



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich



Safety and
Environmental
Technology Group

Institute for Chemical and Bioengineering

Master Thesis

Reducing Complexity — Efficient Modeling for Biorefinery Concepts

Author:

Sudharshan Ravi, B. Tech

Supervisor:

Prof. Dr. Konrad Hungerbühler

“The scientific man does not aim at an immediate result. He does not expect that his advanced ideas will be readily taken up. His work is like that of the planter – for the future. His duty is to lay the foundation for those who are to come, and point the way. He lives and labors and hopes.”

— Nikola Tesla

Acknowledgements

In the first place, I would like to express my sincere gratitude to Prof. Dr. Konrad Hungerbühler for providing me the opportunity to conduct my master thesis within the Safety and Environmental Technology Group. Some of the concise, yet relevant interactions were genuinely enlightening in science and beyond.

To say I have been fortunate to work with Merten Morales, would not do justice to the level of understanding, patience and flexibility shown by him and his omnipresent support. His constant motivation and guidance was the root cause in enabling me to solve problems pertaining to process simulations and modeling during the master thesis. Besides, I was able to improve my presentation and interpersonal skills.

I would like to thank Dr. Stavros Papadokonstantakis for his support with his enormous knowledge on modeling and optimization, and significant discussions which were beyond useful during my master thesis. I am grateful for the time and effort he dedicated in correcting my work. I feel privileged for having had the opportunity to work with him on surrogate modeling. I am much obliged to everyone in the research group for maintaining an excellent and cordial atmosphere. The multi-cultural environment was something I enjoyed. The members of the Waldbox project and fellow international master students in the group deserve a special appreciation for their constant support and advise.

Last but by no means least, I would take this opportunity to thank my entire family and some really special friends without whom none of this would ever be possible. Thank you! I would like to dedicate this to Mum and Dad.

Abstract

An innovative integrated biorefinery for the production of Polyethylene terephthalate (PET) and Polyurethane (PU) was developed with the purview of the ‘Waldbox’ project. PET is an extensively used polymer in the production of food and beverage containers and an essential component in the manufacturing of synthetic fibers. With a far-reaching impact on humanity, PU has found considerable use and emphasis in the modern-day world. Manufacture of high-resilience foam seating, rigid-foam insulation panels, durable elastomeric wheels and tires, and high performance adhesives are some of the inexhaustible applications of the versatile polymer.

Commercially production of the two critical polymers stems through the crude oil reforming. The alternative process, presented within this master thesis, demonstrates the process design for bio-based production of PET and PU in an integrated biorefinery using lignocellulosic-based D-glucose as feedstock. The simulations were performed using Aspen Plus®. Environmental and economic assessment of the alternative process delineates its competitiveness compared to the commercial production of the individual polymers.

The introduction of a new generation of biorefineries into the market presents a complex problem, where questions like environmental sustainability, economic feasibility and conceptual exploration need to be answered beforehand to find an optimal concept. Such an integration requires the use of reliable and robust optimization tools.

Surrogate modeling offers an innovative approach for breaking down complex process design to integrate them into a computer aided platform. This minimalistic approach can lead to an optimized and integrated biorefinery concept regardless of the complexity of the design range. The method was applied to the innovative integrated biorefinery to predict the behavior of the models even outside the simulated range, thereby proving an important tool that can be used for translating complex process design into simplified and accurate models. This would be the first step in building a computer-aided platform for the integration of different bio-based production routes to find the optimal biorefinery concept.

Contents

1. Introduction	17
1.1. Importance of an Integrated Biorefinery	17
1.2. Biomass as a Feedstock	18
1.3. <i>Waldbox</i> – A Biorefinery Project	20
1.3.1. Lignocellulosic biomass in Switzerland	20
1.3.2. Reducing Complexity	21
1.4. Reaction Pathway	22
1.5. Significant Chemicals	24
1.5.1. Sorbitol	24
1.5.2. Polyethylene Terephthalate (PET)	25
1.5.3. Polyurethane (PU)	26
1.6. Summary	27
2. Simulations	29
2.1. Sorbitol Production	29
2.1.1. Catalytic hydrogenation of glucose	29
2.1.2. Evaporation and Crystallization	32
2.1.3. Drying and Purification	33
2.2. Bio-Based synthesis of Glycols	33
2.2.1. Catalytic hydrogenolysis of Sorbitol	35
2.2.2. Purification of Glycols	36
2.3. Polymerization of PET	37
2.3.1. Polymerization reaction mechanism	38
2.3.2. Simulation of the PET process	39
2.4. Polymerization of PU	42
2.4.1. Polymerization Reactions	42
2.4.2. Simulation of PU polymerization	44
2.5. Summary	47
3. Surrogate model – Construction and Beyond	49
3.1. The Modeling Approach	50
3.1.1. Designing a Sampling plan	52
3.1.2. Selecting the type of surrogate model	56
3.1.3. Training/Validation of the Surrogate	56
3.1.4. Prediction using the Model	60
3.1.5. Regressing Kriging	61
3.2. Implementing the Surrogate Model tool	61
3.2.1. Training of the Surrogate	61
3.2.2. Prediction	63

3.2.3. Output Model Parameters	64
3.3. Surrogate Model example	64
3.4. Summary	65
4. Results and discussion	67
4.1. Importance of an Integrated Biorefinery	68
4.1.1. Simulation results	68
4.1.2. Environmental Impact Assessment	74
4.1.3. Economic assessment	81
4.2. Reducing complexity	82
4.2.1. Model specification	83
4.2.2. Training of the surrogate	83
4.2.3. Surrogate Prediction	85
4.2.4. Model Parameters	85
4.2.5. Accuracy of the Surrogate Prediction	87
5. Conclusion and Outlook	89
Appendices	91
A. Reaction and Mechanisms	92
A.1. Sorbitol Production	92
A.2. Glycol synthesis	93
A.3. PET Polymerization	93
A.4. PU Polymerization	96
B. Simulations	97
B.1. Design Specification	97
B.1.1. Sorbitol Production	97
B.1.2. Glycol synthesis	98
B.2. Purification of Glycols	99
C. Surrogate Modeling	101
C.1. Surrogate Modeling example	101
D. Economic and Environmental Metrics	107
D.1. Eco-Indicator 99 Data	107
E. Surrogate Predictions	111
E.1. Dependence of input variables on Revenue	111
E.2. Dependence of input variables on PU Purity	113
References	115

List of Figures

1.1. An Integrated Biorefinery	17
1.2. Global application of biomass	18
1.3. Worldwide acreage of feed, food crops and renewable resources	19
1.4. A complex route map of an integrated biorefinery	21
1.5. PU production and application	27
2.1. Production of Sorbitol	30
2.2. Catalytic hydrogenation of D-glucose	30
2.3. Catalytic hydrogenolysis of sorbitol	34
2.4. Production of Glycols	34
2.5. Simplified flowsheet of the PET melt process	40
2.6. BHET oligomer	40
2.7. Simplified flowsheet of the PU Polymerization	45
3.1. An analogy of the surrogate model	49
3.2. The complex nature involving simulation models	50
3.3. Example of a three-dimensional full factorial sampling plan	52
3.4. Example of a three-dimensional latin hypercube sampling	53
3.5. Space-filling metric	55
3.6. Correlation in Kriging model - Effect of p	58
3.7. Correlation in Kriging model - Effect of θ	58
3.8. The Surrogate training	62
3.9. The Surrogate prediction	63
4.1. Integrated PET and PU biorefinery	67
4.2. Material flows - PET biorefinery	71
4.3. Material flows - PU biorefinery	72
4.4. Material flows - Integrated biorefinery	73
4.5. Global Warming Potential (GWP-100a) of the biorefineries per kg _{Pdt}	76
4.6. Global Warming Potential (GWP-100a) of the biorefineries per kg _{glucose}	77
4.7. Global Warming Potential (GWP-100a) of the biorefineries per kg _{oil-eq} substituted	77
4.8. Cumulative energy demand (CED) of the biorefineries per kg _{Pdt}	78
4.9. Cumulative energy demand (CED) of the biorefineries per kg _{glucose}	79
4.10. Cumulative energy demand (CED) of the biorefineries per kg _{oil-eq} substituted . .	79
4.11. The relative impact of polyol production on the biorefineries	80
4.12. Economic analysis of the biorefineries	81
4.13. Operation costs of the biorefineries	82
4.14. Training of the Surrogate Model	84
4.15. Testing of the Surrogate Model	86
4.16. Surrogate Model Prediction	86

4.17. Accuracy of the Surrogate Model	87
4.18. Accuracy of the Surrogate Model	88
5.1. Wood to chemicals	89
A.1. Reaction mechanism for production of sorbitol	92
A.2. Reaction mechanism for production of glycol	93
C.1. Graphical user interface of the tool	101
C.2. Graphical user interface of the tool showing Input variable form	102
C.3. Completed form containing input variables and their ranges	102
C.4. The Excel® data sheet before obtaining the space-filled training data	103
C.5. The Excel® data sheet after obtaining the space-filled training data	104
C.6. Example of the values for input variables for which the model is required to predict the output	105
D.1. Eco-indicator 99 (EI99) of the biorefineries per kg _{Pdt}	108
D.2. Eco-indicator 99 (EI99) of the biorefineries per kg _{glucose}	108
D.3. Eco-indicator 99 (EI99) of the biorefineries per kg _{oil-eq} substituted	109
E.1. Dependence of D-glucose mass flow on revenue	111
E.2. Dependence of evaporator temperature on revenue	112
E.3. Dependence of crystallizer temperature on revenue	112
E.4. Dependence of D-glucose mass flow on PU purity	113
E.5. Dependence of evaporator temperature on PU purity	114
E.6. Dependence of crystallizer temperature on PU purity	114

List of Tables

1.1. Some of the important figures of the Swiss Forest	20
2.1. Catalytic hydrogenation of D-glucose - An overview	31
2.2. Catalytic hydrogolysis of sorbitol - An overview	36
2.3. Purification of glycols	36
2.4. Segment names and Formulas - PET polymerization	37
2.5. Reaction scheme - PET polymerization	39
2.6. Classification of Functional groups in each reactant	41
2.7. An overview of the PET polymerization	42
2.8. Segment names and structures - PU polymerization	43
2.9. An overview of the PU polymerization	45
2.10. Ascending ranks of boiling points of glycols	46
3.1. Space-filling metric Φ_q values of sampling plans X_1 and X_2	55
4.1. Gate-to-gate inventories – PET Biorefinery	69
4.2. Gate-to-gate inventories – PU Biorefinery	70
4.3. Gate-to-gate inventories – Integrated PET and PU Biorefinery	74
4.4. Parameter ranges of the Surrogate model	84
4.5. Model Parameters of the Surrogate	85
A.1. Complete Reaction Set - PET Polymerization	95
A.2. Reaction Set and rate constants - PU Polymerization	96
B.1. Ascending ranks of boiling points of glycols	99
D.1. Economic and Environmental Markers	107

List of Abbreviations

AA	Acetaldehyde.
BHET	Bis-hydroxyethyl terephthalate.
CAGR	Compound Annual Growth Rate.
CASE	Coatings Adhesives Sealants and Elastomers.
CED	Cumulative Energy Demand.
DMT	Dimethyl terephthalate.
EG	Ethylene glycol.
EI99	Eco-Indicator 99.
EUBIA	European Biomass Industry Association.
EVOP	Evolutionary operation.
GHG	Greenhouse Gases.
GUI	Graphical user interface.
GWP	Global Warming Potential.
LCA	Life cycle assessment.
LCI	Life cycle inventory.
LCIA	Life cycle impact assessment.
LHHW	Langmuir–Hinshelwood–Hougen–Watson.
MDI	Methylene diphenyl diisocyanate.
MLE	Maximum Likelihood Estimates.
PET	Poly(ethylene terephthalate).
PG	Propylene glycol.
PolyNRTL	Polymer non-random two-liquid.
PU	Polyurethane.
SA	Simulated annealing.
TDI	Toluene diisocyanate.
TPA	Terephthalic acid.

1. Introduction

1.1. Importance of an Integrated Biorefinery

The demand for energy in today's world has become synonymous with social and economic development of any country. Contrary to the way sources of energy were dealt with in the past, reality portrays a different picture; availability of non-renewable energy sources is limited. Moreover, exploration, processing and consumption of that energy manifests as expensive, not just restricted to in terms of serious impact to the environment, but sometimes cost as well.^[1] An environmentally favorable response to the demand for energy must respect sustainability. Sustainable development can be defined as development that meets present needs without compromising the ability of future generations to meet their own need.^[2] Beyond the paramount importance directed towards the intelligent use of constrained non-renewable resources, the emphasis on the increased utilization of renewable alternatives is on the rise to achieve this goal.^[3]

A biorefinery is a facility that uses biomass as feedstock for the production of fuels, power and value-added chemicals as shown in Fig. 1.1. It can be visualized as an analogous facility to the one of the present day non-renewable petroleum-based refineries that primarily utilizes crude oil and natural gas as feedstock. Biomass includes any organic matter that is available on a renewable or recurring basis.^[4] The importance of a biorefinery pivots on the use of such a

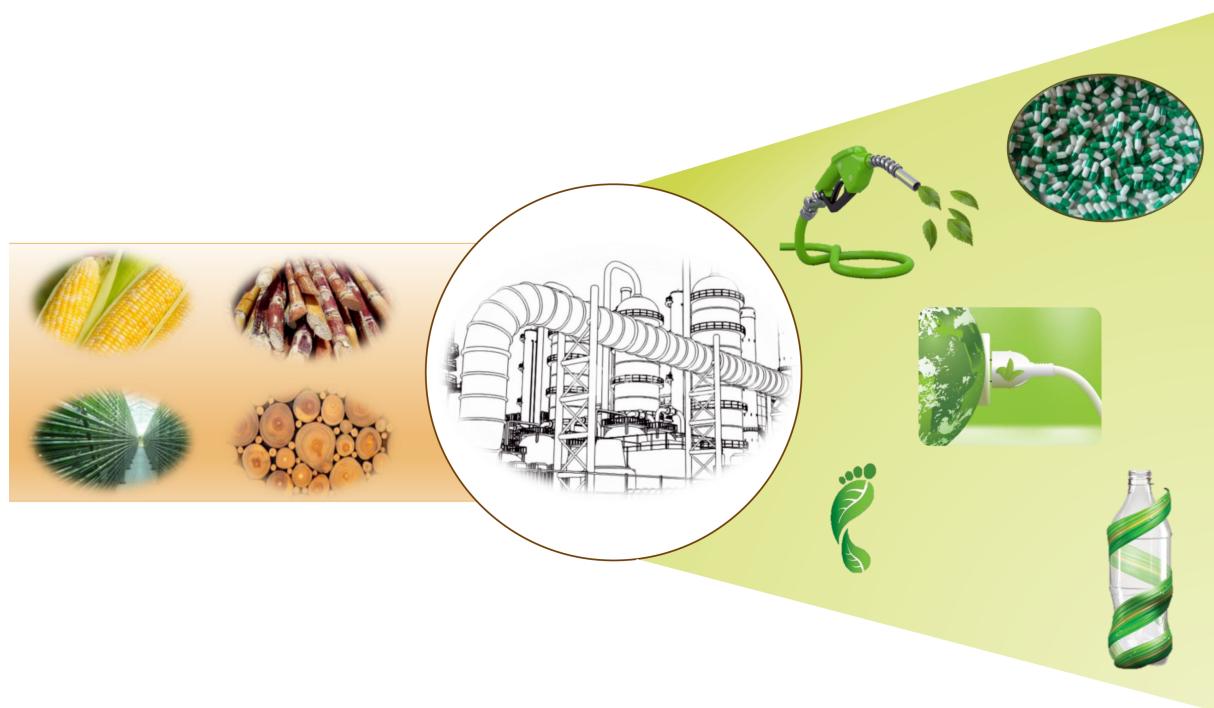


Fig. 1.1. An Integrated Biorefinery: A facility that utilizes biomass from various sources as feedstock for the production of a large portfolio of commodity chemicals, fuels and power.

feedstock. Owing to its abundance and renewable nature, biomass has the potential to replace fossil fuels. Thus, a lot of research has been directed towards the utmost exploitation of biomass components for the commercialization of biorefineries.

An integrated biorefinery takes advantage of the varied and multiple product portfolio that can be derived. Thereby, not only does it curtail the production of biomass waste, but also amplifies the value derived from the feedstock, contributing to sustainable development and global environmental preservation. A vast variety of configurations is possible in the concept of an integrated biorefinery. The unrestricted possibility stems from the wide-ranging source to the extensive list of probable products such as heat, power, fuels, chemicals or even a combination thereof. A biorefinery might produce several low-volume but high-value products, for example pharmaceuticals and functional food ingredients, and low-value but high-volume products, such as biofuels, while additionally generating process heat and electricity, not only for its own use but also for communal advantages such as providing electricity and heating to rural areas. The optimum use of material and energy can only be achieved with appropriate integration.

1.2. Biomass as a Feedstock

The three main non-renewable energy sources, namely petroleum, coal and natural gas, account for over 80% of the global energy utilization.^[1] Remarkably, 98% of the global carbon emission results from the combustion of these fossil fuels.^[1] Along with these, the uneven distribution of fossil fuel reserves also contributes to the global future trend towards using alternative and renewable energy sources, such as biomass.

According to the European Biomass Industry Association (EUBIA), Europe could produce

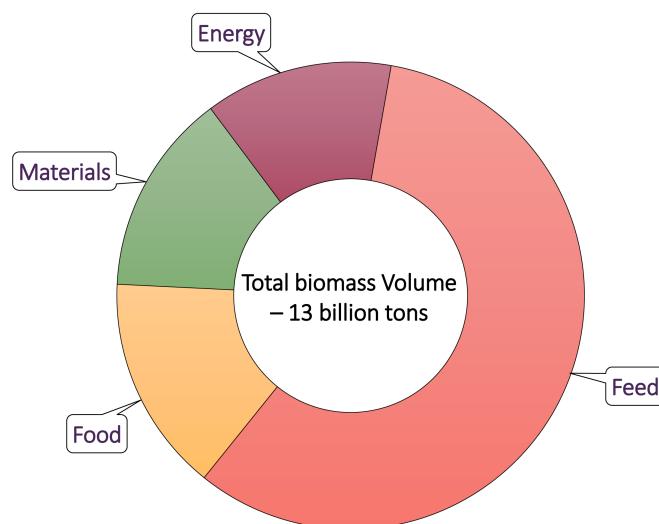


Fig. 1.2. Global application of biomass.^[5] With a total biomass volume of 13 billion tons in 2010, food and animal feed clearly are the priorities for use of biomass. Energy and material use account for a mere 14% and 13% of the annual volume, signifying potential for increased biomass availability for biorefineries, but not without being able to find a balance with the utilization of biomass for food.

8.9 EJ of biomass per year with an energy equivalence of 1.4×10^9 boe and is expected to grow 8–10% each year.^[6] The worldwide usage of biomass for chemical production stands at approximately 35 EJ. However, the majority of biomass is used for food and animal feed. Only a fourth is used for energy and materials^[5] as shown in Fig. 1.2. Clearly, an important issue with socioeconomic impact is the potential competition between food industry and utilization of biomass. First generation biomass is known as the edible biomass, competing directly with the utilization of biomass as an energy resource, eventually resulting in elevated concerns about the future use of biomass.^[7]

“The Bioeconomy strategy and its Action plan aim to pave the way to a more innovative, resource efficient and competitive society that reconciles food security with the sustainable use of renewable resources for industrial purposes, while ensuring environmental protection.” — European Commission 2012

Social challenges like food security and sustainable agriculture may be met by the increasing focus on second generation biomass feedstock such as lignocellulosic biomass. However, a gradual perception that lignocellulosic biomass might be an indirect competition for land is prevalent. Lignocellulosic biomass is well-suited feedstock for biorefinery application. Low-cost and large-scale availability combined with its environmentally benign production makes it an attractive feedstock. Besides, energy production and utilization cycles based on lignocellulosic feedstock have near-zero greenhouse gas emission.^[9] Being one of the most abundant organic source of Earth, only about 4.8% of the annual production of woody biomass in the biosphere is utilized by mankind.^[10]

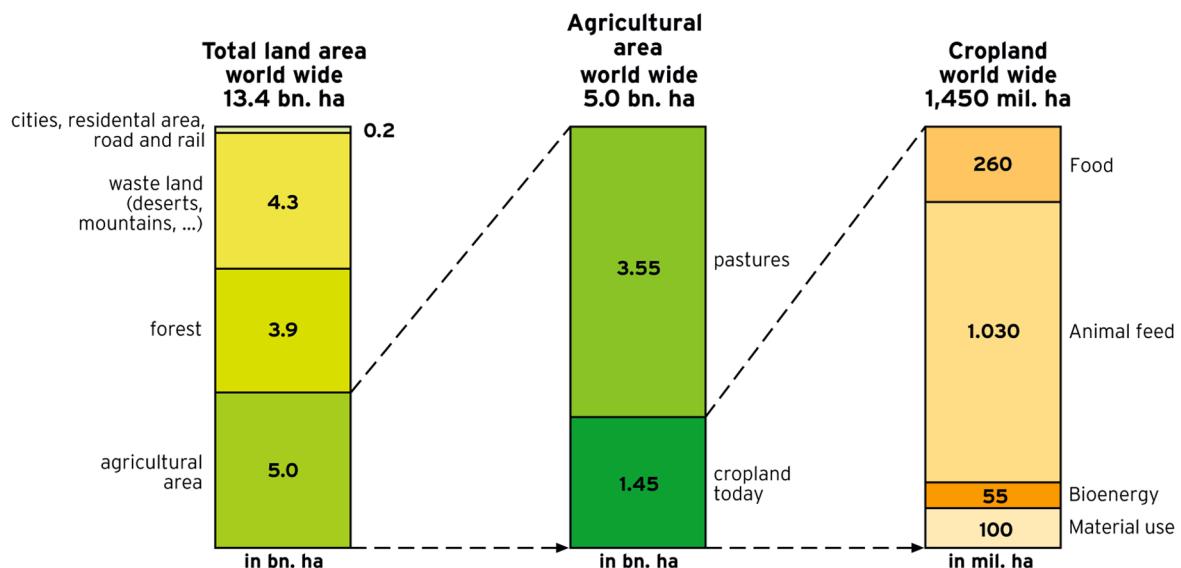


Fig. 1.3. Worldwide acreage of feed, food crops and renewable resources.^[8] A mere 155 million hectares is being utilized for energy and materials. 3.9 billion hectares of land area is covered by forests, signifying the large amount of lignocellulosic residues that could be used as biomass sources.

A minuscule 11% of the worldwide cropland is being utilized for energy and materials. Forests cover a massive 3.9 billion hectares of land area worldwide. The sizable potential in exploitation of lignocellulosic residues is noticeable and is shown in Fig. 1.3. Rich resources in forest are unparalleled to any other type of biomass source. Forest occupy 32% of the land area but account for 89% of the standing biomass.^[11] Tapping into the energy and reclaiming unused potential is imperative to the sustainability of the world economy. Utilization of the woody biomass does not, however, mean destroying the forest nor plant biomass on the planet; it is the carefully controlled utilization of logging and wood residuals, for example, residual woodchips from the construction industry.

1.3. **Waldbox – A Biorefinery Project**

‘Waldbox’ is an innovative project with an objective to take advantage of a computer-aided platform for the process development and design of biorefineries. With an intention to evaluate the numerous configuration of the biorefinery, starting from a diverse use of biomass in terms of composition to generating a multiple product portfolio, Waldbox is designed to help in decision making pertaining to economic and environmental impacts of the biorefinery concept. Currently, the simulations of various processes using lignocellulosic residual feedstock for the production of commercially important polymers and fuels are performed in Aspen Plus®. Evaluations of the process in terms of economic and environmental impact are executed. The long term vision in this direction, would be a stand alone tool that can, with a computer interface, provide an optimal and an integrated biorefinery concept for Swiss industrial partners. However, several fundamental questions would have to be answered before progress can be made.

1.3.1. **Lignocellulosic biomass in Switzerland**

Natural resources such as forests are unevenly spread across the planet. The shortage of such resources hamper the growth of many national economies and persuades the usage of resources of wood as efficiently as possible. Approximately one third of Switzerland is covered by forest with a reservoir of 422 million m³.^[12] Some of the important figures pertaining to the Swiss forest pertaining to the context is shown in Tab. 1.1.

Tab. 1.1. Some of the important figures of the Swiss Forest

The Swiss Forest in figures	
Forest area	1.31 million ha
Wood increment	10 million m ³ yr ⁻¹
Commercially exploitable wood increment	8.2 million m ³ yr ⁻¹
Wood Consumption	10.5 million m ³ yr ⁻¹

Switzerland consumes 10.5 million m³ yr⁻¹ of wood annually:

- 25% is wood products, *e.g.* in the construction sector or furniture production;

- 28% in paper and paperboard;
- 47% for energy production.

If even a mere one tenth of the amount of wood used for the production of energy can be taken as a raw material for the biorefineries, which may end up even producing energy once integrated, a source of dry woody lignocellulosic residues amounting to 50 t h^{-1} would be available. This would be the limiting factor for the biorefinery in terms of size.

1.3.2. Reducing Complexity

The success of an integrated biorefinery concept invariably is established with a fundamentally sound process design. However, the need to incorporate economic and environmental factors along with heat and mass integration of the entire process design and rigorous process modeling, becomes far too complex to be realized, even with today's computational means. In this case, simulations of the individual process designs are performed in Aspen Plus® and are rigorous in order to resemble and replicate the results as that would be achieved from an industrial scale biorefinery. However, it affects negatively the ease of interaction between various individual process models, sustainability assessments and heat and mass integration.

The biorefinery concept relies on seamless integration. Accomplishing this goal seems far-fetched due to the exponential increase in complexity and time, with the growing number of

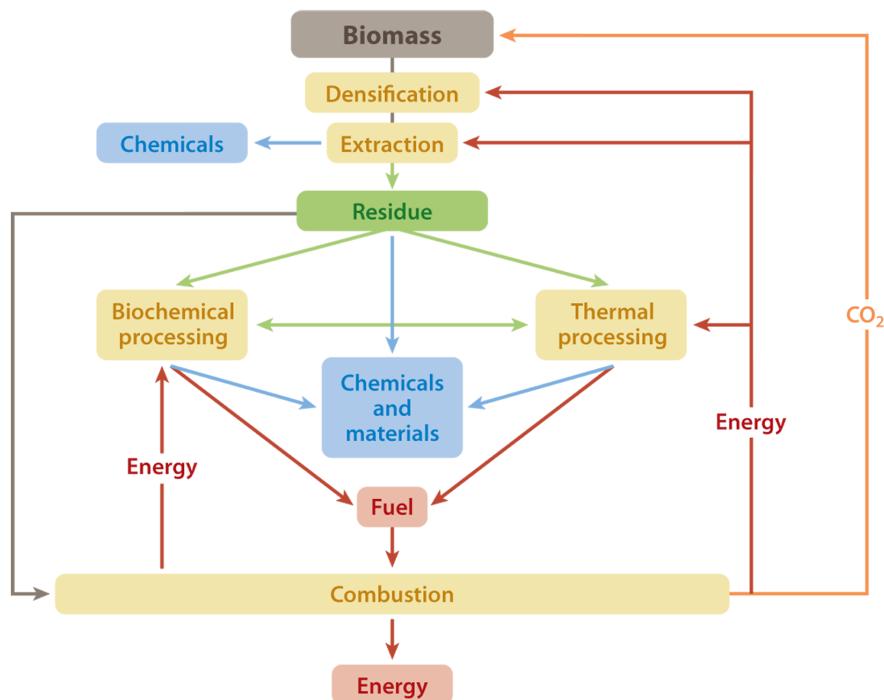


Fig. 1.4. A complex integrated biorefinery pathway. The route map shows the importance of seamless integration and interaction requiring the need for heat and mass integration. The need for reducing complexity with surrogate modeling is evident.

individual process designs signaling a specific pathway for the production of one of the many possible bio-based products. The need to reduce the complexity becomes imperative. Surrogate models seek to provide an answer by being used as a bridge between the limited level of sophistication and accurate prediction of the process. The ease and speed of utilizing such a model to deal with the complex nature of the problem is an advantage much needed.

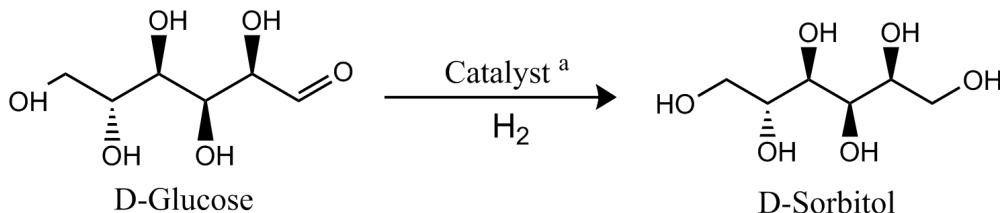
It is not uncommon that an engineering design problem requires the use of a cheap-to-evaluate ‘surrogate’ model that mimics the expensive response of a particular black box. Expensive, in this context, would mean the complexity, time and effort needed to implement such a model to generate a response. The surrogate model typically augments the results coming from the expensive simulation that needs to be run for a range of possible inputs dictated by a design strategy. Once built, the surrogate will be able to understand the behavior and still be faster than its primary source, without compromising on accuracy.^[13]

A tool to build and use surrogate models for integrated biorefinery concept was developed during the master thesis. The tool would play an important role in actualizing the visions of the Waldbox project. The importance of an integrated biorefinery and the use of surrogate models in reducing the complexity is portrayed with a case study dealing with the bio-based production of two commercially important polymers, PET and PU. A concise overview on the reaction pathway is shown in section 1.4. The comprehensive simulations are referred to in chapter 2 and the use of surrogate models is presented in chapter 3.

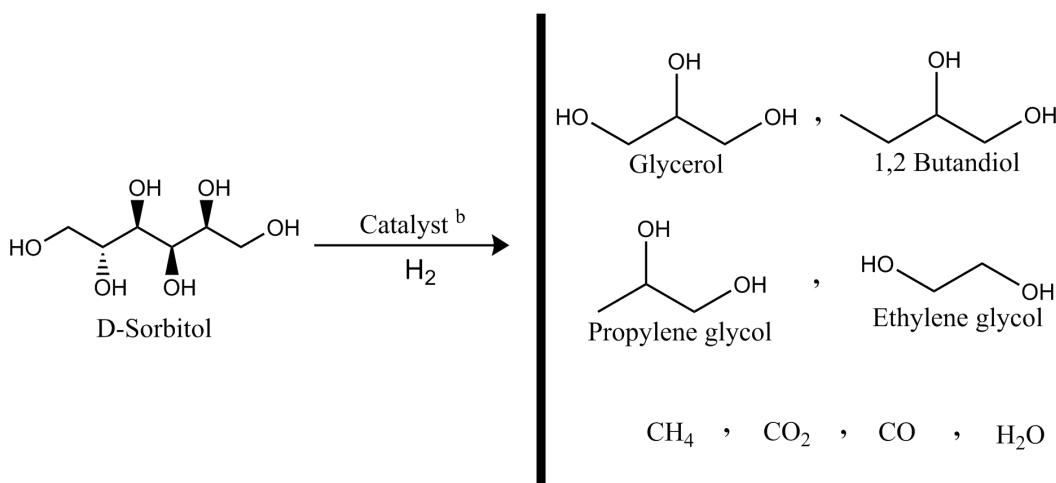
1.4. Reaction Pathway

The conversion of lignocellulosic biomass, whose importance has been established, to a commercially essential platform chemical, D-glucose (or dextrose), has already been implemented in the Waldbox project. For reasons pertaining to simplicity and focus of the current subject, only some of the reaction pathways of the Waldbox project are presented herein. Production of D-glucose is initiated by the extraction of cellulose from lignocellulosic biomass. Subsequently the cellulose is hydrolyzed to yield D-glucose. The master thesis presented focuses on the bio-based production of PET and PU from D-glucose in an integrated biorefinery.

1. Catalytic hydrogenation of D-glucose to sorbitol.

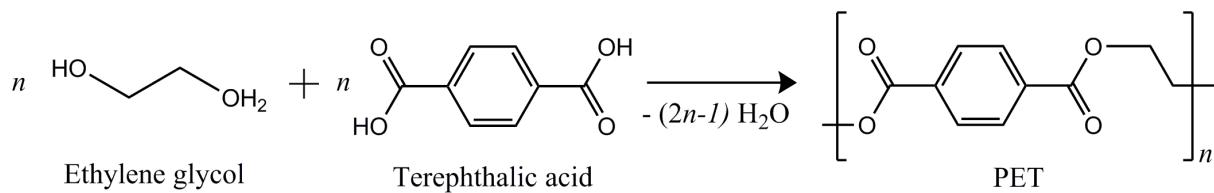


^aConventional process used Raney-Ni based catalysts. Alternative processes was modeled using Ruthenium catalysts on carbon supports.^[14-16]

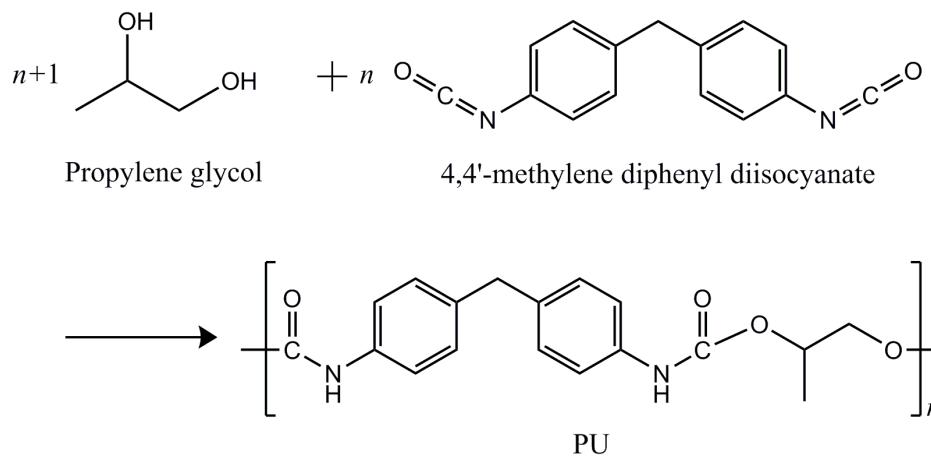
2. Catalytic hydrogenolysis of sorbitol to lower (C_3-C_5) glycols.

