equation:

$$\frac{dM_i}{dt} = \dot{M}_{i,in} - \dot{M}_{i,out} \pm \sum_{j \in J_i} R_j. \tag{B.29}$$

The left side of the equation describes the change of mass of the component over time dM_i/dt with $\dot{M}_{i,in}$ being the mass flow of the component into the system, $\dot{M}_{i,out}$ being the mass flow of the compound out of the system and $\sum_{j \in J_i} R_j$ being a term that sums the set of conversion reactions $J_i$of the component i with R_j as the rate expression for reaction j. Conversion reactions are either production reactions that yield the respective compound as product or consumption reactions in which the component serves as substrate and is subsequently consumed.
The kinetic expression for the reaction rates used in (A-1) can be formulated in various ways [80,95]. For this model, we chose the description as a first-order rate expression. Exemplarily for a reaction j the rate expression R_j is formulated as follows:

$$R_j = A_j \cdot \exp\left(-\frac{E_{A,j}}{\tilde{R} \cdot T}\right) \cdot \prod_{i \in I_j} C_i^{n_i}. \tag{B.30}$$

Here, A_j describes the frequency factor, $E_{A,j}$ the activation energy of the reaction, $\tilde{R}$ the universal gas constant, T the reaction temperature and a term $\prod_{i \in I_j} C_i^{n_i}$ which multiplies over the set of substrate concentrations I_j for the reaction j with C_i as substrate concentration of component i and n_i as exponent describing the reaction order in which the component i participates in the reaction j. The frequency factors, the activation energies, and the reaction orders are usually estimated from experimental data to fit the model to the respective pretreatment method.

Analogously to the mass balance, an energy balance for an open thermodynamic system, in this case, the whole unit operation constitutes as follows in the form of an enthalpy balance:

$$\frac{dH}{dt} = \sum \dot{H}_{in} - \sum \dot{H}_{out} + \sum \dot{Q}_{in} - \sum \dot{Q}_{out} + \sum \dot{H}_{source} - \sum \dot{H}_{sink}. \tag{B.31}$$

With dH/dt being the change of enthalpy in the system, $\dot{H}_{in}$ representing all ingoing enthalpy flows due to mass flow into the system, whereas all outgoing enthalpy flows due

to mass flow out of the system are described with $\dot{H}_{out}$. Ingoing and outgoing heat flows are described with $\dot{Q}_{in}$ and $\dot{Q}_{out}$ respectively, as well as possible heat sources and sinks with $\dot{H}_{source}$ and $\dot{H}_{sink}$ describing e.g. the release of enthalpy of reaction. Equation (B.31) in its presented form is generic, and represented terms can be omitted depending on the respective model.

B.5.1.2 Fermentation Model

The fermentation unit operation serves as the central reaction unit for converting the monomers from the pretreatment into the desired products. The reaction is performed by a cell factory, commonly bacteria or yeast strains, which either produce the desired product from the feedstock naturally or artificially after performing genetic and metabolic modifications on the cell factory a priori. Despite the high complexity of the metabolic network and transcriptional regulations defining the kinetic behavior, the entire cell factory can be considered as black-box, lumping the whole metabolic network into three kinetic equations in the simplest case [96]: the substrate uptake rate, a Herbert-Pirt distribution relation, and a production rate.
The substrate uptake rate q_s is constituted as follows:

$$q_s = q_{s,max} \cdot \frac{C_s}{K_s + C_s}, \tag{B.32}$$

where $q_{s,max}$ relates to the maximum substrate uptake rate under excess conditions, C_s being the concentration of substrate and K_s being the substrate affinity constant.
The Herbert-Pirt distribution relation describes the distribution of the substrate in the cell towards biomass growth, production of the desired component, and cell maintenance and is described in the following way:

$$q_s = a_s \cdot \mu + b_s \cdot q_p + m_s. \tag{B.33}$$

The biomass growth is denoted by the growth rate μ, the product formation as q_p and cell maintenance by m_s. The parameters a_s and b_s as well as the maintenance constant are specific for the employed cell factory and have to be determined experimentally.
The last kinetic equation describes the product formation for catabolic products as a function of the biomass growth for catabolic products:

$$q_p = a_p \cdot \mu + b_p. \tag{B.34}$$

Analogously to equation (B.33), parameters a_p and b_p are specific for the employed cell factory and have to be determined experimentally. Most of the mentioned products (biofuels, biochemicals, and others) are usually catabolic products; nonetheless, there are several formulations of product formation equations for anabolic products, which can be used instead of equation (B.34). For these, the reader is referred to additional literature [96].

In a similar fashion to the pretreatment model, also for the fermentation model, the mass and energy balances are described with equations (A-1) and (B.31) respectively. For the case of a black-box model, exactly three mass balance equations are needed for substrate, cell biomass, and product. The reaction term, which has to be considered in the substrate's mass balance (A-1), is the substrate uptake rate from equation (B.32). The reaction term for the cell biomass derives from equation (B.33), and the reaction term equals equation (B.34).

All rate equations are dependent on biomass growth, so they have to be multiplied with the cell biomass concentration C_X:

$$R_j = q_j \cdot C_X \tag{B.35}$$

B.5.1.3 Evaporation Model

Due to commonly low titers in aqueous fermentation processes, unit operations for upconcentration are employed as part of the downstream processing to obtain the desired product in a sufficiently high concentration. This mainly happens with evaporation or membrane separation units, despite their high operational expenditures [97,98]. Furthermore, the evaporation unit operation mainly also serves to remove undesired compounds through the vapor phase. For the evaporation model, the mass and energy balances described with equations (A-1) and (B.31) apply. The thermodynamic equilibrium between the vapor and the liquid phase is described with vapor pressure equations for the components. Common process simulation software includes implemented evaporation models; therefore, for this paper's scope, instead of developing an own evaporation model, the one implemented in the Aspen Plus software package is utilized.

B.5.1.4 Crystallization Model

Crystallization as a unit operation is commonly employed for recovering solid products out of a liquid phase. Especially for solid products, the isolation and purification are classically done with a crystallization unit as the product size and shape are good to control [99]. In the crystallization process, the thermodynamic driving force is an excess concentration of the product above the respective saturation concentration under the given operational conditions. The process is controlled either by the temperature, which relates to either evaporation or cooling crystallization. Alternatively, an anti-solvent crystallization utilizes a decrease in solubility by adding a second solvent with low solubility for the compound [99]. Like all the other unit operations, this unit operation's mass and energy balances can be described with equations (A-1) and (B.31). The kinetics of the formation of crystals can be described with a population balance equation as the following:

$$\frac{\partial n}{\partial t} + \frac{\partial (G \cdot n)}{\partial L} = B - D, \tag{B.36}$$

where n describes the particle number density, G the crystal growth rate, B the crystal birth rate, D the crystal decay rate, and L the crystal length. Commonly, all rates and the underlying solubility curves are estimated from experimental data in order to quantify the driving force and model the process. There exist several ways of solving a population balance, e.g., by the method of moments or the method of classes, to yield a kinetic expression that describes the mass transfer between the liquid and solid phase of the solute described with equation (A-1).

B.5.2 Uncertainty and Sensitivity Analysis

All of the models introduced in section B.5.1 have in common that they are based on assumptions to a certain extend. Moreover, all of them rely on experimental data for their calibration to the system they are supposed to predict. Both present a significant source of uncertainty for the model prediction, especially when coupling models in a flowsheet

simulation. These errors can propagate and accumulate to an extent that deteriorates the models' predictive quality. Hence, it is crucial to assess these models' robustness by analyzing how uncertainties in the model input propagate to the model output and how the uncertainty in the model output can be apportioned to the model input [100]. The former analysis is called uncertainty analysis, whereas the latter being complementary to the prior is called sensitivity analysis. The purpose of performing uncertainty analysis is to assess the model's robustness and quantify the error propagation due to the estimated model parameters or uncertain input or design variables. The purpose of performing a sensitivity analysis is more diverse: It reaches from a general design space exploration with the underlying model to identifying sensitive input variables or parameters to select a subset of variables for optimization problems and identify crucial parameters to estimate model parameters other tasks.

B.5.2.1 Monte Carlo-Based Uncertainty Analysis

In this framework, the assessment is performed with a Monte Carlo-based method [101]. Therefore, its theory is described briefly:
Monte Carlo methods utilize probability statistics and random numbers [102]. In the case of the uncertainty analysis, the model input space is sampled randomly with a sufficiently high number of random samples [103]. For each of the samples, a simulation is performed in order to calculate the model output. Based on these model outputs, the output uncertainty of the model can be quantified. In practice, this is performed in the following way:

Firstly, ranges and uncertainty distributions are defined for all model parameters m that are considered uncertain. Secondly, a sufficiently high sample number N is chosen, and Latin Hypercube Sampling (LHS) is performed for all the uncertain inputs. The result is a matrix $X_{N\times m}$ over the input space. For each sample n, a model simulation is performed to create the respective output space matrix $Y_{N\times k}$ for all considered outputs k. Based on the output calculations, statistical performance parameters as mean value, standard deviations, and percentiles can be calculated.

B.5.2.2 Variance-Based Sensitivity Analysis

The complementary analysis to the uncertainty analysis is a sensitivity analysis, aiming at apportioning the output uncertainty on the different inputs [104]. The here applied methodology is based on a variance-based sensitivity measure. This method calculates the model's total variance based on Monte Carlo simulations and is usually referred to as Sobol's sensitivity method [105].

Saltelli et al. (2010) describe the methodology as follows: For independent inputs of a model $y = f(\theta_i)$, the variance of the model output $V(y)$ can be partitioned in the following way:

$$V(y) = \sum_{i}^{k} V_i + \sum_{i}\sum_{k} V_{ij} + \ldots + V_{123\ldots k}, \tag{B.37}$$

with the variance V_i as:

$$V_i = \int f(\theta_i)^2 \, d\theta_i. \tag{B.38}$$

Each term serves as a measure of sensitivity of the respective input(s) in equation (B.37). Applying the law of total variance:

$$\mathrm{V}(y) = V\big(E(y|\theta_i)\big) + E\big(V(y|\theta_i)\big), \tag{B.39}$$

for each term in (B.37) yields measures for the first-order sensitivity index S_i and the total sensitivity index S_{Ti}:

$$S_i = V\big(E(y|\theta_i)\big)/\mathrm{V}(y), \tag{B.40}$$

$$S_{Ti} = E\big(V(y|\theta_i)\big)/\mathrm{V}(y). \tag{B.41}$$

The numerical calculation of both sensitivity indices is performed by a Monte Carlo-based procedure: As a first step, Sobol sampling is performed, and two sampling matrices A and B are generated. From those, two mixed matrices A_B^i and B_A^i are generated, where column i from the one matrix is replaced by the same column of the respective other matrix, and all other columns are kept. For all four sampling matrices, the model outputs are calculated. The respective sensitivity measures as in (2.12) and (2.13) can be calculated with those. Applying the method for the first-order sensitivities, these are the following:

$$S_i = V(y) - \frac{1}{2N}\sum_{j=1}^{N}\big(y_B(j) - y_{ABi}(j)\big)^2, \tag{B.42}$$

and the total sensitivities are calculated as follows:

$$S_i = \frac{1}{2N}\sum_{j=1}^{N}(y_A(j) - y_{ABi}(j))^{\,2}. \tag{B.43}$$

The interpretation of both is also twofold: The first order sensitivity explains this parameter's single effect, indicating the expected reduction in the output variance if this particular parameter could be fixed. Hence, the total sensitivity is the expected variance if all the parameters except the respective one could be fixed [106].

Acknowledgments

The authors would like to express their gratitude to the Novo Nordisk Foundation (Grant no: NNF17SA0031362) for funding the Fermentation-Based Biomanufacturing Initiative of which this project is a part.

Bibliography

[1] United Nations, Transforming our world: The 2030 agenda for sustainable development, 2015. https://sdgs.un.org/sites/default/files/publications/21252030 Agenda for Sustainable Development web.pdf (accessed March 1, 2021).

[2] A.T. Ubando, C.B. Felix, W.H. Chen, Biorefineries in circular bioeconomy: A comprehensive review, Bioresour. Technol. 299 (2020) 122585. https://doi.org/10.1016/j.biortech.2019.122585.

[3] A.J.J. Straathof, S.A. Wahl, K.R. Benjamin, R. Takors, N. Wierckx, H.J. Noorman, Grand Research Challenges for Sustainable Industrial Biotechnology, Trends Biotechnol. 37 (2019) 1042–1050. https://doi.org/10.1016/j.tibtech.2019.04.002.

[4] N. Hillson, M. Caddick, Y. Cai, J.A. Carrasco, M.W. Chang, N.C. Curach, D.J. Bell, R. Le Feuvre, D.C. Friedman, X. Fu, N.D. Gold, M.J. Herrgård, M.B. Holowko, J.R. Johnson, R.A. Johnson, J.D. Keasling, R.I. Kitney, A. Kondo, C. Liu, V.J.J. Martin, F. Menolascina, C. Ogino, N.J. Patron, M. Pavan, C.L. Poh, I.S. Pretorius, S.J. Rosser, N.S. Scrutton, M. Storch, H. Tekotte, E. Travnik, C.E. Vickers, W.S. Yew, Y. Yuan, H. Zhao, P.S. Freemont, Building a global alliance of biofoundries, Nat. Commun. 10 (2019) 1038–1041. https://doi.org/10.1038/s41467-019-10079-2.

[5] S.S. Hassan, G.A. Williams, A.K. Jaiswal, Lignocellulosic Biorefineries in Europe: Current State and Prospects, Trends Biotechnol. 37 (2019) 231–234. https://doi.org/10.1016/j.tibtech.2018.07.002.

[6] S.S. Hassan, G.A. Williams, A.K. Jaiswal, Moving towards the second generation of lignocellulosic biorefineries in the EU: Drivers, challenges, and opportunities, Renew. Sustain. Energy Rev. 101 (2019) 590–599. https://doi.org/10.1016/j.rser.2018.11.041.

[7] J. Moncada B, V. Aristizábal M, C.A. Cardona A, Design strategies for sustainable biorefineries, Biochem. Eng. J. 116 (2016) 122–134. https://doi.org/10.1016/j.bej.2016.06.009.

[8] T. Chaturvedi, A.I. Torres, G. Stephanopoulos, M.H. Thomsen, J.E. Schmidt, Developing process designs for biorefineries-definitions, categories, and unit operations, Energies. 13 (2020) 1493. https://doi.org/10.3390/en13061493.

[9] A.C. Kokossis, A. Yang, On the use of systems technologies and a systematic approach for the synthesis and the design of future biorefineries, Comput. Chem. Eng. 34 (2010) 1397–1405. https://doi.org/10.1016/j.compchemeng.2010.02.021.

[10] N.G. Chemmangattuvalappil, D.K.S. Ng, L.Y. Ng, J. Ooi, J.W. Chong, M.R. Eden, A review of process systems engineering (PSE) tools for the design of ionic liquids and integrated biorefineries, Processes. 8 (2020) 1–29. https://doi.org/10.3390/pr8121678.

[11] S.Y. Tey, S.S. Wong, J.A. Lam, N.Q.X. Ong, D.C.Y. Foo, D.K.S. Ng, Extended hierarchical decomposition approach for the synthesis of biorefinery processes, Chem. Eng. Res. Des. 166 (2021) 40–54. https://doi.org/10.1016/j.cherd.2020.11.015.

[12] N.M. Clauser, F.E. Felissia, M.C. Area, M.E. Vallejos, A framework for the design and analysis of integrated multi-product biorefineries from agricultural and forestry wastes, Renew. Sustain. Energy Rev. 139 (2021) 110687.

https://doi.org/10.1016/j.rser.2020.110687.

[13] A.D. Mountraki, B. Benjelloun-Mlayah, A.C. Kokossis, A Surrogate Modeling Approach for the Development of Biorefineries, Front. Chem. Eng. 2 (2020). https://doi.org/10.3389/fceng.2020.568196.

[14] K.A. Pyrgakis, A.C. Kokossis, A Total Site Synthesis approach for the selection, integration and planning of multiple-feedstock biorefineries, Comput. Chem. Eng. 122 (2019) 326–355. https://doi.org/10.1016/j.compchemeng.2018.09.003.

[15] S.I. Meramo-Hurtado, Á.D. González-Delgado, Biorefinery synthesis and design using sustainability parameters and hierarchical/3D multi-objective optimization, J. Clean. Prod. 240 (2019) 118134. https://doi.org/10.1016/j.jclepro.2019.118134.

[16] C. Galanopoulos, A. Giuliano, D. Barletta, E. Zondervan, An integrated methodology for the economic and environmental assessment of a biorefinery supply chain, Chem. Eng. Res. Des. 160 (2020) 199–215. https://doi.org/10.1016/j.cherd.2020.05.016.

[17] K. Ulonska, A. König, M. Klatt, A. Mitsos, J. Viell, Optimization of Multiproduct Biorefinery Processes under Consideration of Biomass Supply Chain Management and Market Developments, Ind. Eng. Chem. Res. 57 (2018) 6980–6991. https://doi.org/10.1021/acs.iecr.8b00245.

[18] V. Aristizábal-Marulanda, C.A. Cardona Alzate, Methods for designing and assessing biorefineries: Review, Biofuels, Bioprod. Biorefining. 13 (2019) 789–808. https://doi.org/10.1002/bbb.1961.

[19] S.I. Meramo-Hurtado, Á.D. González-Delgado, Process Synthesis, Analysis, and Optimization Methodologies toward Chemical Process Sustainability, Ind. Eng. Chem. Res. 60 (2021) 4193–4217. https://doi.org/10.1021/acs.iecr.0c05456.

[20] K. Darkwah, B.L. Knutson, J.R. Seay, A Perspective on Challenges and Prospects for Applying Process Systems Engineering Tools to Fermentation-Based Biorefineries, ACS Sustain. Chem. Eng. 6 (2018) 2829–2844. https://doi.org/10.1021/acssuschemeng.7b03762.

[21] L.T. Biegler, I.E. Grossmann, A.W. Westerberg, Systematic methods for chemical process design, 1st ed., Pearson, London, 1997. https://www.osti.gov/biblio/293030 (accessed December 11, 2020).

[22] Z. Yuan, M.R. Eden, Superstructure optimization of integrated fast pyrolysis-gasification for production of liquid fuels and propylene, AIChE J. 62 (2016) 3155–3176. https://doi.org/10.1002/aic.15337.

[23] Q. Chen, I.E. Grossmann, Recent developments and challenges in optimization-based process synthesis, Annu. Rev. Chem. Biomol. Eng. 8 (2017) 249–283. https://doi.org/10.1146/annurev-chembioeng-080615-033546.

[24] I.E. Grossmann, R.M. Apap, B.A. Calfa, P. García-Herreros, Q. Zhang, Recent advances in mathematical programming techniques for the optimization of process systems under uncertainty, Comput. Chem. Eng. 91 (2016) 3–14. https://doi.org/10.1016/j.compchemeng.2016.03.002.

[25] M. Koutinas, A. Kiparissides, E.N. Pistikopoulos, A. Mantalaris, Bioprocess systems engineering: Transferring traditional process engineering principles to industrial biotechnology, Comput. Struct. Biotechnol. J. 3 (2012) e201210022.

https://doi.org/10.5936/csbj.201210022.

[26] A. Bhosekar, M. Ierapetritou, Advances in surrogate based modeling, feasibility analysis, and optimization: A review, Comput. Chem. Eng. 108 (2018) 250–267. https://doi.org/10.1016/j.compchemeng.2017.09.017.

[27] R. Al, C.R. Behera, K. V. Gernaey, G. Sin, Stochastic simulation-based superstructure optimization framework for process synthesis and design under uncertainty, Comput. Chem. Eng. 143 (2020) 107118. https://doi.org/10.1016/j.compchemeng.2020.107118.

[28] Z. Wang, M. Ierapetritou, Constrained optimization of black-box stochastic systems using a novel feasibility enhanced Kriging-based method, Comput. Chem. Eng. 118 (2018) 210–223. https://doi.org/10.1016/j.compchemeng.2018.07.016.

[29] K. McBride, K. Sundmacher, Overview of Surrogate Modeling in Chemical Process Engineering, Chemie-Ingenieur-Technik. 91 (2019) 228–239. https://doi.org/10.1002/cite.201800091.

[30] M. Friedman, Multivariate Adaptive Regression Splines, Ann. Stat. 19 (1991) 1–67.

[31] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliab. Eng. Syst. Saf. 93 (2008) 964–979. https://doi.org/10.1016/j.ress.2007.04.002.

[32] B.A. Williams, S. Cremaschi, Surrogate Model Selection for Design Space Approximation And Surrogatebased Optimization, Comput. Aided Chem. Eng. 47 (2019) 353–358. https://doi.org/10.1016/B978-0-12-818597-1.50056-4.

[33] H. Janssen, Monte-Carlo based uncertainty analysis: Sampling efficiency and sampling convergence, Reliab. Eng. Syst. Saf. 109 (2013) 123–132. https://doi.org/10.1016/j.ress.2012.08.003.

[34] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer, New York, NY, 2009. https://doi.org/10.1198/jasa.2004.s339.

[35] Z.T. Wilson, N. V. Sahinidis, The ALAMO approach to machine learning, Comput. Chem. Eng. 106 (2017) 785–795. https://doi.org/10.1016/j.compchemeng.2017.02.010.

[36] A. Cozad, N. V. Sahinidis, D.C. Miller, Learning surrogate models for simulation-based optimization, AIChE J. 60 (2014) 2211–2227. https://doi.org/10.1002/aic.14418.

[37] J.C. Eslick, B. Ng, Q. Gao, C.H. Tong, N. V. Sahinidis, D.C. Miller, A framework for optimization and quantification of uncertainty and sensitivity for developing carbon capture systems, Energy Procedia. 63 (2014) 1055–1063. https://doi.org/10.1016/j.egypro.2014.11.113.

[38] D.C. Miller, J.D. Siirola, D. Agarwal, A.P. Burgard, A. Lee, J.C. Eslick, B. Nicholson, C. Laird, L.T. Biegler, D. Bhattacharyya, N. V. Sahinidis, I.E. Grossmann, C.E. Gounaris, D. Gunter, Next Generation Multi-Scale Process Systems Engineering Framework, Comput. Aided Chem. Eng. 44 (2018) 2209–2214. https://doi.org/10.1016/B978-0-444-64241-7.50363-3.

[39] B. Delaunay, Sur la sphère vide, J. Phys. Le Radium. 12 (1934) 793–800. https://doi.org/10.1051/jphysrad:01951001207073500.

[40] B. Žalik, An efficient sweep-line Delaunay triangulation algorithm, CAD Comput. Aided Des. 37 (2005) 1027–1038. https://doi.org/10.1016/j.cad.2004.10.004.

[41] C.B. Barber, D.P. Dobkin, H. Huhdanpaa, The Quickhull Algorithm for Convex Hulls, ACM Trans. Math. Softw. 22 (1996) 469–483. https://doi.org/10.1145/235815.235821.

[42] R. Al, C.R. Behera, A. Zubov, K. V. Gernaey, G. Sin, Meta-modeling based efficient global sensitivity analysis for wastewater treatment plants – An application to the BSM2 model, Comput. Chem. Eng. 127 (2019) 233–246. https://doi.org/10.1016/j.compchemeng.2019.05.015.

[43] C.E. Rasmussen, Gaussian Processes in machine learning, in: O. Bousquet, U. von Luxburg, G. Rätsch (Eds.), Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer Verlag, Berlin, Heidelberg, 2004: pp. 63–71. https://doi.org/10.1007/978-3-540-28650-9_4.

[44] F. Boukouvala, M.G. Ierapetritou, Feasibility analysis of black-box processes using an adaptive sampling Kriging-based method, Comput. Chem. Eng. 36 (2012) 358–368. https://doi.org/10.1016/j.compchemeng.2011.06.005.

[45] J.A. Caballero, I.E. Grossmann, An algorithm for the use of surrogate models in modular flowsheet optimization, AIChE J. 54 (2008) 2633–2650. https://doi.org/10.1002/aic.11579.

[46] E. Davis, M. Ierapetritou, A kriging based method for the solution of mixed-integer nonlinear programs containing black-box functions, J. Glob. Optim. 43 (2009) 191–205. https://doi.org/10.1007/s10898-007-9217-2.

[47] S. Hwangbo, R. Al, G. Sin, An integrated framework for plant data-driven process modeling using deep-learning with Monte-Carlo simulations, Comput. Chem. Eng. 143 (2020) 107071. https://doi.org/10.1016/j.compchemeng.2020.107071.

[48] A.M. Schweidtmann, A. Mitsos, Deterministic Global Optimization with Artificial Neural Networks Embedded, J. Optim. Theory Appl. 180 (2019) 925–948. https://doi.org/10.1007/s10957-018-1396-0.