^bNickel catalysts. Support material varies depending on the product desired. Ni/ Al_2O_3 was utilized in the alternative production route.^[17,18]

3. PET polymerization using the ethylene glycol produced.



4. PU polymerization using the propylene glycol produced.



Catalytic hydrogenation of D-glucose to sorbitol is at the heart of the pathway, presented herein. Commercial production of PET and PU stems from glycols, which are products of crude oil reforming. Replacing these glycols with bio-based counterparts is the most challenging task. Sorbitol is an important platform chemical, which can not only be produced from D-glucose in a sustainable manner, but can also be subjected to catalytic hydrogenolysis for the production of short-chain glycols. These bio-based glycols act as an excellent substitute for the production of bio-based PET and PU in an integrated biorefinery. The following paragraphs deal with the application and market potential of the platform chemical and the bio-based polymers.

1.5. Significant Chemicals

1.5.1. Sorbitol

Applications

Among the numerous explorations for the efficient conversion of biomass resources into valuable chemical materials, the preparation and utilization of Sorbitol presents an attractive route. According to a study conducted by the United States Department of Energy, sorbitol is one the 12 important target chemicals in their biomass program.^[19] It holds the biggest market share among similar polyols and thus is the commonly used sugar alcohol in the production of drugs, cosmetics, toothpastes and so on.^[20] Most notably, sorbitol can be degraded into polyols, such as ethylene glycol and propylene glycol, which are downstream products in the petrochemical industry.^[21]

The main consumers of the sizable annual sorbitol production is the food industry, predominantly as a non-calorific sweetener and as a precursor for the production of L-ascorbic acid (Vitamin-C), a process that consumes upto 15% of the worldwide sorbitol production.^[22] However, it has important non-food applications. Its moisture conditioning, softening and plasticizing properties have resulted in its utilization in adhesives, textiles and pharmaceutical applications. A prominent non-food application of sorbitol arises from its polyol functionality. Esterification and polycondensation reactions provide polyesters, such as polyethylene terephthalate (PET), and biodegradable polyether polyols used for polyurethane foams.^[23,24]

Market Outlook

The current global production volume of sorbitol is 1.8 million MT. It is expected to grow at a compound annual growth rate (CAGR) of 3.4% from 2014 to 2019.^[25] Sorbitol prices vary between 1250 - 1350 USD/t, depending on the physical form of the final product and purity.^[26] Increase in population and the ever growing demand of low calorie sweeteners, especially from the U.S.A and European Union, is expected to increase the demand for sorbitol over the new few years.^[25,26] European sorbitol revenues are expected to grow at a CAGR of 12.5% from 2014 - 2019.^[25] Roquette^[27] maintains the largest stake in the worldwide sorbitol production.^[28]

Production

Commercial sorbitol production is based on the catalytic hydrogenation of D-glucose. Heterogeneous catalysis in a continuously operated packed-bed reactor is the most desirable production process. However, such a method requires a highly active and a selective catalyst operating under permissible reaction conditions. A lot of academic research has been directed towards the search for an optimum catalyst.^[29] Commercially plants use a Raney-Ni type catalyst in slurry reactors at elevated temperature (~ 400 K) for the large scale production of sorbitol. Although they ensure efficient utilization of the catalyst, purification due to catalytic leaching and moderate selectivity becomes cumbersome.^[14]

In spite of low cost, high selectivity and conversion depending on the reaction conditions,^[15] the Raney-Ni catalyst suffers from two major disadvantages: the leaching of nickel prohibits sorbitol production for application relating to food and pharmaceuticals, and the need for steep hydrogen pressure ($\sim 35\text{--}100$ bar) for a high conversion.^[14,15,29] The use of carbon supported ruthenium catalyst has significant advantages over the use of Raney-Ni.

Recent studies have demonstrated the continuous hydrogenation of D-glucose to sorbitol over a carbon supported ruthenium catalyst.^[16] Strikingly, the ruthenium is stable against leaching. Moreover, the catalyst is highly active when compared to the Raney-Ni and thus does not need as high a pressure to obtain high conversion and selectivity. The production of sorbitol using the alternative catalyst and its subsequent downstream processing to produce bio-based glycols propels the integrated biorefinery for the production of PET and PU.

1.5.2. Polyethylene Terephthalate (PET)

Applications

Polyethylene terephthalate or PET is one of the most commonly used plastics worldwide. The thermoplastic polymer resin is used extensively in synthetic fibers, food and beverage containers. The immense demand of PET can be attributed to its applications and recyclable nature. The majority of the global PET production and demand goes to synthetic fibers (in excess of 60%), while packaging accounts for approximately 30% of the global demand. Other minor applications of PET include plastic films and engineered resins.^[30]

Market Outlook

With a production volume of nearly 50 million tonnes a year, it is expected to grow to about 80 million tonnes by 2017. The anticipated market volume and value growth rate from the European region is around 7% annually until 2019.^[31] Jiangsu Sanfangxiang group^[32] and Indorama Ventures^[33] remain one of the biggest players in PET manufacturing. Although the demand for PET is directly affected by the fluctuations in the crude oil production and prices, steady increase in the production of bio-based PET is expected to show a global compound annual growth rate of about 40%,^[34] driven by the focus on sustainable products. Companies such as Coca-Cola,^[35] Procter and Gamble^[36] are focusing on accelerating the development and use of bio-based PET.^[34]

Production

Conventional production of PET relies on the step-growth polymerization of its monomers; ethylene glycol (EG) and terephthalic acid (TPA) or dimethyl terephthalate (DMT). During recent years, a significant shift from DMT to TPA as a dominant raw material occurred for several reasons such as quality, yield and overall production costs.^[37] The direct esterification reaction of purified terephthalic acid (TPA) with ethylene glycol (EG) results in the formation of a mixture of linear oligomers, predominantly bis-hydroxyethyl terephthalate (BHET), which is then subjected to polycondensation reaction resulting in the formation of PET with the desired molecular weight. Bio-based ethylene glycol drives the alternative process for the production of bio-based PET and is presented in the master thesis.

1.5.3. Polyurethane (PU)

Applications

Polyurethanes are among the most versatile polymers in existence today. Part of the versatility is attributed to their existence in a myriad of forms spanning from flexible foams to elastomers, and rigid foams to solid compositions, thus enabling their use in a wide range of consumer and industrial applications. Polyurethane products find end-use in sectors ranging from automotive and furniture as foams, and being used as sealants, adhesives and elastomers in construction, packaging and textile industry. The global PU production is shown in Fig. 1.5. Use of polyurethane foams for end-user application in construction, automotive and furniture reaches 65% of the global polyurethane demand.^[38]

Market Outlook

The global PU demand was 15 million tonnes as of 2014.^[39] A 5% CAGR is expected to raise the demand to well over 22 million tonnes by 2020. A steady growth in the use of polyurethane foams in the automotive, construction and furniture industry is expected to drive the growth of the PU demand globally. PU foams market is expected to sky-rocket to a USD 62 billion worth by 2018.^[40] The current market price is strongly dependent on the type of polyurethane. Bio-based polyurethanes, with polyols from renewable sources, is only estimated to be 28 kilo tonne worldwide,^[40] a rather modest demand when compared to the conventional polyurethane. Regardless, it typifies the potential of bio-based polyurethanes from polyols derived from biomass.

Production

Polyurethanes are a family of polymers with a wide range of characteristics. They are formed by the step growth polymerization by reacting a di- or polyisocyanate with a polyol resulting in a long chain of organic units linked together by urethane links. The choice of raw material, both the polyisocyanates and the polyols significantly impacts the final properties of the polymer. Conventional production of polyurethane may also include cross-linkers, surfactants and blowing agents for control over the eventual defined property of the polymer.

Most commonly used commercial polyisocyanates are the two aromatic diisocyanates, toluene diisocyanate (TDI) and methylene diphenyl diisocyanate (MDI). Nonetheless, the number of commercially used polyols are too many to count. Bio-based diisocyanates have only received

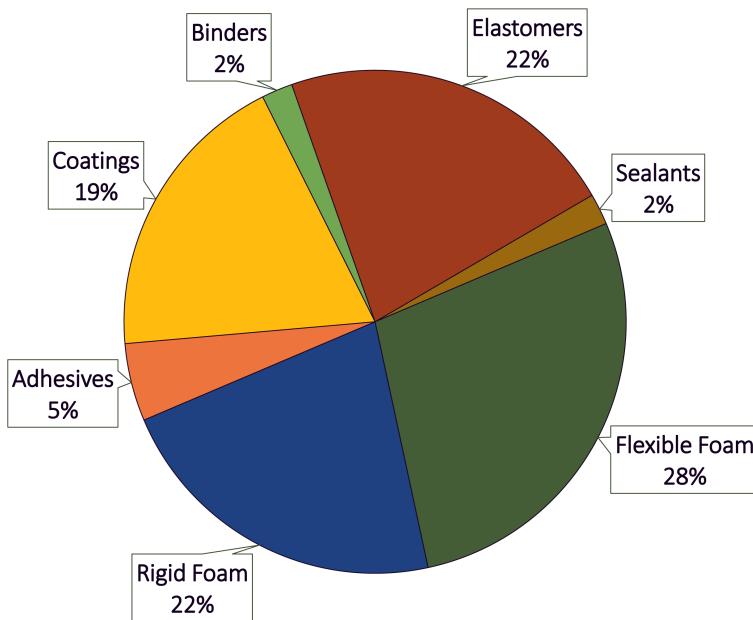


Fig. 1.5. PU production and application.^[39] Global PU production accounts for approximately 15 million tonnes. Majority of the application of polyurethane stems from Coatings, Adhesives, Sealants and Elastomers, CASE marketplace accounting for almost 50% of the global PU production. The contribution of foams also remains commendable, although it doesn't translate in terms of economy.

recent attention. Further exploration is a requisite before utilizing bio-based diisocyanates.^[41,42] On the other hand, the use of bio-based polyols has been extensively investigated over the past few years^[43] for the substitution of petroleum-based starting materials. Bio-based propylene glycol obtained from sorbitol is used as raw material in the alternative production of polyurethane in this master thesis.

1.6. Summary

To compile, the importance of an integrated biorefinery has been delineated. The use of lignocellulosic feedstock makes the biorefinery realistic in terms of availability of renewable biomass resource. The novelty of the master thesis is two-staged. First, alternative bio-based production technologies of two commercially important polymers is investigated and executed, separately and as a part of small biorefinery. The novelty arises from production of bio-based polyols, which are the raw materials of the polymerization processes, thereby replacing its petroleum-derived counterparts. These strategies have been incorporated as drop-in to the Waldbox project.

Secondly and more importantly, a surrogate modeling tool was developed for the seamless integration of various technologies which have been incorporated into the Waldbox project as a drop-in. The surrogate modeling approach helps in realizing the entire potential of the various strategies and technologies (such as heat and mass integration, multicriteria sustainability analysis and optimization) and reducing the complexity of the otherwise complicated and time consuming biorefinery concept. A case study showing the advantage of a surrogate model for a small integrated biorefinery developed using PET and PU is presented.

2. Simulations

The simulations for the production of PET and PU were performed in Aspen Plus® V8.6. The fundamentals of the process designs are based on patents, academic and industrial research. A sensible scale of 1000 kg h^{-1} of input raw material was chosen for two reasons; the availability of raw material in one hand, and an industrially realizable scale that is not too small, on the other. As previously discussed, the simulated processes shown here start with utilizing D-glucose as a raw material feed. The conversion of lignocellulosic biomass to D-glucose has already been implemented as part of the Waldbox project. Approximately 2500 kg h^{-1} of dry woody lignocellulosic residues is required for the production of 1000 kg h^{-1} of D-glucose to be used in the simulations presented herein. D-glucose is converted into bio-based PET and PU through an important platform chemical sorbitol.

2.1. Sorbitol Production

Production of sorbitol from D-glucose has been simulated in a step-wise process including:

- Catalytic hydrogenation of D-glucose
- Evaporation to produce crystal sorbitol
- Drying and Purification of crystal sorbitol

The generalized reaction pathway is shown in Fig. 2.1. The simulated production of sorbitol is based on the academic research by Crezee et al.^[16] and Hoffer et al.^[15] The United States Patent US 6,297,409 B1 by Roquette Freres,^[44] was also used to support simulation procedure.

2.1.1. Catalytic hydrogenation of glucose

The three-phase catalytic hydrogenation of D-glucose was simulated using a ruthenium catalyst with a carbon support. Some commercial producers of sorbitol use a Raney-Ni catalyst. The advantages of utilizing a ruthenium based catalyst have been discussed in the section 1.5.1. The reaction is initiated by feeding the D-glucose at an initial concentration of 1 mol L^{-1} . Hydrolysis of cellulose typically produces D-glucose at a concentration lower than 1 mol L^{-1} ($0.45\text{--}60 \text{ mol L}^{-1}$). Concentrating this solution by solvent evaporation can yield a feed in the operating range of the reactor. Higher concentration is generally not advised due to poor catalyst performance. The drop in performance at increasing D-glucose molar concentration can be attributed to the inability of molecular hydrogen to efficiently diffuse through an increasingly viscous solution of D-glucose to reach the catalyst.

The reaction (see Fig. 2.2) occurs at an elevated temperature of 393 K and at 4.0 MPa hydrogen pressure. The simulated process was made to work for a wide range of input D-glucose

2. Simulations

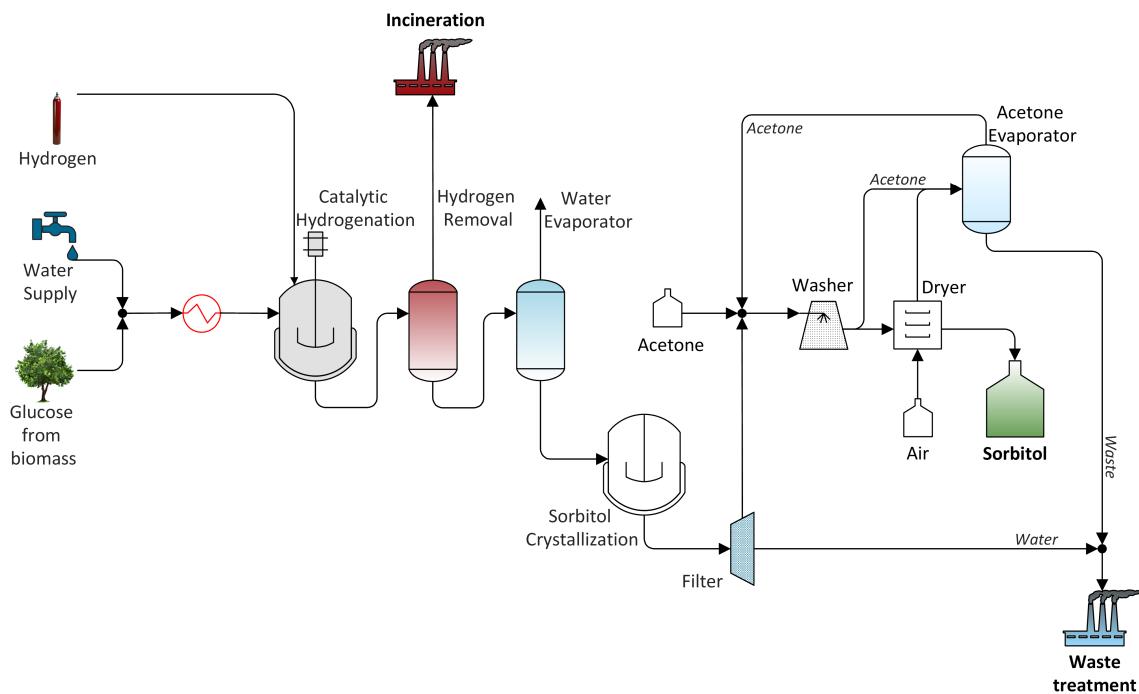


Fig. 2.1. Production of Sorbitol. Initiated by the catalytic hydrogenation of D-glucose followed by evaporation of water and crystallization of sorbitol. The crystal sorbitol is washed and purified.

concentration higher than 1 mol L^{-1} , by using a design specification (for dilution) and subsequently fed into the reactor. Hydrogen was subsequently pressurized and heated to reaction conditions before being fed into the reactor. The reactor functions virtually as a batch reactor in spite of the feeding of D-glucose only after the reactor has reached the pre-determined reaction conditions with the stoichiometric amount of hydrogen present. A batch reactor was used for reasons including cost of operation of a continuous bed reactor, concerns over the conversion of D-glucose and inefficient use of the catalyst.

A 5 wt.% Ru/C catalyst was able to achieve an almost complete conversion of D-glucose. Under the reaction conditions and the catalyst, the selectivity towards sorbitol is virtually 100%. The production of mannitol, an isomer of sorbitol, is negligible. Hydrogen was fed according to the stoichiometric ratio. Typical batch time of the reactor was 5 hours. An overview of the most important parameters of the hydrogenation reaction is presented in Tab. 2.1

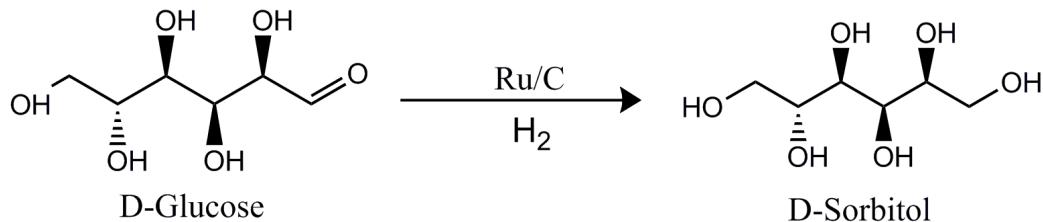


Fig. 2.2. The catalytic hydrogenation reaction of D-glucose to sorbitol over a carbon-supported ruthenium catalyst. The reaction takes place at 393 K and at 4.0 MPa.

Tab. 2.1. Catalytic hydrogenation of D-glucose - An overview

Catalytic Hydrogenation of d-glucose	
D-glucose input mass flow	1000 kg h ⁻¹
D-glucose input concentration	1 mol L ⁻¹
Reaction conditions	393 K; 4.0 MPa
Catalyst	5 wt.% Ru/C
Mass of hydrogen used	11.19 kg h ⁻¹
Sorbitol produced	1001.2 kg h ⁻¹

Reaction mechanism and kinetics

The catalytic reaction was modeled with the Langmuir–Hinshelwood–Hougen–Watson (LHHW) kinetics with the assumption that the surface reaction is the rate determining step. The assumption is justified with the activity of the Ru/C high enough to adsorb D-glucose and hydrogen, yet, low enough to desorb the sorbitol formed. In other words, the characteristic time for adsorption of D-glucose, hydrogen and desorption of sorbitol at the reaction temperature is short in comparison to the characteristic time of the surface reaction. The rate determining step is therefore, the surface reaction of the adsorbed molecular species.^[45]

The hydrogen concentration in the D-glucose solution that affects its adsorption on the catalyst surface is related to its solubility. Therefore, an expected dependence on the partial pressure of hydrogen on reaction rate is evident. The solubility of hydrogen is described by Henry's law.^[46,47] A first order dependence with the partial pressure of hydrogen on the rate of the reaction is mentioned in the literature. Inhibitory effects of sorbitol on the catalyst was neglected in the rate expression. Catalyst deactivation was also neglected.

The rate expression is based on the assumption that ruthenium is a catalyst known to dissociatively chemisorb hydrogen.^[48] A competitive adsorption between the dissociated hydrogen and D-glucose was taken as the basis to construct the rate expression (Eqn. 2.1).

$$r = \frac{k_r K_G C_G K_H P_{H_2}}{(1 + K_G C_G)^3} \quad (2.1)$$

- r – Reaction rate (mol kg⁻¹ s⁻¹)
- k_r – reaction rate coefficient (mol kg⁻¹ s⁻¹)
- K_G – Adsorption constant (m³ mol⁻¹)
- K_H – Constant (Pa⁻¹)
- P – Pressure (Pa)
- C – Concentration (mol m⁻³)

G and H denote the chemical species D-glucose and hydrogen respectively.

The rate of the reaction is first order with respect to D-glucose at low concentration. However, high D-glucose concentration could also hinder rate expression as it takes over catalyst sites, just as one would expect. The partial pressure of hydrogen affects the rate in first order dependence. One would not expect to find the hydrogen term in the denominator due to the low solubility in the D-glucose solution. A competitive adsorption leads to larger extent of reduction in rate of the reaction at higher D-glucose concentrations. The reaction rate coefficients are dependent on temperature . The dependency on temperature allows the calculation of the activation energy needed for the formation of sorbitol.

LHHW kinetic expression in Aspen Plus® V8.6 is modeled as

$$r = \frac{[\text{Kinetic factor}][\text{Driving force}]}{[\text{Adsorption}]} \quad (2.2)$$

In the driving force in the numerator of the Eqn. 2.2 one can define the exponent of the dependence of the components in the rate expression in Eqn. 2.1. The kinetic factor that Aspen Plus® uses, lumps together all the numerator constants in Eqn. 2.1, that is the adsorption and reaction rate coefficients, in an Arrhenius-type expression; thus needing a pre-exponential term and activational energy as the only inputs. The observed activation energy was 55 kJ mol^{-1} .^[15,16] The adsorption term was dependent on cube of the D-glucose concentration. The adsorption constants and the pre-exponential factor were calculated based on the research by Crezee et al.^[16]

2.1.2. Evaporation and Crystallization

Following the catalyzed production of sorbitol, the resulting dilute solution containing produced sorbitol, water, unreacted traces of D-glucose and hydrogen, needs to be processed. Sorbitol is an important platform chemical that can be used for a wide range of applications. Even though, the biorefinery discussed in the master thesis deals with the production of bio-based PET and PU, which require the use of sorbitol in the solution form, the models are designed with a long term goal of integration into a several staged integrated biorefinery platform and sometimes it might be desirable to produce sorbitol in the crystal form. Processing of the produced sorbitol, therefore, becomes important and needs to be implemented in the model.

Evaporation is modeled as a combined process involving a series of flash units for the removal of traces of hydrogen gas and excess water. Design specifications were used to control the temperature of the series of flash columns, which prevented product loss, yet, ensuring maximal removal of the hydrogen gas and then the excess water. The flash unit needed for the removal of excess water to enrich the sorbitol is more energy intensive than the conventional solvent evaporation process. Thus, the simulated process tends to be marginally over estimated in terms of energy consumed. The hydrogen gas removed is then treated in an separate incineration unit, where the gas is burned and thermal energy being generated, can be utilized. The water removed in the form of steam is discarded (*i.e.*, it is not utilized in any energy integration scenarios) and its environmental impact is negligible since it does not contain any harmful components.

The enriched sorbitol is crystallized by reducing the temperature at atmospheric pressure. Reducing the temperature to about 283 K decreases sorbitol solubility in water and results in the formation of sorbitol crystals.^[49] The sorbitol crystals can be separated from the mother

liquor containing only traces of sorbitol by conventional filtration. The uncrystallized sorbitol along with excess water is sent to the waste water treatment plant. The loss of sorbitol in the uncrystallized form is assumed to occur only at the filtration process and is negligible (0.1%). The waste stream contains traces of D-glucose, uncrystallized sorbitol and water. Although the sugars do not alter the pH of the waste stream, it was in the best interest of the environment, that the waste stream were treated with proper pH regulators.

2.1.3. Drying and Purification

The amount of crystal sorbitol generated for every 1000 kg h^{-1} of D-glucose is 978 kg h^{-1} . The final step in the production process of high purity sorbitol crystals involves the washing of the crystals with a solvent that has a low solubility for sorbitol, such as acetone at room temperature and pressure. Using acetone as a washing fluid also renders its drying fairly easy, since the boiling point of acetone is 70°C . The sorbitol crystals are washed with acetone in a washing tank. The amount of acetone needed is set at a liquid to solid mass ratio of 2 to ensure efficient washing of the crystals.^[50]

The washed crystals are subsequently dried by ambient air stream. The dried crystal sorbitol produced can now be either used for further processing or sold to the market as bio-based sorbitol. The used acetone is recycled after boiling the resulting stream to vaporize acetone. The impurities and other unwanted material are mixed with the water stream and sent to a waste water treatment plant.

2.2. Bio-Based synthesis of Glycols

Sorbitol has a considerable potential for the production of versatile chemicals. The ability is retained in its functionality arising from a number of hydroxyl groups. It allows for further processing by a series of reactions such as hydrogenolysis, dehydration and aqueous-phase reforming, thus being able to produce important platform chemicals and commercially important products from renewable sources.^[14] The synthesis of lower alcohols, such as ethylene glycol and propylene glycol is immensely important resulting in chemicals with extensive applications. Today's society is addicted to the use of oil as a fuel and these glycals are invariably produced from crude oil. A gradual attenuation of fossil fuels and a growing global trend to shift to renewable sources has put the focus back on efficient conversion of sorbitol to glycals.

However, the catalytic hydrogenolysis of sorbitol to produce glycals is not a straight-forward process. The reactions for cleavage of C–C and C–O bonds at elevated temperature and pressure yield several mixtures of polyols and the specific selectivity towards a certain product is difficult to achieve. The sorbitol hydrogenolysis reaction is shown in Fig. 2.3.

The simulation of the process is modeled with an input sorbitol feed flow of 978 kg h^{-1} as calculated in section 2.1 for the sorbitol production process from 1000 kg h^{-1} of D-glucose. The entire process flowsheet is shown in Fig. 2.4. The bio-based production of the glycals, specifically ethylene glycol and propylene glycol, which are used as the monomers for the production of bio-based PET and PU, is simulated by a step-wise manner involving:

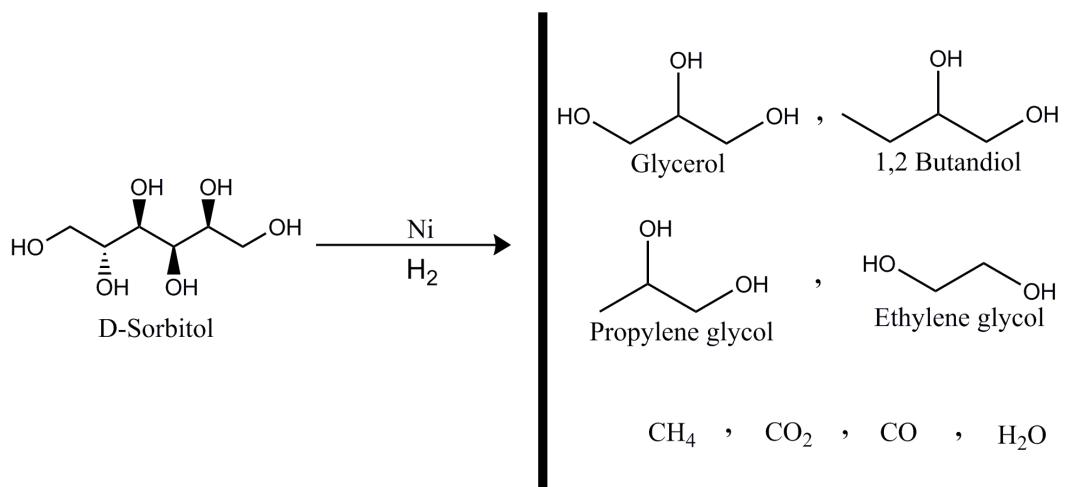


Fig. 2.3. The catalytic conversion of sorbitol to lower glycals. The selectivity of the products depends on the reaction conditions and on the utilization of the catalyst and the support. Nickel catalysts help in the production of ethylene and propylene glycals.

- Catalytic hydrogenolysis of sorbitol.
- Separation of unprofitable by-products.
- Separation via distillation of the various close-boiling glycals.

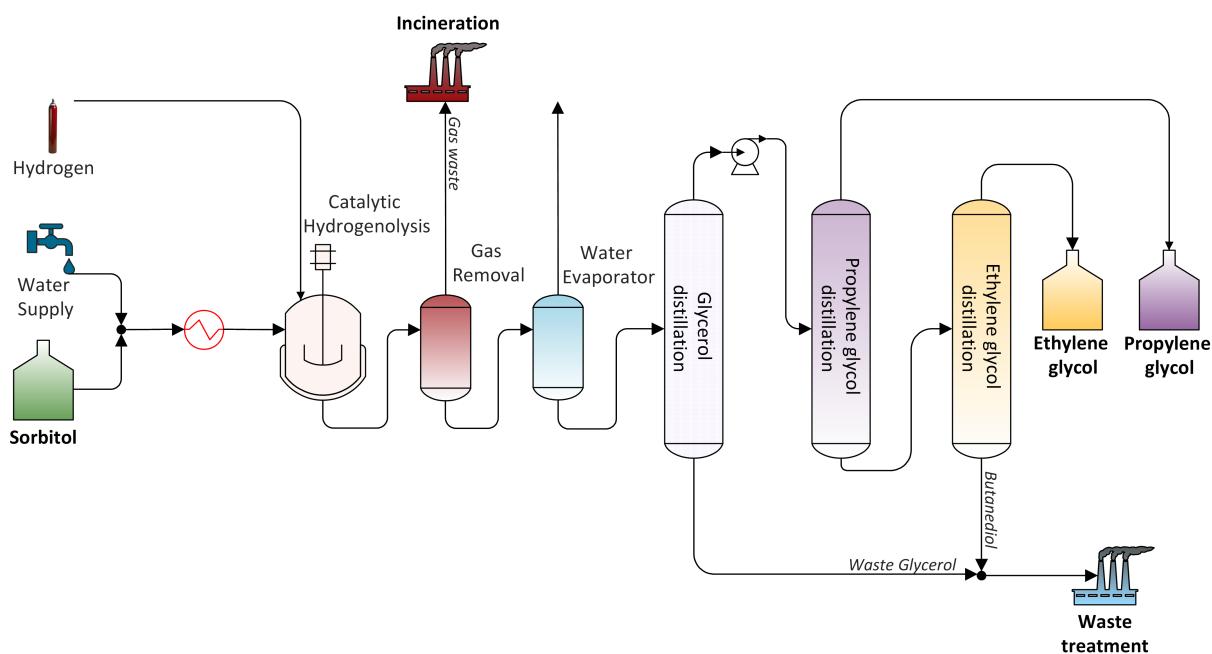


Fig. 2.4. Production of Glycols. The produced sorbitol undergoes a catalytic hydrogenolysis to produce several products which are sequentially separated to yield ethylene glycol and propylene glycol for the bio-based production of PET and PU respectively.

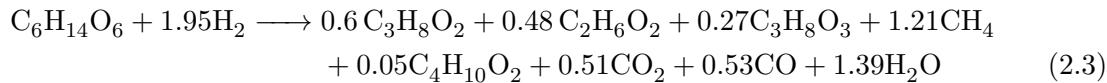
2.2.1. Catalytic hydrogenolysis of Sorbitol

Catalytic hydrogenolysis of sorbitol is based on the academic research of Ye et al.^[17] and Banu et al.^[18] The exact mechanism of the catalytic process is a matter of academic debate. However, the use of Nickel as a catalyst has shown to be advantageous for the production of several polyols. The simulation of the catalytic process is commenced by utilizing the crystallized sorbitol produced from D-glucose. Design specifications are utilized to carefully control the concentration of the sorbitol in the feed solution entering the reactor and to make the simulation robust. A 30 wt.% solution of sorbitol is ideal for the maximum conversion of sorbitol. The solution is subsequently heated to 493 K before being fed into the reactor.

Hydrogen is heated to the reaction condition and pressurized to 7 MPa before fed into the reactor. A batch reactor very similar to the one used for the hydrogenation of D-glucose to sorbitol, was used for the hydrogenolysis reaction with cycle time of 8 hours. Continuous reactors do not reach the goal of stabilized and substantial production of ethylene glycol and propylene glycol. The reactor is pre-loaded with the catalyst. A NI/Al₂O₃ catalyst was used for the hydrogenolysis reaction. The break down of sorbitol in the presence of NI/Al₂O₃ catalyst was modeled using LHHW kinetics similar to the D-glucose hydrogenation. It is safe to approach the reaction as a three-phase catalytic reaction involving the chemisorbed intermediates on the catalyst surface.

Reaction Mechanism

The modeling of the hydrogenolysis presents multiple challenges. The composition of the products at the reactor output is highly dependent on the reaction conditions. Under the simulated conditions, the products are ethylene glycol, propylene glycol, butanediol, glycerol, methane, carbon dioxide, carbon monoxide and water. The selectivity towards one specific product is very difficult to achieve due product similarities. According to the academic research by Ye et al.,^[17] the process always yielded about 2–5% of unknown compounds at the end of the reaction. This was neglected and the balance were restored by distributing the amounts to the above mentioned by-products. Under or over estimation that may be caused by this assumption is expected to be negligible on the bigger picture as they would be separated and end-up in the waste treatments like the rest.