[49] C.A. Henao, C.T. Maravelias, Surrogate-based superstructure optimization framework, AIChE J. 57 (2011) 1216–1232. https://doi.org/10.1002/aic.12341.

[50] H. Yeomans, I.E. Grossmann, A systematic modeling framework of superstructure optimization in process synthesis, Comput. Chem. Eng. 23 (1999) 709–731. https://doi.org/10.1016/S0098-1354(99)00003-4.

[51] L. Mencarelli, Q. Chen, A. Pagot, I.E. Grossmann, A review on superstructure optimization approaches in process system engineering, Comput. Chem. Eng. 136 (2020) 106808. https://doi.org/10.1016/j.compchemeng.2020.106808.

[52] W.R. Huster, A.M. Schweidtmann, J.T. Lüthje, A. Mitsos, Deterministic global superstructure-based optimization of an organic Rankine cycle, Comput. Chem. Eng. 141 (2020) 106996. https://doi.org/10.1016/j.compchemeng.2020.106996.

[53] M. Jones, H. Forero-Hernandez, A. Zubov, B. Sarup, G. Sin, Superstructure Optimization of Oleochemical Processes with Surrogate Models, Comput. Aided Chem. Eng. 44 (2018) 277–282. https://doi.org/10.1016/B978-0-444-64241-7.50041-0.

[54] R. Misener, C.A. Floudas, Piecewise-linear approximations of multidimensional functions, J. Optim. Theory Appl. 145 (2010) 120–147. https://doi.org/10.1007/s10957-009-9626-0.

[55] R. Misener, C.E. Gounaris, C.A. Floudas, Global optimization of gas lifting operations: A comparative study of piecewise linear formulations, Ind. Eng. Chem. Res. 48 (2009) 6098–6104. https://doi.org/10.1021/ie8012117.

[56] E.N. Pistikopoulos, Uncertainty in process design and operations, Comput. Chem. Eng. 19 (1995) 553–563. https://doi.org/10.1016/0098-1354(95)87094-6.

[57] S. Amaran, N. V. Sahinidis, B. Sharda, S.J. Bury, Simulation optimization: a review of algorithms and applications, 4OR. 12 (2014) 301–333. https://doi.org/10.1007/s10288-014-0275-2.

[58] M.C. Fu, C.C. Price, J. Zhu, F.S. Hillier, Handbook of Simulation Optimization Associate Series Editor, Springer New York, New York, NY, 2015. https://doi.org/10.1007/978-1-4939-1384-8.

[59] B. Ankenman, B.L. Nelson, J. Staum, Stochastic kriging for simulation metamodeling, Oper. Res. 58 (2010) 371–382. https://doi.org/10.1287/opre.1090.0754.

[60] D. Bertsimas, M. Sim, The price of robustness, Oper. Res. 52 (2004) 35–53. https://doi.org/10.1287/opre.1030.0065.

[61] C. Ning, F. You, Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming, Comput. Chem. Eng. 125 (2019) 434–448. https://doi.org/10.1016/j.compchemeng.2019.03.034.

[62] G. Hüllen, J. Zhai, S.H. Kim, A. Sinha, M.J. Realff, F. Boukouvala, Managing uncertainty in data-driven simulation-based optimization, Comput. Chem. Eng. 136 (2020) 106519. https://doi.org/10.1016/j.compchemeng.2019.106519.

[63] C.M. Marques, S. Moniz, J.P. de Sousa, A.P. Barbosa-Póvoa, A simulation-optimization approach to integrate process design and planning decisions under technical and market uncertainties: A case from the chemical-pharmaceutical industry, Comput. Chem. Eng. 106 (2017) 796–813. https://doi.org/10.1016/j.compchemeng.2017.04.008.

[64] J.S. Crater, J.C. Lievense, Scale-up of industrial microbial processes, FEMS Microbiol. Lett. 365 (2018) 138. https://doi.org/10.1093/femsle/fny138.

[65] H.J. Noorman, J.J. Heijnen, Biochemical engineering's grand adventure, Chem. Eng. Sci. 170 (2017) 677–693. https://doi.org/10.1016/j.ces.2016.12.065.

[66] S.S. Da Silva, A.K. Chandel, D-Xylitol: Fermentative production, application and commercialization, Springer-Verlag Berlin Heidelberg, 2012. https://doi.org/10.1007/978-3-642-31887-0.

[67] S. Choi, C.W. Song, J.H. Shin, S.Y. Lee, Biorefineries for the production of top building block chemicals and their derivatives, Metab. Eng. 28 (2015) 223–239. https://doi.org/10.1016/j.ymben.2014.12.007.

[68] T.L. De Albuquerque, I.J. Da Silva, G.R. De MacEdo, M.V.P. Rocha, Biotechnological production of xylitol from lignocellulosic wastes: A review, Process Biochem. 49 (2014) 1779–1789. https://doi.org/10.1016/j.procbio.2014.07.010.

[69] L.V. Rao, J.K. Goli, J. Gentela, S. Koti, Bioconversion of lignocellulosic biomass to xylitol: An overview, Bioresour. Technol. 213 (2016) 299–310. https://doi.org/10.1016/j.biortech.2016.04.092.

[70] D. Dasgupta, S. Bandhu, D.K. Adhikari, D. Ghosh, Challenges and prospects of xylitol production with whole cell bio-catalysis: A review, Microbiol. Res. 197 (2017) 9–21. https://doi.org/10.1016/j.micres.2016.12.012.

[71] A.F. Hernández-Pérez, P.V. de Arruda, L. Sene, S.S. da Silva, A. Kumar Chandel, M. das G. de Almeida Felipe, Xylitol bioproduction: state-of-the-art, industrial paradigm shift, and opportunities for integrated biorefineries, Crit. Rev. Biotechnol. 39 (2019) 924–943. https://doi.org/10.1080/07388551.2019.1640658.

[72] Y. Delgado Arcaño, O.D. Valmaña García, D. Mandelli, W.A. Carvalho, L.A. Magalhães Pontes, Xylitol: A review on the progress and challenges of its production by chemical route, Catal. Today. 344 (2020) 2–14. https://doi.org/10.1016/j.cattod.2018.07.060.

[73] A.D. Mountraki, K.R. Koutsospyros, B.B. Mlayah, A.C. Kokossis, Selection of Biorefinery Routes: The Case of Xylitol and its Integration with an Organosolv Process, Waste and Biomass Valorization. 8 (2017) 2283–2300. https://doi.org/10.1007/s12649-016-9814-8.

[74] G. Franceschin, M. Sudiro, T. Ingram, I. Smirnova, G. Brunner, A. Bertucco, Conversion of rye straw into fuel and xylitol: A technical and economical assessment based on experimental data, Chem. Eng. Res. Des. 89 (2011) 631–640. https://doi.org/10.1016/j.cherd.2010.11.001.

[75] A. Giuliano, D. Barletta, I. De Bari, M. Poletto, Techno-economic assessment of a lignocellulosic biorefinery co-producing ethanol and xylitol or furfural, Comput. Aided Chem. Eng. 43 (2018) 585–590. https://doi.org/10.1016/B978-0-444-64235-6.50105-4.

[76] E. Mancini, S.S. Mansouri, K. V. Gernaey, J. Luo, M. Pinelo, From second generation feed-stocks to innovative fermentation and downstream techniques for succinic acid production, Crit. Rev. Environ. Sci. Technol. 50 (2020) 1829–1873. https://doi.org/10.1080/10643389.2019.1670530.

[77] A.J. Ragauskas, G.T. Beckham, M.J. Biddy, R. Chandra, F. Chen, M.F. Davis, B.H. Davison, R.A. Dixon, P. Gilna, M. Keller, P. Langan, A.K. Naskar, J.N. Saddler, T.J. Tschaplinski, G.A. Tuskan, C.E. Wyman, Lignin valorization: Improving lignin processing in the biorefinery, Science (80-.). 344 (2014). https://doi.org/10.1126/science.1246843.

[78] V.K. Ponnusamy, D.D. Nguyen, J. Dharmaraja, S. Shobana, J.R. Banu, R.G. Saratale, S.W. Chang, G. Kumar, A review on lignin structure, pretreatments, fermentation reactions and biorefinery potential, Bioresour. Technol. 271 (2019) 462–472. https://doi.org/10.1016/j.biortech.2018.09.070.

[79] W.C. Wang, L. Tao, Bio-jet fuel conversion technologies, Renew. Sustain. Energy Rev. 53 (2016) 801–822. https://doi.org/10.1016/j.rser.2015.09.016.

[80] R.M. Prunescu, M. Blanke, J.G. Jakobsen, G. Sin, Dynamic modeling and validation of a biomass hydrothermal pretreatment process-a demonstration scale study, AIChE J. 61 (2015) 4235–4250. https://doi.org/10.1002/aic.14954.

[81] W. Tochampa, S. Sirisansaneeyakul, W. Vanichsriratana, P. Srinophakun, H.H.C.

Bakker, Y. Chisti, A model of xylitol production by the yeast Candida mogii, Bioprocess Biosyst. Eng. 28 (2005) 175–183. https://doi.org/10.1007/s00449-005-0025-0.

[82] N. Vollmer, R. Al, S3O, GitHub Repos. (2021). https://github.com/NikolausVollmer/S3O.

[83] R. Al, C.R. Behera, K. V. Gernaey, G. Sin, Towards development of a decision support tool for conceptual design of wastewater treatment plants using stochastic simulation optimization, in: Comput. Aided Chem. Eng., Elsevier B.V., 2019: pp. 325–330. https://doi.org/10.1016/B978-0-12-818634-3.50055-2.

[84] M.R. Kılınç, N. V. Sahinidis, Exploiting integrality in the global optimization of mixed-integer nonlinear programming problems with BARON, Optim. Methods Softw. 33 (2018) 540–562. https://doi.org/10.1080/10556788.2017.1350178.

[85] S. V. Vassilev, D. Baxter, L.K. Andersen, C.G. Vassileva, T.J. Morgan, An overview of the organic and inorganic phase composition of biomass, Fuel. 94 (2012) 1–33. https://doi.org/10.1016/j.fuel.2011.09.030.

[86] J. Eason, S. Cremaschi, Adaptive sequential sampling for surrogate model generation with artificial neural networks, Comput. Chem. Eng. 68 (2014) 220–232. https://doi.org/10.1016/j.compchemeng.2014.05.021.

[87] S.S. Garud, I.A. Karimi, M. Kraft, Smart Sampling Algorithm for Surrogate Model Development, Comput. Chem. Eng. 96 (2017) 103–114. https://doi.org/10.1016/j.compchemeng.2016.10.006.

[88] S.S. Garud, I.A. Karimi, G.P.E. Brownbridge, M. Kraft, Evaluating smart sampling for constructing multidimensional surrogate models, Comput. Chem. Eng. 108 (2018) 276–288. https://doi.org/10.1016/j.compchemeng.2017.09.016.

[89] A. Obermeier, N. Vollmer, C. Windmeier, E. Esche, J.U. Repke, Generation of linear-based surrogate models from non-linear functional relationships for use in scheduling formulation, Comput. Chem. Eng. 146 (2021) 107203. https://doi.org/10.1016/j.compchemeng.2020.107203.

[90] Y. Chen, P. Goetsch, M.A. Hoque, J. Lu, S. Tarkoma, d-Simplexed: Adaptive Delaunay Triangulation for Performance Modeling and Prediction on Big Data Analytics, IEEE Trans. Big Data. (2019) 1–1. https://doi.org/10.1109/tbdata.2019.2948338.

[91] P. Jiang, Y. Zhang, Q. Zhou, X. Shao, J. Hu, L. Shu, An adaptive sampling strategy for Kriging metamodel based on Delaunay triangulation and TOPSIS, Appl. Intell. 48 (2018) 1644–1656. https://doi.org/10.1007/s10489-017-1031-z.

[92] F. Cherubini, The biorefinery concept: Using biomass instead of oil for producing energy and chemicals, Energy Convers. Manag. 51 (2010) 1412–1421. https://doi.org/10.1016/j.enconman.2010.01.015.

[93] M. Galbe, O. Wallberg, Pretreatment for biorefineries: A review of common methods for efficient utilisation of lignocellulosic materials, Biotechnol. Biofuels. 12 (2019) 1–26. https://doi.org/10.1186/s13068-019-1634-1.

[94] H. Rasmussen, H.R. Sørensen, A.S. Meyer, Formation of degradation compounds from lignocellulosic biomass in the biorefinery: Sugar reaction mechanisms, Carbohydr. Res. 385 (2014) 45–57. https://doi.org/10.1016/j.carres.2013.08.029.

[95] J. Shen, C.E. Wyman, A novel mechanism and kinetic model to explain enhanced xylose yields from dilute sulfuric acid compared to hydrothermal pretreatment of corn stover, Bioresour. Technol. 102 (2011) 9111–9120. https://doi.org/10.1016/j.biortech.2011.04.001.

[96] J.J. Heijnen, W.M. van Gulik, Section II - Balances and Reaction Models, in: C.D. Smolke (Ed.), Metab. Pathw. Eng. Handb. Fundam., CRC Press, Boca Raton, 2009: pp. II-1-11–20.

[97] R.C. Kolfschoten, M.E. Bruins, J.P.M. Sanders, Opportunities for small-scale biorefinery for production of sugar and ethanol in the Netherlands, Biofuels, Bioprod. Biorefining. 8 (2014) 475–486. https://doi.org/10.1002/bbb.1487.

[98] H.J. Huang, S. Ramaswamy, U.W. Tschirner, B. V. Ramarao, A review of separation technologies in current and future biorefineries, Sep. Purif. Technol. 62 (2008) 1–21. https://doi.org/10.1016/j.seppur.2007.12.011.

[99] D.J. Kirwan, C.J. Orella, Crystallization in the pharmaceutical and bioprocessing industries, in: Handb. Ind. Cryst., Elsevier, 2002: pp. 249–266. https://doi.org/10.1016/b978-075067012-8/50013-6.

[100] G. Sin, K. V. Gernaey, A.E. Lantz, Good modeling practice for PAT applications: Propagation of input uncertainty and sensitivity analysis, Biotechnol. Prog. 25 (2009) 1043–1053. https://doi.org/10.1002/btpr.166.

[101] H.W. Coleman, W.G. Steele, Experimentation, Validation, and Uncertainty Analysis for Engineers: Third Edition, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2009. https://doi.org/10.1002/9780470485682.

[102] J.M. Hammersley, D.C. Handscomb, Monte Carlo Methods, Springer Netherlands, 1964. https://doi.org/10.1007/978-94-009-5819-7.

[103] J.C. Helton, Treatment of Uncertainty in Performance Assessments for Complex Systems, Risk Anal. 14 (1994) 483–511. https://doi.org/10.1111/j.1539-6924.1994.tb00266.x.

[104] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global Sensitivity Analysis. The Primer, John Wiley & Sons, Ltd, Chichester, UK, 2008. https://doi.org/10.1002/9780470725184.

[105] I.M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simul. 55 (2001) 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6.

[106] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, Comput. Phys. Commun. 181 (2010) 259–270. https://doi.org/10.1016/j.cpc.2009.09.018.

Paper

C

Conceptual Process Design of an Integrated Xylitol Biorefinery with Value-Added Co-Products

Nikolaus I. Vollmer, Krist V. Gernaey, Gürkan Sin,

Frontiers in Chemical Engineering,

Accepted/in press, as of 26th of January 2022

Abstract

This manuscript describes the conceptual process design of an integrated xylitol biorefinery with value-added co-products. Based on an existing three-step framework, the main product of a second-generation integrated biorefinery is chosen in the first stage. Based upon this, other decisions as the feedstock and value-added co-products are made. All relevant unit operations for the process are introduced. An initial superstructure with all potential process alternatives is composed of all introduced models. In the second step of the framework, a global sensitivity analysis is performed, firstly with coarse sampling to determine all viable flowsheet options and secondly with fine sampling to determine the most sensitive operational variables. As a result of the sensitivity analysis, most of the flowsheet options in the initial superstructure are not feasible. Based on these results, flowsheet sampling with the five most sensitive operational variables is performed to create surrogate models. In the scope of this work, three types of surrogate models are benchmarked against each other. Regarding the results of the superstructure optimization, firstly, it becomes apparent that the production of biokerosene does not contribute significantly to the net present value of the biorefinery. Furthermore, reducing the number of unit operations in the downstream processing leads to lower capital expenditures, but it lowers the product yield. Lastly, most flowsheets are economically feasible, indicated by a positive net present value. Based on this result, the most promising candidate process topology is subjected to the third step of the framework, including uncertainty in capital expenditure and operational expenses according to their estimations and uncertainties in the product prices. As a result, the net present value of the flowsheet turns negative, indicating that the high uncertainties for the expenditure and the expenses do not allow for an economically feasible operation. Lastly, the analysis of conceptually designed process flowsheets based on Monte Carlo sampling shows failure rates, with the NPV falling below the break-even point, of around 60% probability or higher. Based on these results, an economically feasible construction and operation of a xylitol biorefinery seems unlikely. Further ways to improve the metrics are elucidated.

C.1 Introduction

Biorefineries are considered an essential part of a strategy towards more sustainable production patterns of fuels and chemicals as demanded by the 2030 Sustainability Agenda of the United Nations [1]. In general, a biorefinery is a production concept for the conversion of biomass into different products, e.g., xylitol [2]. In this regard, a multitude of realizations of this biorefinery concept have been investigated and published in the past decades. As diverse as the realizations are, the classification of biorefinery setups is not standardized and varies depending on the type and generation of utilized feedstock, the utilized conversion technologies, the obtained product type and number, and the integration level [2–4]. Nonetheless, all of the elaborated concepts have in common that their sustainability potential is compelling, but their economic viability is generally described as challenging to achieve [3].

Xylitol as a product has gained attraction throughout research for several decades. Xylitol is an excellent sugar substitute with many beneficial health properties, as around 40% fewer calories than sucrose, anticariogenic properties, and a low glycemic index, making it perfectly suitable for diabetic nutrition [5]. Furthermore, several studies indicate a potential use as building bock chemical [6,7]. Due to the high interest and potentially high product prices, the US Department of Energy declared xylitol one of the top 12 chemicals to be produced in a biorefinery already in 2004 [8].

To this end, all major xylitol producers employ a chemical production process with the hemicellulosic fraction of lignocellulosic biomass as feedstock. According to Delgado Arcaño et al. (2020), the production process consists of four steps, namely 1) the biomass pretreatment, 2) the purification of the obtained xylose, 3) the chemical conversion of xylose to xylitol, and 4) the purification of the produced xylitol. Both the purification steps and the temperature and pressure conditions required for the conversion process induce high costs, which explains the comparatively high product price of xylitol [7].

An alternative production route for xylitol is the fermentation with suitable organisms in a biotechnological process. A significant number of research publications on the biotechnological production with either wild-type or genetically modified microorganisms exist. However, all of them do point out that there is no consensus on any economically viable full-scale production process, and further conceptual research needs to be performed on this, primarily focusing on an efficient pretreatment technology, robust cell factories, and an optimized downstream process [7,9–11]. Both the chemical production route and the biotechnological production route are found on similar pretreatment unit operations, given the fact that also in the biotechnological process, the feedstock is lignocellulosic biomass. However, a significant advantage of the biotechnological route is the resilience towards impurities, reducing the effort to purify the hemicellulosic hydrolysate from the pretreatment [7].

Due to these considerations, this work focuses on a multi-product second-generation integrated biorefinery, referring to lignocellulosic biomass as second-generation feedstock. Both the approach of producing multiple products from biomass to utilize the feedstock to a maximum amount and the optimization of heat and mass integration are prone to augment

the economic viability, which makes this concept most promising both regarding the sustainability and the economic aspect [3,4]. Nonetheless, the question of which products to co-produce in the biorefinery setup, which feedstock to utilize, and how to introduce efficient ways to integrate the process remains challenging. Hence, this demands a conceptual design approach to ensure an optimal process design of the biorefinery to answer the conceptional underlying question regarding the economic feasibility of such biorefinery concepts under given technological and economic conditions.

Concerning conceptual process design, there exist several schools of thinking. In the S3O framework, a synergistic approach is followed to best integrate expert knowledge, optimization approaches, and simulation-based strategies to leverage synergies for the conceptual design of bioprocesses in three steps [12]. The framework is applied for the case study of designing a xylitol biorefinery with value-added co-products in this manuscript. In the first step, the main product (xylitol) and value-added co-products (succinic acid, biokerosene, heat) are selected after thorough reasoning, together with the potential feedstock (wheat straw) and potential unit operations that are suitable for the biotechnological production of xylitol. In the second step, key performance indicators, e.g., the net present value (NPV) of the plant, are used as the objective function to maximize the economic potential of the plant by finding the most suitable operational conditions and process configuration. Lastly, in the third step, the found process and operational conditions are optimized under uncertainty to consolidate the process design. Based on the results of the second and third step of the framework, a techno-economic analysis (TEA) is performed. The TEA involves Monte Carlo-based uncertainty analysis to quantify the influence of different economic and operational factors on the economic feasibility of the process. The major novelty of this work in comparison to similar studies is the integration of mechanistic models via the synergistic optimization-based framework for the conceptual process design and the use of detailed data for the equipment to obtain a realistic estimate of the capital expenditures and operational expenses and detailed market prices for the techno-economic analysis of the biotechnological production of xylitol [13–15].

The remainder of the paper is structured as follows: theoretical background is provided about the product xylitol and its production process in a potential biorefinery setup, followed by an in-depth description of all unit operations and their respective mechanistic model, which is used in the process design framework. Afterward, the structure and functionalities of the framework are elucidated in detail and provide additional considerations regarding the costing and sizing of equipment. Lastly, the specific calculations regarding the techno-economic and the sustainability analysis are introduced. In section 3, the results are presented. In section 4, conclusions are drawn, and an outlook to future research is given.

C.2 Materials and Methods

C.2.1 Feedstock & Products

C.2.1.1 Feedstock

In general, lignocellulosic biomass consists of three main fractions, namely hemicellulose, cellulose, and lignin. All three possess a polymer structure and are composed of typical monomers: Hemicellulose consists mainly of pentose sugars, mostly xylose, and smaller amounts of arabinose [16]. Cellulose consists mainly of glucose monomers, while lignin is a mostly heterogeneous macromolecule consisting of the three monolignols p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. Their fractions largely depend on the type of feedstock, and theoretically, all of them can be converted to different products. Wheat straw, as the considered feedstock in this case study, has a comparatively high fraction of hemicellulose sugars, which indicates a potentially high yield of xylose monomers.

C.2.1.2 Xylitol

Xylitol ($C_5H_{12}O_5$) is the sugar alcohol of xylose, an aldopentose, and naturally occurs in woods and other crop plants [9]. It is highly soluble in water and has a negative heat of solution. As mentioned in section C.1, the increasing interest in xylitol is also reflected by the growing market size of xylitol. With a market size of 670 Mio USD at 161.5 mio MT produced in 2013, this is supposed to grow to a volume of 1.15 Bio USD at 266.5 mio MT in around 2023 at a market price of 4.5-8.1 USD/kg [6,7]. Main producers for the global market are Danisco DuPont (DK, FI), Futaste (CN), Cargill (US) and Mitsubishi Corporation (JP) amongst others [7].

In the chemical production, as mentioned in section C.1, acid hydrolysis is performed for the biomass pretreatment, which mainly fractionates and depolymerizes the hemicellulosic fraction of the lignocellulosic biomass. The obtained hydrolysate consists primarily of xylose monomers, as they constitute the major part of the hemicellulose. The following xylose purification step is necessary due to the formation of various inhibitory substances, e.g., 5-hydroxymethylfurfural and furfural, which otherwise would lead to rapid deactivation of the catalyst in the conversion process. Subsequently, the xylose is converted to xylitol under the presence of hydrogen on a metallic catalyst, classically Raney nickel, with high yields and conversion rates. Lastly, the produced xylitol is purified in several steps [6].

In the biotechnological production, for the employed cell factories in the fermentation, a plethora of different microorganisms has been studied regarding their potential of producing xylitol at a favorable yield, productivity, and titer. Yeast strains are primarily in focus due to their natural ability to assimilate xylose and convert it into xylitol. In particular, *Candida species*, *Debaromyces hansenii,* and *Kluyveromyces marxianus* are suitable cell factories with yields of xylitol from xylose around $Y_{Xyl,Xyo} = 0.5 - 0.6$, volumetric productivities of $q_{Xyo} = 0.2 - 5\ g \cdot L^{-1} \cdot h^{-1}$ and titers of $c_{Xyo} = 10 - 110\ g \cdot L^{-1}$ [11]. Figure 1 shows the pentose assimilation pathway in yeasts for the specific case of xylose with the corresponding co-factors, according to Albuquerque et al. (2014).

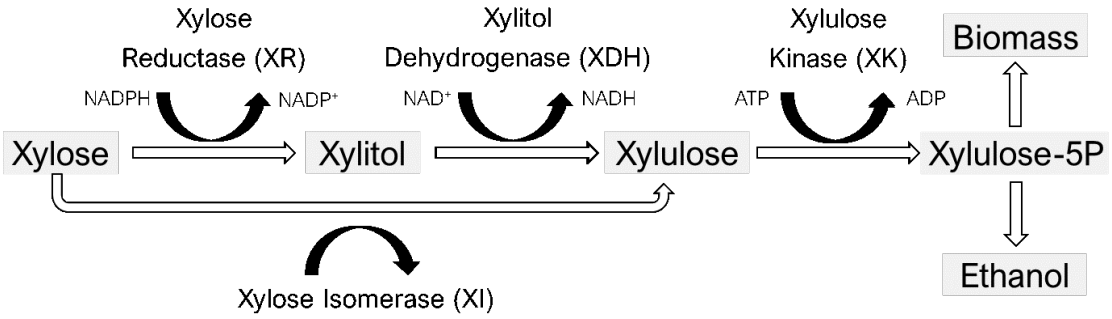

Figure 17: Xylose assimiliation pathways in yeast [9].

As illustrated, xylose is assimilated through the cell wall and converted in the first step to xylitol. Xylitol is then further converted and ultimately ends up in the anabolism to form biomass or converted via the pentose phosphate pathway into metabolites participating in the glycolysis and ending up as pyruvate [9]. In order to maximize the operational yield for xylitol in a fermentation process, the conditions of the process have to be adjusted accordingly to prevent the consumption of xylose for further process steps. A key parameter here is the aeration of the process: several studies have shown that aerobic conditions with a low oxygen availability are advantageous with regards to the availability of co-factors, as shown in Figure 1 [9,10]. Furthermore, the availability of co-substrate as, e.g., glucose at lower concentrations is beneficial for the productivity of the cell factory [7,10].

Those and other prominent candidates as *Saccharomyces cerevisiae* have been subjected to metabolic and genetic engineering strategies to increase productivity and titer for xylitol production [7,10,11]. The main targets are the overexpression of xylose reductase, the suppression of xylitol dehydrogenase, the knockout of the genes for xylose isomerase and xylitol dehydrogenase, as well as more advanced engineering strategies to up- and downregulate the availability or completely shifting co-factors [7,10,11]. Furthermore, evolution techniques have been successfully applied to increase the tolerance to inhibitory components [7]. The performance metrics of these engineered cell factories are yields of xylitol from xylose around $Y_{Xyl,Xyo} = 0.6 - 1.0$, volumetric productivities of $q_{Xyo} = 0.3 - 3.2\ g \cdot L^{-1} \cdot h^{-1}$ and titers of $c_{Xyo} = 20 - 120\ g \cdot L^{-1}$ [7,11]. Overall, this leads to the conclusion that engineered cell factories can significantly contribute to the economic potential of a biotechnological production process of xylitol.

C.2.1.3 Value-Added Co-Products

- Succinic Acid

Succinic acid, also identified as one of the top 12 biobased chemicals by the US Department of Energy, has potential use as a platform chemical [8]. What makes the biotechnological production particularly attractive from a sustainability point of view is the possibility of CO_2 capture in the process, as the process stoichiometry requires a net CO_2 uptake [17]. Succinate is a metabolite in the citric acid cycle and naturally occurs in most microorganisms. The production of succinic acid with cell factories has been investigated thoroughly: the most promising cell factories for production are, amongst others, *Actinobacillus succinogenes* and *Mannheimia succiniciproducens* [17]. Also, engineered cell factories like Escherichia coli show favorable yield, productivity, and titer [17]. A problem

with the biotechnological production is the formation of other organic acids and the induced product inhibition, which can impair the performance of a fermenter and a complex downstream process [17].

- Biokerosene

As lignin consists of a vast mixture of differently polymerized aromatic compounds, several products can be obtained from it, e.g., carbon fibers, thermoplastic and elastomeric polymers, and other fuels and chemicals [18]. Out of this product palette, the research on the production of fuels and chemicals has yielded several potential production strategies with an acceptable technology readiness level that allows for a potential commercial production [18]. The aviation industry currently focuses heavily on improving the sustainability in commercial aviation, of which one approach is the substitution of commercial kerosene with biokerosene from renewable resources [19]. Fuels produced from lignin or lignocellulosic biomass in general naturally contain a high amount of aromatic components, which is particularly important for the properties of aviation fuel [19]. Hence the production of biokerosene from lignin is a potentially suitable valorization strategy [19].

- Heat

Alternatively, the lignin fraction can also be valorized by targeting its high energy content. By combusting the lignin directly, heat can be created, which can be converted into other forms of energy, e.g., steam or electricity [18].

C.2.2 Unit Operations

The following section describes each unit operation that is employed in a possible xylitol biorefinery setup for the production of xylitol and value-added co-products. Furthermore, the respectively developed mechanistic model describing the underlying physical, chemical, and biological phenomena is elucidated for each unit operation. For all model files, the reader is referred to the GitHub repository of this manuscript, where all model and simulation files are provided [20].

All models described in this section are implemented in MATLAB. For the evaporation, the models implemented in ASPEN Plus are used and interfaced with MATLAB through a COM interface. The S3O framework is also implemented in MATLAB. The mechanistic models are all fitted to either proprietary experimental data or data retrieved from literature. The models are all validated within their boundaries for their mass and energy balances, kinetics, and thermodynamic assumptions. They are implemented in the S3O framework and assessed regarding their prediction robustness through an uncertainty and sensitivity analysis. Further information regarding the models and their implementation can be found in the original manuscript describing the S3O framework [12].

C.2.2.1 Pretreatment

For both the chemical and the biotechnological production process of xylitol, the biomass pretreatment as the first unit operation has vital importance for the whole process. With xylitol as the main product, the priority for the pretreatment lies in the fractionation and

depolymerization of the hemicellulosic fraction. The most prominent pretreatment technology for this task are dilute acid, autothermal hydrolysis, and steam explosion pretreatment [21]. Previous studies have shown the excellent metrics of dilute acid pretreatment on wheat straw in terms of high monomer yield, good fractionation, and acceptably low inhibitor formation [16].

Vollmer et al. (2022) developed a mechanistic model describing the pretreatment with mass and energy balances for all major occurring components. The kinetics of the pretreatment are described with first-order reaction equations. For a detailed overview of the work and the specific model equations, the reader is referred to the manuscript and the supplementary material for the model equations and the parameter values [16]. In this manuscript, the pretreatment unit is abbreviated as PT.

C.2.2.2 Enzymatic Hydrolysis

The residue from the pretreatment containing the cellulosic and lignin fraction needs to be processed further to fractionate and depolymerize the cellulosic fraction into glucose monomers. Amongst possibly employable technologies, most biorefinery concepts rely on an enzymatic hydrolysis process for this task [22]. The utilized enzymes are a cocktail of different glucanases, glucosidases, and hydrolases, which efficiently break down the cellulose into cellobiose and finally into glucose but can be inhibited by present sugars like glucose and xylose [23,24].

Kadam et al. (2004) proposed a mechanistic model describing the enzymatic hydrolysis with mass and energy balances for cellulose, cellobiose, glucose, and xylose; the kinetics of the enzymatic hydrolysis are described with Michaelis-Menten kinetics [22]. Prunescu et al. (2013) reviewed the model parameters and updated them accordingly to new commercially available enzymes [24]. For a detailed overview of the work and the specific model equations, the reader is referred to both manuscripts and the supplementary material for the model equations and the parameter values. In this manuscript, the enzymatic hydrolysis unit is abbreviated as EH.

C.2.2.3 Fermentation

Both the xylose-rich stream of the pretreatment and the glucose-rich stream of the enzymatic hydrolysis serve as the substrate for both fermentation steps. Given the elaborations in sections C.1 and C.2.1, the employed cell factory for the fermentation of xylitol is a *Candida mogii* strain, and the employed cell factory for the fermentation of succinic acid is a *Mannheimia succiniciproducens* strain.

The developed mechanistic models are both black-box models with mass and energy balances for all primary components, and the kinetics are set up as described by Heijnen and van Gulik (2009) with substrate uptake rates, Herbert-Pirt distribution relations, and product formation rates [25]. The used data for the parameter estimation of the xylitol fermentation model derives from a paper published by [26]. The used data for the parameter estimation of the succinic acid fermentation model derives from a paper published by [27]. Both models allow for the simulation of batch and fed-batch processes. In this manuscript, the fermentation unit for the production of xylitol is abbreviated as FX, the unit for the production of succinic acid is abbreviated as FS.

C.2.2.4 Evaporation

After the pretreatment, after the enzymatic hydrolysis, and after the fermentations, the process streams contain a certain concentration of the product of interest. For all cases, it can be of interest to increase the concentrations of the stream to achieve higher titers in the fermentation and to remove certain inhibitory compounds. Potential technologies for this unit operation are either evaporation units in various forms or membrane units, with the former ones being widely commercially available and thus easier to implement but having very high energy demands [28]. Mechanistic models for simulating evaporation processes are readily available in different commercial process simulators, e.g., ASPEN Plus. In this manuscript, the upconcentration unit for the hemicellulose hydrolyzate is abbreviated as UH. The unit for the cellulose hydrolyzate is abbreviated as UC, the evaporation unit for the downstream processing of xylitol is abbreviated as EX, the unit for the downstream processing of succinic acid is abbreviated as ES.

C.2.2.5 Crystallization

Xylitol and succinic acid are soluble in water and solid at room temperature. Hence, crystallization is a suitable unit operation to separate and purify those two substances from aqueous streams. Crystallization is performed as either heating, cooling, pH, or antisolvent crystallization [29]. For the xylitol, both a cooling crystallization and an antisolvent crystallization with ethanol are considered, as the high xylitol solubility in water decreases in the presence of ethanol [30]. For succinic acid, a cooling crystallization is considered.
The developed mechanistic model is based on the work of [31,32]. The crystallization is described with mass and energy balances. The kinetics are described with a population balance, describing the nucleation and growth of crystals. The population balance is solved with the method of moments. The solubility and kinetic data for the xylitol crystallization is obtained from experiments, the kinetic data for the succinic acid crystallization is obtained from the literature [33–36]. For a detailed overview of the work and the specific model equations, the reader is referred to both manuscripts and the supplementary material for the model equations and the parameter values. In this manuscript, the crystallization units for the downstream processing of xylitol are abbreviated as CXi, the units for the downstream processing of succinic acid are abbreviated as CSi.

C.2.2.6 Lignin Pyrolysis

After the fractionation and depolymerization of the cellulosic fraction in the enzymatic hydrolysis, the remaining solid fraction consists primarily of lignin. As described in section C.2.1.1, the lignin fraction also consists of monomers, which are supposed to be processed furtherly to sustainable aviation fuel. The first unit operation for this process is a pyrolysis step to break down the lignin's macromolecular structure and create bio-oil as a precursor for the fuel [37]. Fast pyrolysis with time ranges between $t = 1 - 100s$ prove to have the highest yield of liquid compounds and lower yields for gaseous compounds and char, which are side products in the pyrolysis process [38].

The used mechanistic model is based on the work of [39]. The mechanistic model is based on mass and energy balances. First-order reaction equations describe the kinetic behavior. The model describes the degradation of three lignin monomer structures commonly present in different fractions in the biomass into over 50 different products. All parameters for the model are provided in the original publication [39]. For a detailed overview of the

work and the specific model equations, the reader is referred to the original manuscript and the supplementary material for the model equations and the parameter values. In this manuscript, the lignin pyrolysis unit is abbreviated as LP.

C.2.2.7 Lignin Hydrotreatment

The produced bio-oil in the lignin pyrolysis classically contains a comparatively high amount of oxygen and is thus not directly usable as aviation fuel [37]. In order to reduce the oxygen amount, several options have been investigated in research, of which catalytic hydrotreatment is one of the most promising ones [40]. The occurring reactions are quite complex, which is why a lumped kinetic network is introduced that involves eight components [41]. All the components which are formed during the pyrolysis are assigned to one of the lumped components of the hydrotreatment. The reaction itself occurs over a zeolite catalyst and the addition of hydrogen.

The employed and adapted mechanistic model is based on [41], involving mass and energy balances. Reaction-convection-diffusion equations describe the kinetic behavior. All parameters for the model were reestimated based on the kinetic data. For the parameters and a detailed model description, the reader is referred to the original manuscript and the supplementary material. In this manuscript, the lignin hydrotreatment unit is abbreviated as LH.

C.2.2.8 Lignin Fractionation

Lastly, after the hydrotreatment, the present fractions in the stream must be separated. Besides the two products, these are the fraction with gaseous components, phenolics, aromatics, and water. In a two-step separation with two flash drums, first, the gaseous components are removed, and subsequently, the phenolics and aromatics are separated from the water. Depending on the created fractions, a fractionated distillation can also be used as a unit operation. Mechanistic models for simulating fractionation processes are readily available in different commercial process simulators, e.g., ASPEN Plus. In this manuscript, the lignin fractionation unit is abbreviated as LF.

C.2.2.9 Auxiliary Unit Operations

Besides all the listed primary unit operations, the biorefinery setup requires several auxiliary unit operations for full functionality. These are listed shortly in the following section:

- Feedstock Processing:

To process the lignocellulosic biomass, it has to be transported into the process through conveyor belts, milled, and stored before actually entering the pretreatment unit operation. In this manuscript, the feedstock processing unit is abbreviated as Feed.

- Product Storage:

In order to process the produced xylitol, the value-added co-products, and other intermediates, storage capacity is installed to buffer seven days of production, involving solid storages for the xylitol and succinic acid, and liquid storage for the kerosene and

storage for all other chemicals and intermediate steps involved in the process. In this manuscript, the storage units are abbreviated as Store.

- Wastewater treatment:

In order to treat all aqueous effluents from the process, a wastewater treatment plant is considered to be installed to comply with legislative constraints and also to regenerate freshwater for the process in the form of material flows for unit operations or in the form of material flows as cooling water. In this manuscript, the wastewater processing unit is abbreviated as WWT.

- Combustion, power, and steam generation:

Lastly, to recover the organic residues of the process, these gaseous and solid streams are combusted to generate steam and power in a turbine. This process has a significant impact on the overall economics by increasing the biorefinery integration and lowering operational expenditures at the cost of capital investment. In this manuscript, the steam generation unit is abbreviated as Steam.

All the listed auxiliary unit operations are not mechanistically modeled. Instead, based on a report of the National Renewable Energy Laboratory (NREL) on the process design of an ethanol biorefinery, all the ingoing mass and energy streams in the xylitol biorefinery are scaled to the ethanol biorefinery in the report to estimate the outgoing mass and energy streams and the costs [42]. This induces uncertainty in the following analyses, which is why these decisions will be evaluated later on through uncertainty and sensitivity analysis in the techno-economic analysis.

Furthermore, from a practical point of view, additional unit operations in the process can be required. This is related to the feedstock and the production of inhibitory compounds in the pretreatment. Besides the considered major inhibitory compounds, further components are formed in minor amounts that can require, e.g., an adsorption process for their removal [43]. Also, the lignin itself imposes challenges to the process due to the potential of clogging of equipment [44]. This possible requirement of additional unit operation imposes another source of uncertainty, which will be equally addressed in the techno-economic analysis. A detailed description of the calculations can be found in the supplementary material.

C.2.3 Process Design

C.2.3.1 Process Synthesis and Design Framework

As described by Vollmer et al. (2022), the framework consists of three sequential steps: 1) the selection of products, feedstock, and processes, 2) superstructure optimization, and 3) simulation-based optimization. The overall procedure is established with the idea of "having the end in mind." Hence, in the first step, based on expert knowledge, a meaningful selection of products is made and a feedstock from which these products are supposed to be produced. Subsequently, potential process steps are evaluated and composed to a process superstructure in a bottom-up fashion. This aims at keeping the potential number of process realizations low. In the second step, this superstructure is optimized through mathematical optimization to obtain several candidate process topologies. In order to include the different mechanistic models in the superstructure, surrogate models are used

based on different machine learning technologies. For a successful application of these surrogate models, a proper benchmark of different alternatives is vital [45]. After determining the reduced set of candidate process topologies, in the third step, all of these are subjected to simulation-based optimization to consolidate the process design and find a truly optimal solution [12].

C.2.3.2 Sizing of the Equipment

For using the mechanistic models for process design, the overall plant capacity used for the design calculation and the corresponding mass and energy in- and outflows entail a specific capacity for each unit operation. In general, the volumetric capacity of a unit operation V can be calculated as indicated in the following equation:

$$V = \frac{\dot{m}}{\bar{\rho} \cdot \tau}, \tag{C.1}$$

with the hourly capacity $\dot{m}$, the average density of the process medium in the unit operation $\bar{\rho}$ and the residence time τ.

C.2.3.3 Capital Expenditures and Operational Expenses

Based on each unit operation's volumetric or mass-based capacity, the fixed capital investment for all unit operations in the plant can be calculated. Whenever possible, costing is based on the NREL Report on "Process design and economics for conversion of lignocellulosic biomass to ethanol," as the report is based on actual quotations from different equipment manufacturers with a high level of accuracy and detailedness [42]. As both biorefinery setups differ to a certain extent, all fixed capital investment, which is not available in the NREL report, is estimated through a cost estimation tool [46]. All costs retrieved are extrapolated to the capacity of the planned biorefinery by the plant capacity ratio method as follows:

$$\frac{C}{C_0} = \left(\frac{\dot{m}}{\dot{m}_0}\right)^x, \tag{C.2}$$

With C as the capital cost of the unit operation, x as the scaling factor, and all zero-indexed variables referring to the original capacity of the reference plant. The fixed capital investment of the reference plant is adjusted for inflation by multiplying with an average term for the inflation between the current date and the date of the NREL report or the reference date of the costing tool:

$$C_0 = C_{00} \cdot (1 + \varphi_i)^n, \tag{C.3}$$

With C_{00} being the original capital cost of the unit operation, φ_i being the averaged inflation rate and n being the year difference. With the total cost of all defined areas being determined, the total capital investment TCI can be calculated based on the fixed capital investment FCI, and the total direct and indirect costs TDC and TIC. The total direct costs are determined as follows:

$$TDC = 1.0788 \cdot \sum C, \tag{C.4}$$

Being 7.88% higher as the sum of capital costs for all unit operations. The total indirect costs are determined as follows:

$$TIC = 0.6 \cdot TDC. \tag{C.5}$$

Amounting to 60% of the total direct costs. The fixed capital investment is defined as the sum of total direct and indirect costs:

$$FCI = TDC + TIC, \tag{C.6}$$

And lastly, the total capital investment corresponds to:

$$TCI = 1.0547 \cdot FCI, \tag{C.7}$$

Being 5.47% higher than the fixed capital investment and defining the capital expenditures of the biorefinery. The operational expenses of the biorefinery are defined by the total production costs TPC. They consist of fixed and variable operational costs FOC and VOC. The fixed operational costs, consisting mainly of salaries and other fixed payments, are determined with the plant capacity ratio based on the reference plant. The variable operational costs consist mainly of prices for utilities, so all costs are created by the acquisition of chemicals, feedstock, and energy. The total production costs are then defined as follows:

$$TPC = FPC + VOC, \tag{C.8}$$

Being the sum of both operational costs. Lastly, the sales of the products are determined by summing up all individual sales of each product:

$$Sales = \sum_i p_i \cdot m_i, \tag{C.9}$$

With m_i being the produced mass of product i per year and p_i being the price of product i per mass unit.

C.2.4 Techno-Economic Analysis

C.2.4.1 Key Performance Indicators

Ultimately, the economic feasibility of a biorefinery can be evaluated based on capital expenditure and operational expenses by calculating different key performance indicators (KPI). According to Peters et al. (2002), they are calculated as follows: The first KPI considered in this study is the return on investment ROI, evaluating the profitability of the invested capital. It is calculated as follows:

$$ROI = \frac{Sales - TPC}{TCI}. \tag{C.10}$$

The ROI is compared against a set threshold, the so-called minimum acceptable rate of return φ_{mar} and given the case that $ROI > \varphi_{mar}$, the investment is found profitable. The used minimum acceptable rate for the xylitol biorefinery is set to be $\varphi_{mar} = 10\,\%$ [42]. Another way of analyzing the profitability concisely is the calculation of the payback period PBP, indicating after how many years the fixed capital investment is earned back. It is calculated in the following way:

$$PBP = \frac{FCI}{(Sales - TPC) \cdot (1 + \phi) + \phi \cdot \bar{d}}, \tag{C.11}$$

With ϕ being the income tax rate (in this work $\phi = 35\,\%$) and $\bar{d}$ being the average depreciation per year over the plant lifetime. However, both presented KPIs do not consider

the time value of money, which refers to the idea that the value of the money, which is bound in capital and operational expenses, would otherwise increase if invested differently. One KPI which incorporates this is the net present value NPV. It is calculated as follows:

$$NPV = \sum_{i=1}^{y} (1 + \varphi_{mar})^{-i} \cdot \left((Sales - TPC - d_i \cdot FCI) \cdot (1 + \phi) + rec_i + d_i \cdot FCI\right) - \sum_{i=-by}^{y} (1 + \varphi_{mar})^{-i} \cdot TCI_i, \quad \text{(C.12)}$$

With y being the biorefinery lifetime and by being the construction period of the biorefinery. Additionally d_i corresponds to the depreciation of the plant according to the depreciation scheme per year, rec_i are recovery costs from materials and land sales commonly at the end of the plant lifetime and TCI_i indicating the amount of the total TCI being invested in each year in the building period. As a depreciation scheme for the xylitol biorefinery, the MACRS5 scheme is used and a plant life period of 30 years and a building period of two years. The NPV is sought to be positive for an investment to be profitable. Analogously, a discounted cash flow of return $DCFR$ can be calculated, referring to the rate of return where the NPV turns exactly zero:

$$0 = \sum_{i=1}^{y} (1 + DCFR)^{-i} \cdot \left((Sales - TPC - d_i \cdot FCI) \cdot (1 + \phi) + rec_i + d_i \cdot FCI\right) - \sum_{i=-by}^{y} (1 + DCFR)^{-i} \cdot TCI_i. \quad \text{(C.13)}$$

Lastly, the calculation of the $DCFR$ can also be used for calculating a minimum selling price of product $MSEP\ p_i^*$, by fixing the $DCFR$ at the φ_{mar} and instead varying the product price p_i^*:

$$0 = \sum_{i=1}^{y} (1 + \varphi_{mar})^{-i} \cdot \left(\left(\sum_{i \in P} p_i^* \cdot m_i - TPC - d_i \cdot FCI\right) \cdot (1 + \phi) + rec_i + d_i \cdot FCI\right) - \sum_{i=-by}^{y} (1 + \varphi_{mar})^{-i} \cdot TCI_i. \quad \text{(C.14)}$$

On a side note, only one variable can be left free to be optimized to keep this equation solvable. In this case, for xylitol as the primary product, the product price of xylitol is chosen instead of, e.g., a multiplier for all product prices. [46].

C.2.4.2 Economic Risk Analysis

As mentioned earlier, the design of any biotechnological process inherently incorporates uncertainty of various sources. These potentially can be assumptions about yields and productivities, estimates about capacity and capital investment, factors that influence the

scale-up of a process, or external factors like price fluctuations. To assess the impact of uncertainties on the economic feasibility, Monte Carlo-Based Methods are a suitable option. In this case, assessing the effects of the uncertainties on the predicted key performance indicators, an uncertainty analysis is performed as described in [47].

C.3 Results

C.3.1 Selection of Products, Feedstock and Unit Operations

The main product for the biorefinery concept in this study is xylitol. In light of the explanations in sections C.1 and C.2.1.2, an integrated xylitol biorefinery with lignocellulosic biomass as feedstock will utilize the hemicellulosic fraction of the lignocellulosic biomass for the production of xylitol. This leaves the cellulosic and the lignin fraction as possible substrates for value-added co-products in a multi-product biorefinery. The cellulosic fraction consists mainly of glucose monomers, whereas the lignin fraction is an amorphous macromolecule with different aromatic monomers. As discussed in section C.2.1.3, potential value-added co-products, in this case, can be succinic acid, biokerosene, or heat, with the prior one being a product for the cellulosic fraction and the latter two a product for the lignin fraction.

Regarding a potential feedstock, all three fractions should be represented to a sufficient amount to produce all four products possibly. With xylitol being the product with the highest product price, the potential feedstock favorably has a high hemicellulosic fraction. This is the case for most agricultural residues, e.g., wheat straw, which is hence selected for this case study. As plant capacity, an amount of $m_{feedstock} = 150{,}000\ t/a$ of wheat straw is specified. The feedstock selection and composition are based on prior work [16]. The capacity for the biorefinery is based on both the NREL Report regarding the ethanol biorefinery and production data from commercial xylitol producers [7,42]. Regarding the processing units, all relevant unit operations are described in section C.2.2, considering only possibilities with a high technology readiness level to retrieve a realistic process design for the xylitol biorefinery. All possible process routes, which form the superstructure for step 2 of the framework, are displayed in Figure 18.