Since the reaction is known to proceed to a certain extent at a certain time, the data was used for specifying the kinetic parameters. The stoichiometry calculated for the reaction is shown in Eqn. 2.3. The complex reaction produces several products in a mixture very difficult in processing to obtain every single product as a pure component. Consideration was only directed for the purification of ethylene glycol and propylene glycol. Some of the important data pertaining to the chemical reaction has been tabulated in Tab. 2.2.

Tab. 2.2. Catalytic hydrogolysis of sorbitol - An overview

Catalytic Hydrogenolysis of Sorbitol	
Sorbitol input mass flow	978 kg h ⁻¹
Sorbitol input concentration	30 wt.%
Reaction conditions	493 K; 7.0 MPa
Catalyst	NI/Al ₂ O ₃
Hydrogen utilized	21 kg h ⁻¹
Ethylene glycol produced	157 kg h ⁻¹
Propylene glycol produced	244 kg h ⁻¹

2.2.2. Purification of Glycols

The first step in the purification of glycals involves the separation of gases such as CH₄, CO₂ and CO. Among the remaining components, water is the least boiling compound and can be subsequently removed from the mixture as vapor. The vapor stream can be condensed and can be used to dissolve CO₂ from the gas mixture. In spite of the low solubility of CO₂ in water, it is significantly higher than the other two gases. Removing even small amounts of CO₂ increases the net calorific value of the gas stream large enough to be incinerated and obtain energy. The water with small amounts of dissolved CO₂ is sent to be treated in the waste water plant.

The remaining components in the mixture include ethylene glycol, propylene glycol, glycerol and butanediol (EG,PG,Gly and BG respectively). The distillation sequencing follows a traditional approach by ranking the compounds based on their boiling points. The presence of azeotropes in the mixtures would make this impossible. However, binary, tertiary and quaternary analysis was done to make sure that there were no azeotropes in the system. Appendix Tab. B.1 ranks the components in the ascending order of their boiling points. However, this is not the only approach to separate the components. The distillation sequencing was not optimized and presents room for further improvements.

Tab. 2.3. Purification of glycals

Sequence	Temperature (in °C)		Product/Mixtures		Purity (mass %)	
	Condenser	Reboiler	Condenser	Reboiler	Condenser	Reboiler
First	192	289.2	EG/PG/BG	Gly	-	92.95
Second	289.6	317.9	PG	EG/BG	99.93	-
Third	-	-	BG	EG	99	99.9

In the following series of distillation towers (see Tab. 2.3), glycerol leaves the first column as a bottom product being the less volatile compound in the mixture. The removed glycerol can be utilized as crude glycerol in the integrated biorefinery. Before such an integration is realized, the glycerol has to be treated to the desired purity. However, in the present flowsheet the glycerol stream is treated as waste as incinerating it presented a higher environmental burden. The separation of the remaining glycals becomes energy intensive. Separation of propylene glycol is attainable. However, the purification of ethylene glycol from butanediol is not achievable in a simple distillation set-up but only when a solvent is used to facilitate extractive distillation. The method is discussed in United States Patent 4,966,658.^[51] The separation was however modeled as a separator unit due to complexities that arise in the property method estimations. Nonetheless, the simulated process for the production of ethylene glycol and propylene glycol does not deviate by a large extent from reality as the energy consumption of the extractive distillation was taken into consideration.

2.3. Polymerization of PET

PET polymerization takes place by the direct esterification reaction of ethylene glycol (EG) and terephthalic acid (TPA). Simulating the polymerization in Aspen Plus® V8.6 can be performed by using the PolyNRTL (Polymer non-random two-liquid) physical property method. The polymerization of PET is fairly complex and involves the considerations of several side reactions of the reacting segment molecules. Before the actual simulation, the species, polymer segments and the reaction mechanism used will be discussed in the section 2.3.1, followed by the simulation of the polymerization process in section 2.3.2.

Tab. 2.4. Segment names and Formulas - PET polymerization

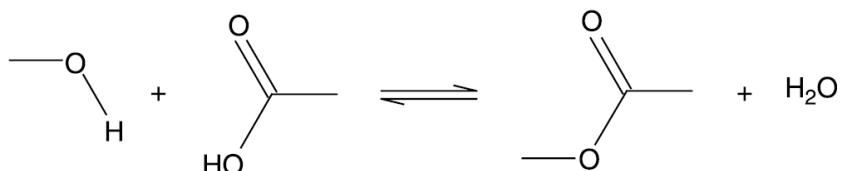
Species	Common Name	Chemical Formula	Molecular Structure
B-DEG	Diethylene glycol repeat segment	C ₄ H ₈ O ₃	
B-EG	Ethylene glycol repeat segment	C ₂ H ₄ O ₂	
B-TPA	Terephthalic acid repeat segment	C ₈ H ₄ O ₂	
T-DEG	Diethylene glycol end segment	C ₄ H ₉ O ₃	
T-EG	Ethylene glycol end segment	C ₂ H ₅ O ₂	
T-TPA	Terephthalic acid end segment	C ₈ H ₅ O ₃	
T-VIN	Vinyl end segment	C ₂ H ₃ O	

2.3.1. Polymerization reaction mechanism

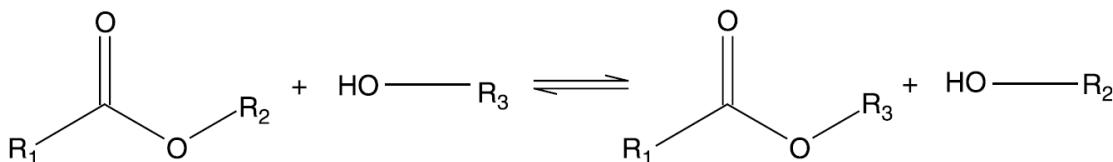
Apart from the conventional species that need to be considered such as ethylene glycol (EG), terephthalic acid (TPA), acetaldehyde (AA) and water (W) several other chain fragments play an important role in the polymerization process. PET is composed of terminal and bound fragments of ethylene glycol (T-EG & B-EG respectively), terminal and bound fragments of terephthalic acid (T-TPA & B-TPA respectively), terminal vinyl segments (T-VIN) and finally, terminal and bound diethylene glycol segments (T-DEG & B-DEG respectively). The Tab. 2.4 summarizes the segmental chemical species involved in the polymerization of PET.

The PET polymerization proceeds through two major reactions:

- Esterification reaction (or also water formation).



- Ester interchange (Transesterification).



The side reactions also affect the polymerization process. The size reactions that are considered are:^[52]

- Degradation of diester groups
- Diethylene glycol formation
- Degradation of ethylene glycol
- Acetaldehyde formation

The reaction scheme for the entire process is shown in Tab. 2.5. For simplification purposes every possible reaction that takes place is only shown in the appendix A.3. The reaction kinetic parameters for each reaction was assigned based on the respective attacking and victim nucleophilic species involved in the reaction.

Tab. 2.5. Reaction scheme - PET polymerization

Reaction Name	Reaction Formula	Description
Esterification	$-\text{OH} + -\text{COOH} \rightleftharpoons -\text{OCO}- + \text{W}$	Alcohol reacts with carboxylic acid to produce an ester and water.
Ester interchange	$\text{R}_3-\text{OH} + \text{R}_1-\text{OCO}-\text{R}_2 \rightleftharpoons \text{R}_2-\text{OH} + \text{R}_1-\text{OCO}-\text{R}_3$	Alcohol reacts with an ester to rearrange groups.
Degradation of diester group	$\text{B-TPA:B-EG} \rightarrow \text{T-VIN} + \text{T-TPA}$	Ester group in B-TPA:B-EG degrades to from T-VIN and T-TPA.
Diethylene glycol formation	$\text{B-TPA:T-EG} + \text{T-EG} \rightarrow \text{B-TPA:T-DEG}$ $\text{T-VIN} + \text{T-EG} \rightarrow \text{B-DEG}$	Formation of T-DEG or B-DEG through the reaction of T-EG with B-TPA:T-EG or T-VIN.
Dehydration of ethylene glycol	$\text{T-EG} + \text{T-EG} \rightarrow \text{B-DEG} + \text{W}$ $\text{T-EG} + \text{EG} \rightarrow \text{T-DEG} + \text{W}$ $\text{EG} + \text{EG} \rightarrow \text{DEG} + \text{W}$	Reaction of 2 ethylene glycol groups to form a diethylene glycol group and water.
Acetaldehyde formation	$\text{B-TPA:T-EG} \rightarrow \text{AA} + \text{T-TPA}$	Formation of acetaldehyde from B-TPA:T-EG.

2.3.2. Simulation of the PET process

Process Description

The simulated model consists of a three-staged melt process as shown in Fig. 2.5. The generalized process consists of an esterifier, a prepolymerizer and a wiped-film evaporator. The ethylene glycol produced from the hydrogenolysis of sorbitol is mixed with the TPA and enters the esterifier which operates at 260 °C and 8 bar. An esterification reaction and the initial formation of oligomer takes place. The polymerization proceeds through the oligomer. The oligomer, bis-hydroxyethyl terephthalate (BHET) (see Fig 2.6), is then fed to the prepolymerizer that operates at a vacuum of 50 mmHg to aid in the build up of the molecular weight of the oligomer. Both the esterifier and the prepolymerizer are modeled as a batch reactor. The polymerization reaction is considerably swift and thus the residence time of the reactants and products in the reactors are short (7 and 3 minutes respectively). The polymerization by-products are evaporated in the wiped-film evaporator. The unreacted ethylene glycol along with the evaporated water from the esterifier and the prepolymerizer are passed through a separation column for the recovery of ethylene glycol. The wiped-film evaporator is modeled as a plug-flow reactor.

A 157 kg h⁻¹ flow of produced ethylene glycol from the hydrogenolysis of sorbitol is mixed with terephthalic acid, in such a way that the stream contains 60 wt.% TPA. The flow of terephthalic acid is controlled with design specifications.

2. Simulations

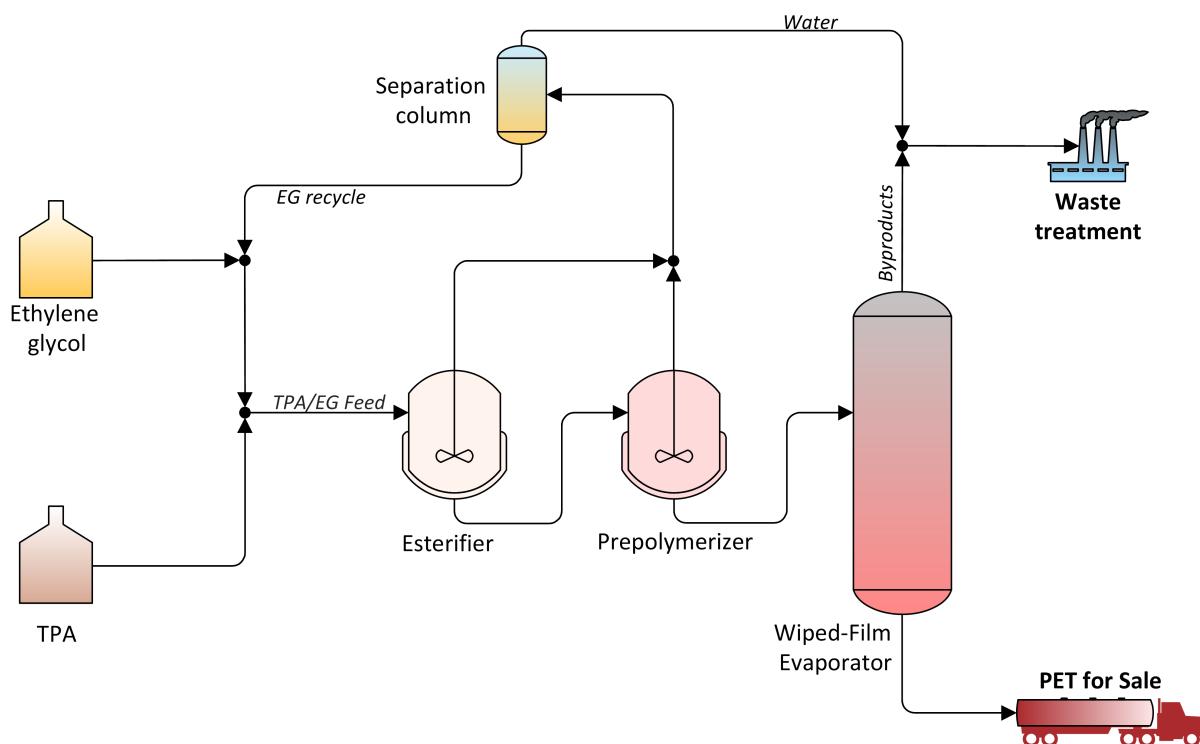


Fig. 2.5. Simplified flowsheet of the PET melt process consisting of an esterifier, a prepolymerizer and wiped-film evaporator for the production of PET.

Simulated Reactions

Aspen Plus® V8.6 has the ability to generate the reaction model provided, the user is able to describe the segments in the polymerization process, the oligomer generated and the functional groups present in each of the reactant. The model generates the reactions based on the reacting species structure. These reactions include the esterification (water formation) and transesterification reactions. The side reactions which the model does not generate were entered manually. The complete list of all the reactions that occur is shown in the appendix A.3. The classification of functional groups in each of the segmented reactants is shown in Tab. 2.6. The polymer database in Aspen Plus® V8.6 already has all these segments defined.

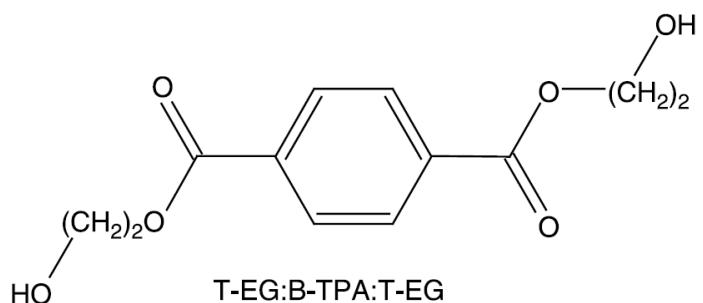


Fig. 2.6. The BHET oligomer consisting of a bound terephthalic acid segment (B-TPA) capped by two terminal ethylene glycol segments (T-EG).

Tab. 2.6. Classification of Functional groups in each reactant

Species	Electrophilic Group H (E-GRP)	Nucleophilic Group OH (N-GRP)	Two-Sided Electrophilic Group BTPA (EE-GRP)	Two-Sided Nucleophilic Group BEG (NN-GRP)	Two-Sided Nucleophilic Group BDEG (NN-GRP)	Nucleophilic Modifier Group VIN (NX-GRP)
TPA		2	1			
EG	2			1		
H ₂ O	1	1				
T-TPA		1	1			
B-TPA			1			
T-EG	1			1		
B-EG				1		
DEG	2				1	
T-DEG	1				1	
B-EG					1	
T-VIN						1
AA	1					1

The reaction rate constants were assigned based on the functional groups involved in the reactions. The reactivity of the functional groups dictate the reaction and thus it makes sense to group all the reactions in the model involving the same functional groups. The reaction rate constants and the side reaction rate constants and the activation energies are entered based on academic research on PET melt polymerization.^[53,54] The 60 reactions are included in the appendix Tab. A.1.

Modeling of the process

The TPA/EG feed mixture containing 60 wt.% TPA enters the esterifier that operates at 260 °C and 8 bar. The batch reactor produces the oligomer. The oligomer is further enriched in the prepolymerizer. The pressure is reduced to 50 mmHg to aid the prepolymerization process. Ethylene glycol and water evaporating from the esterifier and prepolymerizer are passed through a separation column where ethylene glycol is recovered and recycled back to the esterifier. Removal of water is essential in the polycondensation process.

The step growth polycondensation process adheres to Le Chatelier's principle and the removal of water drives the reaction forward. The vaporized water along with ethylene glycol is then separated to recover the unreacted ethylene glycol which is recycled. The water rich stream is sent to a waste water treatment plant.

The wiped-film reactor is modeled as a plug flow reactor in Aspen Plus® operating at 280 °C. The reactor is long enough (40 ft) to complete the polymerization reaction. The PET formed as a result of the polymerization process has a purity of 99.5 wt.%. Some of the important data pertaining to the polymerization reaction has been tabulated in Tab. 2.7. An important point to note is the utilization of TPA as a raw material. One could argue that the bio-based PET is not 100% bio-based as it uses TPA. TPA is produced from non-renewable sources and utilization of

Tab. 2.7. An overview of the PET polymerization

PET polymerization	
Ethylene glycol mass flow	209 kg h ⁻¹
TPA mass flow	406 kg h ⁻¹
Esterifier conditions	433 K; 810.6 kPa
Prepolymerizer conditions	433 K; 50.0 mmHg
Polymerizer conditions	483 K; 101.3 kPa
PET produced	473 kg h ⁻¹

bio-based ethylene glycol is not a guarantee that the process is completely bio-based. However, most of the academic and industrial research was directed towards the production of glycols from biomass. Future research should focus on the bio-based production of TPA to make the process completely dependent only on renewable feedstock.

2.4. Polymerization of PU

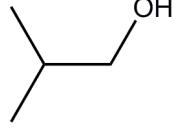
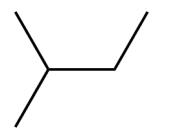
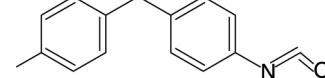
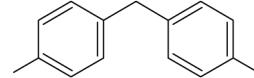
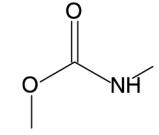
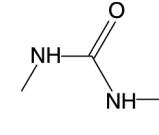
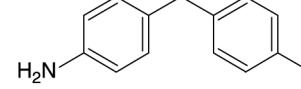
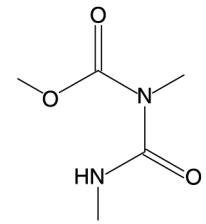
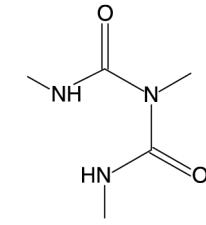
The other commercially important glycol, propylene glycol, is used to produce PU by a polymerization reaction with diisocyanate. The PU polymerization is similar to that of PET, in the sense that they both follow the step-growth polymerization, and they can both be simulated using Aspen Plus® V8.6 with the PolyNRTL physical property method.

2.4.1. Polymerization Reactions

The reaction of propylene glycol with the diisocyanate uses a 1:1 by mole mixtures of the monomers. The diisocyanate used in the simulated production of PU is 4,4-methylene diphenyl diisocyanate (MDI). Apart from the conventional components that are considered such as propylene glycol (PG), 4,4-methylene diphenyl diisocyanate (MDI), water (W), carbon dioxide (CO₂) and methyl diphenyl diamine (MDA), several segments are involved in the polymerization process.

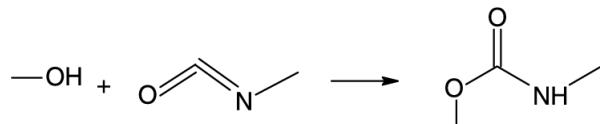
The polymer PU is composed of terminal and bound propylene glycol segments (T-PG and B-PG respectively), terminal and bound MDI segments (T-MDI and B-MDI respectively), terminal MDA segments (T-MDA), bound urethane segments (B-URET), bound urea segments (B-UREA), bound allophone segments (B-ALLO) and bound biuret segments (B-BIU). While all terminal segments are covalently linked once and bound segments are covalently bounded at two points, the B-ALLO and B-BIU are covalently bound at three points. These are units that can introduce cross-linking in the growing polymeric chain. All the segments used and their structures are shown in Tab. 2.8. The single reactions along with their reaction parameters and constants are shown in appendix Tab. A.2.

Tab. 2.8. Segment names and structures - PU polymerization

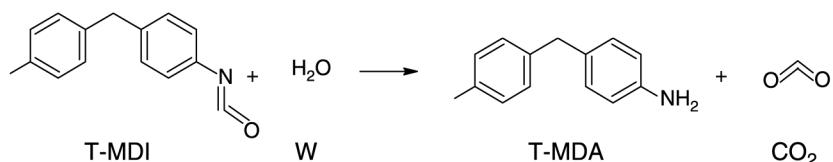
Species	Common name	Molecular structure
T-PG	Propylene glycol end segment	
B-PG	Propylene glycol repeat segment	
T-MDI	Methyl Diphenyl Diisocyanate end segment	
B-MDI	Methyl Diphenyl Diisocyanate repeat segment	
B-URET	Urethane repeat segment	
B-UREA	Urea repeat segment	
T-MDA	Methyl Diphenyl Diamine end segment	
B-ALLO	Allophane repeat segment	
B-BIU	Biuret repeat segment	

The reactions that were included in the PU polymerization process were:

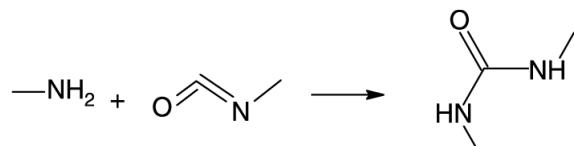
- Urethane formation – the reaction between an alcohol and isocyanate to form a urethane



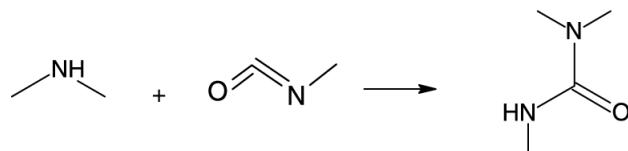
- Amine formation – the reaction between an isocyanate and water to form an amine and CO_2 (foaming reaction)



- Urea formation – the reaction between an amine and isocyanate to form a urea



- Allophane formation – the reaction between an isocyanate and urethane to form a cross-linked allophone group (three-way cross-link)
- Biuret formation – the reaction between an isocyanate and urea to form a cross-linked biuret group (three-way cross-link)



2.4.2. Simulation of PU polymerization

The process simulation for the production of PU is shown in Fig. 2.7. An equimolar feed of propylene glycol and MDI is fed into the reactor. A very small amount of water is added along with MDI in order to make the PU foam. The reactor is modeled as a batch reactor operating at 353 K and atmospheric pressure for a time period of 2 hours for the reaction to near completion. 243 kg h^{-1} of propylene glycol produced from the catalytic hydrogenolysis of sorbitol were used in the production of PU. The amount of MDI needed for the reaction at an equimolar ratio was controlled using design specification.

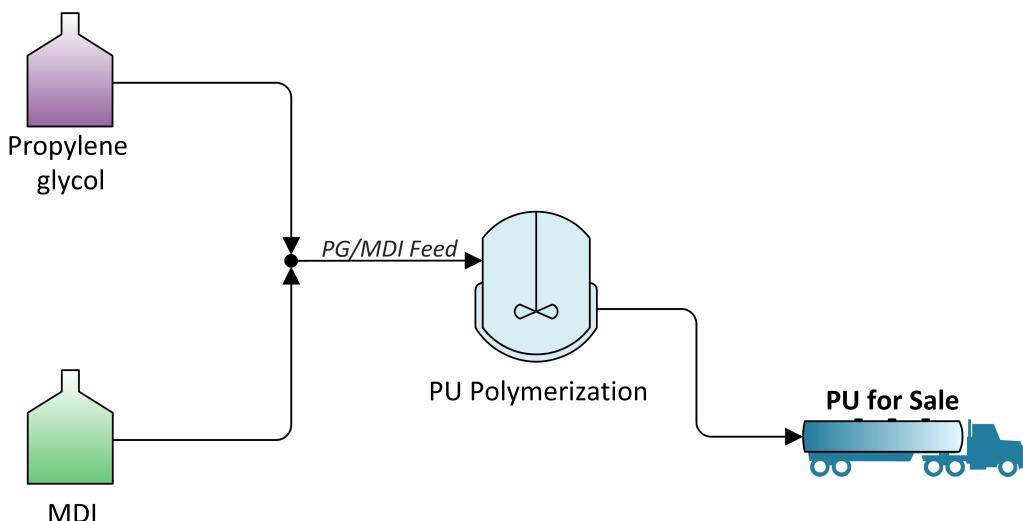


Fig. 2.7. The simulated flowsheet for the PU polymerization process in batch reactor.

Modeling - Components

The segmented components that are used in the PU polymerization reaction are not available in the database of Aspen Plus® V8.6. The components need to be defined in order to characterize the properties and behavior of the reactive system. Their structures can be defined using the Van Krevelen group contribution method,^[55] thus enabling Aspen Plus® to characterize the components. Tab. 2.10 shows the contribution of the groups that occur in the segmented components used in the PU polymerization process.

Process Modeling

The reactions that are considered are grouped according to the type of the reaction. Reaction rate equations and constants have a standard Arrhenius form.^[56] Some of the important data pertaining to the polymerization reaction has been tabulated in Tab. 2.9. Similar to PET polymerization, the production of PU also utilizes a considerable amount of MDI. Although, not 100% bio-based, the production of one of the monomer components, propylene glycol, from lignocellulosic biomass is a significant step forward in realizing a biorefinery concept.

Tab. 2.9. An overview of the PU polymerization

PU polymerization	
Propylene glycol mass flow	243 kg h^{-1}
MDI mass flow	800 kg h^{-1}
Reaction conditions	$353 \text{ K; } 100 \text{ kPa}$
PU produced	1042 kg h^{-1}

Tab. 2.10. Ascending ranks of boiling points of glycols

Group Number	Type	Segment								
		T-PG	B-PG	T-MDI	B-MDI	B-URET	T-MDA	B-UREA	B-ALLO	B-BIU
100	$-\text{CH}_2-$	1	1							
102	$-\text{CH}_3$	1	1							
123	4-MEBZ-1,8BENZENE			1	1			1		
136	$=\text{C}=$				1					
149	$=\text{O}$ or $-\text{O}-$				1					
151	$-\text{O}-\underset{\text{O}}{\overset{=}{\text{C}}}-$								1	
160	$-\text{OH}$	1								
168	$-\text{NH}_2$						1			
169	$-\text{N}=\text{ or }-\text{N}\backslash-$			1					1	1
174	$-\underset{\text{O}}{\overset{=}{\text{C}}}-\text{NH}-$								1	2
175	$-\text{O}-\underset{\text{O}}{\overset{=}{\text{C}}}-\text{NH}$					1				
176	$-\text{NH}-\underset{\text{O}}{\overset{=}{\text{C}}}-\text{NH}-$							1		

2.5. Summary

The importance of utilization of renewable resources for the production of commercially important products is paramount. The utilization of woody biomass for the production of several end products is one of the concepts and ideas of the Waldbox biorefinery project. The conversion of wood biomass to D-glucose has already been implemented in other projects.

D-glucose is catalytically hydrogenated to produce sorbitol, an important platform chemical used in the production of bio-based glycols through the catalytic hydrogenolysis of sorbitol. Polyethylene terephthalate and Polyurethane are invariably produced from glycol monomers derived from crude oil. Replacing the commercial glycol monomers with the bio-based counterparts is an important step forward in realizing an optimum biorefinery. The simulation of the bio-based glycol is described in detail and so is the utilization of these bio-based monomers for the production of bio-based PET and PU.

The environmental and economic impact of these simulated biorefinery will be discussed both as a stand alone and in an integrated layout. Stand-alone biorefinery will concentrate on the production of either bio-based PET or PU. Integrated biorefinery will have a multiple product portfolio and will produce both polymers. The impact on utilizing more raw-materials for the production of both products and how it will compare against the stand-alone production will also be discussed in the chapter 4.

Integrating biorefinery is a vision that may not be achieved without seamless interaction between the various concepts it comprises. The use of surrogate modeling focuses on reducing the complexity to make the search of the optimum biorefinery concept faster and easier. The surrogate models will be discussed in the chapter 3. The impact of these models on the simulated bio-based processes will be shown with a case study example in the chapter 4.

3. Surrogate model – Construction and Beyond

Beyond the obvious analysis of economic and environmental feasibility of a simulated biorefinery concept, the focus, quite often, shifts to the integration of several biorefinery concepts into a single functional computer-aided platform that has the potential to utilize multiple input feedstreams for the production of a variety of products of industrial importance. The simulated environments of bio-based PET and PU have been shown in chapter 2. These concepts are one of the many that can be potentially incorporated in an integrated biorefinery concept such as the Waldbox. As one would expect, this would also need the seamless interaction of economic and environmental markers with process models, heat and mass integration, and generally multi-objective optimization.

However achieving this would result in a complex, dense and a rigid biorefinery which would work against the goal to design a robust design and decision making tool. For example, one can visualize complexities arising in a simulated flowsheet, like the one created by Aspen Plus®. Even though constraints in a flowsheet are dictated by the degrees of freedom of the processes, integrating several individual flowsheets into a single large platform, although can be done automatically, it introduces tremendous computational complexities which is always bound to work against the needs of the user. Increase in computational time and space, and the need for a strong and tenacious machine able to handle the computations are some of the several disad-

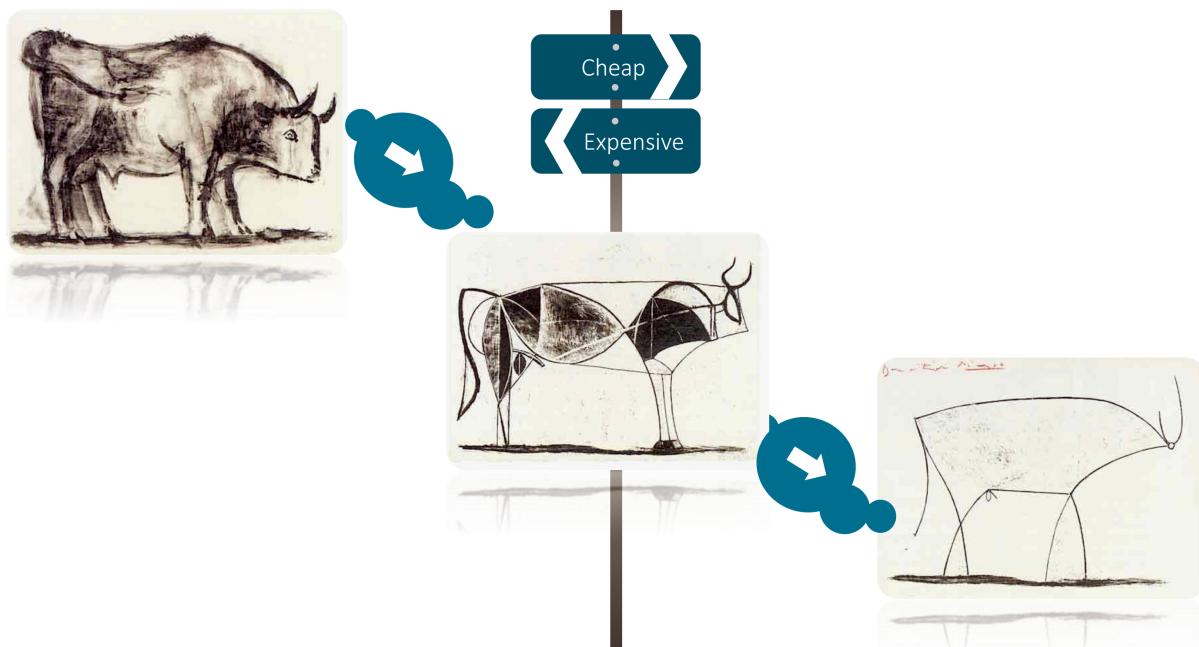


Fig. 3.1. The analogy of the surrogate model is shown with an image of the bull. An ‘expensive’ model is one that is complicated, computationally difficult and time consuming. The ‘cheap’ surrogate model maintains the meaning of the image, yet is far less complicated

vantages that can be expected.

A surrogate, as explained by the analogy in Fig. 3.1, is a stand-in to the original. The models generated with the Aspen Plus® are comparable to the detailed sketch of the bull. Although the complexity is immensely diminished in the ‘surrogate’ sketch, one would still be able to obtain or extract the same quality or type of information according to predefined targets. An analogy is that our mind knows a great deal of information about the obscured parts of the surrogate sketch, thereby enabling us to make an educated guess on the exact nature of the original.

3.1. The Modeling Approach

The core problem is to attempt to learn output responses of a black box model for a wide range of input variables. To translate this into our perspective at hand, it would mean to attempt and learn the output mass flows, concentrations, or even process assessment metrics and so on of a complex model, such as the Aspen Plus® simulated models, in an integrated biorefinery for a wide range of input feed flows or even varying input feed. The problem is depicted in the Fig. 3.2.