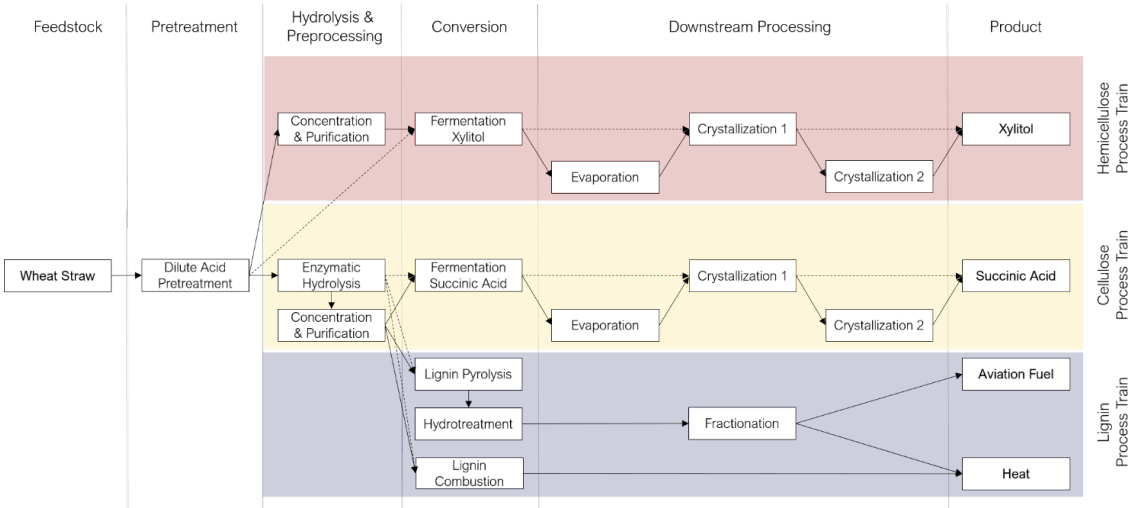

Figure 18: Overview over the postulated superstructure for the xylitol biorefinery in the S3O framework.

As can be seen, the potential options in the superstructure involve the inclusion or exclusion of an upconcentration unit, an evaporation unit, and a second crystallization unit for the xylitol process train, based on previously obtained results [48]. The same superstructure is assumed for the succinic acid process train, also agreeing with other proposed process designs [49]. In the lignin process train, the two possible options are to convert the lignin to biofuel and use the residues of this and other process streams for combustion or to use the lignin entirely for combustion to increase the potential for heat integration in the process.

C.3.2 Design Space Exploration

The outcome of the design space exploration in the scope of this case study is twofold. Firstly, it is used to reduce the size of the superstructure and exclude options that a priori are infeasible due to nontrivial design constraints. It serves to determine the sensitivity of the operational variables to prioritize them for the SSO.

The design space exploration itself is performed in two stages. In the first stage, coarse sampling with $N = 100$ with all operational variables for each flowsheet option is performed. For each sample point, it is analyzed whether the conditions allow for the production of xylitol and succinic acid. The results for all flowsheets are displayed in Figure 19.

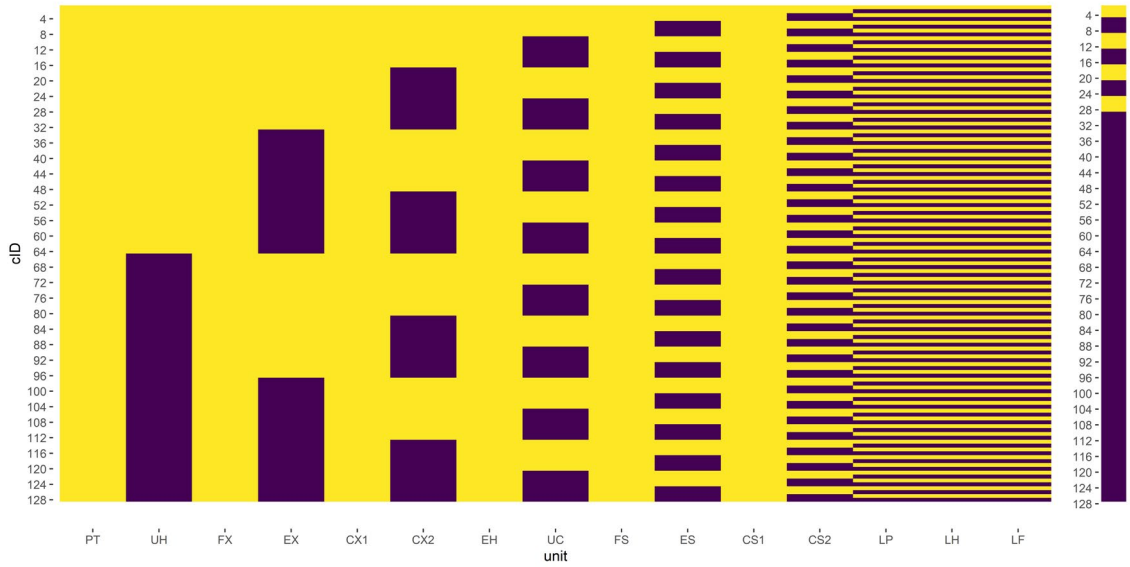

Figure 19: Feasibility matrix of the coarse sampling, displaying the existence (yellow) or absence (blue) of each unit operation in each configuration ID, with the last column indicating the feasibility (yellow) or infeasibility (blue) of the flowsheet.

As illustrated, most flowsheet options actually do not allow for the production of xylitol and succinic acid as all sample points yield infeasible solutions. More precisely, all flowsheet options that do not include either the upconcentration step for the hemicellulose process train (UH) or the evaporation step of the hemicellulose train (EX), or the evaporation step of the cellulose train (ES). In reverse, this indicates that these unit operations are seemingly compulsory for any xylitol biorefinery setup. The cause behind this is the necessary substrate concentration for the xylitol fermentation and the high dilution of the fermentation broth, paired with the high solubility of xylitol. Consequently, this already

indicates a high necessary amount of heat in the downstream process, which will impact the operational expenses of the process.
In the second stage of the design space exploration, fine sampling with $N = 2000$ with all operational variables for all feasible flowsheet options is performed. The input for the sensitivity analysis are the operational variables of each configuration ID, and the output is the calculated NPV. The input and the output from the model are used in the easyGSA toolbox to determine the first-order and total sensitivity index of each operational variable by a neural network-assisted global sensitivity analysis. The results are presented in Figure 20.

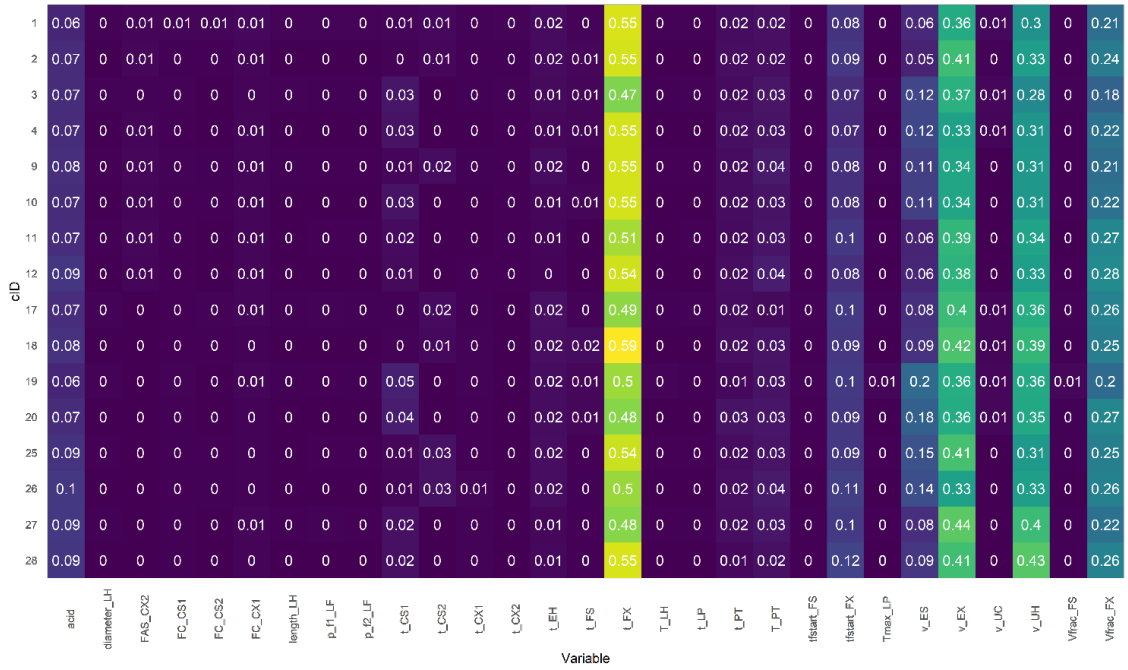

Figure 20: Heatmap of the total sensitivity indices for each flowsheet option and each operational variable (nonexistent variables are indicated with 0).

It is clearly visible that for most flowsheets, particularly variables that influence the xylitol fermentation step show the highest sensitivities. This is explainable since the microorganism can also consume xylitol as a metabolite in the fermentation, indicating an optimum for the xylitol production, which must be met. By meeting this optimum and maximizing the production of xylitol, key performances as the NPV are potentially improved due to the high selling price of the xylitol. In conclusion, the five most sensitive variables for each flowsheet option, as indicated in Figure 20, are used as input variables in the superstructure optimization problem. All other variables are fixed to their set point.

C.3.3 Superstructure Optimization

As the last operation in step (2) of the S3O framework, SSO is performed to determine candidate process topologies consolidated in step (3) of the framework. For this, flowsheet simulations with a reduced input space are performed for all feasible configuration IDs. The reduced input space consists of the five most sensitive input variables concerning the NPV, as presented in Figure 20. For each configuration ID, $N = 500$ samples are simulated, and with the simulation data, three surrogate models, a Delaunay Triangulation Regression (DTR) surrogate, a Gaussian Process Regression (GPR) surrogate, and an Artificial Neural

Network (ANN) surrogate are fitted accordingly. The validation metrics are illustrated in Figure 21. The three surrogates' validation metrics are also listed in the supplementary material.

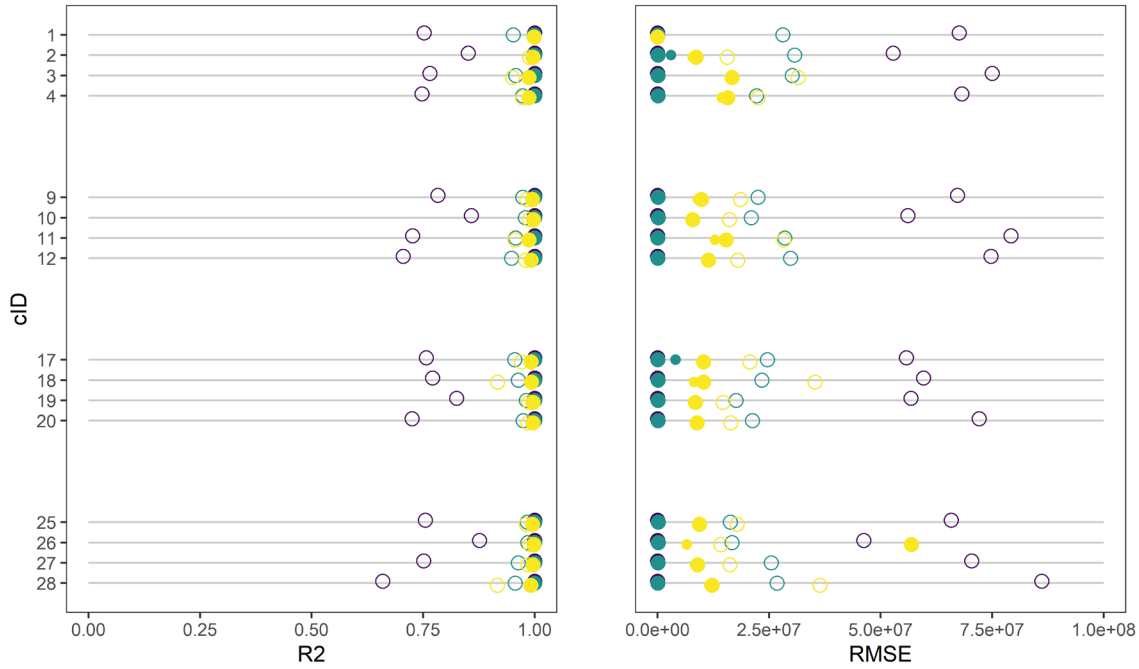

Figure 21: Validation metrics of the surrogate models.

Similar to the results of [12] and [45], the DTR surrogate shows the weakest validation metrics for the test dataset, whereas the GPR surrogate overall excels with respect to the resulting metrics. The ANN surrogate equally performs overall well with slightly weaker metrics than the GPR surrogate. All three models are used in the following SSO. All metrics are also listed in the supplementary material.

For the SSO, the optimization problem is formulated as follows:

$$\begin{aligned} &\max NPV \\ s.t. \quad &m_{xyo} \geq 0.05 \cdot C_{xyo,global} \\ &m_{suc} \geq 0.2 \cdot C_{suc,global} \\ &P_{el} \leq 0.0005 \cdot C_{el,DK} \end{aligned} \quad . \tag{C.15}$$

In the optimization problem, the objective function is the NPV, which is supposed to be maximized. Constraints are imposed on the mass of yearly produced xylitol, which has to be higher than 5% of the global annual production capacity, and on the mass of yearly produced succinic acid, which has to be higher than 20% of the global annual production capacity. Lastly, the required electrical power for the operation of the plant cannot be higher as 0.05% of the electrical power produced through wind energy in Denmark. With the DTR surrogates, the given optimization problem turns into a mixed-integer linear program, which is solved with the GUROBI solver. With the GPR and the ANN surrogate, the given optimization problem turns into a series of nonlinear programs, which are solved with the fmincon solver and a multi-start setup. The results of all optimization runs are presented in the supplementary material.

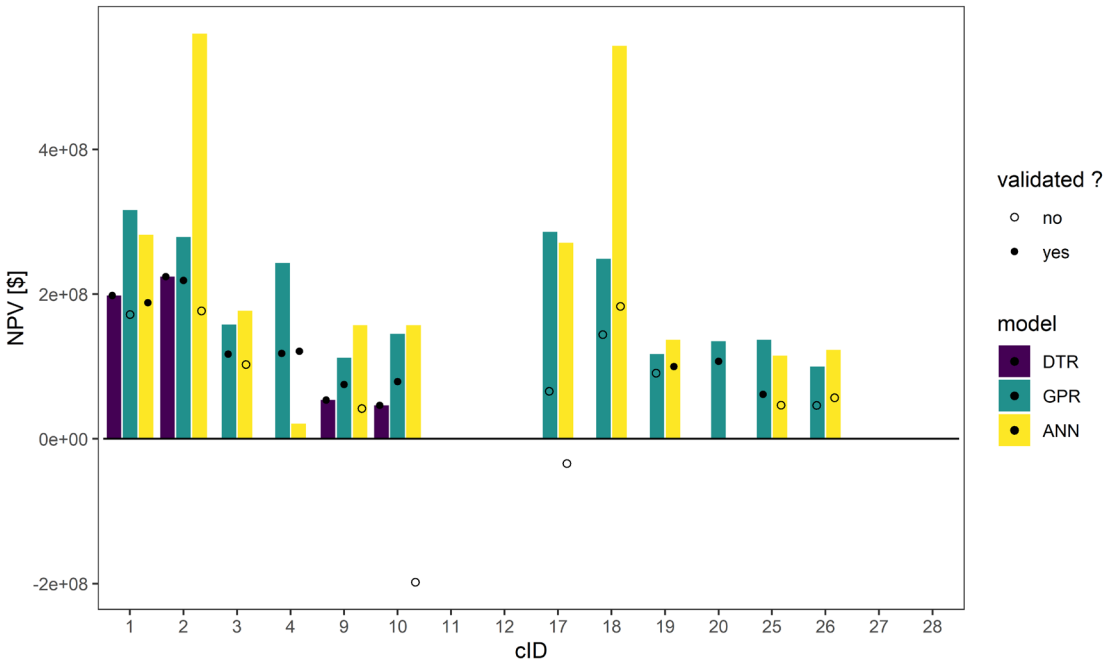

Figure 22: Results of the SSO for all three surrogate model types and the corresponding validation simulation.

The first result displayed in Figure 22 is that all configuration IDs (1,3,9,11,17,19,25,27) that involve lignin production show no significantly higher NPV than their respective counterparts without lignin production. Furthermore, certain missing unit operations, e.g., the upconcentration unit for the cellulose stream, decrease the amounts of xylitol and succinic acid drastically, leading to an infringement of the applied boundary conditions. With respect to the prediction quality of the surrogate models, it is to point out that the DTR surrogate predicted all feasible solutions correctly. However, configuration IDs 3, 4, 19, 20, and 25 were predicted infeasible despite the other surrogates correctly predicting their feasibility. On the other hand, both the GPR and the ANN surrogate overpredicted the amount of produced xylitol for several flowsheets, yielding infeasible solutions. The constraint that most GPR and ANN solutions infringe is the constraint on the minimum xylitol production. For generally infeasible solutions, the constraint on the minimum xylitol production and the maximum electricity consumption are critical. The minimum succinic acid production constraint is surpassed in all found solutions. In conclusion, the same behavior for the surrogate models as discussed in [12] is visible in this case, namely the missing extrapolation ability of the DTR surrogate model and the weak prediction abilities of the GPR and the ANN surrogate. Therefore, selecting an appropriate surrogate model is highly context-specific and should be based on a benchmark after performing the optimization problem instead of relying on the results from the cross-validation of the models before the optimization [12].

In conclusion, configuration ID 2 predicted the highest NPV. Hence it is considered a candidate for step (3) of the framework. For comparative reasons, configuration ID 2 and configuration IDs 4 and 10 are selected for the risk assessment in section C.3.5.3. Despite the infeasibility of configuration ID 18, it is selected as the fourth candidate, as the NPV is also comparatively high and the infringement of the boundary constraint is minimal. The

operational variables are optimized again by subjecting it to step (3) of the framework, potentially leading to improved production, as shown in [12], which renders this configuration ID also feasible. The detailed formulation of the SSO and all results from the SSO are also listed in the supplementary material.

C.3.4 Simulation-Based Optimization

For the third step of the framework, the set of candidate process topologies as found in section C.3.3 is now subjected to simulation-based optimization under uncertainty. The objective for this optimization setup remains to be the NPV of the plant, and the optimizable variables remain the five most sensitive operational variables as found in section C.3.2. Regarding the uncertainties, a variation in the CAPEX, represented by the FCI, and the OPEX, represented by the TPC, is considered. The FCI is assumed to be up to 50% lower or 100% higher than its originally calculated value, following a triangular distribution. Equally, the TPC is assumed to be up to 20% lower or 50% higher than its original value, following a triangular distribution. Furthermore, the product prices for xylitol and succinic acid are considered within a specific range, based on historical price data between 2016 and 2021. The range for xylitol is assumed to be within 4.29 \$/kg and 4.81\$/kg, with a mean value of 4.57 \$/kg, and the succinic acid between 3.18 \$/kg and 3.20 \$/kg, with a mean value of 3.19 \$/kg, both assumed to follow a uniform distribution [50,51]. The MOSKopt solver is set up to run with $k = 100$ iterations, of which $k_0 = 25$ are initial. For each iteration, $N = 100$ Monte Carlo Samples for the realization of the uncertainty are performed. The constraints for the SBO optimization problem remain the same as for the SSO optimization problem. However, they are modified by multiplying the average product price with each constraint to include the uncertainties into the problem. The constraints are hence as follows:

$$\begin{aligned} &\max NPV \\ s.t. \quad &sales_{xyo} \geq 0.05 \cdot C_{xyo,global} \cdot p_{xyo} \\ &sales_{suc} \geq 0.2 \cdot C_{suc,global} \cdot p_{xyo} \\ &P_{el} \leq 0.0005 \cdot C_{el,DK} \end{aligned} \quad . \tag{C.16}$$

The MOSKopt solver hedges against the constraints by the mean value of the realizations of the uncertainty. The chosen infill criterion is the mcFEI criterion, using a particle swarm optimizer as infill solver.

The results from k=100 iterations are shown in Figure 23. Firstly, it is visible that the predicted NPV under uncertainty is significantly lower than predicted in the SSO, and the mean value lies below the break-even point. Despite the optimizer being able to optimize the operational conditions, the uncertainties affect the objective to such an extent that the optimization to values above the break-even point is not feasible.

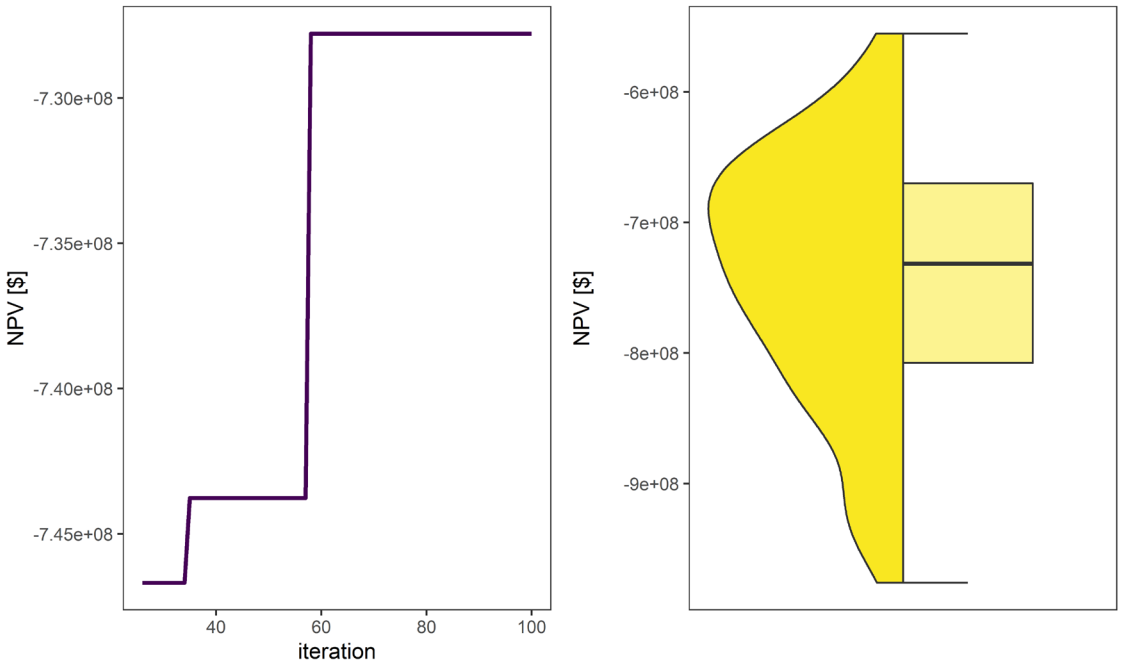

Figure 23: Results from the SBO for configuration ID 2 with the given solver settings for MOSKopt.

C.3.5 Techno-Economic Analysis

C.3.5.1 Analysis of Capital Expenditures and Operational Expenses

For a detailed insight into the techno-economic analysis results, Figure 24 shows the detailed composition of both the FCI and the TPC.

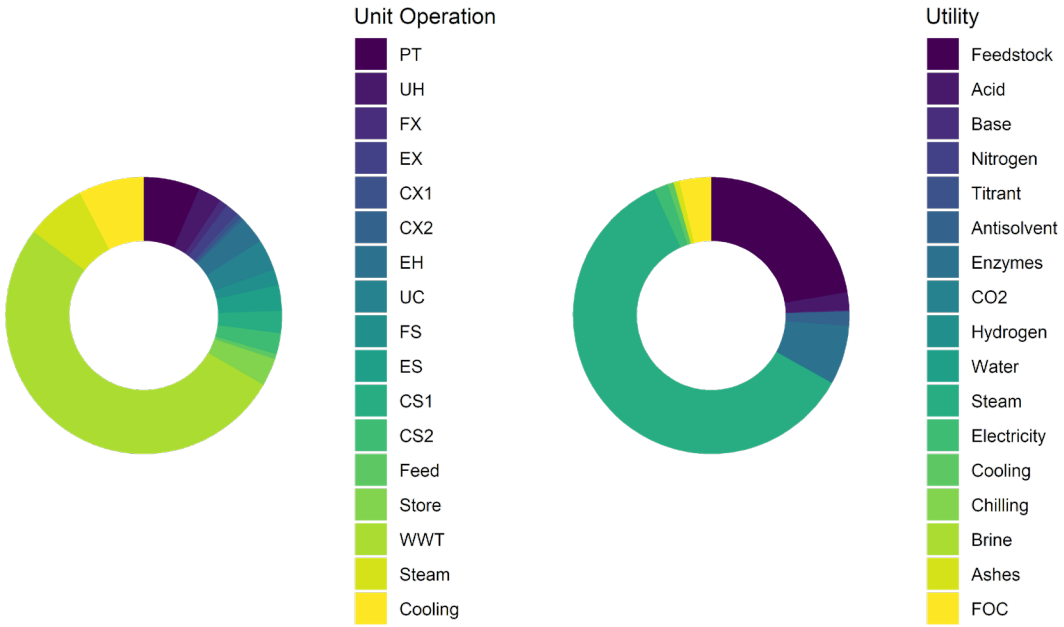

Figure 24: Analysis of the composition of both FCI (left) and TPC (right).