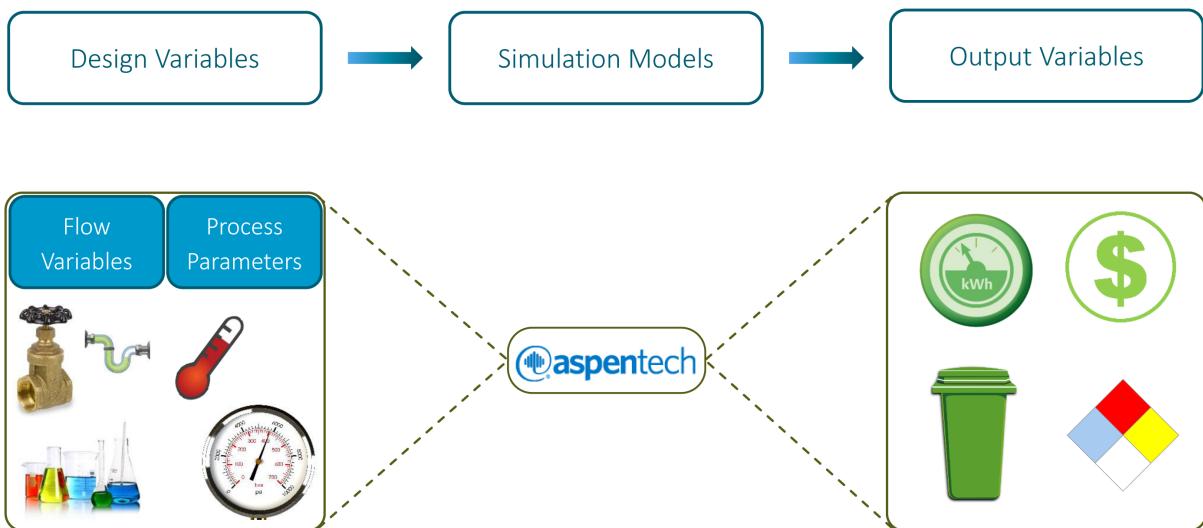


Fig. 3.2. The problem at hand in obtaining output parameters that relate to cost or energy consumption, etc. to varying values of input design parameters using a simulation software Aspen Plus®

Mathematically, we would try and attempt to learn

$$y = f(x) \quad (3.1)$$

where y is a continuous quality, cost or performance metric of the process as calculated by the simulation model, Aspen Plus®, f is the simulation model and x are the input or design parameters. The process is defined by a k -vector of design variables.

$$x \in D \subset \mathbb{R}^k \quad (3.2)$$

D is referred to as the design space. Sometimes the only insight that we have on the computa-

tionally costly function f is a discrete sample of observations.

$$\{\mathbf{x}^{(i)} \rightarrow y^{(i)} = f(\mathbf{x}^{(i)}) | i = 1, \dots, n\} \quad (3.3)$$

The task is thus to use an approximation, such as a surrogate model, \hat{f} , which can replace f and thus can make a computationally cheap performance prediction. It is evident that the approximation of the model comes from learning from sample data such as,

$$\begin{aligned} \mathbf{X} &= \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}\}^T \\ \mathbf{y} &= \{y^{(1)}, y^{(2)}, \dots, y^{(n)}\}^T \end{aligned} \quad (3.4)$$

However, obtaining such a data could be difficult without the proper aids. To explain this with an example, consider a simulated Aspen Plus® flowsheet of a simulated process producing a certain product. Let us say you are interested in the output flow of the product for a varying range of input feed flow, reactor temperature and reflux ratio of a distillation column. Now, the product output flows would denote \mathbf{y} , f is the simulated model, k is your number of design variables which in this example would be 3 and \mathbf{X} denotes your sample data. Having run the simulation once, *i.e.*, $i = 1$, you have one sample response for the original model. However, for the construction of a surrogate model, you might need many more so (*i.e.*, n) that the model is able to ‘learn’ from the expensive simulated model.

In the previous example, one could also understand the computation difficulties one would have to go through if the model had to predict the behavior for thousands of individual integrated flowsheet calls. A single flowsheet call in itself may contain one or more optimization problems and process integration. Every call of the flowsheet will therefore solve one of more optimization problems (e.g., design specifications, loop converging), especially when the model is in the form of a modular simulator. The number of variables of interest may start growing exponentially in the case of integrated biorefineries and so will the number of design variables. This is when a single simulation run may end up taking a lot of computational time and effort. However, a surrogate model of the small flowsheets before integration may retain its robust nature even after integration, proving essential for realizing the goal of an optimal biorefinery. The surrogate model concept is particularly adept when the model behavior can be reduced to be represented in algebraic functions, computing which is significantly faster.

The steps involved in the construction of a surrogate model involve:

- Designing a sampling plan
- Selecting the type of surrogate model
- Training/Validation of the surrogate
- Prediction using the model

3.1.1. Designing a Sampling plan

In some scenarios, experimental data would dictate the initial data (X & y , see Eqn. 3.4). Simulation concepts do not have this limitation and one would need to design a sample plan to create the data based on which the surrogate model will be trained. The accuracy of the surrogate is directly dependent on the quality of the sampling plan. A uniform spread of points in the design space is widely considered to be the best approach since the knowledge and training of the surrogate with a wide spread across the design space is thought to impose better accuracy in the final prediction. A sampling plan possessing a uniform spread of points is said to be space-filling and intuitively one can expect a uniform level of model accuracy throughout the design space.

Random Latin Hypercubes

The most straightforward method of sampling a design space uniformly is by means of a rectangular grid of points or a full factorial sampling. For example, Fig. 3.3 shows a three-dimensional sampling plan. Such a design, in spite of satisfying the uniformity criterion, has flaws. When projected on to the axes, sets of points will overlap and it can be argued that the sampling of any individual variable could be improved by making sure that these projections are as uniform as possible.

Stratification of a sampling plan aims to generate points whose projections onto the variable axes are uniform. Latin hypercube sampling generates a sampling plan that is stratified in all dimensions. Constructing a latin hypercube can be achieved by splitting the design space in equal sized hypercubes (bins) and subsequently placing points in the bins, one in each, making sure that from each occupied bin we could exit the design space along any direction parallel with any of the axes without encountering any other occupied bins. This is illustrated for three

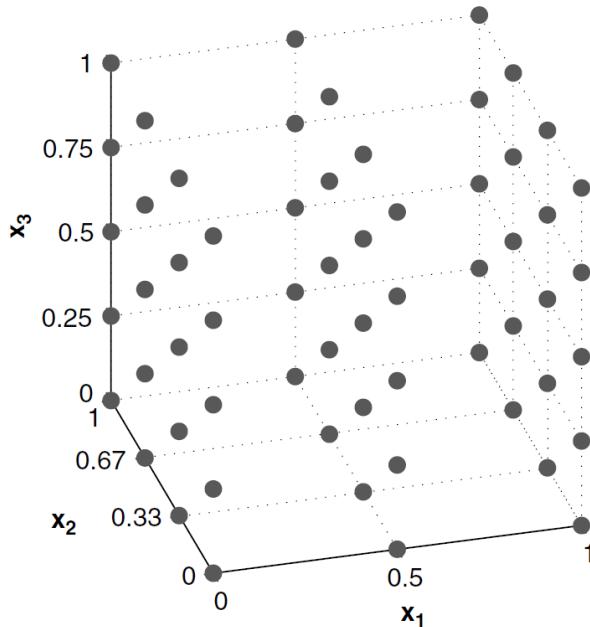


Fig. 3.3. Example of a three-dimensional full factorial sampling plan

dimensions in Fig. 3.4. If X denotes the $n \times k$ matrix in which we wish to build our sampling plan of n points in k dimensions (each row represents a point), one would begin by filling up X with random permutations of $\{1, 2, \dots, n\}$ in each column and normalizing the plan into a $[0, 1]^k$ box to make some of the subsequent mathematics easier while dealing with a plethora of multidimensional scaling issues.

Although the sampling plan ensures multidimensional stratification, it does not guarantee that the sampling plan will be space-filling. Ultimately, placing all of the points on the main diagonal of the design space will fulfill the multidimensional stratification criterion, but, intuitively, will not fill the available space uniformly. One of the widely-used sampling plan that ensures multidimensional stratification, randomized sampling and uniformity is the space-filling latin hypercube.

Space-filling latin hypercubes

Johnson et al. (1990)^[57] introduced the maximin metric to evaluate the uniformity ('space-fillingness') of a sampling plan and to distinguish between 'good' and 'bad' latin hypercubes. The number of distinct latin hypercubes increases rather sharply with both the number of points (n) and the number of dimensions (k). The possible number of latin hypercubes is given by,

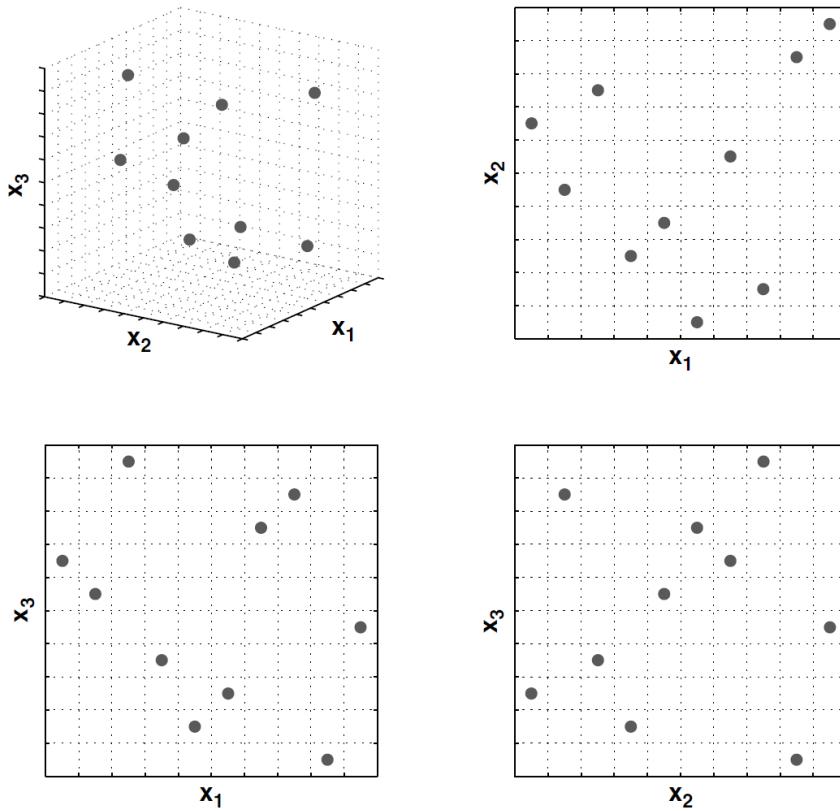


Fig. 3.4. Three-variable, ten point latin hypercube sampling plan shown in three dimensions (top-left), along with its two-dimensional projections. All ten points are visible on each of the projections and each row and column of bins contain exactly one point.

$$\left(\prod_{i=0}^{n-1} (n-i) \right)^{k-1} = (n!)^{k-1} \quad (3.5)$$

Johnson et al. (1990)^[57] described the maximin criterion to assess the competing latin hypercubes. Let d_1, d_2, \dots, d_m be the list of unique values of distances between all possible pairs of points in the sampling plan X sorted in the increasing order. Further, let J_1, J_2, \dots, J_m be defined such that J_j is the number of pairs of points in X separated by a distance d_j .

We call X the maximin plan among all available plans if it maximizes d_1 , among plans for which this is true, minimizes J_1 , among plans for which this is true, maximized d_2 , among plans for which this is true, minimizes J_2, \dots , minimizes J_m . The ‘distance’ in the definition is defined as,

$$d_p(x^{(i_1)}, x^{(i_2)}) = \left(\sum_{j=1}^k |x_j^{(i_1)} - x_j^{(i_2)}|^p \right)^{1/p} \quad (3.6)$$

This p-norm metric (1 is rectangular and 2 yields euclidean) can be at times very time-consuming to calculate. The vector contains the distances between all possible pairs of points and can especially become cumbersome for large sampling plans. A optimization of the pairwise comparison becomes quite deceptive and therefore difficult to search reliably. The reason being that the comparison process will stop as soon as one finds a nonzero element in the comparison array and therefore the remaining values in d_1, d_2, \dots, d_m and J_1, J_2, \dots, J_m will be lost. Since the pairwise comparisons becomes unrealistic for large sampling plans, Morris and Mitchell (1995)^[58] defined a scalar-valued criterion function used to rank competing sampling plans.

$$\Phi_q(X) = \left(\sum_{j=1}^m J_j d_j^{-q} \right)^{1/q} \quad (3.7)$$

The smaller the value of Φ_q , the better the space-filling properties of X will be. This equation distills the cumbersome definition of the maximin criterion into a rather neat and compact form. Nonetheless, it still raises the question on how to choose the value of q . Morris and Mitchell’s method suggested minimizing Φ_q at fixed values of $q = 1, 2, 5, 10, 20, 50$ and 100 and then choosing the best of the resulting plans according to the actual maximin definition.

An example of two-variable, ten point latin hypercube sampling plans X_1 and X_2 can be seen in Fig. 3.5. The space-filling metric Φ_q values of the two competing hypercubes for varying values of q are tabulated in Tab. 3.1. X_1 will have better space-filling properties as the values of Φ_q across all q ’s is lower than the sampling plan of X_2 .

Optimizing the space-filling quality metric Φ_q is an important and a necessary step after establishing a criterion to evaluate the quality of latin hypercube sampling plan. Indeed, the optimization procedure would need to take into account the computational time budget in trying to locate the best possible sampling plan. One of the defining features of a latin hypercube X is that each column is a permutation of the list of the possible levels of the corresponding variable. Swapping two of the elements within any of the columns of X will bring about the smallest alter-

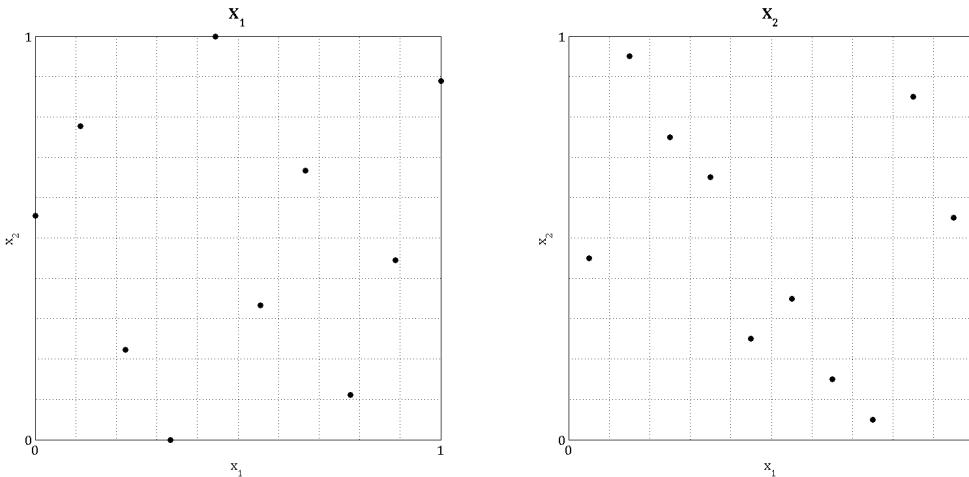


Fig. 3.5. Two-variable, ten point latin hypercube sampling plans, X_1 and X_2 .

Tab. 3.1. Space-filling metric Φ_q values of sampling plans X_1 and X_2

Sampling plan	Φ_q						
	$q = 1$	$q = 2$	$q = 5$	$q = 10$	$q = 20$	$q = 50$	$q = 100$
X_1	64.35	10.37	4.01	3.27	3.11	3.04	3.02
X_2	78.46	13.76	6.44	5.59	5.28	5.11	5.06

ation possible without impairing the key multidimensional stratification property. Morris and Mitchell (1995)^[58] used a simulated annealing (SA) algorithm to introduce ‘mutations’ to the latin hypercube to create offspring hypercubes. The hypercube sampling plan with the smallest Φ_q is selected.

As an alternative an evolutionary operation (EVOP) was implemented. The evolutionary operation is based on the concept introduced by Box (1957),^[59] to optimize chemical processes. A variable scope strategy was applied for the purposes of latin hypercube search. One starts with a long step length (a relatively large number of alteration within the columns) and as the search gradually progresses, the step is reduced to a single change. In each generation the parent hypercube is mutated a number of times and a sampling plan with the smallest Φ_q value among all offspring and the parent is selected. The EVOP based search for space-filling latin hypercube is an evolutionary process where the optimized sampling plan is a result of a nonrandom survival of random variations.

The sequential process for a space-filled latin hypercube generator is initiated by the random hypercube generator. The space-filling metric is then optimized for all the competing hypercubes that are generated as a result of random mutations by an evolutionary operation. Finally, the comparison function based on the initial maximin criterion developed by Johnson et al. (1990)^[57] is used to select the best of the optima found for the various q ’s.

3.1.2. Selecting the type of surrogate model

The suitability of a surrogate method depends entirely on its intended use. Surrogate methods vary in capability, computational prowess and ease of applicability. Each method has its own limitations and advantages. Some of the commonly used methods include the polynomial models, the radial basis function, kriging method and support vector regression. The master thesis presented focuses on the use of kriging method for its ease of implementation, accuracy and reliability. Radial basis function was initially implemented, but owing to the advanced accuracy of the kriging method for the specific applications, it is presented in detail.

3.1.3. Training/Validation of the Surrogate

A prerequisite to the training of a surrogate model is to start with a set of sample data, as shown in Eqn. 3.4. Although, we could obtain the input space in latin hypercube form (X), we would still need to obtain the output response for those inputs (y) in order to have a complete set of sampling data. Once that is available, our target with a surrogate model would be to use it to predict (y) at various unknown values of (X).

The core problem is to attempt to learn a mapping of $y = f(x)$ that converts the vector x into a scalar output y . Collecting the output values $y^{(1)}, y^{(2)}, \dots, y^{(n)}$ that result from the set of inputs $x^{(1)}, x^{(2)}, \dots, x^{(n)}$ and finding a best guess $\hat{f}(x)$ based on these known observation presents the most generic solution to the problem. To be capable of accomplishing this learning process, one would initially proceed by identifying the set of observations of the inputs that have a significant impact on f . In other words, identifying the shortest possible design variable vector $x = \{x_1, x_2, \dots, x_k\}^T$ by sweeping the ranges of all of its variables that can elicit the behavior of the model. Subsequent recruitment of n of these k -vectors into a list $X = \{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}^T$ ensures that this represents the design space as thoroughly as possible.

$$\psi^{(i)} = \exp \left(- \sum_{j=1}^k \theta_j |x_j^{(i)} - x_j|^{p_j} \right) \quad (3.8)$$

This basis function is used in the *Kriging* method.^{[60]a} In the process of building a kriging model, one would start with a set of sample data $X = \{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}^T$ and its corresponding observed responses $y = \{y^{(1)}, y^{(2)}, \dots, y^{(n)}\}^T$ and try to find an expression for a predicted value at a new point x . Although the observed responses may be coming from a deterministic process, we begin with a rather abstract concept by viewing the observed responses as being derived from a stochastic process. This is denoted by the set of random vectors

$$\mathbf{Y} = \begin{pmatrix} Y(x^{(1)}) \\ Y(x^{(2)}) \\ \vdots \\ Y(x^{(n)}) \end{pmatrix} \quad (3.9)$$

^aThe basis function is similar to a Gaussian radial basis function. The only difference is that the Kriging method allows the width of the basis function and the exponent to vary from variable to variable and in each dimension of x respectively.

The random field has a mean of $\mathbf{1}\mu$ ($\mathbf{1}$ is a $n \times 1$ column vector of ones). The correlation between the random variables with each other depends on the spacial distance between them and can be described using the basis function expression

$$\text{cor}[Y(\mathbf{x}^{(i)}), Y(\mathbf{x}^{(l)})] = \exp\left(-\sum_{j=1}^k \theta_j |x_j^{(i)} - x_j^{(l)}|^{p_j}\right) \quad (3.10)$$

The number of variables and hence the number of dimensions of the hypercube is denoted by k . A complete set of correlation between all possible observed data would result in a symmetric $n \times n$ matrix.

$$\boldsymbol{\Psi} = \begin{pmatrix} \text{cor}[Y(\mathbf{x}^{(1)}), Y(\mathbf{x}^{(1)})] & \cdots & \text{cor}[Y(\mathbf{x}^{(1)}), Y(\mathbf{x}^{(n)})] \\ \vdots & \ddots & \vdots \\ \text{cor}[Y(\mathbf{x}^{(n)}), Y(\mathbf{x}^{(1)})] & \cdots & \text{cor}[Y(\mathbf{x}^{(n)}), Y(\mathbf{x}^{(n)})] \end{pmatrix} \quad (3.11)$$

The matrix $\boldsymbol{\Psi}$ describes the correlated set of our observed data. This correlation depends of the data depends on the absolute distance between the sample points $|x_j^{(i)} - x_j^{(l)}|$ and the parameters θ_j and p_j . It is intuitive to understand how changes in the distance between the points affects the correlation.

- When two points move close to each other, $x_j^{(i)} - x_j \rightarrow 0$, $\exp\left(-|x_j^{(i)} - x_j^{(l)}|^{p_j}\right) \rightarrow 1$. That is the points show very close correlation and $Y(\mathbf{x}_j^{(i)}) = Y(\mathbf{x}_j)$.
- Similarly when the points move wide apart, $x_j^{(i)} - x_j \rightarrow \infty$, $\exp\left(-|x_j^{(i)} - x_j^{(l)}|^{p_j}\right) \rightarrow 0$. Thus, the points have no correlation.

Fig. 3.6 shows this particular nature of correlation being affected by the separation. The parameter p_j affects the smoothness of the curve. The ‘smoothness’ parameter determines how quickly the correlation drops as two points move away from each other. Lower value of p_j , as shown in the figure, leads to steep drop of the correlation with even a small separation.

Another parameter that affects the correlation is θ_j . It can be visualized as a width parameter that determines how ‘active’ a function is to a variable. The effect is shown in Fig. 3.7. A low θ_j would mean that all the points would have a high correlation and thus they would be no significant changes in $Y(\mathbf{x}_j)$, which would remain similar across the sample plane. A higher value of θ_j , however, would lead to significant differences between the $Y(\mathbf{x}_j)$ ’s θ_j .

One could clearly understand how this affects the correlation by considering a simple example of reactor converting reactant A to B. We would like to measure the output product flow for varying reaction temperature (x_1), input feed flow (x_2) and the color of the walls of the reactor (x_3). Since one would expect the color of the walls of the reactor to have no influence on the product flow, $\theta_3 = 0$. Since, the input product flow will have the biggest impact on the output flow, an impact more than the temperature of the reactor, $\theta_2 > \theta_1 > 0$. The ‘activity’ parameter proves immensely helpful in higher dimensional problems where it becomes difficult to know the

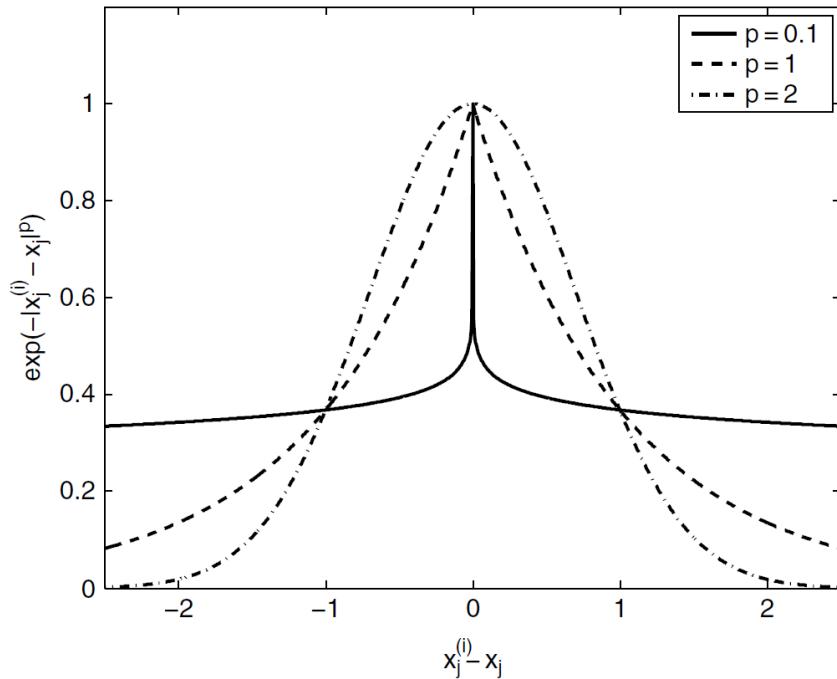


Fig. 3.6. The effect of varying p on the correlation

effect of certain variables. Careful consideration of values of θ would lead to the most important variables and sometimes even the elimination of unnecessary ones.

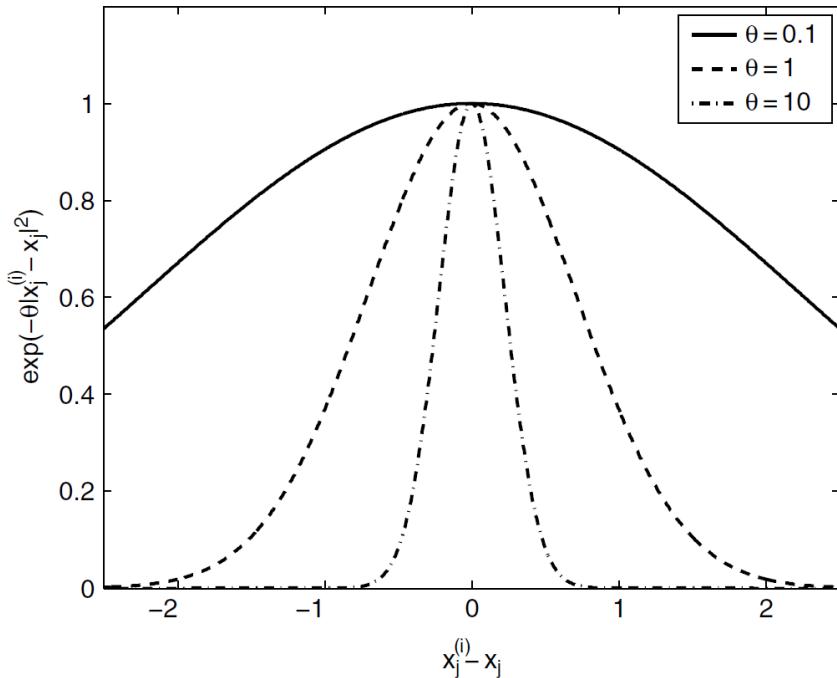


Fig. 3.7. The effect of varying θ on the correlation

The values of θ and p cannot be estimated directly. We therefore, choose to set their values in order to maximize the likelihood^b of \mathbf{y} . That is, we are asking the model to minimize the error in evaluating the \mathbf{y} for the observed data. Therefore, we are trying to build a model based on the observed data \mathbf{y} . The likelihood estimations must therefore eliminate all error in the observed data. That is the equivalent as to say that the expected performance of the surrogate model in its training points should be 100%. The likelihood expressed in terms of the sample data will be

$$L = \frac{1}{(2\pi\sigma^2)^{n/2}|\boldsymbol{\Psi}|^{1/2}} \exp \left[-\frac{(\mathbf{y} - \mathbf{1}\mu)^T \boldsymbol{\Psi}^{-1} (\mathbf{y} - \mathbf{1}\mu)}{2\sigma^2} \right] \quad (3.12)$$

To simplify the likelihood maximization we take the natural logarithm

$$\ln(L) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2} \ln |\boldsymbol{\Psi}| - \frac{(\mathbf{y} - \mathbf{1}\mu)^T \boldsymbol{\Psi}^{-1} (\mathbf{y} - \mathbf{1}\mu)}{2\sigma^2} \quad (3.13)$$

Taking the derivatives of this equation to μ and σ and equating to zero, we can end up with maximum likelihood estimates (MLEs) for μ and σ^2 :

$$\hat{\mu} = \frac{\mathbf{1}^T \boldsymbol{\Psi}^{-1} \mathbf{y}}{\mathbf{1}^T \boldsymbol{\Psi}^{-1} \mathbf{1}} \quad (3.14)$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{1}\mu)^T \boldsymbol{\Psi}^{-1} (\mathbf{y} - \mathbf{1}\mu)}{n} \quad (3.15)$$

These MLEs can now be substituted back in the Eqn. 3.13 and constant terms removed to get a *concentrated ln-likelihood function*:

$$\ln(L) \approx -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2} \ln |\boldsymbol{\Psi}| \quad (3.16)$$

The concentrated ln-likelihood function depends on the values of the parameters of θ and p . An optimization was performed to maximize the value of Eqn. 3.16. Since the concentrated ln-likelihood function is easy to compute, we use a global search method such as a genetic algorithm search for tuning the values of θ and p within a certain range in order to maximize Eqn. 3.16. θ was searched on a logarithmic scale, as shown in Fig. 3.7. The suitable bounds of the search were between 10^{-3} to 10^2 . However, the user has the control to change the bounds if necessary. Although, tuning \mathbf{p} is an advantage, it was preferred to set a constant value of $p=2$.

^bA simple analogy to understand likelihood can be attained by taking the example of a coin toss. Let us assume that we have a fair coin and we flip it 10 times. The probability that one would end up with heads up every time is the same as to say, if one would flip a coin 10 times and end up with heads every single time, and ask the question what is the likelihood that the coin was fair.

3.1.4. Prediction using the Model

So far, we have chosen correlation parameters in such a way that the likelihood of the observed data is maximized. To use the model for a prediction, say \hat{y} at \mathbf{x} , it makes sense to think that it will be consistent with correlation parameters that have been found from the observed data. Hence we can choose a prediction which will maximize the likelihood of the observed data along with the prediction, for the set of correlation parameters.^[61] That is, we augment the observed data with the new prediction and also define the correlation vector between the observed data and the new prediction.^[62]

$$\tilde{\mathbf{y}} = \{\mathbf{y}^T, \hat{y}\}^T \quad (3.17)$$

$$\psi = \begin{pmatrix} \text{cor}[Y(\mathbf{x}^{(1)}), Y(\mathbf{x})] \\ \vdots \\ \text{cor}[Y(\mathbf{x}^{(n)}), Y(\mathbf{x})] \end{pmatrix} = \begin{pmatrix} \psi^{(1)} \\ \vdots \\ \psi^{(n)} \end{pmatrix} \quad (3.18)$$

The corresponding augmented correlation matrix would be:

$$\tilde{\Psi} = \begin{pmatrix} \Psi & \psi \\ \psi^T & 1 \end{pmatrix} \quad (3.19)$$

It is worthy to note here that the last element of the matrix in the augmented correlation matrix $\tilde{\Psi}$ is $\mathbf{1}$. That is the leading diagonal of the correlation matrix is one. The ln-likelihood of this augmented data would be:

$$\ln(L) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2} \ln |\tilde{\Psi}| - \frac{(\tilde{\mathbf{y}} - \mathbf{1}\hat{\mu})^T \tilde{\Psi}^{-1} (\tilde{\mathbf{y}} - \mathbf{1}\hat{\mu})}{2\hat{\sigma}^2} \quad (3.20)$$

Only the last term of this equation depends on \hat{y} and hence only this term is considered in the maximization. Substituting in expressions for $\tilde{\mathbf{y}}$ and $\tilde{\Psi}$ gives

$$\ln(L) \approx \frac{-\left(\mathbf{y} - \mathbf{1}\hat{\mu}\right)^T \begin{pmatrix} \Psi & \psi \\ \psi^T & 1 \end{pmatrix}^{-1} \left(\mathbf{y} - \mathbf{1}\hat{\mu}\right)}{2\hat{\sigma}^2} \quad (3.21)$$

Thus, to find the MLE for \hat{y} , we would need differentiate Eqn. 3.21 with respect to \hat{y} and equate it to 0.

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \psi^T \Psi^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu}) \quad (3.22)$$

This equation gives the prediction at new point \hat{y} at \mathbf{x} based on the observed data \mathbf{y} .

3.1.5. Regressing Kriging

Yet another a possible scenario is when a noise in the training data or testing input variables would need to be filtered. A simple and efficient approach is regressing the model in order to eliminate certain points that are inconsistent with the correlations generated between all other data points. This simply means adding a regression constant λ to the leading diagonal of the correlation matrix Ψ . A similar method of prediction can be used for the regressed kriging predictor.

$$\hat{y}_r = \hat{\mu}_r + \psi^T(\Psi + \lambda\mathbf{I})^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}_r) \quad (3.23)$$

where,

$$\hat{\mu}_r = \frac{\mathbf{1}^T(\Psi + \lambda\mathbf{I})^{-1}\mathbf{y}}{\mathbf{1}^T(\Psi + \lambda\mathbf{I})^{-1}\mathbf{1}} \quad (3.24)$$

\mathbf{I} is a $n \times n$ identity matrix. The regression constant λ can be determined by the same maximum likelihood estimates as for the other correlation parameters. The user can choose whether or not to predict the model using a regression parameter. Introduction of the regression parameter as a single constant and not as a parameter for every variable saves considerable computation time. Regressed models do not necessary pass exactly over all the points (*i.e.*, preventing over-fitting) and provides a smooth function by filtering noise.

3.2. Implementing the Surrogate Model tool

The modeling approach to build the surrogate model has been discussed in section 3.1. The modeling approach needs to be translated into an executable code that can be implemented for a process simulation environment. Training of a surrogate model relies on the need for observed data based on which the training can be done. The process simulation can be used to generate the training data, even if it is in a small range of the input variables. Once trained the model is able to replace the simulation model with acceptable accuracy, yet being simple, reliable and significantly faster responding.

3.2.1. Training of the Surrogate

Generally, a user has a working environment of a process in the form of a flowsheet. The aims of potentially building and utilizing a surrogate model arises so that the user would not need to use the flowsheet model developed in software such as Aspen Plus®. However, to use such a surrogate satisfying the needs would require training the model. The mathematical concept behind the model discussed in section 3.1 needs to be implemented in a way the user can easily take advantage of its potential.