It is prominently visible that the FCI is dominated by around 50% through the investment in the wastewater treatment facilities, and the TPC is dominated mainly by the required steam and naturally also through the feedstock. These results again emphasize two prominent issues with (second-generation) biorefineries in particular. As mentioned in section 1, a big issue for these processes is the costs in the downstream processing. By choosing evaporation units in the downstream processing, the requirement for steam is naturally high. This is intertwined with the results for the FCI, as the process runs on an aqueous basis and low concentrations induce both high costs for the downstream processing to remove the liquid but also high costs for the wastewater treatment capacity, as the aqueous streams, being very high, need to be treated before being recycled or released to the environment.

C.3.5.2 Capacity, Yield and Size Analysis of the Plant

For the process design with cID 2 and a given annual feedstock capacity of $m_{feedstock} = 150{,}000\ t/a$, the achieved mass of xylitol per year is $m_{xyo} = 12{,}200\ t/a$ and the achieved mass of xylitol per year is $m_{suc} = 19{,}200\ t/a$. With a given composition of wheat straw, containing 33.2% hemicellulose and 44.5 % cellulose, this corresponds to process yields of $Y_{xyo} = 24.5$ % for the xylitol and $Y_{suc} = 28.8$ %. Comparing this to existing processes that produce xylitol via the chemical production route or succinic acid via fermentation, both amounts of products are comparable to those plants that produce commercially [7,49]. Nonetheless, the overall achieved yields are comparatively low, indicating two potential optimization targets, as the monomer yield in the pretreatment unit and the enzymatic hydrolysis unit are already considerably high [16,24]. Firstly, the yields of the fermentation units are inherently bound to the yield of product over substrate, which for both used cell factories lies around $40 - 60$ %, and the rest of the substrate is used for cell growth and cell maintenance. A potential way to increase product yields is the application of cell factory optimization strategies, e.g., metabolic engineering, genetic engineering, or other approaches. Secondly, the remaining yield loss is consequently attributed to the downstream process. Despite both downstream processes being viable and having the potential to be operated in commercial processes, as explained in section C.2.2.4, further optimization and the investigation on the potential use of alternative unit operations can be further explored. An increased process integration beyond the current level is a further aspect worth investigating. For the chemical production process of, e.g., Dupont, such process integration with a pulp and paper mill is already performed to centralize unit operations, e.g., the steam and power generation and also the heat integration across different plants [6,52,53]. Such levels of integration offer the potential to significantly decrease the CAPEX and OPEX for the plant. All capacities and operational conditions are also listed in the supplementary material.

C.3.5.3 Risk-Based Economic Evaluation

For the risk analysis, the consolidated process design of configuration IDs 2,4, 10, and 18 are chosen as a basis with the optimized operational conditions as determined in section C.3.3. The considered uncertainties are the same as for the simulation-based optimization. To quantify the uncertainty in the market data of both products, different distributions are analyzed concerning their quality of fit to historical price data for xylitol and succinic acid. The investigated distributions are a normal distribution, a lognormal distribution, a nonparametric distribution, an extreme-value distribution, and a uniform distribution. The

results are summarized in Figure 25. On the x-axis, the price x is indicated as ma min-max normalized random variable, while the y-axis shows the probability of observing a price is higher than a threshold $P(x \leq X)$, commonly known as cumulative distribution function.

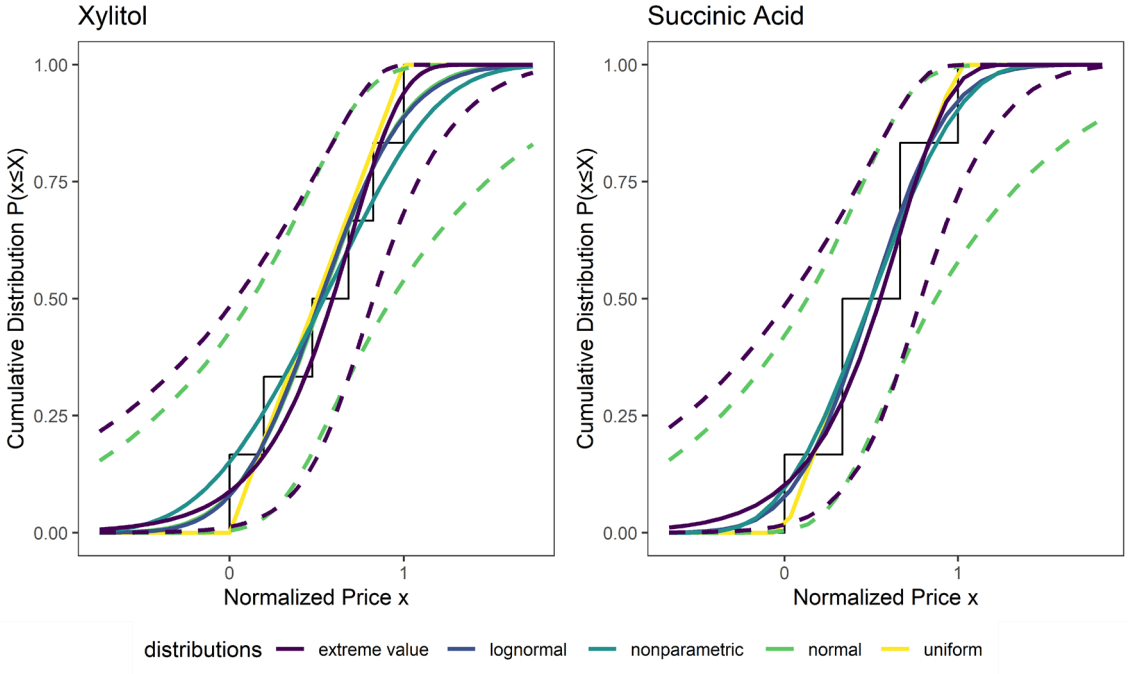

Figure 25: Cumulative density functions of all investigated distributions for the xylitol and the succinic acid price data.

It becomes apparent that all investigated distributions fit equally well, which is also confirmed by the coefficient of determination and the root mean squared error, both listed in the supplementary material. This implies that the historical data is not conclusive enough to identify the actual probability distribution function that describes the price uncertainties. Hence, a heuristic approach is employed by selecting a normal distribution and an extreme-value distribution as a thin-tailed and fat-tailed distribution. The parameters and metrics of all distributions are listed in the supplementary material. The Monte Carlo analysis is performed with $N = 1000$ samples, created by Latin Hypercube sampling. The resulting MSEP of xylitol for all simulations for the normal distribution is presented in Figure 10.

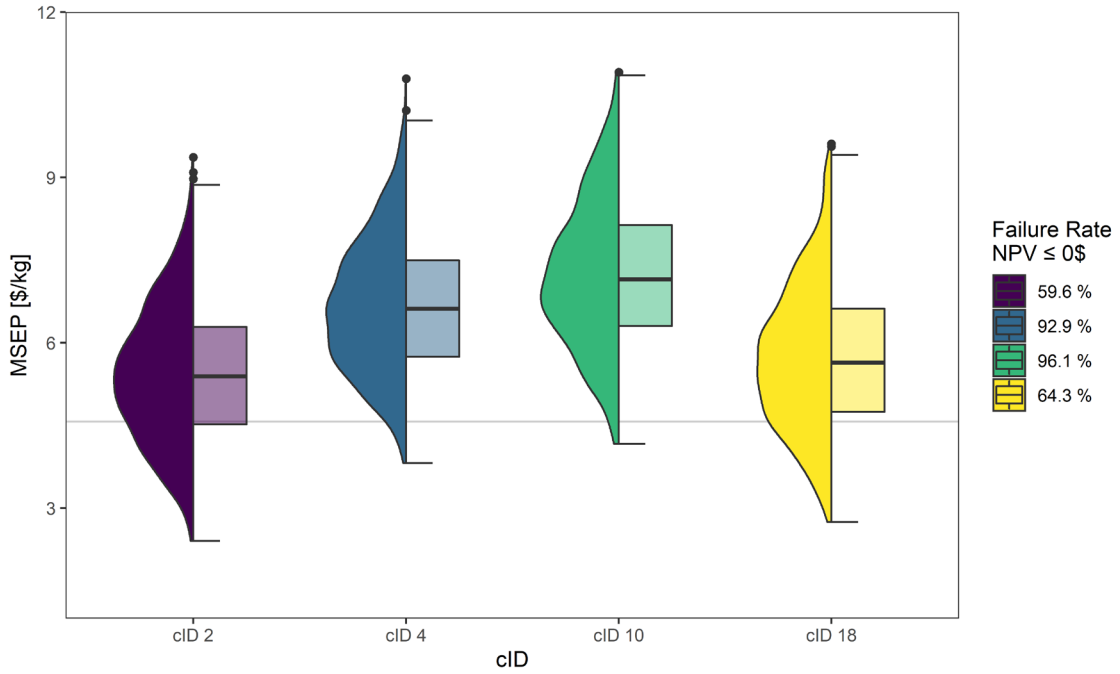

Figure 26: Violin plots of the MSEP of xylitol for N=1000 Monte Carlo samples for the risk analysis of the consolidated process design.

For all four evaluated configuration IDs, it becomes apparent that their economic feasibility is highly uncertain. The average selling price for xylitol of 2021, indicated at 4.57 $\$/kg$ is, is reached for all four configuration IDs. However, particularly for cID 4 and 10, this is only the case for overestimated FCI and TPC and a favorable succinic acid price. Also, for cID 2 and 18, the break-even point ($NPV = 0\$$)is only reached in around 40 % of all cases. The process design with cID 2 has the lowest failure rate. Knowing that all other presented configuration IDs refer to flowsheets that do not possess a second crystallization step in the xylitol or succinic acid process train, a clear tendency towards the downstream process with two crystallization units each is advised, despite the higher CAPEX. Similar results are found in [48]. The mean value of the minimum selling price of xylitol, the confidence interval, the mean selling price of xylitol between 2016 and 2021, and the corresponding confidence interval are also listed in Table 18.

Table 18: Overview over the Minimum Selling Price of Xylitol based on Monte Carlo Simulations compared to the average selling price in the reference period 2016 – 2021.

All in [$/kg]	cID 2	cID 4	cID 10	cID 18
Mean	5.43	6.67	7.25	5.73
CI +-95%	2.94 – 7.93	4.26 – 9.08	4.58 – 9.92	3.11 – 8.36
Average price	4.57	4.57	4.57	4.57
CI+-95%	4.17 - 4.97	4.17 - 4.97	4.17 - 4.97	4.17 - 4.97

Given this picture, a profitable construction and operation of a xylitol biorefinery is a high-risk venture. As the uncertainties in the product prices are determined through the development of their respective markets, which do not show significant volatilities over the past five years, a profitable operation seems highly unlikely. Furthermore, due to the low volatility in prices, the large overall uncertainty bounds are highly attributable to the uncertainties in the CAPEX and OPEX of the plant. Further investigation and a detailed

design are crucial to minimize these and perform the risk analysis with updated costs to provide a confident conclusion on the economic viability.

C.4 Conclusions and Future Research

As the overarching goal in this manuscript, the process design of a biorefinery with xylitol as the main product was investigated to find a biotechnological process alternative to the existing chemical production process for xylitol. As several questions arise, like the decision on potential value-added co-products and considerations regarding which feedstock to use and how to best integrate the process, a conceptual design approach was required. For this, the S3O framework from a previous publication of the authors was utilized. Compared to other studies, mechanistic models were used, and detailed data regarding the costing of the equipment and current market prices of the product.

In a three-stage approach, products, feedstock, and process units are selected, and mechanistic unit operation models are built. Subsequently, an initial superstructure is formulated, and its design space is evaluated by both a coarse and a global sensitivity analysis. The coarse sampling serves the investigation on the feasibility of the flowsheets. As the first result in this study, only 16 out of 128 initial process flowsheet options are considered feasible. The global sensitivity analysis was used to investigate which operational variables are the most sensitive ones regarding the net present value as a key performance indicator of the plant. For all feasible flowsheets, the most sensitive variables influence the operation unit of the xylitol fermentation. This is explained by the fact that xylitol can be metabolized by the cell factory again, indicating a defined global optimum for the production of xylitol regarding the operational conditions.

Flowsheet sampling with the five most sensitive operational variables is performed for all feasible flowsheets. Three different surrogate model types are fitted to it (DTR, GPR, ANN) and benchmarked against each other regarding their validation metrics (coefficient of determination, root mean squared error). The results agree with those found earlier in [12], showing that the DTR surrogates have worse performance metrics than the GPR and the ANN. However, utilizing them in the surrogate-assisted superstructure optimization, the DTR models perform equally well as the other two. The results indicate, also in agreement with [12], that the DTR lacks the possibility of interpolating highly nonlinear functional relationships with an insufficient amount of data points, which leads to suboptimal solutions. In contrast, the GPR and the ANN surrogates tend to overpredict the objective function value and constraint function values, resulting in infeasible solutions. From a process point of view, the conclusion prevails that the utilization of lignin for the production of biokerosene does not involve any economic advantage over using it for combustion and the generation of steam and electricity. Furthermore, excluding downstream process operations like the additional crystallization units or the upconcentration units before the fermentation units reduces the CAPEX of the potential process design, but this does not result in an increased NPV as the amount of recovered product is reduced significantly.
Based on these conclusions, the process configuration that involves all unit operations and utilizes the lignin for combustion (cID 2) is selected for simulation-based optimization in the third step. The considered uncertainties are both in the CAPEX and OPEX, expressed by a variation in the FCI and the TPC according to the used class 5 estimates. In addition, the

product prices are considered uncertain within a range according to their global market prices between 2016 and 2021. The results indicate that when hedging the uncertainties against the mean value of the predictions, the operational conditions are further optimized. However, the uncertainties impact the objective stronger than the improvement of the objective, which decreases the NPV even below the break-even point. An additional Monte Carlo analysis with 1000 points based on the operational conditions found by the SSO and the SBO shows a failure rate of almost 50% for the configuration ID 2, and – in comparison – failure rates of 55-90% for other potential candidates from the SSO. Similar results are seen throughout the literature for other realized biorefinery setups:

Mountraki et al. (2017) obtain similar results, stating that the chemical production of xylitol is more profitable than their biorefinery setup [13]. On the other hand, Franceschin et al. (2011) state that their xylitol biorefinery setup can be feasible, but only with estimated capital investment data and a significantly higher xylitol market price [14]. Giuliano et al. (2018) do not conclude a clear answer regarding the profitability but rather indicate the potential MSEP of ethanol for an integrated production with xylitol in a biorefinery setup is lower than the production of ethanol alone [15]. While numerous companies, e.g., Dupont, hold patents for the biotechnological production of xylitol, all these companies still rely on its chemical production of it [3]. Regarding biotechnological succinic acid production, earlier joint ventures of different companies failed in the past years. However, the market for biotechnologically produced succinic acid from sugar directly – and not from lignocellulosic biomass – is supposed to grow in the future [17,49,51].

Given these considerations, the economic feasibility is not overall given and highly depends on the uncertainties in CAPEX and OPEX. Only a detailed design with concrete values for CAPEX and OPEX would allow for a clear decision for or against an investment as the uncertainty in the decision decreases. This comprises both more detailed sizing and costing of the considered equipment and potential additional equipment regarding the inhibitory compounds and lignin, as mentioned in section C.2. With those uncertainties reduced, a stable market situation, as a projection based on historical price data, can potentially lead to an economically feasible xylitol biorefinery. Ultimately, the global price for xylitol is dictated by global trends towards healthier nutrition or the use of bioplastics, so an increase in the price in the future could be realistic [50]. However, the price for succinic acid as a product that is already produced biotechnologically depends much more on the global price for fossil oil, which has – against expectations – not significantly increased over the past ten years, which has been the primary reason for economic infeasibilities of other biorefinery projects in the past [3,17,51]. A comprehensive analysis of these and further factors would go beyond the scope of this thesis, hence, but is concisely discussed in a report by, e.g., McKinsey about "the future of second-generation biomass" [54].

Putting these findings in a future perspective, further research on several aspects could lead to improvements in the xylitol biorefinery process itself and, subsequently, on the economic feasibility. Firstly, the utilization of the economies of scale is a considerable possibility to improve the economic performance of the plant, as the CAPEX does not scale linearly with the plant size and hence can lead to higher KPIs for larger plant capacities [55].

Secondly, both fermentation units currently utilize wild-type cell factories. As presented in sections C.3.3 and C.3.5, the average product yields for both products range between 50% and 60%. Using engineered or optimized cell factories can significantly increase this value, together with the achievable productivities and titers for the fermentation [56]. The optimization of this has an immediate effect on the KPIs of the plant. Further research needs to investigate the impacts of engineered cell factories on the downstream processing and other requirements. Potential modifications of the downstream process can either increase or decrease the CAPEX and OPEX for the biorefinery, which ultimately determines the full effect of engineered/optimized cell factories. Lastly, the aspect of sustainability has not been investigated in the scope of this work. However, the quantitative assessment of the sustainability impact of a new biorefinery, not only focusing on the reduced emissions of CO_2 but also other environmental aspects, is necessary to fully assess the potential of these biorefineries and their contribution to more sustainable value chains and production patterns for the future.

Acknowledgments

The authors would like to express their gratitude to the Novo Nordisk Foundation (Grant No. NNF17SA0031362) for funding the Fermentation-Based Biomanufacturing Initiative, of which this project is a part.

References

[1] United Nations, Transforming our world: The 2030 agenda for sustainable development, 2015. https://sdgs.un.org/sites/default/files/publications/21252030 Agenda for Sustainable Development web.pdf (accessed March 1, 2021).

[2] F. Cherubini, The biorefinery concept: Using biomass instead of oil for producing energy and chemicals, Energy Convers. Manag. 51 (2010) 1412–1421. https://doi.org/10.1016/j.enconman.2010.01.015.

[3] A.T. Ubando, C.B. Felix, W.H. Chen, Biorefineries in circular bioeconomy: A comprehensive review, Bioresour. Technol. 299 (2020) 122585. https://doi.org/10.1016/j.biortech.2019.122585.

[4] J.-R. Bastidas-Oyanedel, J.E. Schmidt, eds., Biorefinery: Integrated Sustainable Processes for Biomass Conversion to Biomaterials, Biofuels and Fertilizers, 2019. https://doi.org/10.1007/978-3-030-10961-5.

[5] S.S. Da Silva, A.K. Chandel, D-Xylitol: Fermentative production, application and commercialization, Springer-Verlag Berlin Heidelberg, 2012. https://doi.org/10.1007/978-3-642-31887-0.

[6] Y. Delgado Arcaño, O.D. Valmaña García, D. Mandelli, W.A. Carvalho, L.A. Magalhães Pontes, Xylitol: A review on the progress and challenges of its production by chemical route, Catal. Today. 344 (2020) 2–14. https://doi.org/10.1016/j.cattod.2018.07.060.

[7] A.F. Hernández-Pérez, P.V. de Arruda, L. Sene, S.S. da Silva, A. Kumar Chandel, M. das G. de Almeida Felipe, Xylitol bioproduction: state-of-the-art, industrial paradigm shift, and opportunities for integrated biorefineries, Crit. Rev. Biotechnol. 39 (2019) 924–943. https://doi.org/10.1080/07388551.2019.1640658.

[8] T. Werpy, G. Petersen, Top Value Added Chemicals from Biomass Volume I, 2004. https://doi.org/10.2172/15008859.

[9] T.L. De Albuquerque, I.J. Da Silva, G.R. De MacEdo, M.V.P. Rocha, Biotechnological production of xylitol from lignocellulosic wastes: A review, Process Biochem. 49 (2014) 1779–1789. https://doi.org/10.1016/j.procbio.2014.07.010.

[10] L.V. Rao, J.K. Goli, J. Gentela, S. Koti, Bioconversion of lignocellulosic biomass to xylitol: An overview, Bioresour. Technol. 213 (2016) 299–310. https://doi.org/10.1016/j.biortech.2016.04.092.

[11] D. Dasgupta, S. Bandhu, D.K. Adhikari, D. Ghosh, Challenges and prospects of xylitol production with whole cell bio-catalysis: A review, Microbiol. Res. 197 (2017) 9–21. https://doi.org/10.1016/j.micres.2016.12.012.

[12] N.I. Vollmer, R. Al, K. V. Gernaey, G. Sin, Synergistic optimization framework for the process synthesis and design of biorefineries, Front. Chem. Sci. Eng. 16 (2022) 251–273. https://doi.org/10.1007/s11705-021-2071-9.

[13] A.D. Mountraki, K.R. Koutsospyros, B.B. Mlayah, A.C. Kokossis, Selection of

Biorefinery Routes: The Case of Xylitol and its Integration with an Organosolv Process, Waste and Biomass Valorization. 8 (2017) 2283–2300. https://doi.org/10.1007/s12649-016-9814-8.

[14] G. Franceschin, M. Sudiro, T. Ingram, I. Smirnova, G. Brunner, A. Bertucco, Conversion of rye straw into fuel and xylitol: A technical and economical assessment based on experimental data, Chem. Eng. Res. Des. 89 (2011) 631–640. https://doi.org/10.1016/j.cherd.2010.11.001.

[15] A. Giuliano, D. Barletta, I. De Bari, M. Poletto, Techno-economic assessment of a lignocellulosic biorefinery co-producing ethanol and xylitol or furfural, Comput. Aided Chem. Eng. 43 (2018) 585–590. https://doi.org/10.1016/B978-0-444-64235-6.50105-4.

[16] N.I. Vollmer, J.L.S.P. Driessen, C.K. Yamakawa, K. V. Gernaey, S.I. Mussatto, G. Sin, Model development for the optimization of operational conditions of the pretreatment of wheat straw, Chem. Eng. J. 430 (2022) 133106. https://doi.org/10.1016/j.cej.2021.133106.

[17] E. Mancini, S.S. Mansouri, K. V. Gernaey, J. Luo, M. Pinelo, From second generation feed-stocks to innovative fermentation and downstream techniques for succinic acid production, Crit. Rev. Environ. Sci. Technol. 50 (2020) 1829–1873. https://doi.org/10.1080/10643389.2019.1670530.

[18] A.J. Ragauskas, G.T. Beckham, M.J. Biddy, R. Chandra, F. Chen, M.F. Davis, B.H. Davison, R.A. Dixon, P. Gilna, M. Keller, P. Langan, A.K. Naskar, J.N. Saddler, T.J. Tschaplinski, G.A. Tuskan, C.E. Wyman, Lignin valorization: Improving lignin processing in the biorefinery, Science (80-.). 344 (2014). https://doi.org/10.1126/science.1246843.

[19] D. Chiaramonti, M. Prussi, M. Buffi, D. Tacconi, Sustainable bio kerosene: Process routes and industrial demonstration activities in aviation biofuels, Appl. Energy. 136 (2014) 767–774. https://doi.org/10.1016/j.apenergy.2014.08.065.

[20] N.I. Vollmer, Xylitol Biorefinery, (2021). https://github.com/NikolausVollmer/Xylitol-Biorefinery.

[21] S.I. Mussatto, G.M. Dragone, Biomass Pretreatment, Biorefineries, and Potential Products for a Bioeconomy Development, in: Biomass Fractionation Technol. a Lignocellul. Feed. Based Biorefinery, Elsevier Inc., 2016: pp. 1–22. https://doi.org/10.1016/B978-0-12-802323-5.00001-3.

[22] K.L. Kadam, E.C. Rydholm, J.D. McMillan, Development and Validation of a Kinetic Model for Enzymatic Saccharification of Lignocellulosic Biomass, Biotechnol. Prog. 20 (2004) 698–705. https://doi.org/10.1021/bp034316x.

[23] Novozymes, Novozymes Cellic® CTec3 HS application sheet - secure your plant's lowest cost, 2017.

[24] R.M. Prunescu, G. Sin, Dynamic modeling and validation of a lignocellulosic enzymatic hydrolysis process - A demonstration scale study, Bioresour. Technol. 150 (2013) 393–403. https://doi.org/10.1016/j.biortech.2013.10.029.

[25] J.J. Heijnen, W.M. van Gulik, Section II - Balances and Reaction Models, in: C.D. Smolke (Ed.), Metab. Pathw. Eng. Handb. Fundam., CRC Press, Boca Raton, 2009: pp. II-1-11–20.