The need for training the model in a space-filled sample space is obvious. The mode of training is developed in a MATLAB® vR2014b. For reasons pertaining to simplicity and ease of use, the graphic-user interface was developed in Microsoft Excel® 2013. The interfaces acts as a communication bridge, connecting and conveying user needs to the training code implemented in MATLAB®. The user-friendly platform is able to aid the user in building a training range for

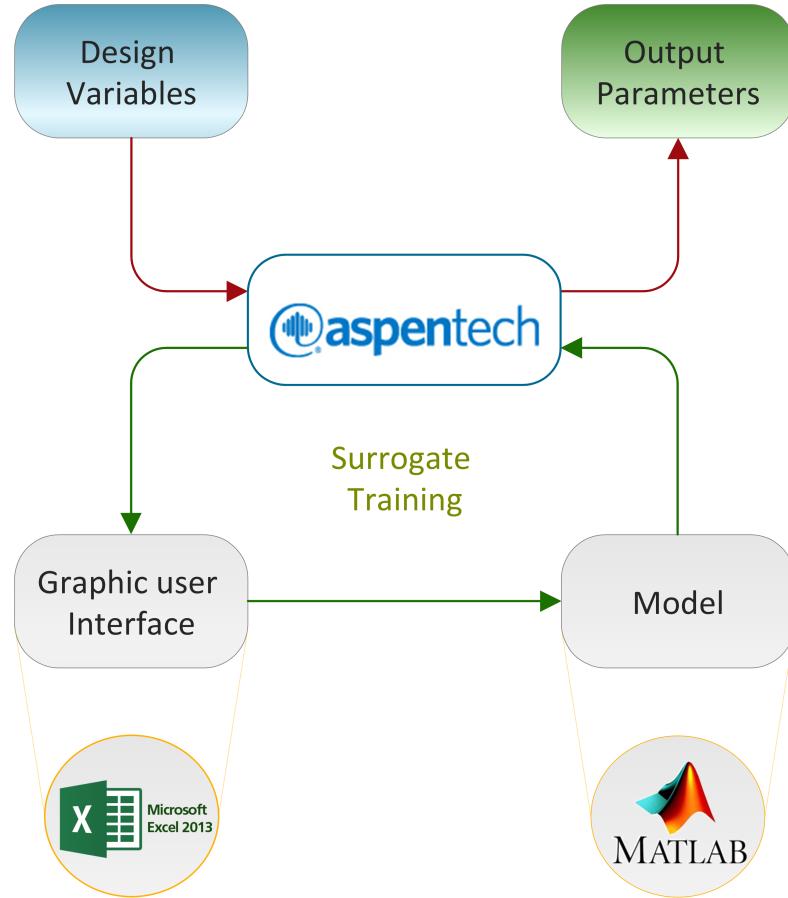


Fig. 3.8. Training of the surrogate model. The varying input design parameters are fed to the simulated Aspen Plus® flowsheet to realize output estimates. A graphic–user interface in the form of Microsoft Excel® 2013 is able to derive the observed data from Aspen Plus® and feed it to the surrogate model coded in MATLAB®.

which he can easily run the original simulation to obtain the observed data that the model uses as its basis in predicting at new set of input variables without the need for a simulation model. The training of the model is shown in Fig. 3.8.

The following steps are implemented in training the simulated Aspen Plus® flowsheet:

- Identifying and specifying the number of design variables that are being varied and their range.
- Identifying the output parameters that needs to be trained and visualized
- Using the graphical user interface to generate a space-filled training range.
- Using the simulated Aspen Plus® flowsheet to generate the observed data to be used for prediction.

The MATLAB® code is based on the methods detailed in the section 3.1. Implementing the methods in a simpler, swift and a user-friendly code, is important for the way that the maximum likelihood estimates (MLE) are calculated (see Eqn. 3.14–3.16). The matrix algebra involved in

likelihood estimates is time consuming and thus efficient usage of the code is important. The approach used to make it effective consists of the following points:

- Calculation of ln-likelihood needs several matrix inversions. For a computationally efficient calculation, a Cholesky factorization^[62] was performed. Backward and forward substitutions are then used for each inverse.
- Regressions were used when Ψ becomes too close to being singular^[63], as Cholesky factorization only works for positive-definite matrix.^[62]

3.2.2. Prediction

Once the model is trained based on the observed data, the implementation of the model to predict at unknown input variable values is fairly simple. The trained surrogate model uses the correlation parameters, augments the new unknown input vector with the already observed data and predicts the output values. The user only would need to give the unknown input values for which he would like to know the output value for. The important thing to be kept in

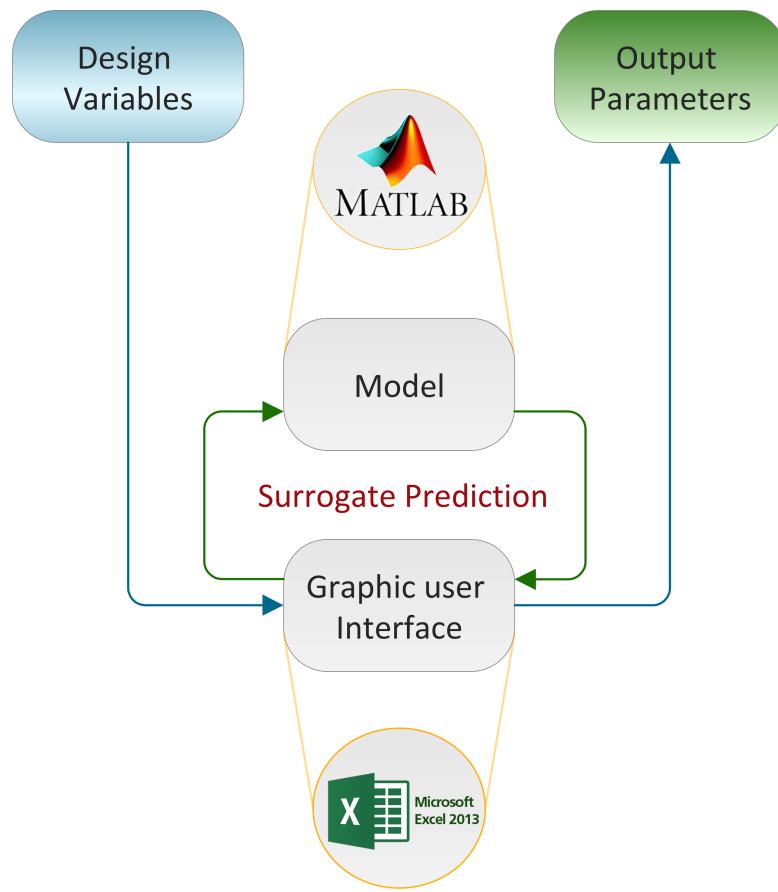


Fig. 3.9. Prediction using the surrogate model. Once the surrogate model is trained, it can replace the Aspen Plus® model. Microsoft Excel® 2013 acts as the graphic-user interface to input data and visualizing the results generated by the model built in MATLAB®.

mind is that the model will only be able to predict for output parameters it had been trained for.

The trained surrogate model can then successfully replace the Aspen Plus® simulated model, provided it is trained for the desired output vectors. Fig. 3.9 shows the surrogate model replacing the original model and now the graphic-user interface acts as a bridge to input data and obtain the results of the model.

3.2.3. Output Model Parameters

Apart from the prediction itself, there are several key parameters that may well be of equal importance for the user. θ and p are important model parameters that can convey the behavior of the model. Although the user works with a fixed value of p , the values of θ comprise information pertaining to the importance of the individual variables used in the modeling process. The user might find such information essential in understanding the relative importance of the various input variables. Even though the model is not able to convey the actual effect of every variable, it gives a relative importance between the variables that build the correlation vectors.

Regressing kriging is sometimes preferred and in such cases, the regression constant, λ , can throw light on the amount of deviation that the model has been regressed in order to reduce the noise that is caused by an outlier. These parameters can become crucial in defining a surrogate model. Even though the user, as a result, would care about his output variable values, deeper understanding of the model can be achieved by paying a closer attention to the model parameters. The graphical user interface, along with providing the results of the output variables, also provides the values of the correlation parameter θ for every variable and the regression constant, λ , should the user choose to regress the model.

3.3. Surrogate Model example

Visualizing the surrogate model becomes a whole lot easier with an example. Let us take an example of a working simulation in the form of sorbitol production explained in chapter 2. We would want to train a surrogate model to visualize two of the output parameters, such as, the amount of sorbitol produced and the energy needed for the process. In a process such as the one for sorbitol production, the majority of changes in the process arise from changes in the D-glucose raw material, in terms of mass flow and temperature of the crystallizer and evaporator in the process. The example is presented with a few screen captures in appendix C.1.

- The user starts by specifying the input variables in the graphic-user interface, namely the D-glucose raw material flow, the evaporator temperature and the crystallizer temperature.
 - D-glucose raw material - 800 kg h^{-1} — 1000 kg h^{-1}
 - Evaporator temperature - 170°C — 190°C
 - Crystallizer temperature - 5°C — 15°C
- The user now has to specify the range of the input variables that he wishes the model to be trained for.

- The user has the discretion to determine and state how many training data points he would like to have. It is a rule of thumb that more the number of points that the model is trained for, the better it will be able to understand the behavior of the process. An efficient sampling (*i.e.*, a space-filled latin hypercube) avoids the computational explosion that would happen from all the combinations when defining too many data points or too many variables.
- The training input data is generated by the MATLAB® code and the user now will have to obtain the output parameters for the input training data.
- This can be achieved by Aspen Plus® from which the user can obtain the output values of the training data. The training of the surrogate using Aspen Plus® is a manual step and proves critically time consuming. However, further developments can make this step automatic to save considerable time and improve computational speed.
- Once this is generated, the training information (both input and output) is available to the surrogate, which can then understand the behavior. From now on, the surrogate model can be able to stand-in.
- The user can now use the graphic-user interface to predict the values of the output parameters for any changes in his input variables. Even values of input variables beyond the range that was given for the training process, can be realized using the surrogate.

The graphical user interface provides a platform for communicating with the model. The prediction results along with the important model parameter values are displayed for the user to understand and improvise if needed. The interface also allows for customization in choosing the range for genetic algorithm search of the model parameters or in opting to regress the model. Such intricate controls were introduced for deeper understanding of the model by the user. The output model parameters also convey the importance, in terms of the level of importance, of the individual variables on each and every output parameter.

3.4. Summary

Surrogate models can serve as a valid replacement for complex simulated flowsheets thus reducing the complexity. The gain in speed and simplicity is attained without having to compromise on accuracy or dependability. The steps involved in the construction of a surrogate model were discussed.

The use of space-filled latin hypercubes is an efficient method to avoid computational explosion that would happen from the combinations of large number of data points or variables. It provides a uniform spread of points in the design space necessary for high accuracy in the prediction. Space-filled latin hypercube sampling was implemented using MATLAB®. Kriging method was implemented for its ease of implementation and reliability. The training of the surrogate model and its use in prediction were executed with the help of a graphical user interface built in Microsoft Excel® that interacts with the mathematical concepts coded in MATLAB®.

Regression is desired when there is noise in the training data. Regressing kriging is a method to provide smooth kriging models not exactly passing through all the data points and avoiding

the noisy data. The graphical user interface allows for additional customization apart from communicating with the model, such as specifying the range for the genetic algorithm search. Training of the surrogate model with the help of Aspen Plus® is done manually. This step is significantly time consuming and future developments will be focused on making it automatic to improve computation speed. Surrogate models were constructed for the joint bio-based production of PET and PU. The model parameters and the accuracy of the surrogate predictions are discussed in chapter 4.

4. Results and discussion

The impact of the simulated models were analyzed to present a clearer picture in terms of their potential advantage. Production of the bio-based PET and PU have been modeled as a step-wise process progressing through sorbitol. The integrated biorefinery for the production of these two commercially essential polymers is shown in Fig. 4.1. The term ‘integrated biorefinery’ throughout this text means the joint PU-PET production system. Other aspects such as heat integration and mass integration have not been explored and presents room for further developments. As one might notice, it is very important to analyze whether the integration is essential and beneficial. Although the production of two important polymers seems lucrative, the polymerization of PET utilizes TPA and MDI is used in the PU polymerization, both of which are expensive raw materials that are derived from non-renewable resources. Therefore, it is justifiable that one might want to produce only one of the polymers. Since the bio-based monomers ethylene glycol and propylene glycol are also commercially important, they can be sold as well. The scenarios that were considered are:

- Production of PET and selling propylene glycol to the market – *A PET biorefinery*
- Production of PU and market out ethylene glycol – *A PU biorefinery*
- Integrating and producing both the polymers in a multiple product portfolio – *An Integrated biorefinery*

The integrated biorefinery is synonymous to the problem at hand in terms of incorporating several platforms into a single optimum biorefinery. It is essential to analyze the processes both economically and environmentally to answer questions that are directed at the use of more raw materials which may not be bio-based. The importance of the integrated biorefinery will be

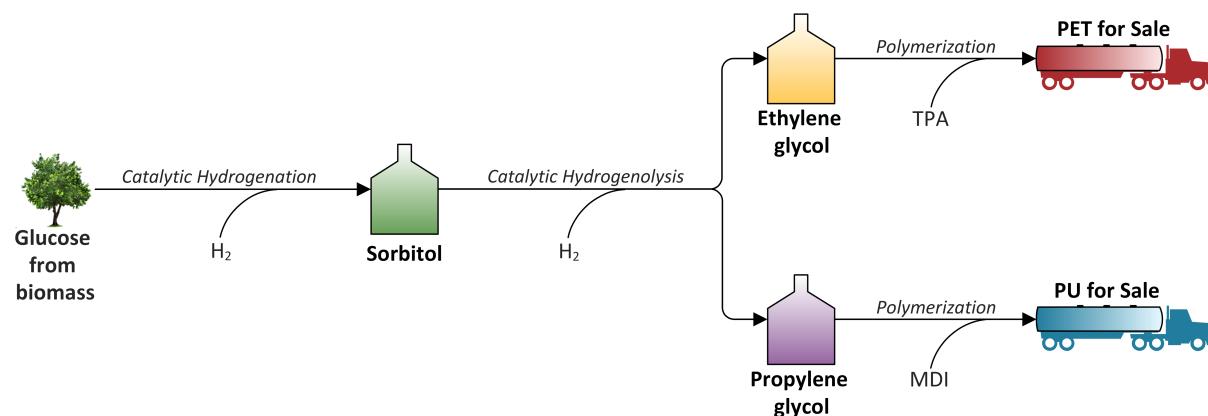


Fig. 4.1. Road map showing the integrated PET and PU biorefinery. The monomers ethylene glycol and propylene glycol are bio-based as they are produced from D-glucose and through sorbitol. However, the monomers TPA and MDI are derived from non-renewable sources.

shown with the example of integrating PET and PU productions and its advantage over producing just one of the polymers. Achieving such an integrated biorefinery might seem simple with two polymer products. But the vision of integrating several products and simulations is challenging involving great innovation potential. Surrogate models can pave the way of and overcome simulation and optimization limits and, therefore, will be essential in realizing this vision. The importance of the surrogate modeling will be shown with a case study example with the integrated PET and PU biorefinery.

4.1. Importance of an Integrated Biorefinery

A crucial aspect in this integration is that the production of bio-based ethylene glycol (monomer in the production of PET) and the production of bio-based propylene glycol (monomer in the production of PU) are not independent. Specifically, the relative output volume of the two glycol co-products is fixed by the stoichiometry of the catalytic hydrogenolysis of sorbitol. Thus, the biorefinery represents a joint co-production system, where system expansion leading to the identification of the displaced or avoided process or product and the utilization of the dependent co-product becomes relevant.

On complete utilization of the dependent co-product, the environmental impact of the entire process will be shouldered or ascribed to the determining co-product and credited for displacing the reference production of the dependent co-product. Therefore, identification of the product the determines the volume of the co-production process (determining co-product) becomes important and is based on economic model assumptions. The credit given to the determining co-product in terms of cost is not necessarily referring to a specific displayed process. Rather, the market price of the dependent co-product is subtracted from the operational costs incurred for the determining co-product.

4.1.1. Simulation results

PET biorefinery

The PET biorefinery focuses on the production of the bio-based polymer PET. The monomer used for the production, ethylene glycol, is produced from sorbitol as a result of catalytic hydrogenolysis. Propylene glycol, which is also co-produced in the same reaction, is assumed to be sold in the market and thus, displaces its conventional production. Thereby, the environmental impact of producing propylene glycol with a conventional process^[64] will be subtracted from that of PET. The economic displacement will be manifested by subtracting the market price of propylene glycol, not necessarily specific to the conventional production process it displaces, rather, trims its demand. PET shoulders the burden for the entire production process however. The analysis of the biorefinery is concisely shown in Tab. 4.1.

The simulated PET biorefinery is initiated by the catalytic hydrogenation of 1000 kg h^{-1} of D-glucose. The product of the hydrogenation, pure crystal sorbitol, is obtained at a yield of 97.8%. Sorbitol is subsequently broken down to lower glycals, most importantly, ethylene glycol and propylene glycol.

Tab. 4.1. Gate-to-gate inventories – PET Biorefinery

Classification	Component	Value / kg _{PET}
Input Material	D-glucose	2.11 kg
	Water	15.81 kg
	Hydrogen	0.07 kg
	TPA	0.86 kg
Dependent co-product	Propylene glycol	0.51 kg
Utilities	Steam	18.51 kJ
	Electricity	4.28 kJ
	Cooling Water	1461 kg
Waste	Incineration	1.34 kg
	Waste treatment	0.29 kg

The simulation results in the production of 473 kg h^{-1} of PET as the determining co-product and 243 kg h^{-1} of propylene glycol as the dependent co-product. A considerable amount of water is utilized for the process. The control on the composition of the inlet feed of D-glucose for its catalytic hydrogenation and diluting the sorbitol before its break down are the major consumers of water in the process. Yet, most of the water is evaporated and does not impact the waste water treatment of the process. One might expect water evaporation to be an energy intensive process. It could also be considered as a source of low grade waste heat, which in a heat integrated biorefinery could be utilized elsewhere. However, whether this demanding water evaporation affects the overall economic balance of the process will be further analyzed in this chapter of the thesis. Another interesting observation is the amount of cooling water required for the process. The impact of the utilities will also be analyzed. The material flows in the PET biorefinery is shown in the form of Sankey diagrams in Fig. 4.2.

PU Biorefinery

In the case of PU biorefinery, propylene glycol produced as a result of the catalytic conversion of sorbitol, is polymerized through the step growth polycondensation with the other monomer 4,4-methylene diphenyl diisocyanate (MDI) to produce Polyurethane (PU). In this case, ethylene glycol is considered as the dependent co-product to displace the conventionally produced ethylene glycol^[64] and PU as the determining co-product. The simulation results of the process are shown in Tab. 4.2.

Similar to the PET biorefinery, the PU refinery utilizes 1000 kg h^{-1} of D-glucose input feed. The reaction pathway remains the same until the production of glycols. The propylene glycol

Tab. 4.2. Gate-to-gate inventories – PU Biorefinery

Classification	Component	Value / kg _{PU}
Input Material	D-glucose	0.96 kg
	Water	7.17 kg
	Hydrogen	0.03 kg
	MDI	0.77 kg
Dependent co-product	Ethylene glycol	0.15 kg
Utilities	Steam	8.22 kJ
	Electricity	1.94 kJ
	Cooling Water	653.4 kg
Waste	Incineration	0.61 kg
	Waste treatment	0.04 kg

produced is then reacted with MDI for the production of 1042 kg h^{-1} of PU. 156 kg h^{-1} of ethylene glycol as the dependent co-product.

The amount of water required for the reaction processes and the amount of cooling water per kg_{PU} is lower than that of the PET biorefinery. This can be mainly explained by the significantly higher amount of PU being produced. The impact of the utilities of the process and those of the raw materials will be analyzed as a stand-alone unit and in comparison with the other biorefinery concepts discussed. The Sankey diagram showing the material flows in the biorefinery is shown in Fig. 4.3

Integrated PET and PU Biorefinery

To reiterate, the integrated biorefinery presented involves the incorporation of multiple products co-produced in a joint production. Aspects such as heat and mass integration haven't been dealt with. In such an integrated biorefinery, both the commercially important polymers are produced from ethylene glycol and propylene glycol. The bio-based glycals produced are polymerized for the production of 1042 kg h^{-1} of PU and 473 kg h^{-1} of PET. Identifying the product that determines the volume of the co-producing process is critical in such an integrated biorefinery. In this biorefinery, PU is the determining co-product. A potential larger market trend for bio-based PU and significantly higher revenue generated justify culling PU as the determining co-product and PET as the dependent co-product in the joint production. The results of the simulation per kg_{PU} is shown in Tab. 4.3 and in the Sankey diagram in Fig. 4.4.

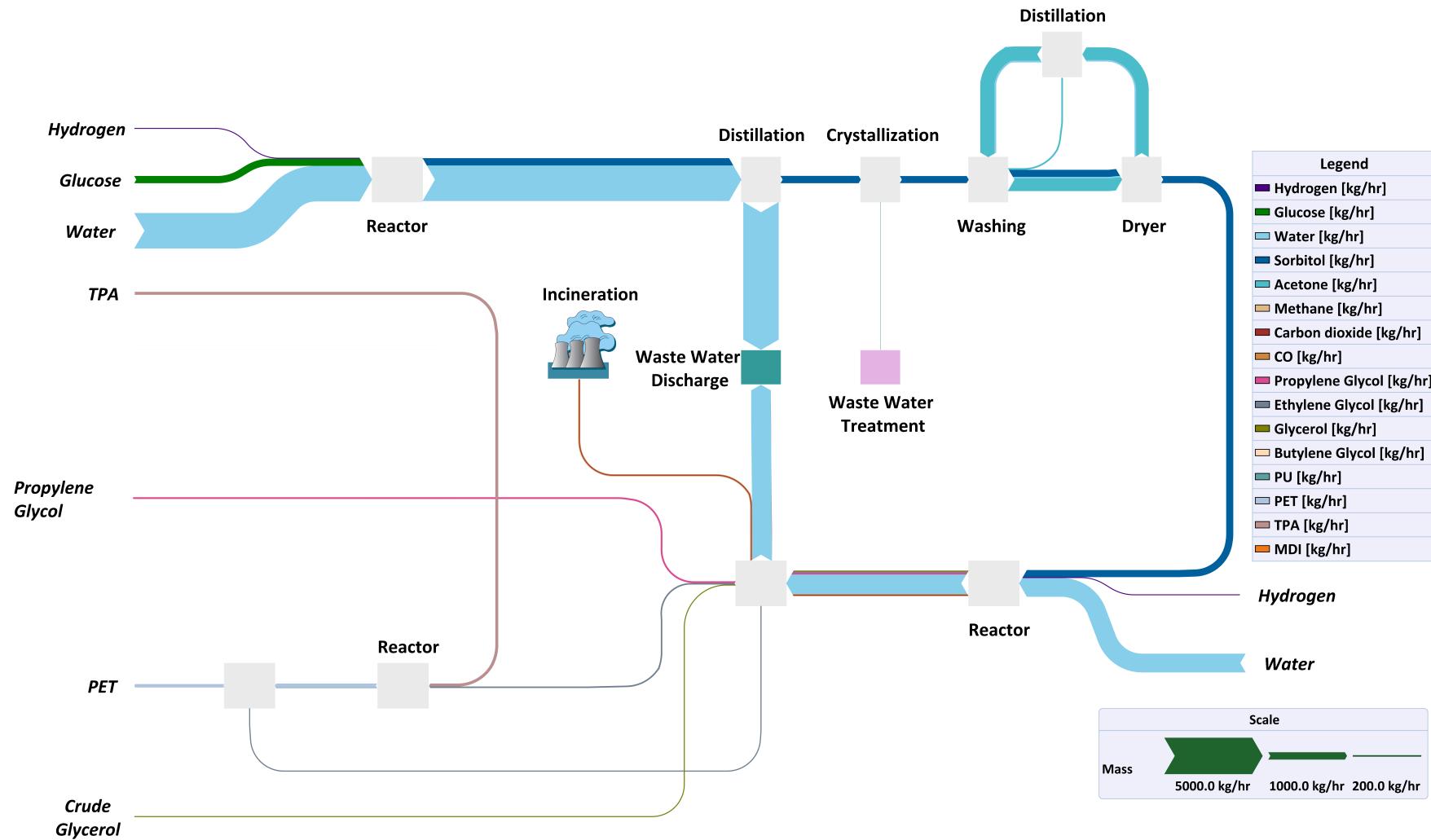


Fig. 4.2. Sankey diagram showing the material flows in the PET biorefinery

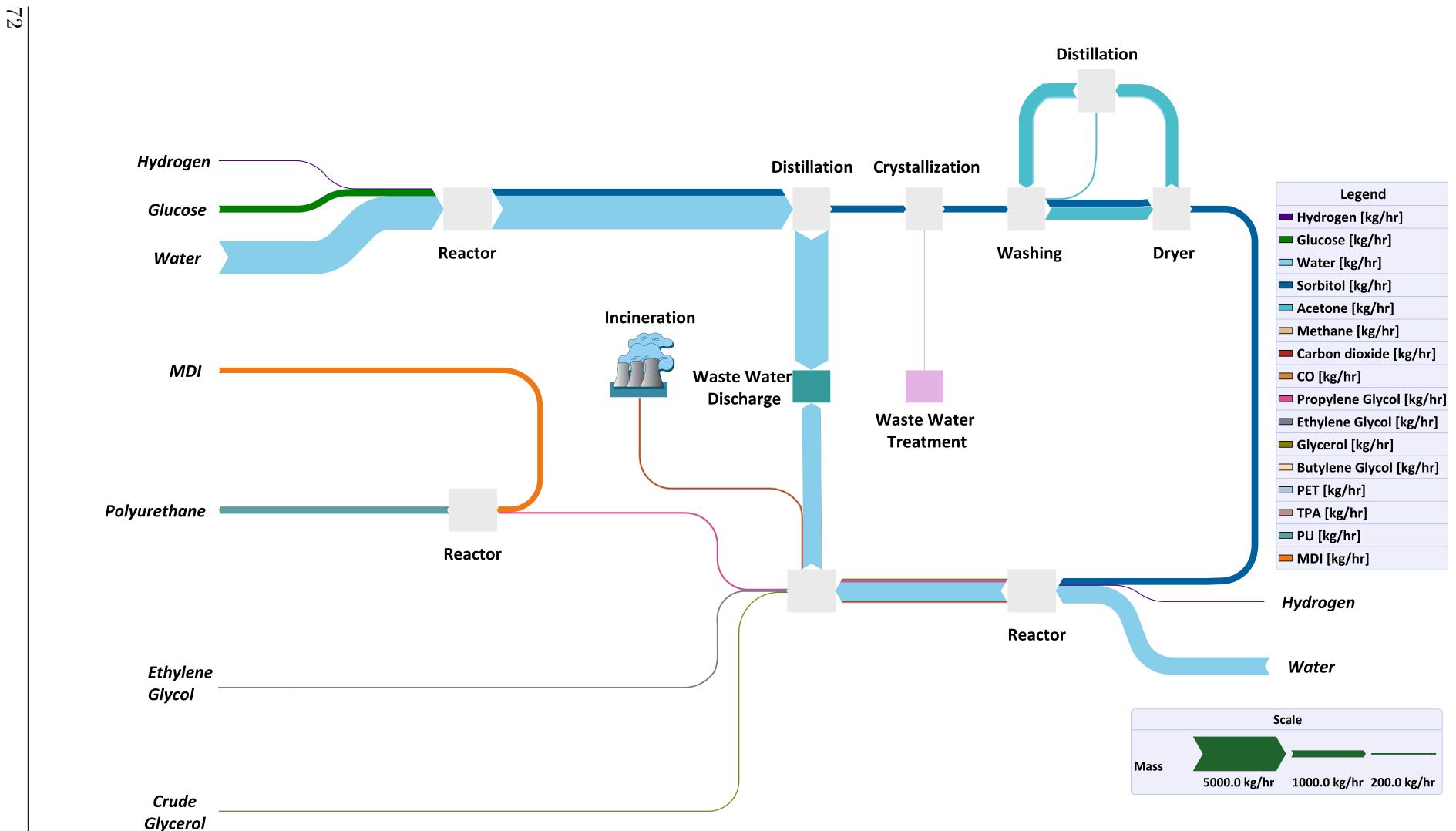


Fig. 4.3. Sankey diagram showing the material flows in the PU biorefinery

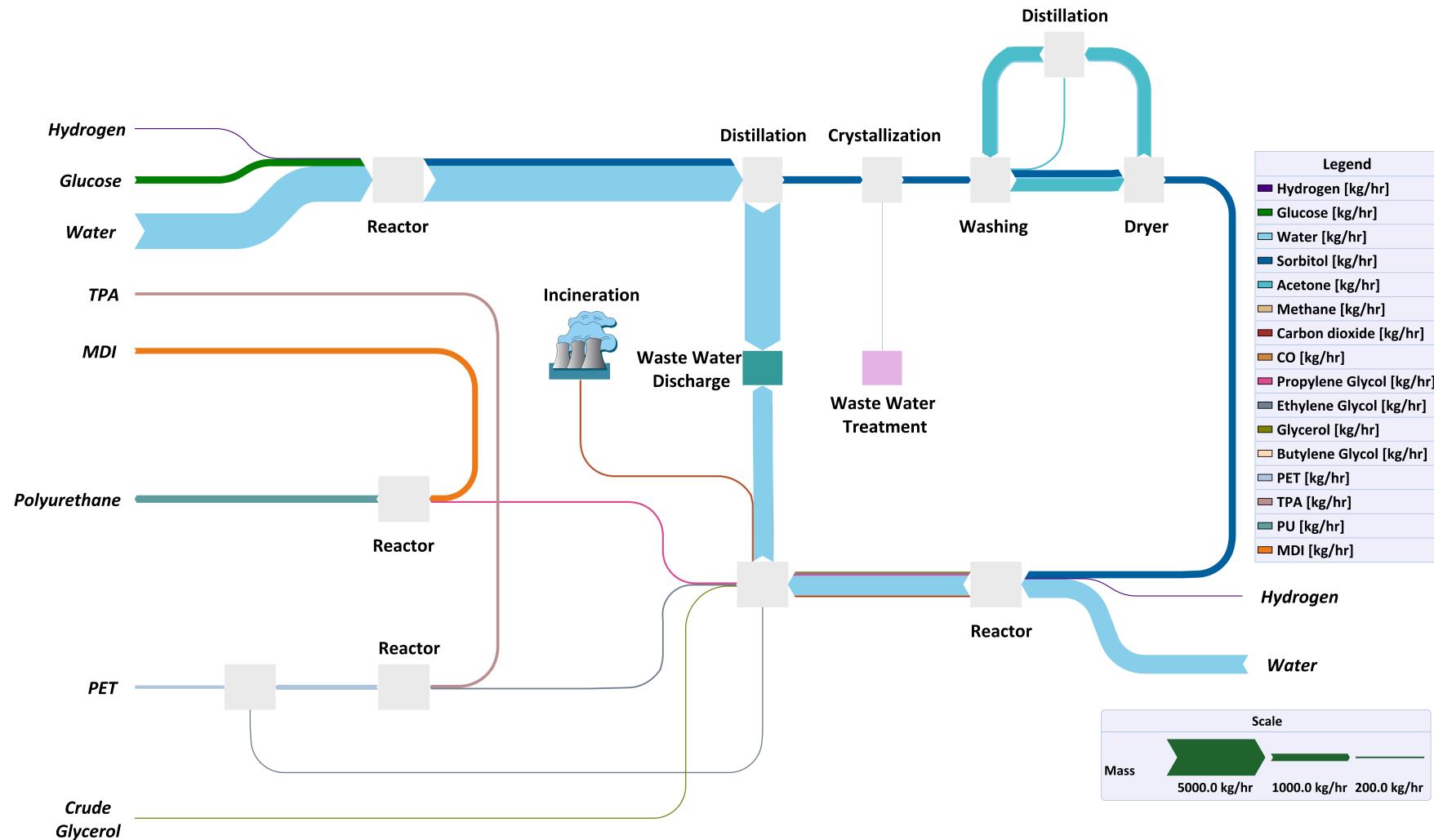


Fig. 4.4. Sankey diagram showing the material flows in the Integrated PET and PU biorefinery

Tab. 4.3. Gate-to-gate inventories – Integrated PET and PU Biorefinery

Classification	Component	Value / kgPU
Input Material	D-glucose	0.95 kg
	Water	7.17 kg
	Hydrogen	0.03 kg
	TPA	0.39 kg
	MDI	0.77 kg
Dependent co-product	PET	0.45 kg
Utilities	Steam	8.52 kJ
	Electricity	1.94 kJ
	Cooling Water	668.5 kg
Waste	Incineration	0.61 kg
	Waste treatment	0.13 kg

Comparison of the integrated biorefinery with the PU biorefinery gives similar results in terms of utilities consumed. The integrated biorefinery has an added utility and material consumption due to the polymerization reactors of both the polymers, and as well as the use of both TPA and MDI monomers. Whether this increase in the amount of raw materials consumed and utilities will have a larger impact on the process will be analyzed in comparison with the PET and PU biorefineries to pin point the advantage of the integrated biorefinery.

4.1.2. Environmental Impact Assessment

The assessment of the processes in terms of environmental impact is imperative. Comparison of the process in terms of impact on the environment is essential in isolating the integrated biorefinery's advantage. Life cycle impact assessment (LCIA) aims to describe or indicate, the impacts of the environmental loads quantified in the inventory analysis. The relevant life cycle inventory (LCI) data (*i.e.*, consumption of resources and process emissions) were estimated in a cradle-to-gate approach using the process models.^[65] The environmental impacts in the usage phase and the environmental fate of the polymers were not considered. The environmental impact is assessed using three well-known life cycle assessment (LCA) metrics: Global warming potential (GWP-100a), Cumulative energy demand (CED) and Eco-Indicator 99 (EI99). GWP is a LCA metric that accounts for the equivalence in terms of greenhouse gas emission. It is measured in $\text{kgCO}_2\text{-eq}/\text{functional unit}$ taking into account every single process, the raw materials and the waste treatments. The amount of equivalents of CO_2 released to the atmosphere for every kg of product is a direct way in measuring the environmental impact of the process.