[26] W. Tochampa, S. Sirisansaneeyakul, W. Vanichsriratana, P. Srinophakun, H.H.C. Bakker, Y. Chisti, A model of xylitol production by the yeast Candida mogii, Bioprocess Biosyst. Eng. 28 (2005) 175–183. https://doi.org/10.1007/s00449-005-0025-0.

[27] H. Song, S.H. Jang, J.M. Park, S.Y. Lee, Modeling of batch fermentation kinetics for succinic acid production by Mannheimia succiniciproducens, Biochem. Eng. J. 40 (2008) 107–115. https://doi.org/10.1016/j.bej.2007.11.021.

[28] A.A. Kiss, J.P. Lange, B. Schuur, D.W.F. Brilman, A.G.J. van der Ham, S.R.A. Kersten, Separation technology–Making a difference in biorefineries, Biomass and Bioenergy. 95 (2016) 296–309. https://doi.org/10.1016/j.biombioe.2016.05.021.

[29] D.J. Kirwan, C.J. Orella, Crystallization in the pharmaceutical and bioprocessing industries, in: Handb. Ind. Cryst., Elsevier, 2002: pp. 249–266. https://doi.org/10.1016/b978-075067012-8/50013-6.

[30] E.A. Martínez, M. Giulietti, J.B. de Almeida e Silva, S. Derenzo, Kinetics of the xylitol crystallization in hydro-alcoholic solution, Chem. Eng. Process. Process Intensif. 47 (2008) 2157–2162. https://doi.org/10.1016/j.cep.2007.11.004.

[31] M. Giulietti, M.M. Seckler, S. Derenzo, M.I. Ré, E. Cekinski, Industrial crystallization and precipitation from solutions: State of the technique, Brazilian J. Chem. Eng. 18 (2001) 423–440. https://doi.org/10.1590/S0104-66322001000400007.

[32] M. Öner, C. Bach, T. Tajsoleiman, G.S. Molla, M.F. Freitag, S.M. Stocks, J. Abildskov, U. Krühne, G. Sin, Scale-up Modeling of a Pharmaceutical Crystallization Process via Compartmentalization Approach, in: Comput. Aided Chem. Eng., Elsevier B.V., 2018: pp. 181–186. https://doi.org/10.1016/B978-0-444-64241-7.50025-2.

[33] Y. Qiu, Å.C. Rasmuson, Growth and dissolution of succinic acid crystals in a batch stirred crystallizer, AIChE J. 36 (1990) 665–676. https://doi.org/10.1002/aic.690360504.

[34] Y. Qiu, Å.C. Rasmuson, Nucleation and growth of succinic acid in a batch cooling crystallizer, AIChE J. 37 (1991) 1293–1304. https://doi.org/10.1002/aic.690370903.

[35] Y. Qiu, Å.C. Rasmuson, Estimation of crystallization kinetics from batch cooling experiments, AIChE J. 40 (1994) 799–812. https://doi.org/10.1002/aic.690400507.

[36] J.W. Mullin, M.J.L. Whiting, Succinic Acid Crystal Growth Rates in Aqueous Solution, Ind. Eng. Chem. Fundam. 19 (1980) 117–121. https://doi.org/10.1021/i160073a020.

[37] W.C. Wang, L. Tao, Bio-jet fuel conversion technologies, Renew. Sustain. Energy Rev. 53 (2016) 801–822. https://doi.org/10.1016/j.rser.2015.09.016.

[38] J. Zakzeski, P.C.A. Bruijnincx, A.L. Jongerius, B.M. Weckhuysen, The catalytic valorization of lignin for the production of renewable chemicals, Chem. Rev. 110 (2010) 3552–3599. https://doi.org/10.1021/cr900354u.

[39] P. Debiagi, G. Gentile, A. Cuoci, A. Frassoldati, E. Ranzi, T. Faravelli, A predictive model of biochar formation and characterization, J. Anal. Appl. Pyrolysis. 134 (2018) 326–335. https://doi.org/10.1016/j.jaap.2018.06.022.

[40] A.H. Zacher, M. V. Olarte, D.M. Santosa, D.C. Elliott, S.B. Jones, A review and perspective of recent bio-oil hydrotreating research, Green Chem. 16 (2014) 491–515. https://doi.org/10.1039/c3gc41382a.

[41] T. Cordero-Lanzac, I. Hita, F.J. García-Mateos, P. Castaño, J. Rodríguez-Mirasol, T. Cordero, J. Bilbao, Adaptable kinetic model for the transient and pseudo-steady states in the hydrodeoxygenation of raw bio-oil, Chem. Eng. J. 400 (2020) 124679. https://doi.org/10.1016/j.cej.2020.124679.

[42] D. Humbird, R. Davis, L. Tao, C. Kinchin, D. Hsu, A. Aden, P. Schoen, J. Lukas, B. Olthof, M. Worley, D. Sexton, D. Dudgeon, Process design and economics for conversion of lignocellulosic biomass to ethanol, 2011. http://www.nrel.gov/docs/fy11osti/51400.pdf%5Cnpapers2://publication/uuid/49A5007E-9A58-4E2B-AB4E-4A4428F6EA66.

[43] S.K. Bhatia, S.S. Jagtap, A.A. Bedekar, R.K. Bhatia, A.K. Patel, D. Pant, J. Rajesh Banu, C. V. Rao, Y.G. Kim, Y.H. Yang, Recent developments in pretreatment technologies on lignocellulosic biomass: Effect of key parameters, technological improvements, and challenges, Bioresour. Technol. 300 (2020) 122724. https://doi.org/10.1016/J.BIORTECH.2019.122724.

[44] E. Pienihäkkinen, C. Lindfors, T. Ohra-Aho, J. Lehtonen, T. Granström, M. Yamamoto, A. Oasmaa, Fast Pyrolysis of Hydrolysis Lignin in Fluidized Bed Reactors, Energy & Fuels. 35 (2021) 14758–14769. https://doi.org/10.1021/ACS.ENERGYFUELS.1C01719.

[45] N.I. Vollmer, R. Al, G. Sin, Benchmarking of Surrogate Models for the Conceptual Process Design of Biorefineries, in: Comput. Aided Chem. Eng., Elsevier, 2021: pp. 475–480. https://doi.org/10.1016/B978-0-323-88506-5.50075-9.

[46] M. Peters, K. Timmerhaus, M. Peters, Plant Design and Economics for Chemical Engineers: Peters, Max, Timmerhaus, Klaus, West, Ronald, Peters , Max: 0639785503897: Amazon.com: Books, 5th ed., McGraw-Hill Education, 2002. https://www.amazon.com/Plant-Design-Economics-Chemical-Engineers/dp/0072392665 (accessed November 11, 2021).

[47] G. Sin, K. V. Gernaey, A.E. Lantz, Good modeling practice for PAT applications: Propagation of input uncertainty and sensitivity analysis, Biotechnol. Prog. 25 (2009) 1043–1053. https://doi.org/10.1002/btpr.166.

[48] N.I. Vollmer, K. V. Gernaey, S.I. Mussatto, G. Sin, Surrogate Modelling Based Uncertainty and Sensitivity Analysis for the Downstream Process Design of a Xylitol Biorefinery, Comput. Aided Chem. Eng. 48 (2020) 1663–1668. https://doi.org/10.1016/B978-0-12-823377-1.50278-0.

[49] M.L.A. Jansen, W.M. van Gulik, Towards large scale fermentative production of succinic acid, Curr. Opin. Biotechnol. 30 (2014) 190–197. https://doi.org/10.1016/j.copbio.2014.07.003.

[50] IMARC, Xylitol Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2021-2026, 2021.

[51] Orion Market Research, Global Succinic Acid Market Forecast, 2016-2026, (2020).

[52] DuPont, XIVIA™ Xylitol White Paper, 2012. http://www.danisco.com/fileadmin/user_upload/danisco/documents/products/2e_XIVIA_White_Paper.pdf.

[53] K. Özdenkçi, C. De Blasio, H.R. Muddassar, K. Melin, P. Oinas, J. Koskinen, G. Sarwar, M. Järvinen, A novel biorefinery integration concept for lignocellulosic biomass, Energy Convers. Manag. 149 (2017) 974–987. https://doi.org/10.1016/J.ENCONMAN.2017.04.034.

[54] S. Alfano, F. Berruti, N. Denis, A. Santagostino, The future of second-generation biomass, (2016) 1–5. https://www.mckinsey.com/business-functions/sustainability/our-insights/the-future-of-second-generation-biomass.

[55] N.I. Vollmer, K. V. Gernaey, G. Sin, Value Chain Optimization of a Xylitol Biorefinery with Delaunay Triangulation Regression Models, Comput. Aided Chem. Eng. (2022).

[56] N.I. Vollmer, K. V. Gernaey, G. Sin, Sensitivity Analysis and Risk Assessment for the In-Silico Design and Use of Optimized Cell Factories in a Xylitol Biorefinery, Comput. Aided Chem. Eng. (2022).

Paper

D

Life Cycle Analysis of an Integrated Xylitol Biorefinery with Value-Added Co-Products

Nikolaus I. Vollmer, Carina L.C.L. Gargalo, Krist V. Gernaey, Stig I. Olsen, Gürkan Sin,

Journal of Cleaner Production,

under review as of 31st of January 2022

Abstract

This manuscript comprises a detailed life cycle assessment of an integrated xylitol biorefinery with value-added co-products. The biorefinery utilizes wheat straw as lignocellulosic feedstock and employs bio-based processes to produces xylitol as main product and succinic acid as co-product. The biorefinery was conceptually designed in an optimization-based framework and assessed through a techno-economic analysis, published in prior publications of the authors. The presented life cycle assessment was performed according to the standardized ISO procedure. The goal is to determine the environmental impacts of the xylitol biorefinery and to compare the impacts of the bio-based production in the biorefinery to the current chemical production processes of xylitol. The scope is set as cradle-to-gate in order to allow a direct comparison to the chemical processes. The reference unit is related to the feedstock as multiple products are produced, and an economic allocation is chosen. The life cycle inventory is based on secondary data from process simulations, stemming from the earlier published work. The impact assessment is performed with the ReCiPe 2016 Midpoint H V1.05 method, and with the IMPACT2002+ method, since the available data of the life cycle assessment for the chemical processes was obtained with the latter. The characterization of the impacts show high impacts for the terrestrial, marine and human carcinogenic toxicity impact categories, and a comparatively low impact on the global warming. The results are interpreted and assessed with an additional sensitivity analysis. The results are furthermore compared with the two chemical production processes. The comparison shows lower impacts of the xylitol biorefinery compared to the standard chemical production process but slightly higher impacts compared to the proprietary production process of Dupont, which employs a high level of process integration. These results are further discussed and contextualized.

D.1 Introduction

One of the overarching political, industrial, and also societal aims of our generation is the sustainable transition. On a global level, the United Nations have agreed upon 17 sustainable development goals, reaching from the eradication of hunger and poverty, over the global access to clean freshwater and education up to the sustainable production of energy and sustainable production patterns, as well as general climate action [1]. Particularly, more sustainable production patterns, comprising concepts such as circular economy and resource recovery, and bio-based processes are considered vital elements in the sustainable transition [1–4].

In comparison to existing chemical production processes with fossil resources as feedstock, bio-based processes are deemed to be more sustainable due to milder process conditions, lower use of potentially harmful chemicals, and lower CO_2 emissions [5–7]. A particular concept in this context are biorefineries, as analogue to a fossil refinery, producing multiple products as chemicals, fuels, and energy based on renewable feedstocks [8]. The palette of feedstocks is vast; hence biorefineries are commonly classified into generations, depending on the feedstock: while first-generation biorefineries utilize food crops containing starch or glucose, second-generation biorefineries utilize crop residues or non-food crops [8–10].
Regarding the sustainability aspects of such biorefineries, while their sustainability potential, in general, seems promising, there are several aspects to consider. Despite around a thousand of them being implemented worldwide, first-generation biorefineries significantly impact the food-vs-fuel nexus [11,12]. Hence, considering an increased implementation of such biorefineries for the large-scale production of biofuels does stand diametrically to the SDGs, e.g., to the goal of eradicating hunger [12]. In addition, their sustainability potential in terms of, e.g., greenhouse gas emissions is not as prevailing as for higher-generation biorefineries [11].

This particular issue shows that sustainability and the transition towards it are multi-faceted challenges and that changes in a specific part of a system can lead to subsequent effects in other parts of the system. Despite this, most sustainability analyses, mainly referred to as life cycle assessments (LCAs), are commonly focused on specific impacts and have unclear allocation methods and system boundaries, and do not analyze the entire system [13–15].

A potential example of a chemical production process that a biotechnological process can possibly replace is the production of xylitol. Xylitol is one of the top 12 chemicals to be produced in a biorefinery, according to an evaluation of the US Department of Energy [16]. Itcan be used either as a building block for plastics or as a product for nutritional purposes, e.g., as a sugar substitute with several beneficial health properties [17,18]. The current chemical production process utilizes lignocellulosic biomass as feedstock and is based on a chemical conversion process that requires high purities, making the upstream process quite complex and the product itself rather expensive. [19,20].

Alternatively, xylitol could be produced in a biotechnological process, using lignocellulosic biomass as feedstock, but employing a fermentation unit instead of a chemical conversion, tolerating higher levels of impurities with suitable microorganisms. Up to this date, most

experimental studies proved this process to be feasible at laboratory scales, while conceptual studies regarding a technologically and economically viable process at a commercial scale are scarce. Vollmer et al. (2022) have performed a conceptual process design for a xylitol biorefinery process with value-added co-products based on a synergistic optimization-based framework, including as much detailed knowledge about the process as possible [21,22]. While the process is technologically viable, its economic feasibility depends heavily on the present uncertainties regarding capital expenditures, operational expenses, and the market conditions, particularly the product prices.

For a proper assessment regarding the sustainability potential of this xylitol biorefinery, an LCA on the xylitol biorefinery will be performed in the scope of this paper. This facilitates a decision whether implementing such a process as an alternative to a chemical equivalent can be considered positive from an environmental sustainability perspective. Based on the obtained LCA results, a comparison with the LCA results of the chemical process will allow for a profound comparison of the different processes and facilitate conclusions regarding the potentials of the implementation of biotechnological processes as an alternative to chemical ones in this specific case, but also in general.

The remainder of this paper is structured as follows: in section D.2, first, a detailed overview of the current chemical production process of xylitol is given, as well as a detailed explanation of the mentioned xylitol biorefinery based on a biotechnological process. Furthermore, the standard four-step methodology of LCA according to ISO norm 14040 is explained together with the used programs, databases, and methods. In section D.3, the detailed results of all four steps of the LCA for the xylitol biorefinery are presented. Subsequently, the results are compared to those of the chemical process. Lastly, in section D.4, conclusions for the xylitol biorefinery in specific and also general conclusions are drawn.

D.2 Materials and Methods

D.2.1 Xylitol Biorefinery

D.2.1.1 Current Production Process of Xylitol & Succinic Acid

As of today, xylitol is exclusively produced in a chemical conversion process. Delgado Arcaño et al. (2020) describe it as follows: The single reaction step is the catalytic hydrogenation of xylose into xylitol over a Raney nickel catalyst at a temperature of around $T = 400\ K$ and $p = 50\ bar$ hydrogen pressure [23]. The yield of this conversion is around 98% which allows for a simple downstream process, including a filtration, ion exchange, and crystallization unit. The xylose monomers are obtained from lignocellulosic biomass in a pretreatment unit; the hemicellulosic fraction of the lignocellulosic biomass is fractionated and depolymerized [24]. Due to the high sensitivity of the catalyst and in order to prevent an accelerated degradation, after the pretreatment step, an intensive purification of the process stream is necessary to remove undesired by-products. The necessary purification before the reaction and the high temperature and pressure requirements have a substantial economic and significant sustainability impact regarding the entire production process and the final product price [19]. Commonly, the used lignocellulosic biomass is either corncob or derives from birch trees, depending on the producer: Several Chinese

companies employ corncob as their feedstock, while Dupont Danisco utilizes the side stream of a paper mill, which utilizes the named birch trees [17].

A potential alternative to the chemical production process of xylitol is the production via a biotechnological process route. Hernández-Perez et al. (2019) describe it as follows: The feedstock for the biotechnological process route is also lignocellulosic biomass and involves the same biomass pretreatment unit to obtain xylose monomers as hydrolysate. As opposed to the chemical production route, an extensive purification of the obtained monomers is not required due to the higher resilience of microorganisms towards impurities compared to the catalyst. However, higher concentrations of certain inhibitory compounds, which are formed in the pretreatment, can negatively influence the performance of the microorganisms. Hence, either detoxification of the hydrolysate or an adaption of the cell factory through engineering strategies might be necessary. The following biotechnological conversion reaction through microorganisms, also called fermentation, occurs under much milder conditions – i.e. lower pressures and temperatures – than the chemical process. Commonly used microorganisms are different yeasts, e.g., *Candida, Debaromyces or Kluyveromuces* species, or engineered cell factories, e.g., engineered *Saccharomyces cerevisiae*. The yield of xylitol from xylose reaches from 45% in wild-type microorganisms to up to 90% in engineered microorganisms. In the downstream processing, after removing the cell biomass, the water content is commonly reduced before the purification of the xylitol. For the final purification step, crystallization units are considered a viable option [17]. As mentioned, the biotechnological production of xylitol is not yet commercialized. Existing studies primarily focus on investigating the pretreatment, microorganisms, and initial studies regarding the entire process, pointing out a need for more conceptual investigations [17,23,25,26].

D.2.1.2 Xylitol Biorefinery

Based on this general evaluation, in a previous publication, the authors have conducted a study on the conceptual process design of a xylitol biorefinery – the biotechnological production of xylitol in an integrated biorefinery setup with value-added co-products. In Vollmer et al. (2022), the conceptual process design is described: xylitol is the primary product, the hemicellulosic fraction of the used lignocellulosic biomass – in this case, wheat straw – is used for the production of it. In order to utilize the feedstock to the maximum extent, the cellulosic and lignin fractions are also used for the production of respectively succinic acid and combustion for the generation of steam and electricity. The aims of this process integration step are twofold: firstly, it aims at improving the economic performance of the plant by reducing the net amount of steam and electricity needed, and secondly, to improve the sustainability potential.

After a biomass pretreatment unit, the xylitol process train is composed similarly, as described in section D.2.1: after the pretreatment follows an evaporation unit that serves the upconcentration of the hydrolysate and the removal of volatile inhibitory compounds. After that, the fermentation unit follows. In the downstream process, after the cell removal, firstly, an evaporation unit reduces the amount of water of the process stream and is followed by two crystallization units to purify the product. Also, after the biomass pretreatment, the remaining solid fraction is transferred towards an enzymatic hydrolysis unit, which fractionates the cellulosic fraction and depolymerizes it. The following process

steps are the same as for the xylitol process train. The only difference is that the produced product is succinic acid and the operation mode of the second crystallization is another cooling crystallization instead of an antisolvent crystallization. After the enzymatic hydrolysis, the lignin is transferred to combustion with an included steam and power generation unit. Furthermore, the xylitol biorefinery comprises a wastewater treatment unit [22]. The entire conceptual process flowsheet, partly excluding the auxiliary unit operations, is displayed in Figure 27.

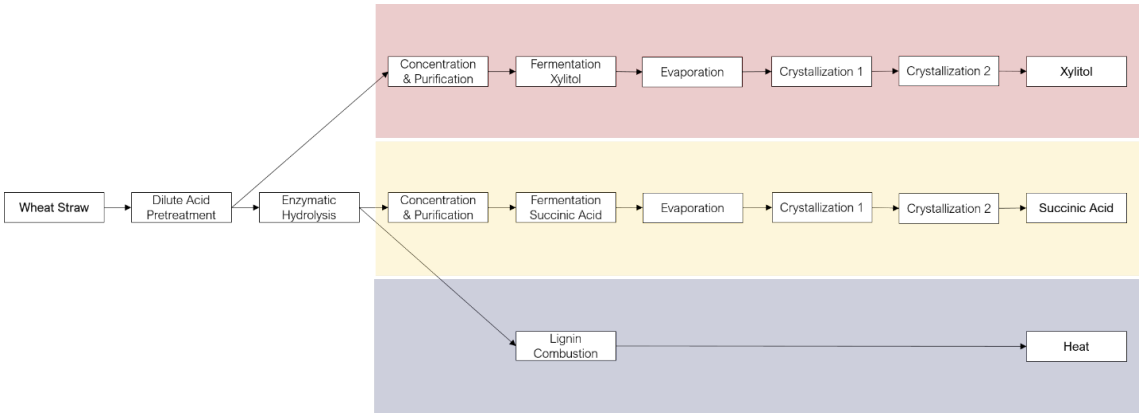

Figure 27: Conceptual process flowsheet of the xylitol biorefinery.

The entire flowsheet has been designed in a synergistic optimization-based framework for the conceptual process design of biorefineries and is thus composed of different mathematical models [21]. The framework and the models for the xylitol biorefinery are made available through a GitHub repository [27].

Noteworthy from a sustainability perspective besides the implemented process integration is that the microorganisms in the succinic acid fermentation take up CO_2 based on the process stoichiometry. Technically, this is achieved by using CO_2 from potentially fossil sources and sparging it in the fermentation reactor. Hence, the process has a net negative CO_2 balance. As mentioned in section D.1, while an extensive techno-economic analysis regarding the economic viability of the biorefinery has been performed, the sustainability potential of this biorefinery in specific, or such biorefinery concepts in general, has hardly been investigated comprehensively with several impact factors.

D.2.2 Life Cycle Assessment

According to the ISO 14040 norm, an LCA is performed in four consecutive steps. The entire procedure is described by Hauschild, Rosenbaum and Olsen (2018) as follows:

- Step 1: Definitions of Goal and Scope

Firstly, the goal of the LCA needs to be defined and formulated. That includes the intended application of the LCA, the reasons for the analysis, the target audience, and the intended way of disseminating the results. With the defined goal of the LCA at hand, also its scope needs to be defined. The scope of an LCA contains several specific elements and definitions: these are namely the analyzed system and its boundaries, its functional units, a reference flow to scale all other flows to, an overview of assumptions and limitations, the specification of allocation principles, the applied impact assessment method and information about

required data quality and the intended way of storing data. Common system boundaries are applied in different LCAs, spanning over different lengths of the life cycle. They reach from "gate-to-gate," "cradle-to-gate," and "cradle-to-grave" up to "cradle-to-cradle." The allocation procedure and the impact assessment will be specified in steps 2 and 3.

- Step 2: Life Cycle Inventory Analysis

The life cycle inventory (LCI) analysis serves to quantify all input and output flows to and from the analyzed system. Through the quantification, a comprehensive inventory of the system is obtained. The flows comprise both mass flows in all states and energy flows. Also, the allocation is performed in this step: if several products are produced, a decision on how the impacts are weighed among the products has to be taken. Typical examples are mass-based or economy-based allocation.

- Step 3: Life Cycle Impact Assessment

The third step comprises the life cycle impact assessment (LCIA), where potential impacts of the investigated system on the defined impact categories are analyzed most commonly by using an LCA software, e.g., Simapro. Firstly, these impact categories are selected, by choosing a particular LCIA method, e.g., ReCiPe [28]. Subsequently, all elements in the inventory are classified over all impact categories based on their contribution to respective environmental effects. Lastly, the LCI results are characterized via the characterization factors through multiplication, answering how much each individual element in the inventory contributes in each category. These results can be summed up for each impact category. As an additional step, the summed results are commonly normalized, based on specific reference points, e.g., an average person's contribution to the impact category, to quantitatively express if the characterization results are comparatively high or low.

- Step 4: Interpretation

Lastly, the obtained results are interpreted to conclude the sustainability impact of the analyzed system, the LCI, the LCIA, and they are assessed regarding their validity concerning the scope of the study. Additionally, complementary sensitivity analyses or other assessments can be performed to obtain a more detailed picture for the interpretation [29].

D.3 Results

D.3.1 Definitions of Goal and Scope

For the xylitol biorefinery, the goal of the LCA is to evaluate the environmental sustainability impacts in order to identify the potential environmental improvements of the xylitol production via a biotechnological process route over the existing chemical one. The function of the system is the annual production of $m = 12186\ t$ of food-grade xylitol. Since succinic acid is a value-added co-product, its production is also included in the scope of the LCA. For the sake of simplicity regarding the calculations and to include all co-products, the reference flow is hence defined as the feedstock and accounted for as $m = 1\ t$. As the xylitol biorefinery processes $m = 150{,}000\ t$ of wheat straw on an annual basis, all other streams

are normalized with that number. The scope is set as “cradle-to-gate” to enable a precise comparison to the chemical process and exclude other auxiliary factors.

D.3.2 Life Cycle Inventory Analysis

Firstly, for the LCI, a flowsheet simulation with the existing models is performed to determine all flows into and out of the system. The principal input is the wheat straw as feedstock. For the pretreatment unit, a flow of sulfuric acid is accounted. For the enzymatic hydrolysis, a flow of sodium hydroxide and the flow of the enzymes are accounted. Subsequently, additional sodium hydroxide for the neutralization and titration and ammonium sulfate as the nitrogen source for the fermentation are accounted for both fermentations. In addition, for the succinic acid fermentation, a stream of fossil CO_2 is considered. Furthermore, for the downstream process of the xylitol process train, ethanol as the antisolvent in the second crystallization step is included. Moreover, the overall process requires a net stream of process water, in addition to the internal recovery, as well as a net stream of medium-pressure steam in addition to the internally generated steam and a certain amount of electricity in addition to the internally generated electricity. Lastly, make-up streams for the cooling and chilling water in addition to the internal recovery are included.

For the transport of the wheat straw to the factory, an average distance of 100 km is assumed for a fictional location of the biorefinery in Denmark. The flows out of the biorefinery that are accounted for are the products xylitol and succinic acid. Furthermore, waste brine as a residual outflow from the wastewater treatment and ashes from the combustion process are assumed, together with biogenic CO_2 as a by-product from the fermentations. The flows are retrieved from a flowsheet simulation with the existing models and hence are secondary data and do not stem from a biorefinery directly, as this study is conceptual and no commercial process exists yet. All flows are listed in Table 19.

Table 19: Annual inputs and outputs to the system.

Inputs from technosphere			**Outputs to technosphere**		
Materials			Products		
Wheat straw (dry)	150,000	t	Xylitol	12,186	t
Sulfuric acid	17,756	t	Succinic acid	19,220	t
Enzymes	2,591	t			
Ammonium sulfate	13	t			
Sodium hydroxide	827	t			
Process water	21,419	t			
Ethanol	2,592	t			
CO_2 (fossil)	1,551	t			
Transport			Waste		
To plant	15,000,000	$t \cdot km$	Waste brine	59,576	t
			Ashes	911	t
Energy			Emissions		
Steam (MP)	1,611,300	t	CO_2 (biogen)	523	t
Electricity	78,732,000	MJ			
Cooling Water	2,407,300	t			
Chilling Water	6.77	t			

The allocation of the products for the LCA is chosen to be economy-based. All flows are normalized with the reference stream. Figure 28 illustrates all flows in and out of the system according to their unit operations

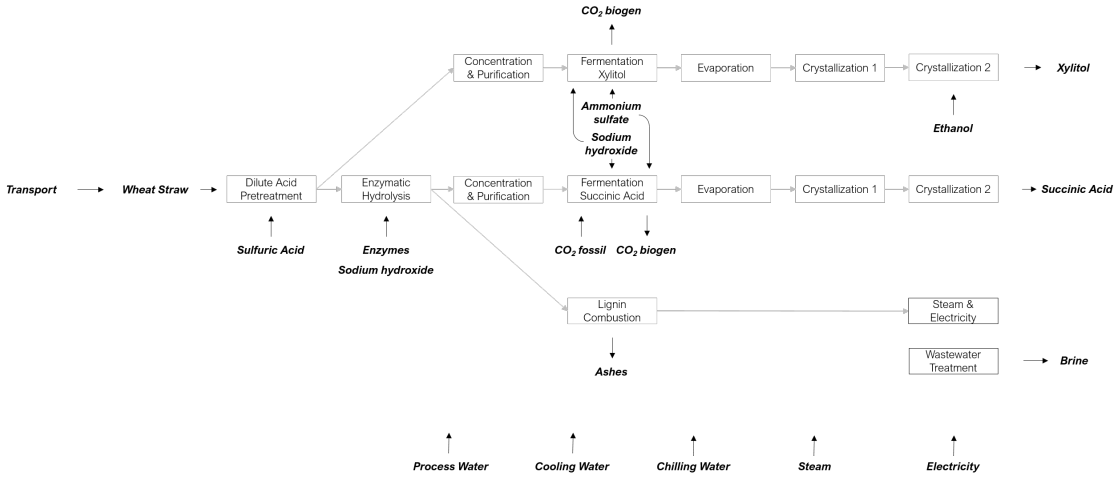

Figure 28: Conceptual process flowsheet (grey) and flows into and out of the system (black).

D.3.3 Life Cycle Impact Assessment

For the LCIA, the ReCiPe 2016 Midpoint H V1.05 method is chosen [28]. The ReCiPe method considers 18 impact categories. They are listed in Table 20.

Table 20: Impact categories of the ReCiPe 2016 Midpoint H V1.05 database.

Impact category	Unit	Category abbreviation
Global Warming	$kg\ CO_2\ eq$	GW
Ozone Depletion	$kg\ CFC-11\ eq$	OD
Ionising Radiation	$kBq\ Co-60\ eq$	IR
Ozone Formation, Human Health	$kg\ NO_x\ eq$	OH
Fine Particular Matter Formation	$kg\ PM2.5\ eq$	FP
Ozone Formation, Terrestrial Ecosystems	$kg\ NO_x\ eq$	OT
Terrestrial Acidification	$kg\ SO_2\ eq$	TA
Freshwater Eutrophication	$kg\ P\ eq$	FT
Marine Eutrophication	$kg\ N\ eq$	MT
Terrestrial Ecotoxicity	$kg\ 1{,}4-DCB$	TE
Freshwater Ecotoxicity	$kg\ 1{,}4-DCB$	FE
Marine Ecotoxicity	$kg\ 1{,}4-DCB$	ME
Human Carcinogenic Toxicity	$kg\ 1{,}4-DCB$	HC
Human Non-Carcinogenic Toxicity	$kg\ 1{,}4-DCB$	HN
Land Use	$m^2a\ crop\ eq$	LU
Mineral Ressource Scarcity	$kg\ Cu\ eq$	MR
Fossil Ressource Scarcity	$kg\ oil\ eq$	FR
Water Consumption	m^3	WC

All inventory items for the processes in Table 21 are retrieved from the Ecoinvent 3.6 database, except process numbers 3, 16, and 17. The characterized results for the enzymes

stem from a proprietary study of Novozymes [30]. The fossil and biogenic CO2 values for the GW impact are set to a standard definition ($CO_{2,fossil} = -1, CO_{2,biogenic} = 0$), as CO_2 is the reference component for the global warming impact category in the ReCiPe methodology. The respective processes for the flows are listed in Table 21.

Table 21: Processes from the Ecoinvent3 database with their according names.

Process Number	Units	Ecoinvent 3.6 database processes
1	kg	Wheat Straw at Farm/DK
2	kg	Sulfuric Acid (RER), production of sulfuric acid cut-off (S)
3	kg	Enzyme, Cellulase, Novozymes Celluclast/kg/RER
4	kg	Ammonium Sulfate (RER), ammonium sulfate production cut-off (S)
5	kg	Sodium hydroxide, without water, in 50% solution state {RER}| chlor-alkali electrolysis, diaphragm cell | Cut-off, S
6	kg	Water, deionised {Europe without Switzerland}| market for water, deionised | Cut-off, S
7	kg	Ethanol, without water, in 99.7% solution state, from ethylene {RER}| ethylene hydration | Cut-off, S
8	kg	Carbon dioxide, liquid {RER}| production | Cut-off, S
9	kg	Sodium chloride, brine solution {RER}| production | Cut-off, S
10	kg	Ash, from combustion of bagasse from sugarcane {GLO}| market for ash, from combustion of bagasse from sugarcane | Cut-off, S
11	$t \cdot km$	Transport, freight, lorry >32 metric ton, EURO4 {RER}| transport, freight, lorry >32 metric ton, EURO4 | Cut-off, S
12	kg	Steam, in chemical industry {RER}| production | Cut-off, S
13	MJ	Electricity, high voltage {DK}| electricity production, wind, >3MW turbine, onshore | Cut-off, S
14	kg	Tap water {RER}| market group for | Cut-off, S
15	kg	Tap water {RER}| market group for | Cut-off, S
16	kg	CO2 fossil
17	kg	CO2 biogen

The contribution of each process to each impact category is calculated and then scaled to the reference flow. The results of calculating the impacts for each process number with the characterization factors are listed in Table 22. The results of the LCIA are displayed in Figure 29. It becomes apparent, that only six out of seventeen processes have a significant relative impact in these categories, namely the wheat straw, the transport, the waste brine, the steam, the electricity and the cooling water.

The normalization of the LCIA results is done by the ReCiPE Midpoint World (2010) H (V1.05), corresponding to the sustainability impact of one global citizen. The results are displayed in Figure 30.It becomes visible that for most significant impacts when normalized are the ones for freshwater ecotoxicity, marine ecotoxicity, human carcinogenic toxicity and water consumption. The economic allocation with a market price of 4.81 $\$/kg$ and 3.20 $\$/kg$ for xylitol and succinic acid, the economic allocation is distributed 48.8% of the impact on the xylitol production and 51.2% on the succinic acid production.

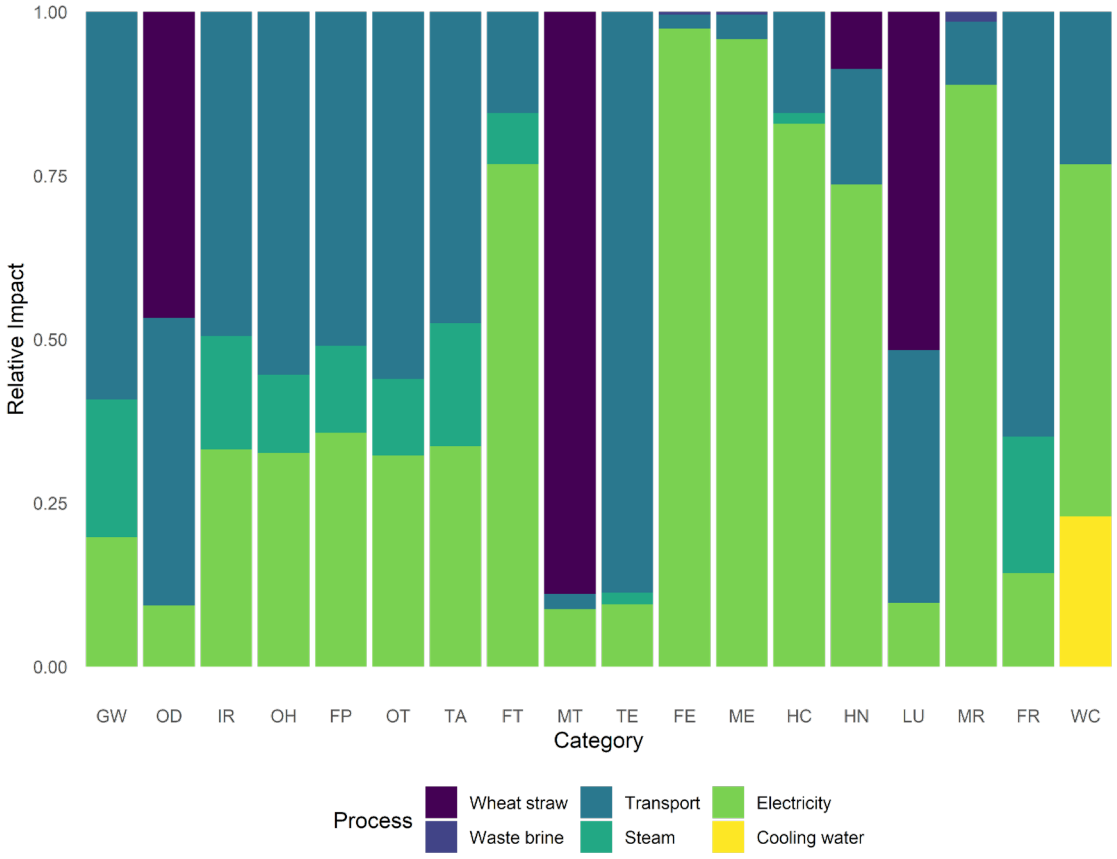

Figure 29: LCIA results of the characterization given in percentages for each category.

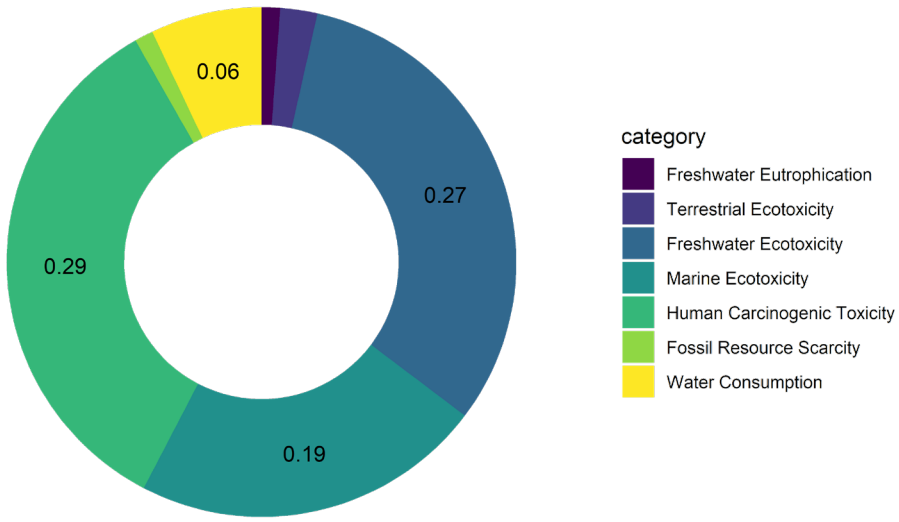

Figure 30: Results of the LCIA with applied normalization.

Table 22: Results for the characterization in the LCIA

	GW	OD	IR	OH	FP	OT	TA	FT	MT	TE	FE	ME	HC	HN	LU	MR	FR	WC
	$kg\ CO_2eq$	$kg\ CFC - 11\ eq$	$kBq\ Co - 60\ eq$	$kg\ NO_xeq$	$kg\ PM2.5\ e$	$kg\ NO_x\ eq$	$kg\ SO_2\ eq$	$kg\ P\ eq$	$kg\ N\ eq$	$kg\ 1,4 - DCB$	$kg\ 1,4 - DCB$	$kg\ 1,4 - DCB$	$kg\ 1,4 - DCB$	$kg\ 1,4 - DCB$	$m^2a\ crop\ e$	$kg\ Cueq$	$kg\ oil\ eq$	m^3
1	2,84 e-1	6,73 e-6	6,21 e-4	5,32 e-4	5,39 e-4	5,36 e-4	3,92 e-3	1,05 e-4	2,10 e-3	2,26 e-1	1,75 e-2	3,61 e-3	1,15 e-4	2,81	9,17 e-1	1,21 e-4	2,82 e-2	4,47 e-3
2	1,98 e-2	1,21 e-8	1,01 e-3	1,30 e-4	2,21 e-4	1,33 e-4	7,12 e-4	2,33 e-5	9,65 e-7	4,81 e-1	3,92 e-3	5,75 e-3	1,91 e-3	1,99 e-1	9,51 e-4	4,71 e-4	1,70 e-2	2,50 e-3
3	7,06 e-2	0,00	0,00	1,26 e-5	7,65 e-5	2,03 e-5	2,64 e-4	5,34 e-6	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,98 e-2	3,60 e-4
4	1,03 e-4	3,19 e-11	9,01 e-6	1,84 e-7	2,75 e-7	1,94 e-7	9,67 e-7	3,99 e-8	1,03 e-7	5,46 e-4	1,85 e-5	2,15 e-5	1,77 e-5	2,47 e-4	3,52 e-6	1,27 e-6	3,88 e-5	3,08 e-6
5	7,24 e-3	7,70 e-9	7,91 e-4	1,90 e-5	1,56 e-5	1,92 e-5	2,62 e-5	3,55 e-6	3,62 e-7	1,38 e-2	4,01 e-4	5,22 e-4	5,96 e-4	8,43 e-3	1,66 e-4	2,30 e-5	1,80 e-3	1,90 e-4
6	6,49 e-5	6,24 e-11	5,90 e-6	1,35 e-7	1,67 e-7	1,38 e-7	4,20 e-7	2,39 e-8	2,76 e-9	2,18 e-4	4,04 e-6	5,32 e-6	9,13 e-6	8,27 e-5	1,44 e-6	4,34 e-7	1,83 e-5	1,50 e-4
7	1,94 e-2	1,77 e-7	5,80 e-4	7,75 e-5	5,58 e-5	8,09 e-5	1,99 e-4	6,11 e-6	3,81 e-5	4,67 e-2	1,08 e-3	9,30 e-4	1,16 e-3	2,98 e-2	2,69 e-2	8,71 e-5	4,04 e-3	2,74 e-3
8	7,62 e-3	1,53 e-9	8,97 e-4	7,15 e-6	4,89 e-6	7,30 e-6	1,21 e-5	2,33 e-6	4,88 e-7	4,39 e-2	4,08 e-4	5,17 e-4	4,84 e-4	6,71 e-3	1,28 e-4	2,71 e-5	1,34 e-3	5,55 e-5
9	9,95 e-2	3,92 e-8	1,15 e-2	2,60 e-4	2,14 e-4	2,65 e-4	4,22 e-4	7,23 e-5	3,97 e-6	4,79 e-1	2,79 e-2	3,54 e-2	2,94 e-2	3,98 e-1	6,63 e-3	2,05 e-3	2,44 e-2	-1,11 e-4
10	4,07 e-6	1,84 e-12	7,92 e-8	2,21 e-8	5,83 e-9	2,26 e-8	1,35 e-8	3,46 e-10	3,05 e-11	7,64 e-5	8,19 e-8	1,50 e-7	2,41 e-7	2,92 e-6	2,25 e-7	7,66 e-9	1,41 e-6	7,64 e-9
11	8,66	6,32 e-6	2,09 e-1	1,73 e-2	9,05 e-3	1,83 e-2	1,76 e-2	5,90 e-4	5,43 e-5	2,18 e2	1,43 e-1	3,06 e-1	4,63 e-1	5,69	6,85 e-1	1,33 e-2	3,21	1,64 e-2
12	3,08	7,49 e-7	7,32 e-2	3,72 e-3	2,36 e-3	3,82 e-3	6,97 e-3	2,98 e-4	2,04 e-5	4,41	1,42 e-2	2,46 e-2	4,89 e-2	5,28 e-1	2,66 e-2	7,20 e-4	1,03	3,22 e-3
13	2,89	1,34 e-6	1,40 e-1	1,02 e-2	6,34 e-3	1,05 e-2	1,25 e-2	2,93 e-3	2,08 e-4	2,34 e1	6,50	7,87	2,48	2,38 e1	1,73 e-1	1,23 e-1	7,08 e-1	3,78 e-2
14	5,47 e-3	2,37 e-9	1,73 e-3	1,29 e-5	9,16 e-6	1,33 e-5	2,03 e-5	3,97 e-6	3,78 e-7	1,25 e-2	2,80 e-4	3,85 e-4	4,30 e-3	6,63 e-3	1,41 e-4	6,34 e-5	1,45 e-3	1,61 e-2
15	1,54 e-8	6,67 e-15	4,86 e-9	3,64 e-11	2,58 e-11	3,74 e-11	5,72 e-11	1,12 e-11	1,06 e-12	3,51 e-8	7,88 e-10	1,08 e-9	1,21 e-8	1,86 e-8	3,97 e-10	1,78 e-10	4,07 e-9	4,54 e-8
16	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
17	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Total	**25.3**	**0.00**	**2.68**	**0.08**	**0.04**	**0.08**	**0.01**	**0.00**	**6.67**	**82.3**	**2.57**	**0.40**	**7.62**	**16.1**	**25.3**	**0.00**	**2.68**	**0.08**

D.3.4 Sensitivity Analysis on the LCIA Results

For the sensitivity analysis of the characterization results of the LCIA, all inputs, as indicated in Table 19, are considered randomly distributed by $\pm 15\%$ around their nominal value. With the flowsheet model, $N = 1000$ simulations are performed to obtain the characterization results with the according characterization factors. Using the easyGSA toolbox, applying a variance-based sensitivity analysis, the first-order and the total sensitivity indices are calculated. A detailed description of the sensitivity analysis methodology and the presented toolbox can be found in [31]. The results for the total sensitivity index are illustrated in Figure 31Figure 31.

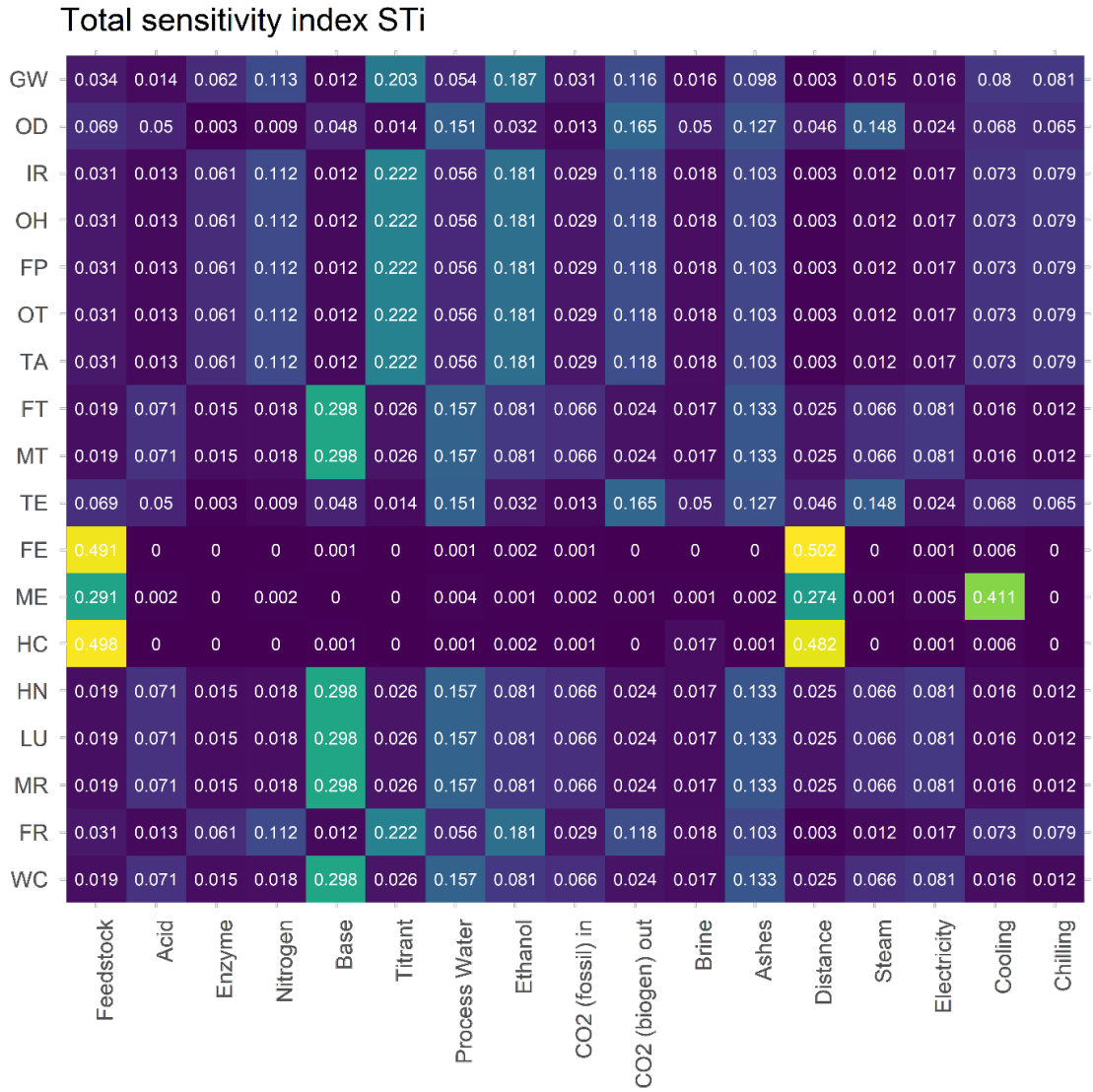

Figure 31: Heat map of the first-order and total sensitivity indices for all process numbers and all impact categories.

It is visible that the electricity, which influences the results of the LCIA the most, shows only low sensitivity indices. The amount of feedstock, the transportation distance, and the amount of cooling water influence the freshwater and marine ecotoxicity and the human carcinogenic potential most. For the feedstock and the transportation, this is explainable

due to the impact that the agricultural process of growing and harvesting wheat straw has, equally to longer transportation routes. The effect on the marine ecotoxicity of the higher use of cooling water can also be attributed to the use of electricity in the process.