Cumulative energy demand (CED) measures the primary energy sources that were utilized during the entire production process. This includes the energy that is required for the production of the raw materials in the biorefinery, energy used for the treatment of waste and required for gate-to-gate heating, cooling and pumping utilities. Noteworthy is the fact that renewable primary sources were not taken into consideration. This is justifiable as the motivation of the master thesis is to delineate the importance of a biorefinery over the conventional processes for the reduction in the utilization of petroleum based non-renewable energy sources.

Eco-Indicator 99 (EI99) encompasses a wide set of sustainability aspects from human health to the quality of the ecosystem in its purview. It is measured in a dimensionless form as Points/*functional unit*. Even though expressed in an abstract manner, it includes a variety of factors that contribute towards sustainability and can be used as an overall environmental damage oriented metric.

The cradle-to-gate production of raw materials not modeled in the master thesis was estimated using the Ecoinvent database.^[66] The choice of the functional unit is essential in comparative studies involving products with different functionality. 1 kg_{Pdt} (kilogram of product) is not the most appropriate functional unit when comparing PET and PU biorefineries. The products have different set of applications. Expressing their function in quantitative terms, as a functional unit, that corresponds to a reference flow to which all other modeled flows of the system are related is difficult to achieve.^[67] However, per kg_{Pdt} is at times useful to visualize productivity and efficiency of the processes which may be lost with using 1 kg_{glucose} (raw material) as a functional unit. Primary energy of the substituted fossil fuel can be used as functional unit to compare processes striving to avoid consumption of fossil fuels.^[68] In this respect, 1 kg_{oil-eq} (kilogram of oil equivalents) substituted is also used as a functional unit.

Global Warming Potential

The biorefineries are compared with each other and that of the conventional production of PET and PU. Fig. 4.5 shows the GWP-100a of the biorefineries with 1 kg_{Pdt} as the functional unit. The first column details the GWP-100a of the conventional production of PET. It is compared with the PET biorefinery producing PET and selling propylene glycol. Similarly the third and fourth column compare the conventional PU production with that of the PU biorefinery that produces PU as the determining co-product and ethylene glycol as the dependent co-product. Finally, the last column shows the integrated biorefinery producing both the polymers, PET and PU. PU is the determining co-product in the integrated biorefinery and therefore, 1 kg_{PU} will be used as the functional unit for the integrated biorefinery..

It is very distinct to observe that the conventional processes for the production of the PET and PU has a much lower impact when compared to that of the biorefineries. The PET biorefinery releases 9.63 kgCO₂-eq/kg_{Pdt} which is three times more than that of the conventional PET production. A similar trend is observed in the production of PU by a commercially producer when compared to the biorefinery. The PU biorefinery emits approximately 170% more CO₂ than the conventional process. The integrated biorefinery has the lowest total impact among the biorefineries. An interesting observation is the performance of the integrated biorefinery, which is only slightly better than the PU biorefinery. In spite of utilizing more raw materials for the production of both the polymers in the integrated biorefinery, the difference may only arise

4. Results and discussion

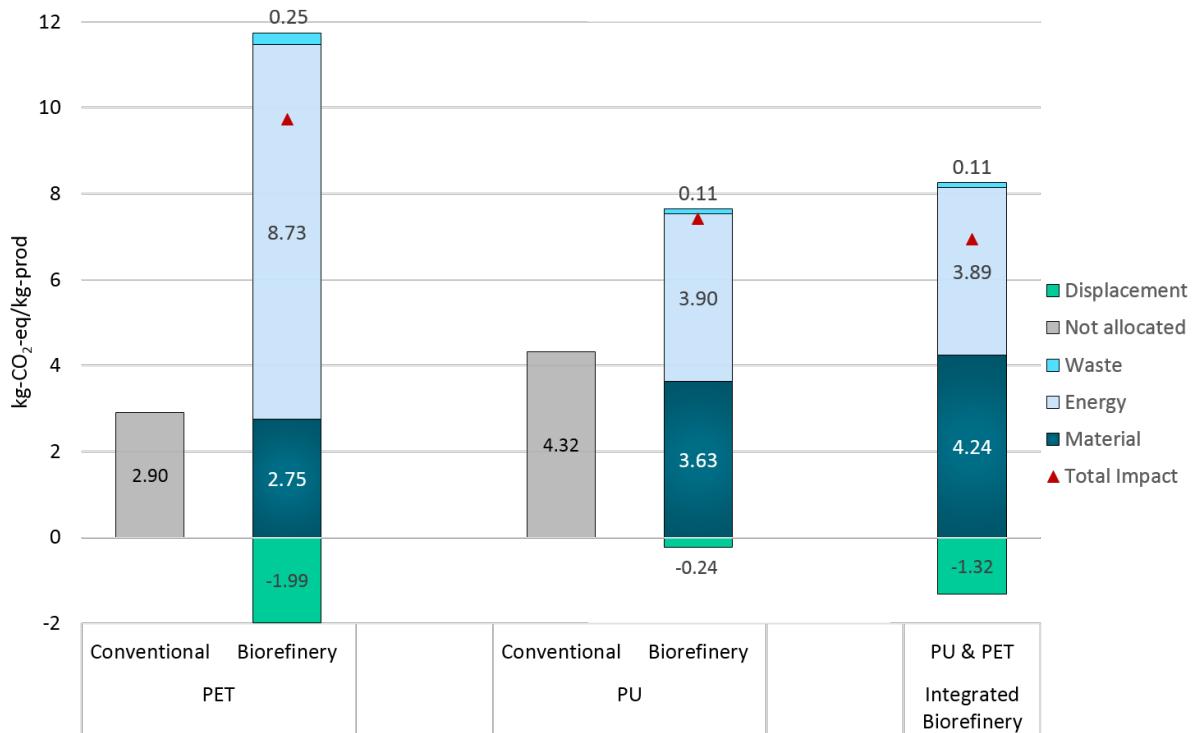


Fig. 4.5. The global warming potential (GWP-100a) of the biorefineries with 1 kg_{Pdt} as the functional unit. The first column corresponds to the production of PET by a conventional industry and the second column shows the impact of the simulated biorefinery producing only PET. A similar comparison is shown in the third and the fourth column, comparing the conventional PU production with that of the PU production only in the PU biorefinery. The last column shows the integrated biorefinery producing both the polymers. 1 kg_{Pdt} refers to PU in the integrated biorefinery as it is the determining co-product.

from the difference in the displaced co-product for which the determining process is credited (ethylene glycol in the case of PU biorefinery and PET in the integrated biorefinery). If indeed the ethylene glycol produced as a dependent co-product in the PU biorefinery were to be used in the production of PET in a conventional unit, the difference between the cases may be negligible.

A change in the functional unit to 1 kg_{glucose} (Fig. 4.6) or to 1 kg_{oil-eq} substituted (Fig. 4.7) does not improve the resolution to differentiate the PU biorefinery from the integrated scenario. Nevertheless, the use of 1 kg_{glucose} as a functional unit fails to convey the productivity of the processes. PET biorefinery produces the least amount of products in terms of volume and thereby has a lowered impact. The negligible difference between PU biorefinery and the integrated biorefinery is merely dependent on the fate of the dependent co-product. Multiple product portfolio generally improves utilization of raw materials and enhances productivity. Even though the difference may be small, an effective integration involving heat and mass integration will further enhance the advantage of an integrated biorefinery.

The raw material impact is equally important and since the mass percentage of MDI is higher in PU than the mass percentage of TPA in PET, and MDI has a higher impact, significant increase in greenhouse gases (GHG) emissions is witnessed with the use of MDI as a monomer in the production of PU.

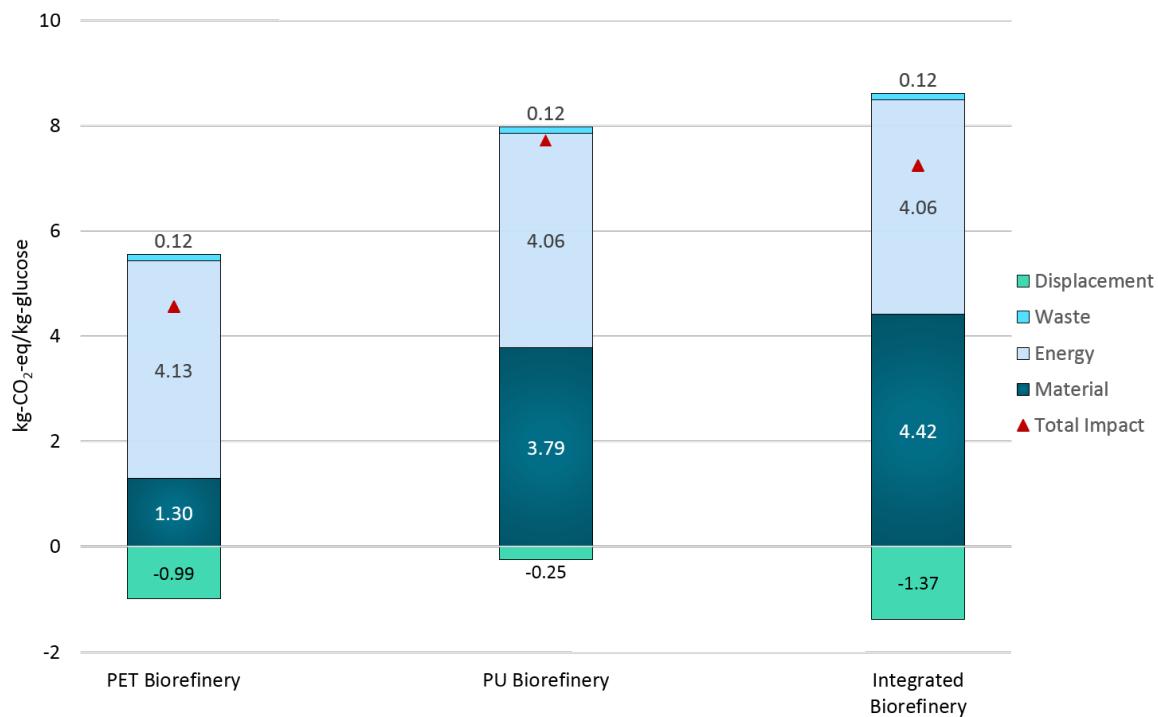


Fig. 4.6. The global warming potential (GWP-100a) of the biorefineries with 1 kg_{glucose} as the functional unit.

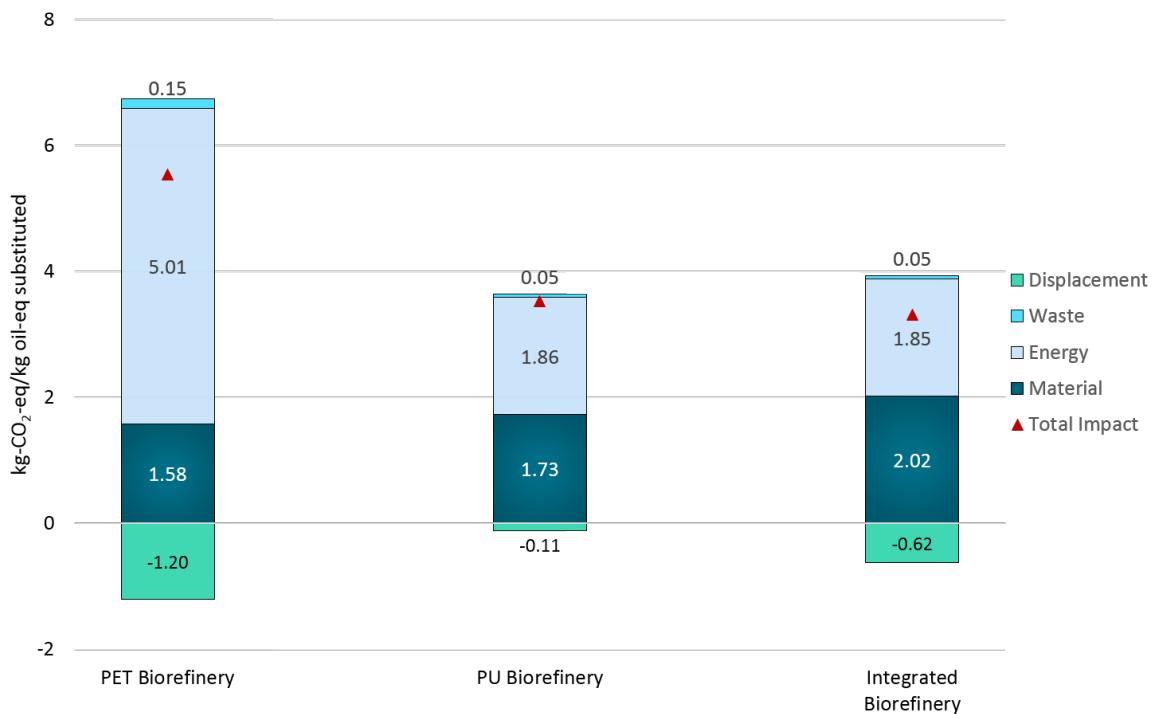


Fig. 4.7. The global warming potential (GWP-100a) of the biorefineries with 1 kg_{oil-eq} substituted as the functional unit.

Cumulative Energy Demand

The gate-to-gate energy utility consumption in the PET production played a crucial part in significantly increasing the GHG emissions. It is necessary that the results are analyzed bearing in mind the functional unit. The contribution of gate-to-gate energy utilities towards greenhouse gases (GHG) per kg_{Pdt} is higher for the PET biorefinery owing to its meager production volume. One can notice this when changing the functional unit to 1 $\text{kg}_{\text{glucose}}$. Fig. 4.8 and Fig. 4.9 show the cumulative energy demand of the biorefineries with 1 kg_{Pdt} and 1 $\text{kg}_{\text{glucose}}$ as functional units respectively.

Using equivalents of the fossil fuels conserved by producing a bio-based product as a functional unit, can couple productivity and comparability. The cumulative energy demand using 1 $\text{kg}_{\text{oil-eq}}$ substituted as the functional unit is shown in Fig. 4.10. The bio-based processes have a higher impact than the conventional counterparts. Among the bio-based processes, PET biorefinery not only has a lower productivity but also has a significantly higher GHG emissions arising from excessive consumption of the gate-to-gate energy utilities. No appreciable differences were seen between the PU biorefinery and the integrated biorefinery. The small differences that arise are due to the displaced process of the dependent co-product.



Fig. 4.8. The cumulative energy demand (CED) of the biorefineries with 1 kg_{Pdt} as the functional unit. Similar to the graph showing the GWP-100a, the first two columns compare the conventional PET production with that of the PET biorefinery in terms of the energy demand required for the entire process. The third and the fourth column deal with the PU production and the comparison is made between the commercial and the biorefinery. The last column shows the integrated biorefinery. 1 kg_{Pdt} refers to PU in the integrated biorefinery as it is the determining co-product.

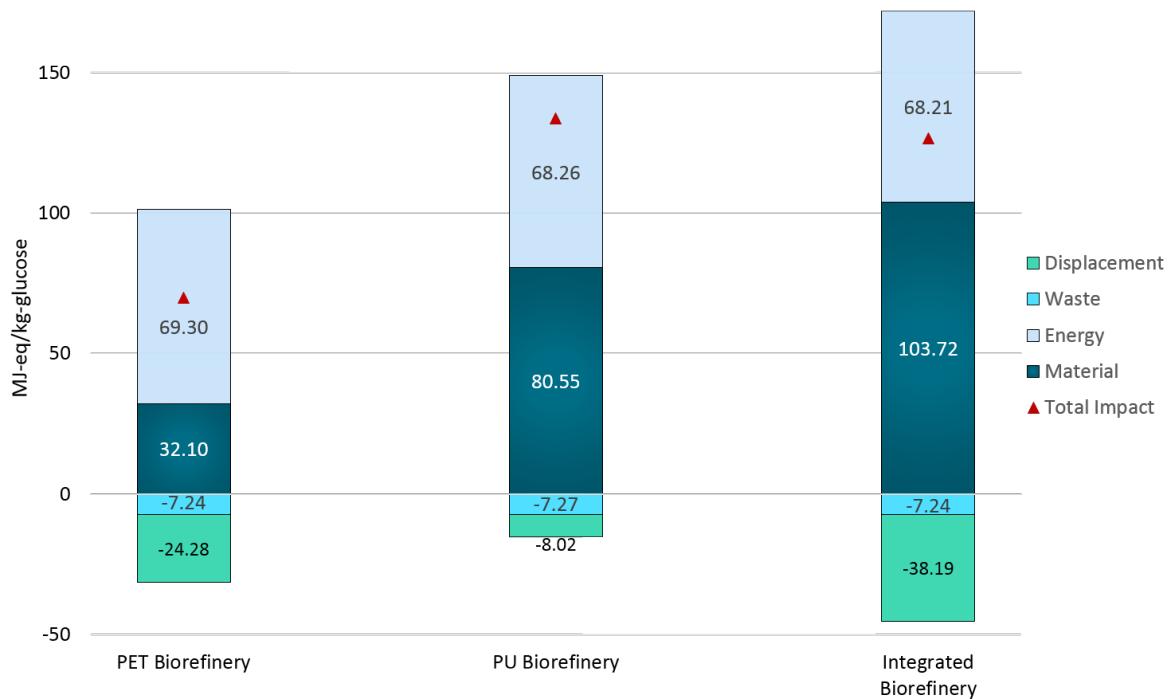


Fig. 4.9. The Cumulative energy demand (CED) of the biorefineries with 1 kg_{glucose} as the functional unit.

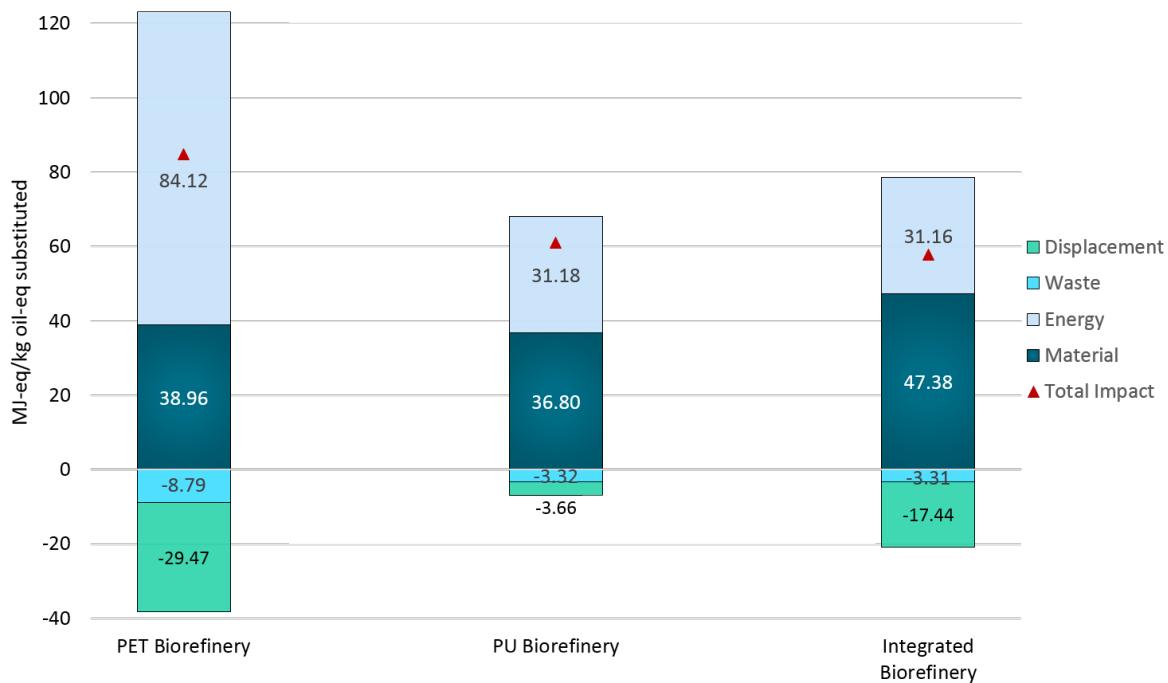


Fig. 4.10. The Cumulative energy demand (CED) of the biorefineries with 1 kg_{oil-eq} substituted as the functional unit.

4. Results and discussion

The displaced process in PET biorefinery^[64] (propylene glycol is the dependent co-product) and PU biorefinery^[64] (ethylene glycol is the dependent co-product) do not have similar LCA metrics. Propylene glycol is marginally more environmentally impacting when compared to ethylene glycol. Although, this difference in LCA metrics does not warrant the large difference seen in the credit given to the determining co-product in the PET and PU biorefinery. The low productivity of PET also plays a role. The waste treatments turned out to be reasonable small in all cases.

Strikingly noticeable are the material impact in all the processes. Fig. 4.11 shows the relative impact of the processes till the production of ethylene glycol and propylene glycol with respect to the entire process. The production of the polyols is common to all the biorefineries and paying closer attention reveals added information. In each of the biorefinery cases, more than 95% of the gate-to-gate energy utility consumption was due to the polyol production. Further research must be directed on producing bio-based polyols competitively. Furthermore, the majority of the raw material comes from contributions after polyol production, namely, the usage of TPA and MDI monomeric feeds. Yet again, future research and implementation of the bio-based production of TPA and MDI will enhance the integrated biorefinery.

Similar results were obtained using Eco-indicator 99 (EI99) and the results are shown in appendix D.1. Moving towards a multiple product portfolio has its advantages by efficiently utilizing raw material resources. Although not clearly evident in this case study, the difference between the integrated biorefinery and PU biorefinery may arise from the difference in the credit given for the displaced process. The importance of integrating the production processes



Fig. 4.11. The relative impact of polyol production on the biorefineries presented in terms of Cumulative energy demand (CED) with 1 kg_{Pdt} as the functional unit.

of the two polymers however, should not be understated. An effective integration would require the production of the non-renewable TPA and MDI monomeric feeds from bio-based resources. Improvements to the bio-based polyol production and future heat and mass integration would make the process competitive.

4.1.3. Economic assessment

An economic assessment was also performed to further access the different utilization of D-glucose to PET and PU respectively. A general assumption was made that the operating cost of conventional commercial processes for the production of PET and PU in the industry is approximately 25% lower than the market price of the products. The profit generated (US dollars/kg_{Pdt}) for each of the cases is shown in Fig. 4.12. The economic assessment was only based on operating costs. Capital cost estimations were not considered in this study. It is evident that the integrated biorefinery is the only biorefinery concept that generates a positive value.

In spite of the better performance of the integrated biorefinery, the profitability values are not close to the conventional processes. However, integrating several biorefinery concepts, for example, the utilization of glycerol, which is one of the by-products produced in the biorefinery, will add an another product that can increase profit and reduce environmental burden. A change in

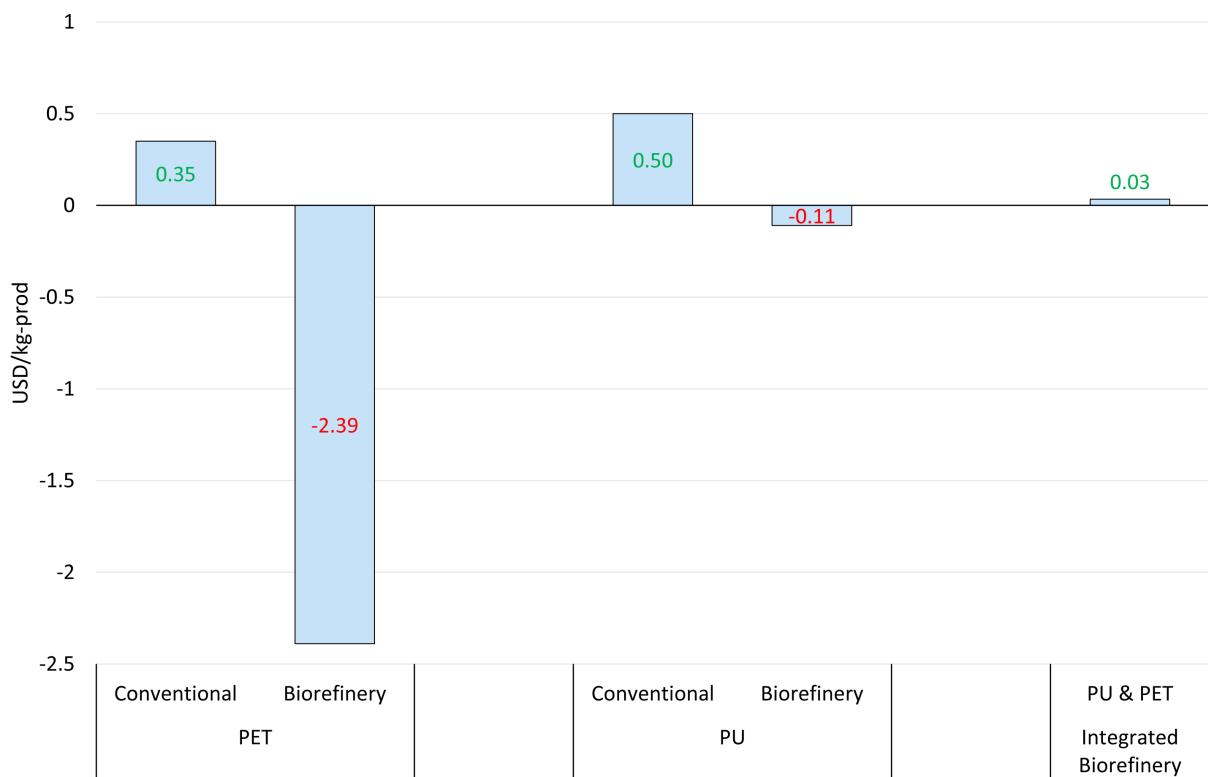


Fig. 4.12. The economic analysis of the biorefineries. The first column shows the profit generated by a conventional PET production plant and is compared to the PET biorefinery. The third and fourth columns show similar results with respect to PU biorefinery. The last column shows the integrated biorefinery profit which is the only scenario providing a positive profit.

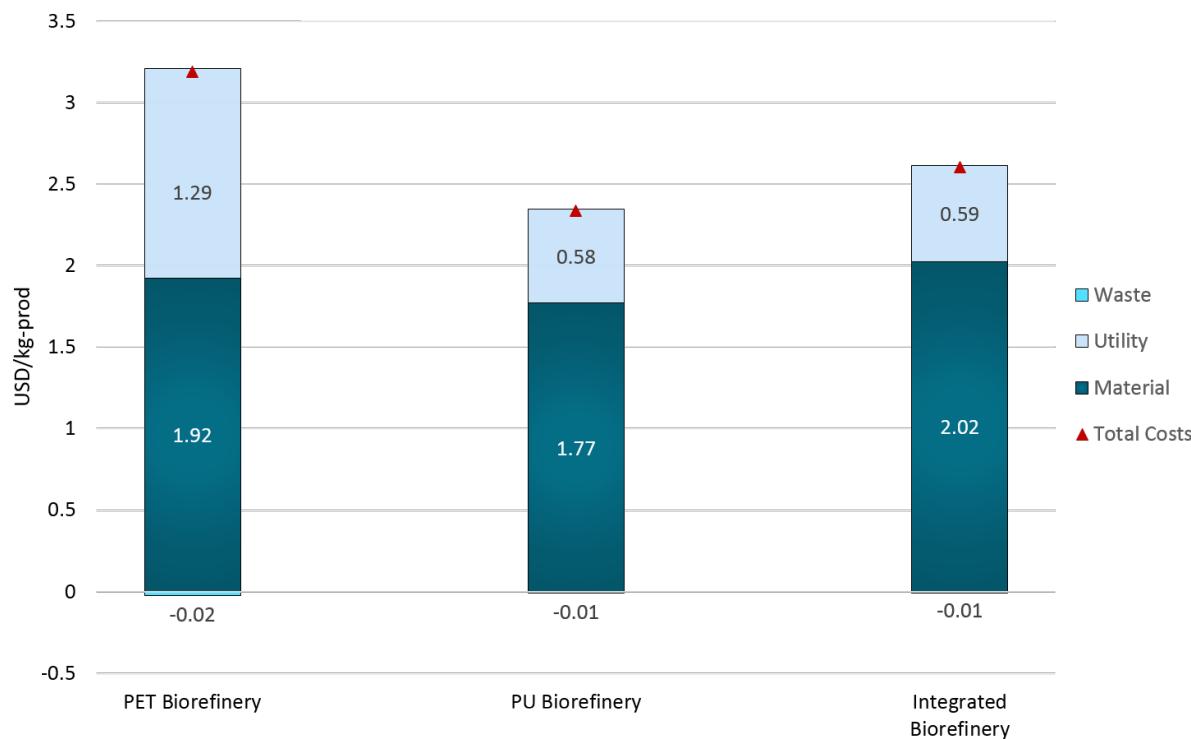


Fig. 4.13. The breakdown of the operational costs of the biorefineries.

the functional unit did not bring about any relative differences between the cases and integrated biorefinery was again the only bio-based production system that was marginally profitable.

Fig. 4.13 shows the breakdown of the operational costs of the biorefineries. Raw material costs have a fairly large contribution towards the operational costs of all the biorefinery cases. A significant portion of the raw material costs stems from the large volume of expensive TPA and MDI monomeric feeds needed for the polymerization reactions. Material costs for the PU and the integrated biorefinery accounts for more than 75% of the total operational costs, as MDI is needed in larger volumes and is significantly expensive than TPA. Production of these monomeric raw materials from bio-based resources will remarkably improve the economic performance of the biorefineries. The incorporation and addition of other technologies such as heat and mass integration can make it even more competitive. Another scenario is the increase in the cost of crude oil in the near future which will make the biorefinery increasingly competitive.

4.2. Reducing complexity

The importance of an integrated biorefinery is presented and achieving it will be one of the primary goals of any biorefinery concept. To build a computer-aided platform for the biorefinery concept that will be able to incorporate various concepts and technologies and seamlessly interact with one and other, reduction in the complexities of the process modeling becomes a prerequisite. Rigorousness and fast responsiveness are often opposing characteristics of simulation models and a balance between them is not always straightforward to achieve. Surrogate models are one interesting option to achieve the balance. Development and utilization of surro-

gate models according to the methods presented in chapter 3 will be demonstrated with a case study example of the integrated biorefinery producing both PET and PU.

4.2.1. Model specification

In the case study example, a surrogate model will be constructed for the simulated models of the integrated biorefinery producing PET and PU. An objective of the surrogate first needs to be specified. Generally the most common objective functions that arises in a production process is the revenue generated due to the sales. The input variables also needs to be specified before a surrogate model is built. Input variables may arise from a number of several degrees of freedom of the process. However, most of these are fixed in industrial processes according to predefined, and perhaps optimized set points. Some variables that can be freely changed in a biorefinery such as the one producing PET and PU, are those that refer to the size of the production system, such as the flow of the raw material, D-glucose.

Process parameters such as temperatures are also sometimes free (or not optimized) within some boundaries. The evaporator temperature and that of the crystallizer are taken as two design variables in the construction of the model. For visualization reasons in order to explain the surrogate modeling, only three variables, were chosen in the case study example. The number of output variables similarly is not constrained. The output parameters shown in this model are the revenue generated and the purity of the polymer products. Although the output is defined as a single scalar value, it is composed of various components which were the initial values of the output prediction from the surrogate model. For example, the revenue is dependent on sales, energy consumption, waste generated, treatment and so on. These values were the initial output parameters which after being predicted by the model can be lumped together and calculated to show the parameter of interest.

4.2.2. Training of the surrogate

Once the objective and the input variables have been set, the surrogate can be trained to understand the behavior of the simulation model for the prediction of the output objective as a function of the input variables. Surrogate training depends on the range of the input variables that the user wishes to train the model. In the specific case study for the production of PET and PU (see chapter 2 for simulations and ranges), the parameters and their ranges are specified in Tab. 4.4

The training of the surrogate model can be done in two ways. Either the user can provide his own training data, or the user can rely on the tool that was created to generate the training data with a working simulation model. The user has control over the number of data points in the variable dimensional space one would like to train the model for. The training data of the case study model (*i.e.*, 40 generated points in the input variable domain according to the specific ranges) is shown in Fig. 4.14. The output parameter illustrated according to a color scale. It can be observed that the location of the points generated with the help of the tool is stratified, meaning dispersed in the space defined by the range of the variables. This leads to the model being able to understand the behavior of the simulation. Once the model is trained with the observed data, it is equipped to predict the output parameter, in this case, the revenue generated for values of input variables.