D.3.5 Interpretation of Results

Given the results displayed in Figure 4, it becomes evident that the impact of the xylitol biorefinery on global warming is minimal. This is in agreement with most studies regarding biorefineries and also one of the main assumptions regarding bio-based processes in general, which is why they are promoted as more sustainable in the first place. However, while most impacts are relatively low, the impacts on freshwater toxicity, marine ecotoxicity, and human carcinogenic potential are substantially higher. This is a finding, which is often seen.[29,32]. Analyzing those three impact categories in detail in Figure 3, it is visible that more than 80% of the impact is associated with electricity. Given that the electricity is chosen to be generated by wind turbines, which are supposedly more sustainable and seen as a crucial element as part of national and global renewable energies, this seems initially contradictory. Assessing the process of creating electric energy through wind turbines in the Ecoinvent3 database shows that the major sustainability impact derives from the use of copper in the wind turbine generators. Although copper itself is a micronutrient for many species, the mining process of copper involves significant amounts of toxic chemicals, which, if not discarded correctly, can damage aquatic organisms [33–35]. This issue is known and is a problem for the environment [36–38].

In order to illustrate these results, an example shall be given: taking into account the total characterized impact and the economic allocation, the production of 1 kg of xylitol equals 12.3 kg of CO_2 emissions, which is approximately half the emissions of the production of 1 kg of beef, which equals $20 - 40$ kg of CO_2 emissions, and around four times the emission of the production of 1 kg of a chocolate bar, which equals 3.5 kg of CO_2 emissions [39].

D.3.6 Comparison to the Current Production Process

Ultimately, to compare the biorefinery process with the existing chemical process, an LCA of Danisco Dupont regarding their process based on birch trees is taken from a whitepaper [39]. In the white paper, both process setups are described and compared to the other commercial processes that utilizes corncob instead of the side stream of a paper mill. This is a realization of process integration performed for the process of Dupont, improving the sustainability metrics [39]. The used LCIA methodology in (Dupont 2012) is IMPACT2002+ [40]. Hence, the results of the LCIA for the xylitol biorefinery are also generated through this methodology in addition to the previously shown results by the ReCiPe methodology. The summed impact factors for each category for the entire process are listed for Dupont's process, the corncob-based process, and the xylitol biorefinery are reported in Table 5. The set reference flow for the study conducted by Dupont is $m = 1t$ of xylitol, with the same purity requirements, and accounting also for the material and energy flows and the emissions caused by its production as cradle-to-gate [39].

Firstly, it becomes apparent that the impacts of the xylitol biorefinery show similar magnitudes for all impact categories compared to both chemical processes. Secondly, for all impact categories, the impacts of the xylitol biorefinery are higher than for the Dupont process, which again shows lower impacts in all categories than the corncob process. This can be attributed to the fact that Dupont's process is integrated into a pulp and paper mill

and can decrease the sustainability impact through this significantly by this. For a fair comparison to the xylitol biorefinery, this integration is investigated in specific by assessing the impacts only for the xylitol process train and disregarding the impacts associated with the succinic acid. For the energy flows, it is assumed that the net flows are reduced as the process integration for steam and electricity is still yielding identical amounts regardless of the succinic acid production but are not used for the succinic acid production. The flows that are exclusively needed for the succinic acid production are subtracted for the material flows. For the cooling water, process water, and waste brine, the flows are reduced by 50% as they are only calculated for the entire biorefinery and cannot be split up precisely.

Table 23: LCIA of the results for the impacts in all categories for the Dupont process, the corncob process, and the xylitol biorefinery.

Impact category	**Unit**	**Dupont process** [38]	**Corncob process** [38]	**Xylitol biorefinery**
Aquatic acidification	$kg\ SO_2\ eq$	0.00873	0.334	**0.0551**
Aquatic ecotoxicity	$kg\ TEG\ water$	599	60600	**2463**
Aquatic eutrophication	$kg\ PO_4\ P\ limited$	0.00119	0.0512	**0.0023**
Carcinogens	$kg\ C_2H_3Cl\ eq$	0.04119	0.283	**0.3075**
Global warming	$kg\ CO_2\ eq$	3.59	38.6	**14.548**
Ionising radiation	$Bq\ C-14\ eq$	51.1	477	**108.6**
Land occupation	$M2Org.Arable$	0.0487	9.1	**1.8854**
Mineral extraction	$MJ\ surplus$	0.0623	0.435	**1.7066**
Non-carcinogens	$kg\ C_2H_3Cl\ eq$	0.0335	1.18	**1.2547**
Non-renewable energy	$MJ\ primary$	66.8	454	**238.655**
Ozone layer depletion	$kg\ CFC-11\ eq$	0.000563	0.00417	**2.24e-6**
Respiratory inorganics	$kg\ PM2.5\ eq$	0.00152	0.0433	**0.0134**
Respiratory organics	$kg\ C_2H_4\ eq$	0.000991	0.00606	**0.0084**
Terrestial acid/nutria	$kg\ SO_2\ eq$	0.034	1.06	**0.2283**
Terrestrial ecotoxicity	$kg\ TEG\ soil$	150	2660	**1760**

Table 24: LCIA of the results for the impacts in all categories for the Dupont process and the xylitol process train of the xylitol biorefinery.

Impact category	Unit	Dupont process [38]	Xylitol process train	Dupont/Xylitol
Aquatic acidification	$kg\ SO_2\ eq$	0.00873	0.0342	**25.5%**
Aquatic ecotoxicity	$kg\ TEG\ water$	599	1391	**43.1%**
Aquatic eutrophication	$kg\ PO_4\ P\ limited$	0.00119	0.0013	**91.5%**
Carcinogens	$kg\ C_2H_3Cl\ eq$	0.04119	0.0774	**53.2%**
Global warming	$kg\ CO_2\ eq$	3.59	10.7783	**33.3%**
Ionising radiation	$Bq\ C-14\ eq$	51.1	83.3113	**61.3%**
Land occupation	$M2Org.Arable$	0.0487	1.7175	**2.84%**
Mineral extraction	$MJ\ surplus$	0.0623	0.0838	**74.3%**
Non-carcinogens	$kg\ C_2H_3Cl\ eq$	0.0335	0.5223	**6.41%**
Non-renewable energy	$MJ\ primary$	66.8	186.369	**35.8%**
Ozone layer depletion	$kg\ CFC-11\ eq$	0.000563	1.94e-6	**-290%**
Respiratory inorganics	$kg\ PM2.5\ eq$	0.00152	0.0082	**18.5%**
Respiratory organics	$kg\ C_2H_4\ eq$	0.000991	0.0062	**16.0%**
Terrestial acid/nutria	$kg\ SO_2\ eq$	0.034	0.1512	**22.5%**
Terrestrial ecotoxicity	$kg\ TEG\ soil$	150	1491	**10.0%**

Although the metrics for the xylitol process train only improved the overall sustainability impact, the impacts of the Dupont process are still generally lower, but all in comparable orders of magnitude. However, it has to be pointed out that the biotechnological process route for xylitol production, as investigated in this study, is still conceptual. In the scale-up and commercialization process of new plants, potentials for optimizing the process both from an economic and environmental perspective are commonly leveraged [41,42].

D.4 Conclusions

In the scope of this manuscript, a systematic LCA is performed for an integrated xylitol biorefinery with value-added co-products. The goal of the LCA is to evaluate the environmental impacts of the xylitol production via a biotechnological process route in an integrated biorefinery setup with value-added co-products, as well as to compare the impacts to the existing chemical xylitol production process. The LCA is performed according to the four-step procedure as defined in ISO 14040. The xylitol biorefinery utilizes wheat straw as renewable feedstock. It involves a pretreatment unit and an enzymatic hydrolysis unit in the upstream process, two fermentation units for the biotechnological conversion of the feedstock to xylitol and succinic acid, respectively, and in each downstream process an evaporation unit and two crystallization units for the purification of the products. Additionally, auxiliary unit operations, e.g., a combustion unit for lignin to generate steam and electricity as a process integration measure, and a wastewater treatment unit are considered. The results of the LCA show that while the process generally shows low impacts, particularly with regards to greenhouse gas emissions, other categories show higher impacts. These results are compared to two types of chemical production processes. While the xylitol biorefinery process has lower impacts than the standard chemical conversion process, it has higher impacts than a process integrated into a pulp and paper mill operated by Dupont.

This leads to the conclusion that the biotechnological production process per se is not more sustainable than the chemical one. When comparing it to the existing chemical process, the sustainability potential becomes apparent, while this fact is not directly visible compared to the Dupont process. Compared to the classic chemical one, the Dupont process employs the same reaction system but heavily employs process integration measures by being located adjoint to a pulp and paper mill. The white paper of Dupont does not clearly indicate to which extent the process integration and parts of the existing pulp and paper mill are allocated to the system that is analyzed in the LCA. Since the presented LCA fully includes the process integration, it can be assumed that the presented LCA yields higher impacts than the LCA of Dupont [39]. Furthermore, the LCA of Dupont uses primary data, while the presented LCA uses secondary data, as the process is still conceptual and not commercialized. This needs further investigation during the commercialization phase to optimize the flows and reduce the sustainability impact, as suggested in the previous chapter. Nonetheless, it can be concluded that process integration shows a significant effect on the sustainability potential. Furthermore, all three mentioned processes use lignocellulosic biomass as feedstock.

As a general conclusion, it is also important to point out that while the commonly regarded impact on greenhouse gas emissions for the xylitol biorefinery is marginal, other impacts are significantly higher, namely freshwater ecotoxicity, marine ecotoxicity, and human carcinogenic toxicity. This is due to the use of electric energy from windmills induced by copper in the windmills. The mining of copper has a considerable impact on these impact categories, which translates directly into the results of the LCA of the biorefinery, despite wind energy being considered as renewable and hence more sustainable due to its low impact on global warming. In order to reduce this impact directly, the use of other sources of electricity, e.g., solar or hydroelectricity, can be a solution, but this depends on the

location of the plant and can possibly lead to other increased impacts in different categories. As a general conclusion, it can be stated that sustainable processes are a multi-layered issue, and a straightforward sustainable solution does not commonly exist; hence further research on biorefinery concepts in the different impact categories is necessary.

Lastly, the presented analysis shows the importance of systematic LCAs for the sustainability assessment of processes. While the apparent lower impact on greenhouse gas emissions is visible, other priorly unexpected impacts can change the entire sustainability assessment of such processes. Hence, the systematic impacts of all processes are crucial to compare the improved sustainability impact to existing processes. Complementarily, the combination with a techno-economic assessment gives additional perspectives for decision-making on the overall feasibility and potential benefits of the investigated process [22,43,44]. Particularly for novel biotechnological processes, or biorefineries, such comparisons allow for a quantified statement and assist the expedited transition towards genuinely sustainable processes, as postulated by the 2030 agenda for sustainable development of the United Nations [1].

Acknowledgments

The authors would like to express their gratitude to the Novo Nordisk Foundation (Grant No. NNF17SA0031362) for funding the Fermentation-Based Biomanufacturing Initiative, of which this project is a part.

References

[1] United Nations, Transforming our world: The 2030 agenda for sustainable development, 2015. https://sdgs.un.org/sites/default/files/publications/21252030 Agenda for Sustainable Development web.pdf (accessed March 1, 2021).

[2] M. Lieder, A. Rashid, Towards circular economy implementation: A comprehensive review in context of manufacturing industry, J. Clean. Prod. 115 (2016) 36–51. https://doi.org/10.1016/j.jclepro.2015.12.042.

[3] J. Singh, I. Ordoñez, Resource recovery from post-consumer waste: important lessons for the upcoming circular economy, J. Clean. Prod. 134 (2016) 342–353. https://doi.org/10.1016/j.jclepro.2015.12.020.

[4] E. De Jong, H. Stichnothe, G. Bell, H. Jorgensen, Bio-Based Chemicals: A 2020 Update, 2020. https://task42.ieabioenergy.com/wp-content/uploads/sites/10/2020/02/Bio-based-chemicals-a-2020-update-final-200213.pdf.

[5] M. Gavrilescu, Y. Chisti, Biotechnology - A sustainable alternative for chemical industry, Biotechnol. Adv. 23 (2005) 471–499. https://doi.org/10.1016/j.biotechadv.2005.03.004.

[6] Y. Lokko, M. Heijde, K. Schebesta, P. Scholtès, M. Van Montagu, M. Giacca, Biotechnology and the bioeconomy—Towards inclusive and sustainable industrial development, N. Biotechnol. 40 (2018) 5–10. https://doi.org/10.1016/j.nbt.2017.06.005.

[7] J.M. Woodley, Towards the sustainable production of bulk-chemicals using biotechnology, N. Biotechnol. 59 (2020) 59–64. https://doi.org/10.1016/j.nbt.2020.07.002.

[8] F. Cherubini, The biorefinery concept: Using biomass instead of oil for producing energy and chemicals, Energy Convers. Manag. 51 (2010) 1412–1421. https://doi.org/10.1016/j.enconman.2010.01.015.

[9] F. Cherubini, G. Jungmeier, M. Wellisch, T. Willke, I. Skiadas, R. van Ree, E. de Jong, Toward a common classification approach for biorefinery systems, Biofuels, Bioprod. Biorefining. 3 (2009) 534–546. https://doi.org/10.1002/bbb.172.

[10] A.J.J. Straathof, S.A. Wahl, K.R. Benjamin, R. Takors, N. Wierckx, H.J. Noorman, Grand Research Challenges for Sustainable Industrial Biotechnology, Trends Biotechnol. 37 (2019) 1042–1050. https://doi.org/10.1016/j.tibtech.2019.04.002.

[11] S.N. Naik, V. V. Goud, P.K. Rout, A.K. Dalai, Production of first and second generation biofuels: A comprehensive review, Renew. Sustain. Energy Rev. 14 (2010) 578–597. https://doi.org/10.1016/j.rser.2009.10.003.

[12] M.W. Rosegrant, S. Msangi, T. Sulser, R. Valmonte-santos, Bioenergy and Agriculture: Promises and Challenges. Biofuels and the Global Food Balance, 2020 Vis. Briefs. (2006) 2005–2006. https://ideas.repec.org/p/fpr/2020br/14(3).html (accessed January 4, 2022).

[13] Y. Liu, Y. Lyu, J. Tian, J. Zhao, N. Ye, Y. Zhang, L. Chen, Review of waste biorefinery

development towards a circular economy: From the perspective of a life cycle assessment, Renew. Sustain. Energy Rev. 139 (2021) 110716. https://doi.org/10.1016/J.RSER.2021.110716.

[14] F. Cherubini, G. Jungmeier, LCA of a biorefinery concept producing bioethanol, bioenergy, and chemicals from switchgrass, Int. J. Life Cycle Assess. 15 (2010) 53–66. https://doi.org/10.1007/S11367-009-0124-2/FIGURES/8.

[15] F. Cherubini, S. Ulgiati, Crop residues as raw materials for biorefinery systems – A LCA case study, Appl. Energy. 87 (2010) 47–57. https://doi.org/10.1016/J.APENERGY.2009.08.024.

[16] T. Werpy, G. Petersen, Top Value Added Chemicals from Biomass Volume I, 2004. https://doi.org/10.2172/15008859.

[17] A.F. Hernández-Pérez, P.V. de Arruda, L. Sene, S.S. da Silva, A. Kumar Chandel, M. das G. de Almeida Felipe, Xylitol bioproduction: state-of-the-art, industrial paradigm shift, and opportunities for integrated biorefineries, Crit. Rev. Biotechnol. 39 (2019) 924–943. https://doi.org/10.1080/07388551.2019.1640658.

[18] S.S. Da Silva, A.K. Chandel, D-Xylitol: Fermentative production, application and commercialization, Springer-Verlag Berlin Heidelberg, 2012. https://doi.org/10.1007/978-3-642-31887-0.

[19] Y. Delgado Arcaño, O.D. Valmaña García, D. Mandelli, W.A. Carvalho, L.A. Magalhães Pontes, Xylitol: A review on the progress and challenges of its production by chemical route, Catal. Today. 344 (2020) 2–14. https://doi.org/10.1016/j.cattod.2018.07.060.

[20] IMARC, Xylitol Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2021-2026, 2021.

[21] N.I. Vollmer, R. Al, K. V. Gernaey, G. Sin, Synergistic optimization framework for the process synthesis and design of biorefineries, Front. Chem. Sci. Eng. 16 (2022) 251–273. https://doi.org/10.1007/s11705-021-2071-9.

[22] N.I. Vollmer, K. V. Gernaey, G. Sin, Conceptual Process Design of an Integrated Xylitol Biorefinery with Value-Added Co-Products, Front. Chem. Eng. (2022).

[23] T.L. De Albuquerque, I.J. Da Silva, G.R. De MacEdo, M.V.P. Rocha, Biotechnological production of xylitol from lignocellulosic wastes: A review, Process Biochem. 49 (2014) 1779–1789. https://doi.org/10.1016/j.procbio.2014.07.010.

[24] N.I. Vollmer, J.L.S.P. Driessen, C.K. Yamakawa, K. V. Gernaey, S.I. Mussatto, G. Sin, Model development for the optimization of operational conditions of the pretreatment of wheat straw, Chem. Eng. J. 430 (2022) 133106. https://doi.org/10.1016/j.cej.2021.133106.

[25] L.V. Rao, J.K. Goli, J. Gentela, S. Koti, Bioconversion of lignocellulosic biomass to xylitol: An overview, Bioresour. Technol. 213 (2016) 299–310. https://doi.org/10.1016/j.biortech.2016.04.092.

[26] D. Dasgupta, S. Bandhu, D.K. Adhikari, D. Ghosh, Challenges and prospects of xylitol production with whole cell bio-catalysis: A review, Microbiol. Res. 197 (2017) 9–21. https://doi.org/10.1016/j.micres.2016.12.012.

[27] N.I. Vollmer, Xylitol Biorefinery LCA, (2022).

https://github.com/NikolausVollmer/Xylitol-Biorefinery-LCA.

[28] M.A.J. Huijbregts, Z.J.N. Steinmann, P.M.F. Elshout, G. Stam, F. Verones, M. Vieira, M. Zijp, A. Hollander, R. van Zelm, ReCiPe2016: a harmonised life cycle impact assessment method at midpoint and endpoint level, Int. J. Life Cycle Assess. 22 (2017) 138–147. https://doi.org/10.1007/s11367-016-1246-y.

[29] M.Z. Hauschild, R.K. Rosenbaum, S.I. Olsen, Life Cycle Assessment: Theory and Practice, Life Cycle Assess. Theory Pract. (2017) 1–1216. https://doi.org/10.1007/978-3-319-56475-3.

[30] P.H. Nielsen, K.M. Oxenbøll, H. Wenzel, Enzyme Products LCA Case Studies 432 LCA Case Studies Cradle-to-Gate Environmental Assessment of Enzyme Products Produced Industrially in Denmark by Novozymes A/S, Int J LCA. 12 (2007) 432–438. http://dx.doi.org/10.1065/lca2006.08.265.1.

[31] R. Al, C.R. Behera, A. Zubov, K. V. Gernaey, G. Sin, Meta-modeling based efficient global sensitivity analysis for wastewater treatment plants – An application to the BSM2 model, Comput. Chem. Eng. 127 (2019) 233–246. https://doi.org/10.1016/j.compchemeng.2019.05.015.

[32] S. Bello, C. Ríos, G. Feijoo, M.T. Moreira, Comparative evaluation of lignocellulosic biorefinery scenarios under a life-cycle assessment approach, Biofuels, Bioprod. Biorefining. 12 (2018) 1047–1064. https://doi.org/10.1002/BBB.1921.

[33] M. Olivares, R. Uauy, Copper as an essential nutrient, Am. J. Clin. Nutr. 63 (1996) 791S-796S. https://doi.org/10.1093/AJCN/63.5.791.

[34] M. Fuentes, M. Negrete, S. Herrera-León, A. Kraslawski, Classification of indicators measuring environmental sustainability of mining and processing of copper, Miner. Eng. 170 (2021) 107033. https://doi.org/10.1016/J.MINENG.2021.107033.

[35] Y.M. Nor, Ecotoxicity of copper to aquatic biota: A review, Environ. Res. 43 (1987) 274–282. https://doi.org/10.1016/S0013-9351(87)80078-6.

[36] Z. Lyu, J. Chai, Z. Xu, Y. Qin, Environmental Impact Assessment of Mining Activities on Groundwater: Case Study of Copper Mine in Jiangxi Province, China, J. Hydrol. Eng. 24 (2018) 05018027. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001739.

[37] W.P. Covre, S.J. Ramos, W.V. da S. Pereira, E.S. de Souza, G.C. Martins, O.M.M. Teixeira, C.B. do Amarante, Y.N. Dias, A.R. Fernandes, Impact of copper mining wastes in the Amazon: Properties and risks to environment and human health, J. Hazard. Mater. 421 (2022) 126688. https://doi.org/10.1016/J.JHAZMAT.2021.126688.

[38] J.C. Castilla, E. Nealler, Marine environmental impact due to mining activities of El Salvador copper mine, Chile, Mar. Pollut. Bull. 9 (1978) 67–70. https://doi.org/10.1016/0025-326X(78)90451-4.

[39] DuPont, XIVIA™ Xylitol White Paper, 2012. http://www.danisco.com/fileadmin/user_upload/danisco/documents/products/2e_XIVIA_White_Paper.pdf.

[40] O. Jolliet, M. Margni, R. Charles, S. Humbert, J. Payet, G. Rebitzer, R. Rosenbaum, IMPACT 2002+: A new life cycle impact assessment methodology, Int. J. Life Cycle Assess. 2002 86. 8 (2003) 324–330. https://doi.org/10.1007/BF02978505.

[41] F. Piccinno, R. Hischier, S. Seeger, C. Som, From laboratory to industrial scale: a scale-up framework for chemical processes in life cycle assessment studies, J. Clean. Prod. 135 (2016) 1085–1097. https://doi.org/10.1016/j.jclepro.2016.06.164.

[42] J.A. Bergerson, A. Brandt, J. Cresko, M. Carbajales-Dale, H.L. MacLean, H.S. Matthews, S. McCoy, M. McManus, S.A. Miller, W.R. Morrow, I.D. Posen, T. Seager, T. Skone, S. Sleep, Life cycle assessment of emerging technologies: Evaluation techniques at different stages of market and technical maturity, J. Ind. Ecol. 24 (2020) 11–25. https://doi.org/10.1111/jiec.12954.

[43] Ó. Ögmundarson, S. Sukumara, M.J. Herrgård, P. Fantke, Combining Environmental and Economic Performance for Bioprocess Optimization, Trends Biotechnol. 38 (2020) 1203–1214. https://doi.org/10.1016/j.tibtech.2020.04.011.

[44] E.T. Grasa, Ó. Ögmundarson, H.N. Gavala, S. Sukumara, Commodity chemical production from third-generation biomass: a techno-economic assessment of lactic acid production, Biofuels, Bioprod. Biorefining. 15 (2021) 257–281. https://doi.org/10.1002/bbb.2160.

Afterword

After having spent three years and three months with the preceding 222 pages, I want to share some personal thoughts about the thesis and the entire PhD for this very last page of the thesis – to those who did enjoy reading this up to here and for myself, in case I ever should dare to touch this book again after submitting it in a couple of minutes.

Being inspired as a little kid by the infamous German TV show "Löwenzahn," which literally was the most important part of my Sundays, the idea of pursuing a career as an engineer was rather evident. The inspiring curiosity and the simplicity with which a funny old man in dungarees living in a trailer questioned the world and vigorously pointed out how important it is to take care of the environment was certainly a good time investment.

While any PhD project, and in particular also mine, lives off the idea to "do something to improve the world," it takes a little bit more than just mere inspiration to commit to it for three years and let fall short of many other things, even home and friends.

Even worse, when academic reality really hits after a while and one notices that one's immediate work most certainly will change the world for the better. Rather the opposite is true, namely that academia in its most extreme forms could be paraphrased as a pool of people where nobody really cares. It is also rather evident that this is not the entire truth, and many scientists, also the vast majority of scientists and friends in my surroundings, about whom I can say this with certainty and confidence and certainty, are, in fact, integer and true scientists.

Nonetheless, the immediate consequences of this – to not take science as religion and to not fall too deeply in love with one's own research project – take some time to be processed. However, once accepted, this actually means freedom for oneself and which time in one's life would actually be better to explore, learn and try out other researchers' ideas, tools, and concepts than a PhD. This liberty in interest will most likely never come back.

Hence, from my very personal point of view, I can say with confidence that I have to the largest extent enjoyed this time, despite some long workdays and imminent deadlines, uncertainties regarding the COVID-19 pandemic, and others. In fact, apart from the obvious knowledge gain, the personal development that one undergoes when doing a PhD is at the end certainly equally valuable, if not even more.

While the outcomes of my PhD might not immediately revolutionize the world of biomanufacturing, the personal development will surely be helpful in the future, and maybe in a hundred years from now, somebody opens the first page of this book, finds inspiration in it, and comes up with another scientific contribution that means a small step towards the right direction.

In that regard, it might be best to follow the advice of Hannah Arendt, one of the most famous German political philosophers: "Prepare for the worst, expect for the best, and take what comes."