Tab. 4.4. Parameter ranges of the Surrogate model

	Variables	Range	
		Low	High
Input	D-glucose inlet mass flow	800 kg h ⁻¹	1200 kg h ⁻¹
	Evaporator Temperature	175 °C	190 °C
	Crystallizer Temperature	5 °C	15 °C
Output	Revenue	2190 USD h ⁻¹	3310 USD h ⁻¹
	PU Purity (Mass %)	99.90	99.93
	PET Purity (Mass %)	99.84	99.88

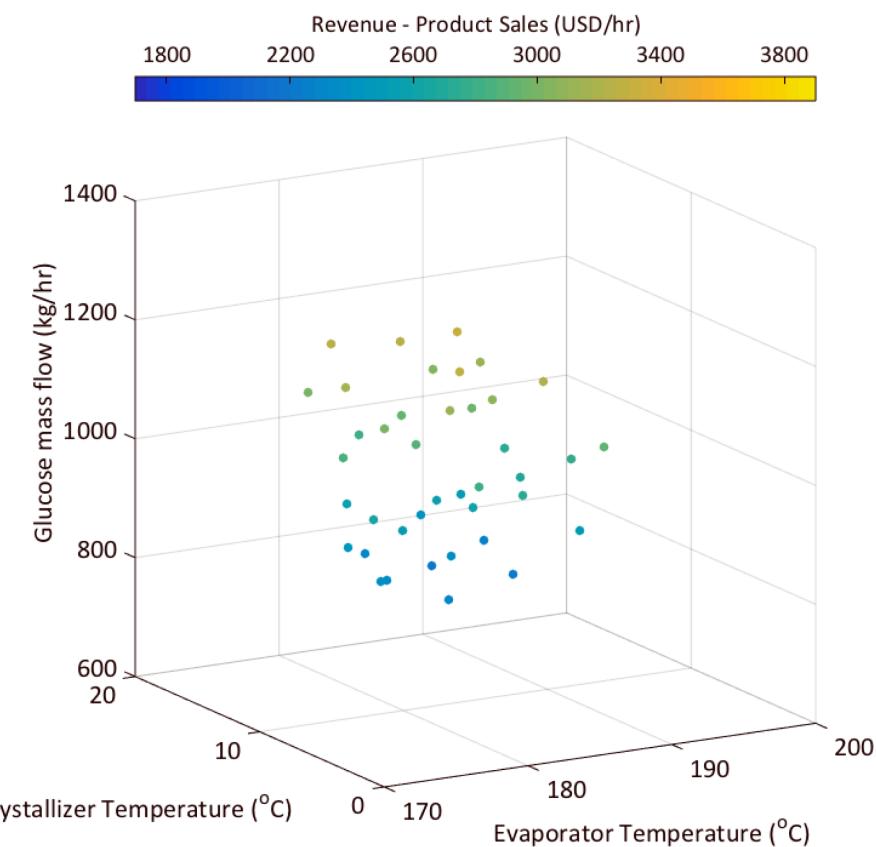


Fig. 4.14. The training space of the surrogate model. The x-axis shows the Evaporator temperature, y-axis the crystallizer temperature and the z-axis shows the D-glucose input flow. The location of the point in the variable dimensional space is determined by the input variable values. The color of the point determines its output parameter.

4.2.3. Surrogate Prediction

The trained surrogate model can now be used to predict values of the input variables with very high accuracy in the training range (*i.e.*, interpolation problem) and with satisfactory accuracy outside this range (*i.e.*, extrapolation problem) to some extent. Fig. 4.15 shows the points at which the surrogate model was asked to predict (*i.e.*, points with no color). The points are both in the interpolation and extrapolation region.

Fig. 4.16 shows the ability of the surrogate model to be able to predict the output variable. The hollow points are now filled with the color that corresponds to the value of the revenue. The prediction is based on the correlation parameters that were generated by the surrogate during the training phase. The prediction is made for 80 points which are within and beyond the range that was used to train the model.

4.2.4. Model Parameters

The graphical user interface, apart from providing a platform to visualize the results of the prediction made by the surrogate model, also is a tool to relatively understand the importance of the input variables towards the output parameter being measured. In the case study, the relative importance of the input variables, D-glucose mass flow, the crystallizer temperature and the evaporator temperature for the prediction of the process revenue is related by studying the model parameters generated. The model parameters, θ and λ , are tuned using a maximum likelihood estimate approach as discussed in chapter 3. θ was tuned within a range of 10^{-3} to 10^2 , while, λ was searched between 10^{-6} and 1 respectively.

Tab. 4.5. Model Parameters of the Surrogate

Input variables	Model Parameters	
	θ	λ
D-glucose inlet mass flow	0.28	10^{-6}
Evaporator Temperature	0.05	10^{-6}
Crystallizer Temperature	0.04	10^{-6}

Tab. 4.5 shows the model parameter results towards the output revenue. Interpreting such information can produce significant inferences. θ is measured in exponent of 10. The value of the parameter signifies the relative importance of the variable. As one would expect the revenue depends mainly on the amount of input D-glucose. The values of the θ portrays exactly the same inference. The large difference between D-glucose and the remaining two variables signify a larger impact of D-glucose. The closeness in the values of the crystallizer and evaporator temperature can be understood as similar dependence of the two on the output.

The regression was also used to test its performance. Regression is desired only when sample data points are suspected to contain outliers. Therefore, in the presence of consistent data with

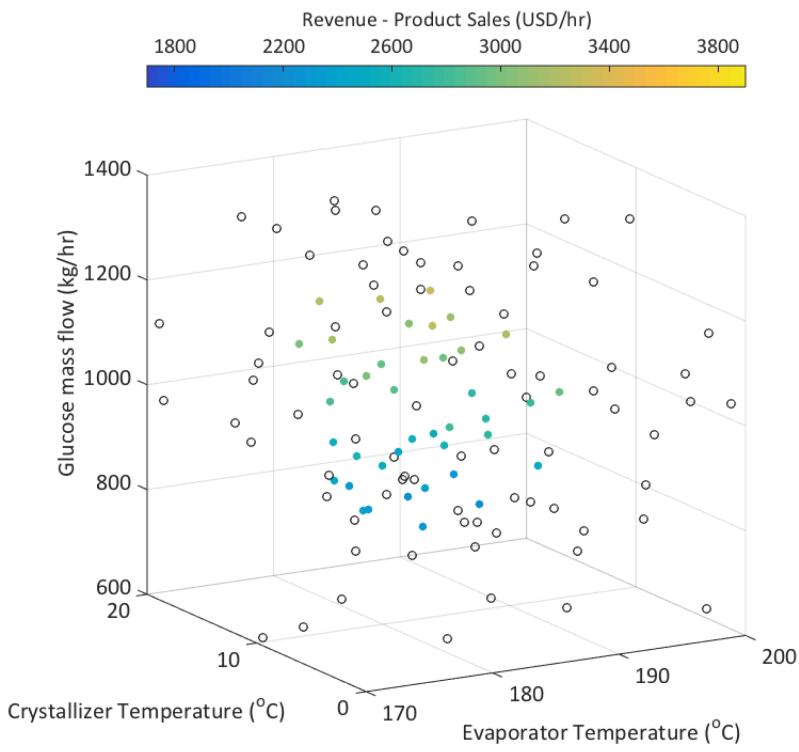


Fig. 4.15. The testing of the surrogate model. The hollow points depict the combination of input values of variables for which the model is asked to predict the output parameter

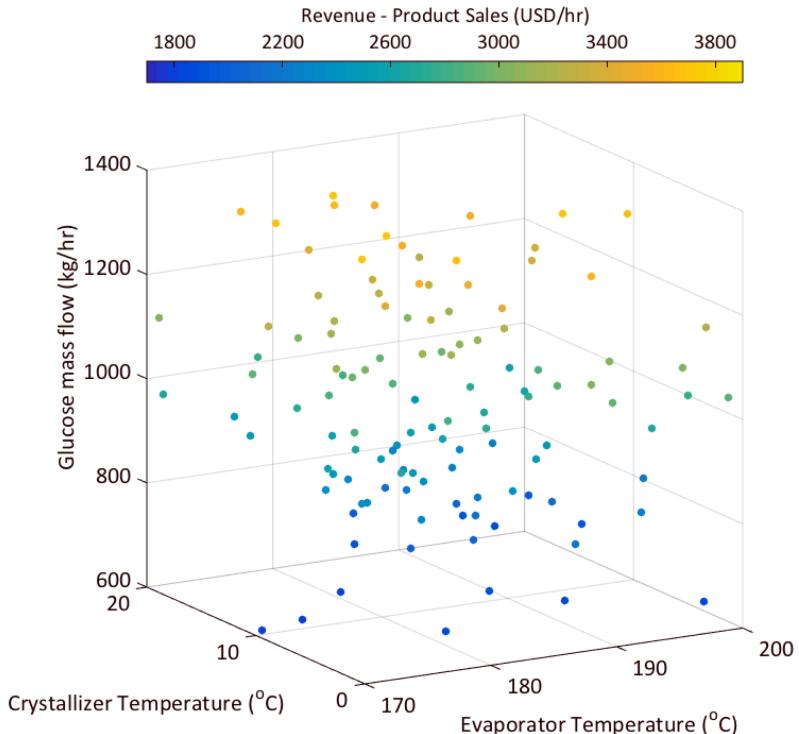


Fig. 4.16. The surrogate model prediction. The hollow points shown in Fig. 4.15 are now colored according to the prediction of the output parameter value using the surrogate model.

no outliers, the regression parameter, λ , should ideally take the least value in its bounds of search. Consistent with the expected value, the regression parameter choose the lowest attainable value since the sample training data did not have any inconsistent outlying data point.

4.2.5. Accuracy of the Surrogate Prediction

Fig. 4.17 shows the accuracy of the prediction results for revenue as the output parameter. The red circles denote the training data that is given to the surrogate for it to understand the behavior of the simulation model. The blue circles denote the prediction results which is plotted with the real simulated revenue on the x-axis and prediction results on the y-axis. The black 45° line denotes the ideal scenario when the surrogate prediction is the same as those obtained by the simulation models.

The accuracy of the surrogate model is discernible. The model is able to predict with extremely high accuracy both in the interpolation as well as extrapolated regions. The blue points correspond to hollow circle in Fig. 4.15 which were predicted by the model. The relationship of the simulated revenue to the individual input variables are shown in appendix E.1. The high precision of predictions can be argued to arise from the linearity of the most influential input variable (*i.e.*, D-glucose mass flow) to the output parameter (see appendix Fig. E.1).

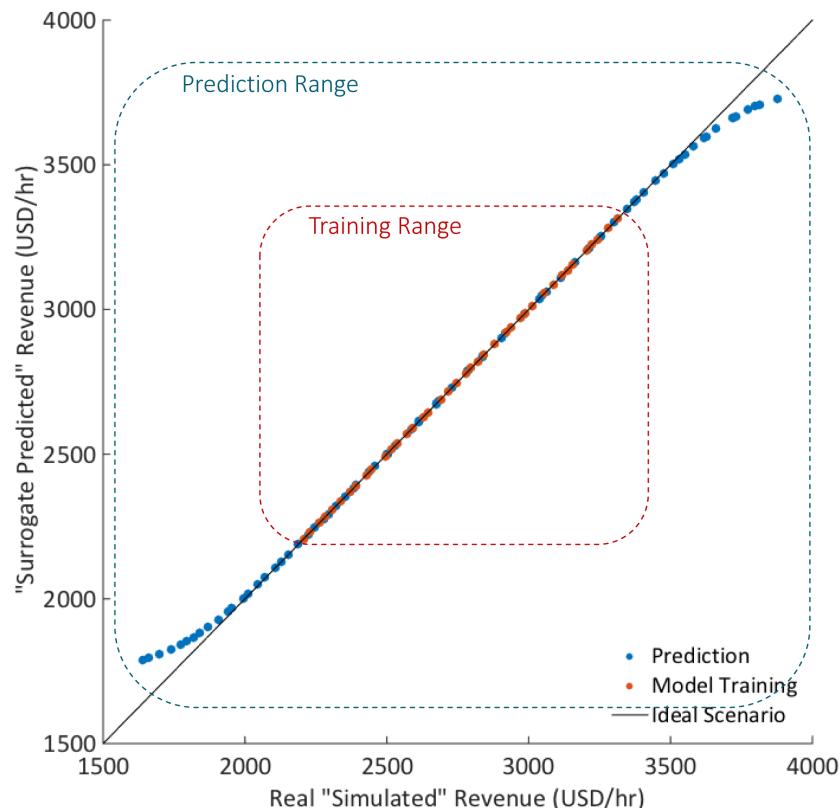


Fig. 4.17. The accuracy of the surrogate model. The red circles denote the training data while the blue circles denote the predictions which is compared with the real data. The black 45° line depicts ideal case scenario.

However, the surrogate model is also accurate for prediction involving non-linear parameters, such as the purity of PU. The non-linear dependence of the purity of PU with the individual input variables is shown in appendix E.2. Fig. 4.18 shows the accuracy of the prediction results compared to the simulated value. Except for a few data points at the far end of the prediction range, the surrogate proves that it will be able to handle complex simulation designs and mimic its performance by making computation quicker, simpler and most importantly without compromising on the accuracy of the prediction, proving an extremely valuable tool in realizing an optimum biorefinery concept.

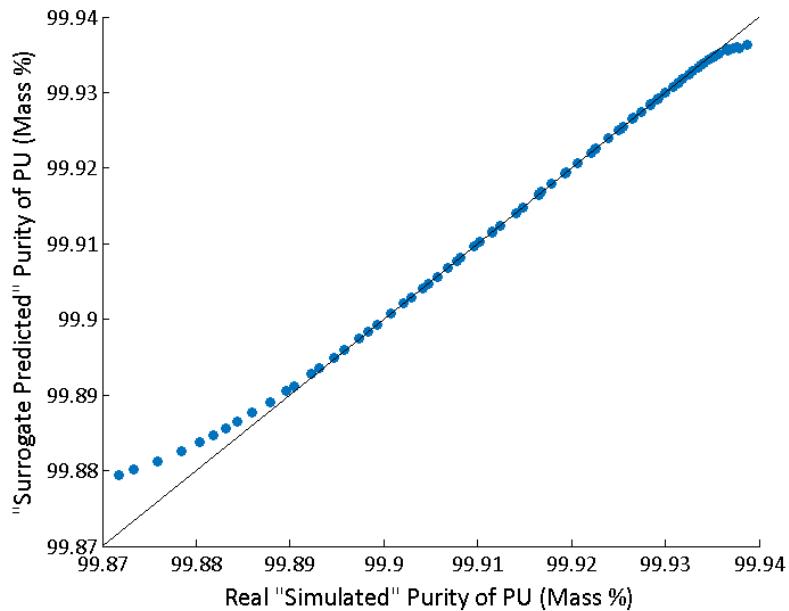


Fig. 4.18. The accuracy of the surrogate model for predicting purity of PU. The black 45° line depicts ideal case scenario.

5. Conclusion and Outlook

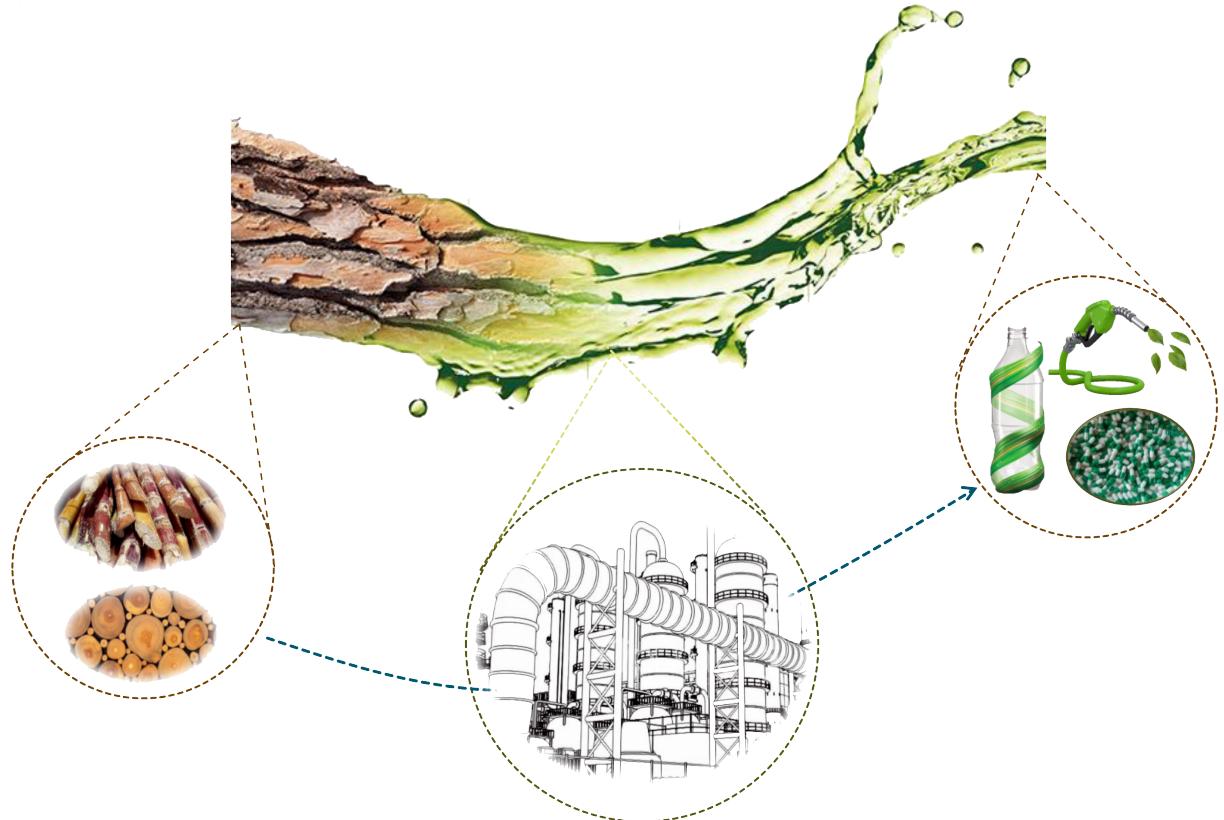


Fig. 5.1. The wood to chemicals concept of the Waldbox project.

The vision to face the challenging task of the society to transform from the abundant utilization of fossil fuels to a sustainable industry needs to be realized rather quickly. The usage of biomass as a renewable resource for the production of commercially important chemicals and valuable end-products and services are being explored in an increasing scale since the steady decline in fossil fuel reserves across the world. Biomass, being the only renewable carbon source, alleviates the problems commonly associated with the use of non-renewable resources, such as the even distribution across the globe. This particular master thesis concentrated on the utilization of wood based lignocellulosic biomass towards the production of two valuable polymers, Poly(ethylene terephthalate) and Polyurethane.

The alternative process for the bio-based production of PET and PU are part of the Waldbox project, the focus of which is the bigger picture of realizing a computer-aided platform for the search of an optimal biorefinery concept. Finding an optimal biorefinery requires the need for several concepts to be built and integrated. The production of PET and PU were drop-in strategies of the several others built in the project. The simulated alternative production of PET

5. Conclusion and Outlook

and PU utilized bio-based D-glucose obtained from lignocellulosic wood residues as input raw material. The novelty of the process arises from the alternative routes taken for the production of the polyol monomers required for PET and PU. The importance of an integrated biorefinery although not exactly evident with the example of PET and PU, it cannot be understated for maximizing raw material utilization and minimizing energy consumption. The processes were assessed both in terms of environment and economy.

Environmental markers such as Global warming potential (GWP-100a), Cumulative energy demand (CED) and Eco-Indicator 99 (EI99) were used to evaluate the novel simulated production process. The biorefineries were not as environmentally friendly as their commercial counterparts were. A multiple product portfolio producing both PU and PET in a joint production (referred in the text as ‘integrated biorefinery’) did not show appreciable difference in terms of environmental impact compared to stand-alone bio-based PU production process. The small difference that does exist between the two biorefineries in terms of environmental burden is a result of the difference in credit given for the disparate displaced processes. The importance of the integrated biorefinery, however, is shown in maximizing the utilization of the available raw materials for better bio-based productivity without compromising on economic benefit nor environmental burden. The integrated biorefinery showed a marginal profitability. Future research focus on heat and mass integration concepts along with incorporating several drop-in strategies from the Waldbox project can lead to a truly integrated biorefinery concept challenging commercial practices.

Realizing an integrated biorefinery is rarely straightforward. It requires the careful and effective interactions between various concepts and tools to start with. Process designs are a prerequisite in moving forward. In spite of good process designs, an optimal concept is evasive due to enormous complexities that arise when several component designs try to interact with each other and trying to integrate heat and mass. An innovative barrier that is reached with the use of process designs alone can be breached with the use of surrogate models.

The second section of the master thesis was the creation of an innovative surrogate model tool that was incorporated for the biorefinery strategies. The surrogate model is a mimic of the process designs but yet powerful enough to handle the rigorousness. A space-filling latin hypercube sampling is essential in avoiding computational explosion. Kriging method was utilized for its ease of control, flexibility and implementation and for its accuracy and reliability. They are computationally effective in understanding the behavior of complex models and are able to break it down to be integrated in a computer aided platform. The minimalistic approach is important in realizing an optimal biorefinery concept irrespective of the complexity of the design range. The performance of the tool for a case study example dealing with the production of PET and PU in an integrated biorefinery was shown along with its advantage that it inherently brings with it due to its reliable and robust nature. The surrogate built was applied for prediction even outside the simulated range, thereby proving to be an important tool that can be used for translating complex process design into simplified and accurate models. This would be the first step in building a computer-aided platform for the integration of different bio-based production routes to find the optimal biorefinery concept.

Appendices

A. Reaction and Mechanisms

A.1. Sorbitol Production

The catalytic hydrogenation from D-glucose to sorbitol was performed using a ruthenium catalyst. The exact reaction mechanism is shown in the Fig. A.1. A similar approach was employed for modeling the process using a 5 wt.% Ru/C catalyst.

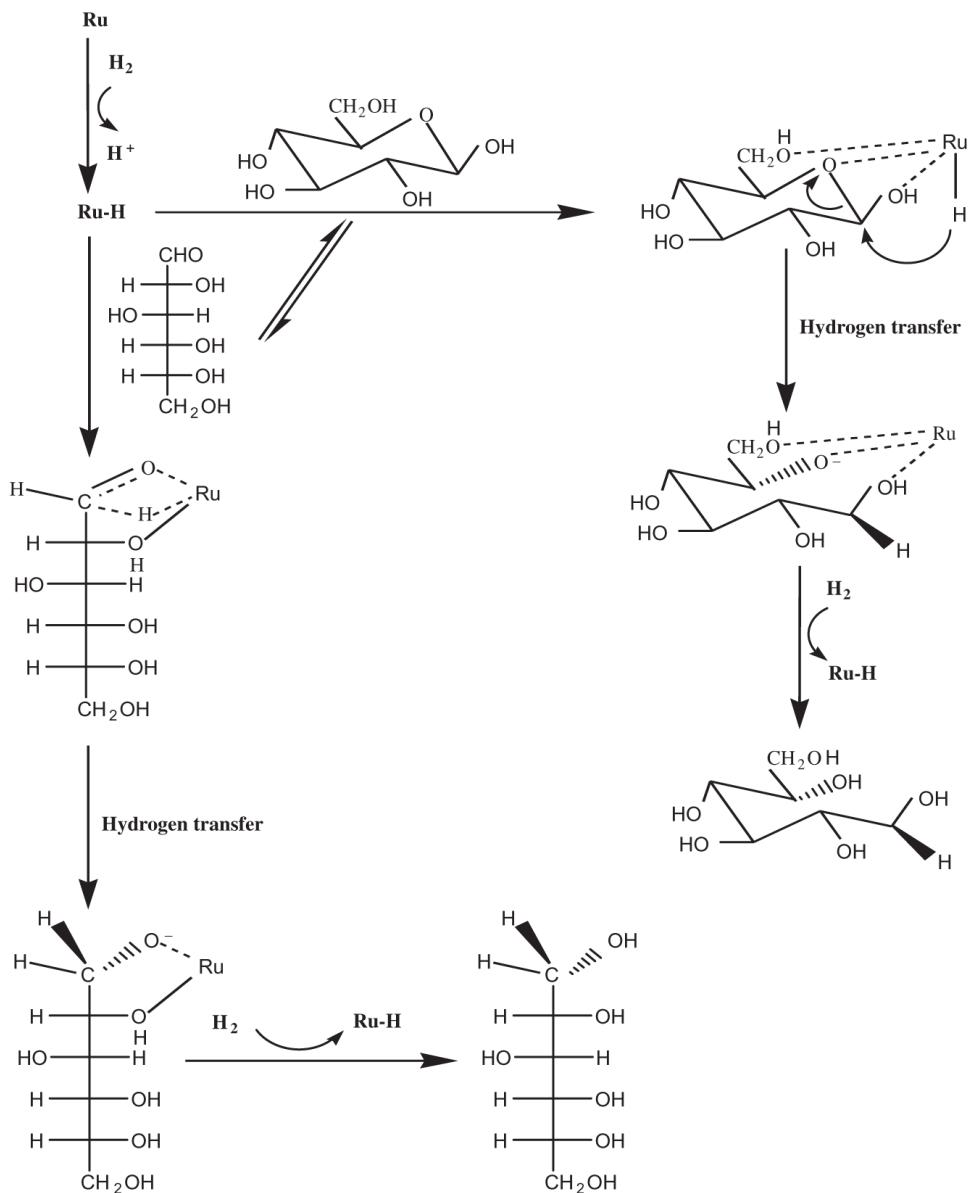


Fig. A.1. The reaction mechanism for the production of sorbitol with a ruthenium catalyst.

A.2. Glycol synthesis

Sorbitol is catalytically hydrogenolyzed to lower glycols. The reaction is simulated using Nickel as the metal catalyst. The general reaction mechanism for sugar and sugar alcohol hydrogenolysis is shown in Fig. A.2.

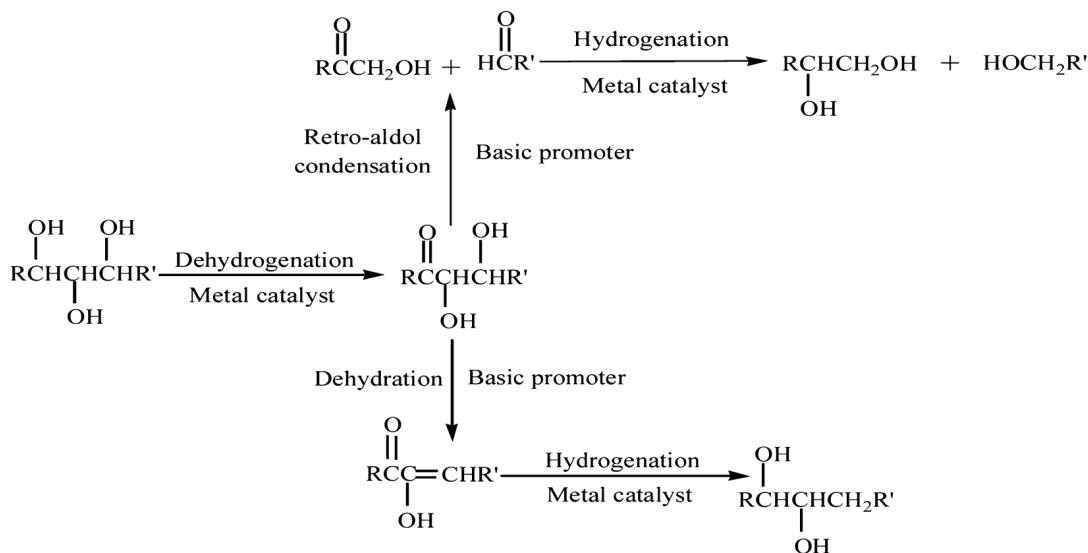
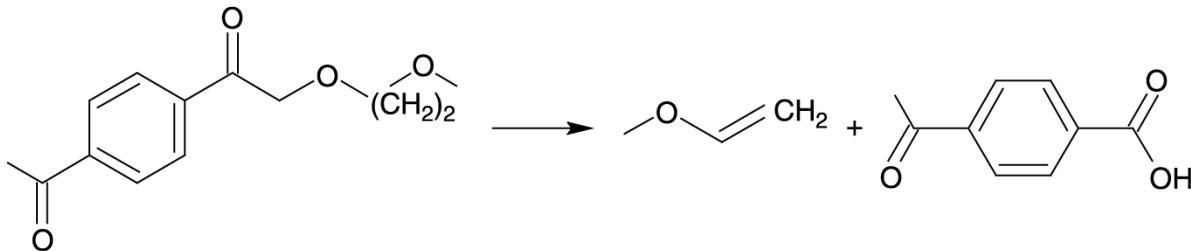


Fig. A.2. The reaction mechanism for the production of lower glycol by the catalytic hydrogenolysis of sorbitol with a metal (Nickel) catalyst.

A.3. PET Polymerization

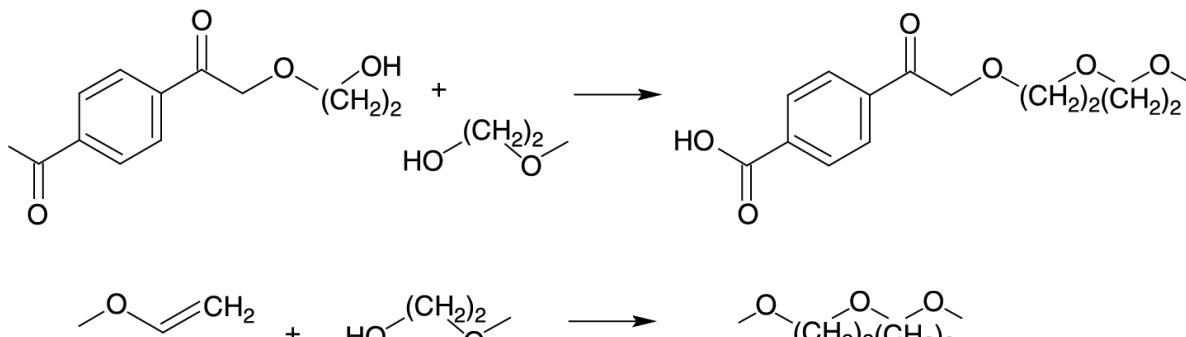
The polymerization of ethylene glycol and terephthalic acid proceeds through two primary reactions. Additionally, the side reactions also play a major role. The side reaction discussed in section 2.3.1 are shown in detail below.

- Degradation of Diester groups.

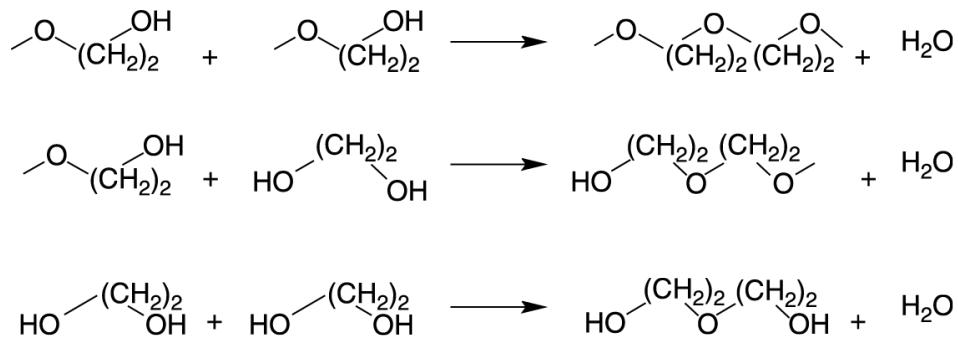


A. Reaction and Mechanisms

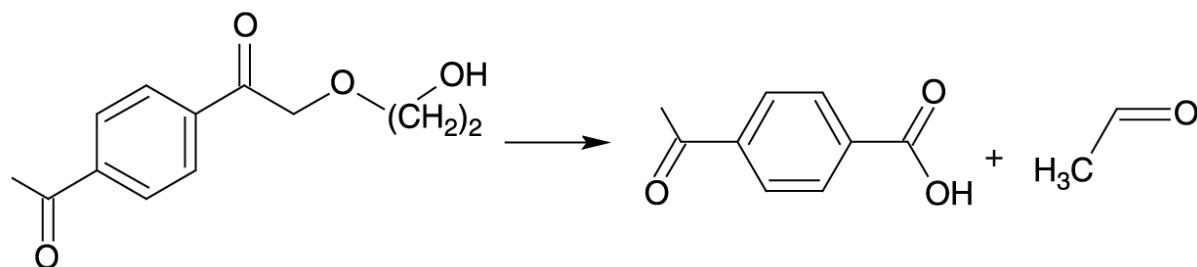
- Diethylene glycol formation.



- Dehydration of EG and T-EG.



- Acetaldehyde formation.



The complete list of reactions that we generated by the model and the user modeled side reactions are shown in Tab. A.1.

Tab. A.1. Complete Reaction Set - PET Polymerization

Water Formation (Model-generated)
EG = TPA ⇌ T-EG+T-TPA + W
EG + T-TPA ⇌ T-EG + B-TPA + W
DEG + TPA ⇌ T-DEG + T-TPA + W
DEG + T-TPA ⇌ T-DEG + B-TPA + W
T-EG + TPA ⇌ B-EG + T-TPA + W
T-EG + T-TPA ⇌ B-EG + B-TPA + W
T-DEG + TPA ⇌ B-DEG + T-TPA + W
T-DEG + T-TPA ⇌ B-DEG + B-TPA + W
T-VIN + T-TPA + W → AA + TPA
T-VIN + B-TPA + W → AA + T-TPA
Ester Interchange (Model-generated)
EG + B-EG ⇌ T-EG + T-EG
EG + T-DEG ⇌ DEG + T-EG
EG + B-DEG ⇌ T-DEG + T-EG
EG + B-EG ⇌ T-EG + T-EG
EG + T-DEG ⇌ DEG + T-EG
EG + B-DEG ⇌ T-DEG + T-EG
DEG + T-EG ⇌ EG + T-DEG
DEG + B-EG ⇌ T-EG + T-DEG
DEG + B-DEG ⇌ T-DEG + T-DEG
DEG + T-EG ⇌ EG + T-DEG
DEG + B-EG ⇌ T-EG + T-DEG
DEG + B-DEG ⇌ T-DEG + T-DEG
EG + T-VIN → AA + T-EG
EG + T-VIN → AA + T-EG
DEG + T-VIN → AA + T-DEG
DEG + T-VIN → AA + T-DEG
T-EG + T-VIN → AA + B-EG
T-EG + T-VIN → AA + B-EG
T-DEG + T-VIN → AA + B-DEG
T-DEG + T-VIN → AA + B-DEG
Degradation of Diester Group (User Reactions)
B-TPA + B-EG → T-VIN + T-TPA
Diethylene Glycol Formation (User Reactions)
B-TPA + T-EG + T-EG → T-TPA + T-DEG
T-VIN + T-EG → B-DEG
Ethylene Glycol Dehydration (User Reactions)
2T-EG → B-DEG + W
T-EG + EG → T-DEG + W
2EG → DEG + W
Acetaldehyde Formation (User Reactions)
B-TPA + T-EG → AA + T-TPA

A.4. PU Polymerization

The complete set of reaction considered along with their rate constants are shown in the Tab. A.2. The Arrhenius constants **A₀** and **E_a** are measured in L mol⁻¹ s⁻¹ and kcal mol⁻¹ respectively.

Tab. A.2. Reaction Set and rate constants - PU Polymerization

Reaction	Stoichiometry	A ₀ , E _a
Urethane formation	PG + MDI → T-PG + B-URET + T-MDI	
	PG + T-MDI → T-PG + B-URET + B-MDI	2500, 10
	T-PG + MDI → B-PG + B-URET + T-MDI	
	T-PG + T-MDI → B-PG + B-URET + B-MDI	
Amine formation	T-MDI + W → T-MDA + CO ₂	1000, 10
Urea formation	T-MDA + MDI → B-MDI + B-UREA + T-MDI	5000, 10
	T-MDA + T-MDI → B-MDI + B-UREA + B-MDI	
Allophane formation	MDI + B-URET → T-MDI + B-ALLO	10, 10
	T-MDI + B-URET → B-MDI + B-ALLO	
Biuret formation	MDI + B-UREA → T-MDI + B-BIU	100, 10
	T-MDI + B-UREA → B-MDI + B-BIU	

B. Simulations

B.1. Design Specification

The process flowsheets simulated in Aspen Plus® were controlled using several design specifications. These make the process flowsheets robust and automated to work in wide range of input variables. The various design specification used in the flowsheets are detailed.

B.1.1. Sorbitol Production

The design specification detailed here are based on the academic research by Crezee et al.^[16] and Hoffer et al.^[15] The United States Patent US 6,297,409 B1 by Roquette Freres,^[44] was also used to support the data used in the simulation process.

Input glucose concentration

The input D-glucose concentration was maintained at a final concentration flowing into the reactor at 1 mol L^{-1} . The design specification was used to achieve this.

$$C_G = \frac{\dot{n}_{glu}}{m_{tot}\rho_{sol}} \quad (\text{B.1})$$

C_G – D-glucose concentration (mol m^{-3})

\dot{n}_{glu} – Mole flow of input D-glucose(mol h^{-1})

m_{tot} – Total mass flow (kg h^{-1})

ρ_{sol} – Mass density of the solution (g m^{-3})

C_G is measures in mol m^{-3} and is set to 0.001 such that the concentration of D-glucose in the reactor is 1 mol L^{-1} .

Hydrogen gas inlet flow

The inlet molar flow of hydrogen was set at an equimolar ratio to that of D-glucose respecting the reaction stoichiometry.

$$\dot{n}_{glu} = \dot{n}_{H_2} \quad (\text{B.2})$$

\dot{n} signifies the molar flow of the components in mol h⁻¹. A calculator block was used to achieve this equality.

Pre-Exponential Reaction factor

The Pre-exponential reaction factor used in the reactor for the catalytic conversion of D-glucose to sorbitol was given a range that it can iterate and find the best suited value depending on a set conversion that was to be achieved. A design specification was used to give the pre-exponential factor for the reaction rate a small regressive range for operation to meet a 99% conversion.

$$\frac{C_{In}^G - C_{Out}^G}{C_{In}^G} = 0.99 \quad (\text{B.3})$$

Acetone input for washing crystals

The liquid to solid mass ratio in the crystal washing unit was set to 2. The acetone flow into the crystal washing unit was controlled to meet the requirement. The recycled acetone is also considered and taken into account when computing the amount of fresh acetone needed for the washing process.

B.1.2. Glycol synthesis

Similar design specifications and control were employed in the synthesis of glycols from sorbitol. The catalytic hydrogenolysis of sorbitol is based on the academic research of Ye et al.^[17] and Banu et al.^[18] and the purification of glycols was in coherence with the United States Patent 4,966,658.^[51].

Hydrogen gas flow

The hydrogen gas inflow was calculated and controlled to respect the stoichiometry of the reaction. The reaction is shown in Eqn. 2.3. Similar reaction based controls were employed to control various sections of the simulation.

B.2. Purification of Glycols

Tab. B.1 ranks the components in the ascending order of their boiling points.

Tab. B.1. Ascending ranks of boiling points of glycols

Component	Boiling Point (in °C)
Propylene glycol	188.2
Butanediol	195
Ethylene glycol	197.3
Glycerol	290

C. Surrogate Modeling

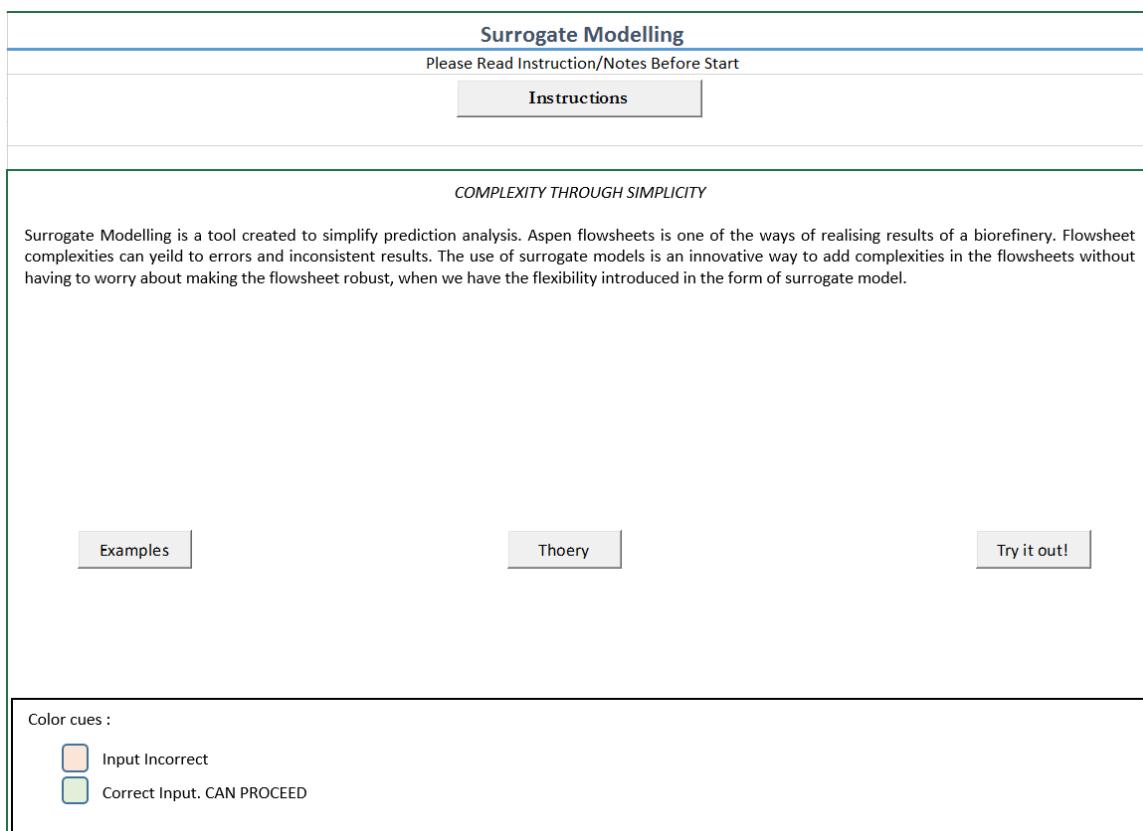


Fig. C.1. The graphical user interface of the tool developed in Microsoft Excel® 2013.

Fig. C.1 shows the graphical user interface built in Microsoft Excel® 2013. The interface allows for step wise understanding of the problem rendering it easier to work with. The tool also allows for customization, the controls for which are also incorporated in the interface.

C.1. Surrogate Modeling example

Fig. C.2 shows the graphical user interface for the input variable data built in Microsoft Excel® 2013. Once the variables and their ranges have been entered (see Fig. C.3), the interface guides the user to obtain the space-filled latin hypercube sampling using MATLAB®. Fig. C.4 and Fig. C.5 show the graphical user interface before and after obtaining the space-filled training data. Aspen Plus® is used to obtain the output parameter values for the space-filled training data using a sensitivity analysis. Finally, the values for input variables for which the model is required to predict the output is entered (see Fig. C.6). The prediction is realized with the help of the interface between Microsoft Excel® and MATLAB®.

C. Surrogate Modeling

Fig. C.2. The graphical user interface of the tool developed in Microsoft Excel® 2013 for entering the input variables and their ranges.

Fig. C.3. Completed form containing input variables and their ranges.

Surrogate Modelling															
Step 2: Training Data															
		User category	2												
Enter the number of training data points you want the model to predict.															
Save the excel as Main_File.xlsx and run space_filled_input.m															
Note: Your Training Data affects your prediction output. A good training data will and should have all points in the range for better interpolation.															
			Number of training data points (Default 100)			25									
<< Back					Clear		Continue >>								
# Data points	Variables														
	Glu_Mflow	E_Temp	C_Temp												
1															
2															
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
13															
14															
15															
16															
17															
18															
19															
20															
21															
22															
23															
24															
25															

Fig. C.4. The Excel® data sheet before obtaining the space-filled training data.

Surrogate Modelling									
Step 2: Training Data									
	User category	2							
Enter the number of training data points you want the model to predict.									
<i>Save the excel as Main_File.xlsx and run space_filled_input.m</i>									
Note: Your Training Data affects your prediction output. A good training data will and should have all points in the range for better interpolation.									
	Number of training data points (Default 100)	25							
<< Back		Clear							Continue >>
# Data points	Variables								
	Glu_Mflow	E_Temp	C_Temp						
1	916.6667	185.8333	10						
2	1166.667	188.3333	10.41667						
3	983.3333	183.3333	7.5						
4	816.6667	172.5	5						
5	1000	173.3333	7.916667						
6	866.6667	170.8333	8.333333						
7	1016.667	177.5	5.833333						
8	1133.333	179.1667	9.166667						
9	1100	182.5	14.58333						
10	1116.667	186.6667	6.666667						
11	1050	178.3333	11.66667						
12	833.3333	190	12.5						
13	1200	180.8333	13.75						
14	900	180	8.75						
15	966.6667	185	12.91667						
16	850	175.8333	11.25						
17	950	171.6667	12.08333						
18	1066.667	184.1667	9.583333						
19	1183.333	175	10.83333						
20	933.3333	187.5	6.25						
21	1083.333	189.1667	13.33333						
22	800	176.6667	7.083333						
23	883.3333	181.6667	14.16667						
24	1033.333	170	15						
25	1150	174.1667	5.416667						

Fig. C.5. The Excel® data sheet after obtaining the space-filled training data using MATLAB®.

Fig. C.6. The values for input variables for which the model is required to predict the output

D. Economic and Environmental Metrics

The economic and environmental data used for the calculations in the Life cycle assessments (LCA) are shown in Tab. D.1.

Tab. D.1. Economic and Environmental Markers

Component	Formula	CED ^{a,b} MJ eq/kgPdt	GWP-100a ^a kgCO ₂ eq/kgPdt	EI99 ^a Points/kgPdt	Price ^c USD t ⁻¹
Acetone	C ₃ H ₆ O	67.12	2.23	0.24	1800
D-glucose	C ₆ H ₁₂ O ₆	6.48	0.51	0.05	525
EG	C ₂ H ₆ O ₂	51.21	1.57	0.18	1500
Hydrogen	H ₂	69.70	1.67	0.23	4000
MDI	C ₁₅ H ₁₀ N ₂ O ₂	89.79	4.03	0.32	1500
PET	(C ₁₀ H ₈ O ₄) _n	80.72	2.89	0.30	1400
PG	C ₃ H ₈ O ₂	99.86	4.06	0.33	1550
PU	(C ₃ H ₈ N ₂ O) _n	100.02	4.03	0.37	2000
TPA	C ₈ H ₆ O ₄	57.47	1.82	0.21	650

^aData obtained from the Eco-Invent^[66] database.

^bNon-renewable CED - The contribution from only non-renewable sources.

^cPrices obtained from Alibaba.^[69]

D.1. Eco-Indicator 99 Data

The Eco-indicator 99 shows results similar to those presented with cumulative energy demand and global warming potential. Fig. D.1, Fig. D.2 and Fig. D.3 show the results with the functional unit as 1 kg_{Pdt}, 1 kg_{glucose} and 1 kg_{oil-eq} respectively.

D. Economic and Environmental Metrics



Fig. D.1. Eco-indicator 99 (EI99) of the biorefineries with 1 kg_{Pdt} as the functional unit.

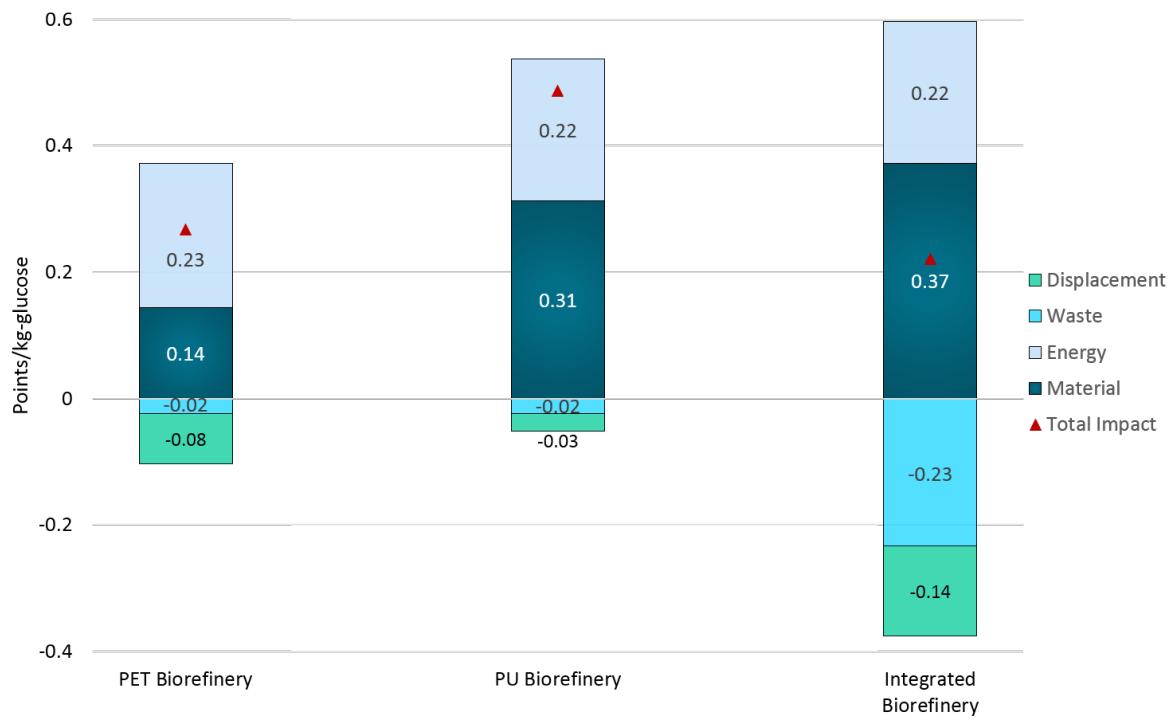


Fig. D.2. Eco-indicator 99 (EI99) of the biorefineries with 1 kg_{glucose} as the functional unit.

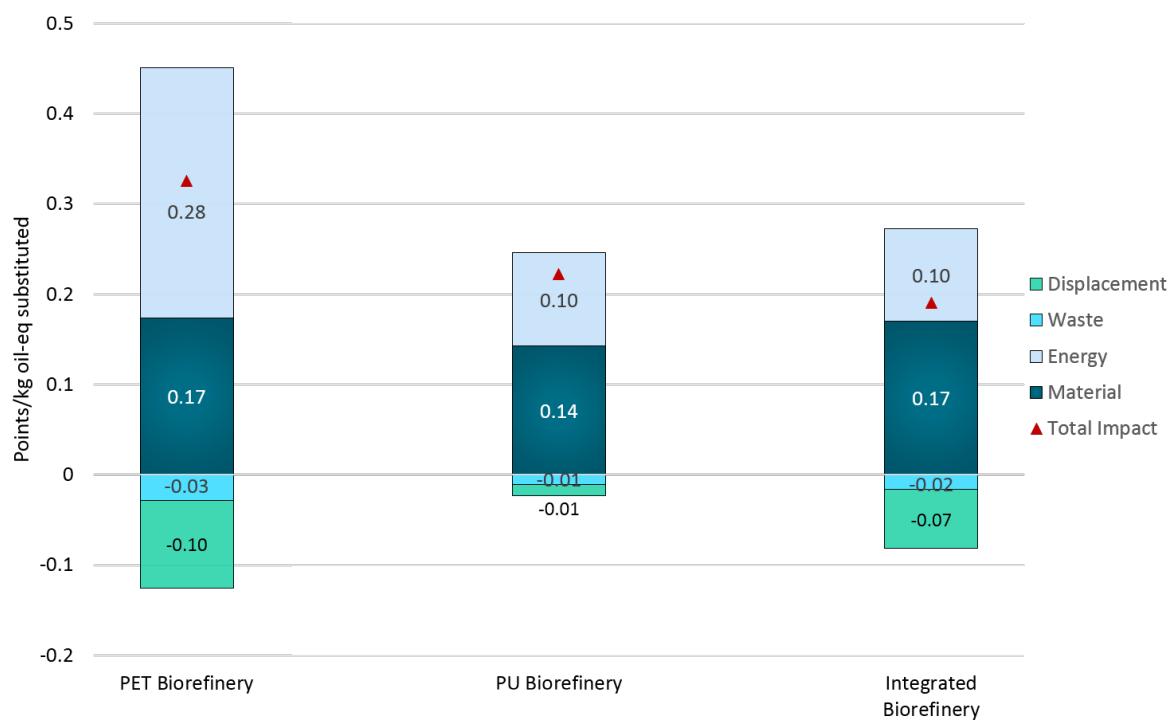


Fig. D.3. Eco-indicator 99 (EI99) of the biorefineries with 1 kg_{oil-eq} substituted as the functional unit.

E. Surrogate Predictions

E.1. Dependence of input variables on Revenue

The dependence of the individual input variables, *i.e.*, D-glucose mass flow, evaporator temperature and the crystallizer temperature on the simulated revenue are shown in Fig. E.1, Fig. E.2 and Fig. E.3 respectively.

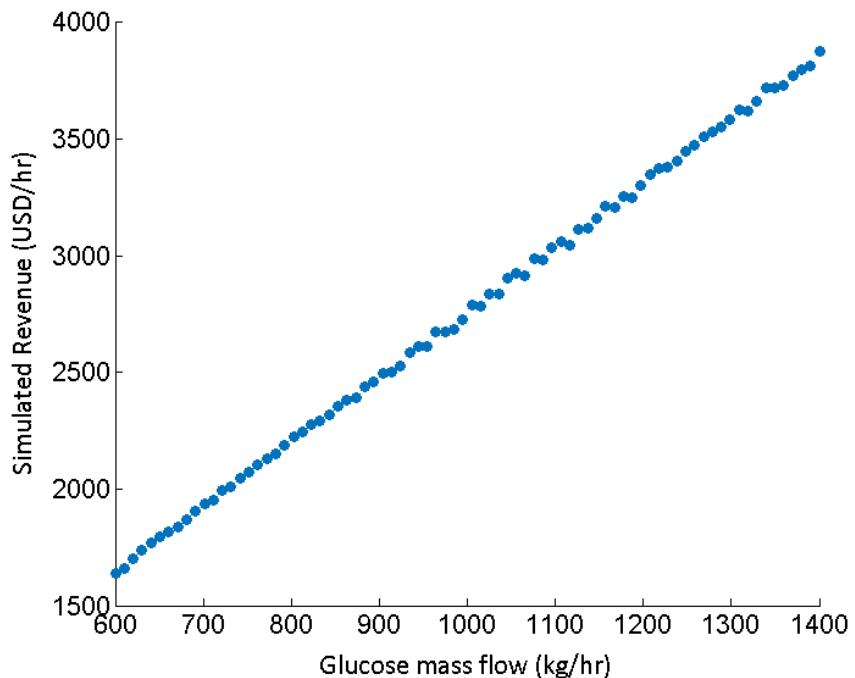


Fig. E.1. Dependence of D-glucose mass flow on the simulated revenue.

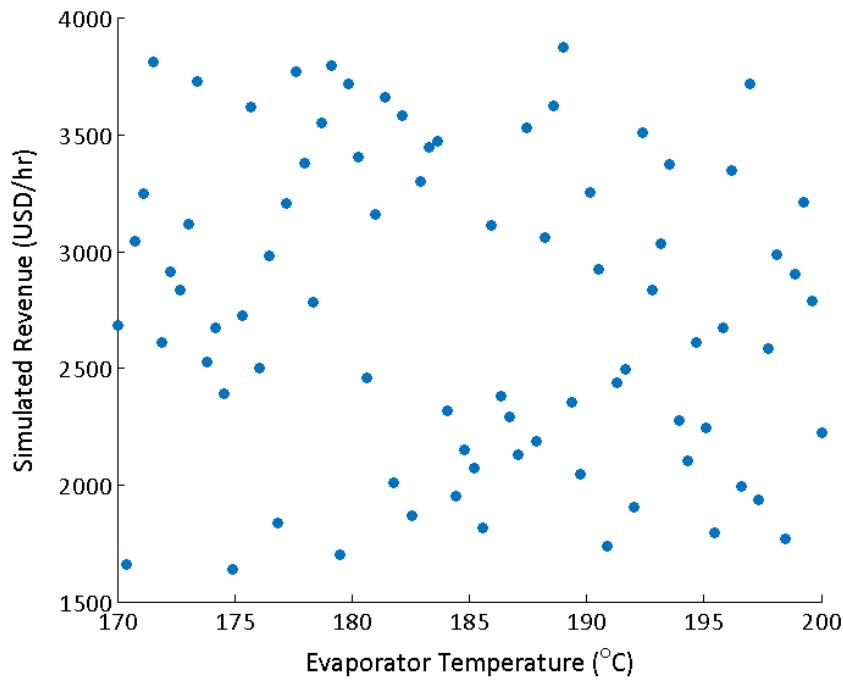


Fig. E.2. Dependence of evaporator temperature on the simulated revenue.

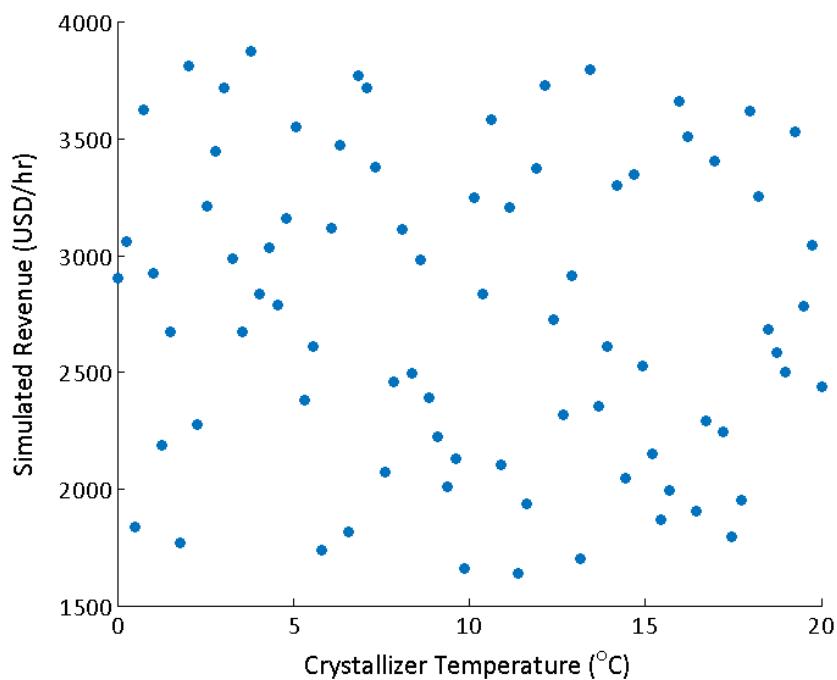


Fig. E.3. Dependence of crystallizer temperature on the simulated revenue.

E.2. Dependence of input variables on PU Purity

The dependence of the individual input variables, *i.e.*, D-glucose mass flow, evaporator temperature and the crystallizer temperature on the simulated PU purity are shown in Fig. E.4, Fig. E.5 and Fig. E.6 respectively.

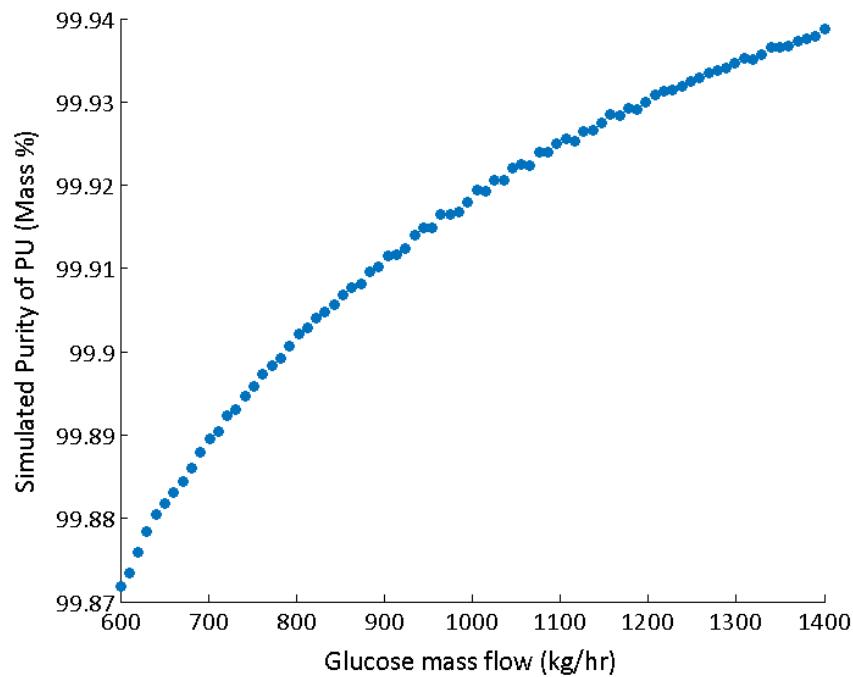


Fig. E.4. Dependence of D-glucose mass flow on the simulated PU purity.

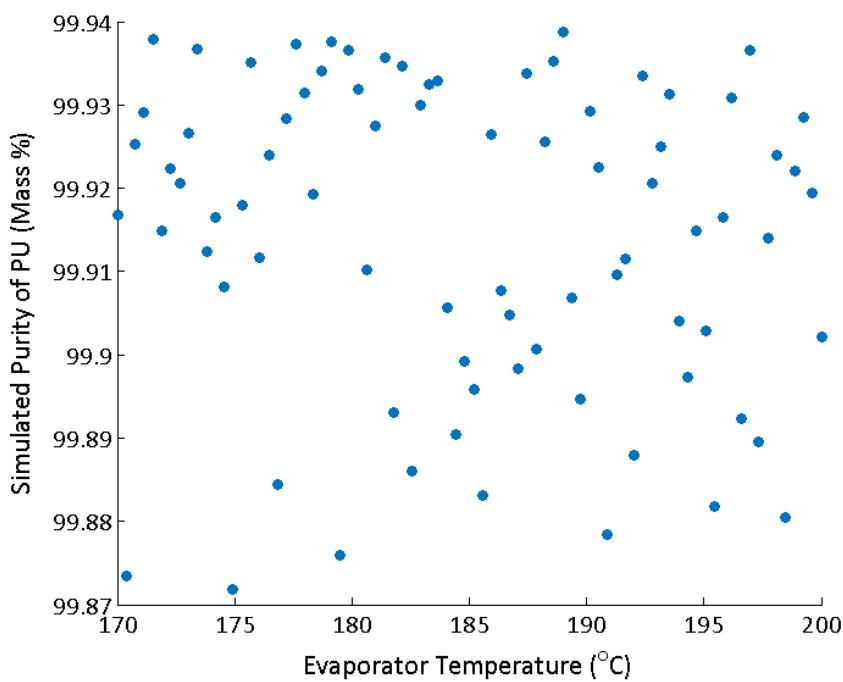


Fig. E.5. Dependence of evaporator temperature on the simulated PU purity.

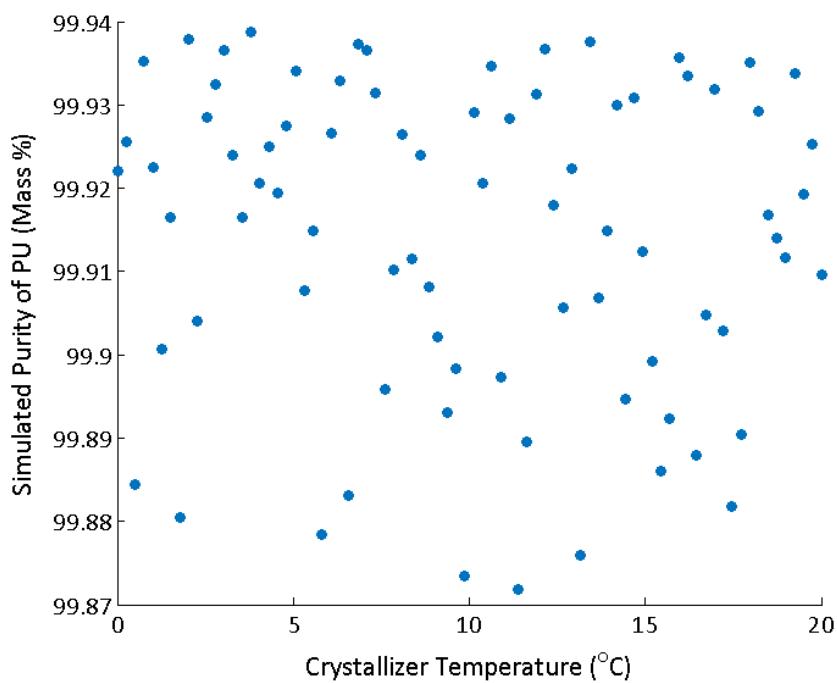


Fig. E.6. Dependence of crystallizer temperature on the simulated PU purity.

References

- [1] A. Demirbas, *Biorefineries, For Biomass Upgrading Facilities*, 1st Ed., Springer, Verlag London, **2010**.
- [2] World Commission on Environment and Development, *Our Common Future*, 1st Ed., Oxford University Press, **1987**.
- [3] *Integrated Biorefineries: Design, Analysis, and Optimization*, 1st Ed., (Eds.: P. R. Stuart, M. M. El-Halwagi), CRC Press, **2012**.
- [4] *Bioprocessing Technologies in Biorefinery for Sustainable Production of Fuels, Chemicals, and Polymers*, 1st Ed., (Eds.: S.-T. Yang, H. A. El-Ensashy, N. Thongchul), Wiley-AIChE, **2013**, Chapter Integrated Biorefinery for Sustainable Production of Fuels, Chemicals and Polymers, pp. 1–27.
- [5] M. Carus, L. Dammer, Food or Non-Food Which Agricultural Feedstocks are Best for Industrial Uses?, Research papers, nova-Institut GmbH, Hürth, Germany, **2013**.
- [6] G. W. Huber, S. Iborra, A. Corma, *Chem. Rev.* **2006**, *106*, 4044–4098.
- [7] C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, *Science* **1998**, *281*, 237–240.
- [8] A. Raschka, M. Carus, Industrial material use of biomass. Basic data for Germany, Europe and the world, report on the R&D project – Environmental Innovation Policy: more efficient resource use and climate protection through sustainable material use of biomass, nova-Institut GmbH, Hürth, Germany, **2012**.
- [9] A. Baral, B. Bakshi, *Ecological Modelling* **2010**, *221*, 1807–1818.
- [10] J. H. Clark, R. Luque, A. S. Matharu, *Annu. Rev. Chem. Biomol. Eng.* **2012**, *3*, 183–207.
- [11] S. Liua, H. Luc, R. Hua, A. Shupe, L. Line, B. Liangc, *Biotechnology Advances* **2012**, *30*, 785–810.
- [12] Bundesamt für Umwelt, Jahrbuch Wald und Holz 2014, Yearly Report, Schweizerische Eidgenossenschaft, **2014**.
- [13] A. Forrester, A. Sobester, A. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*, 1st Ed., Wiley, **2008**.
- [14] J. Zhang, J.-b. Li, S.-B. Wu, Y. Liu, *Ind. Eng. Chem. Res.* **2013**, *52*, 11799–11815.
- [15] B. Hoffer, E. Crezee, P. Mooijmana, A. van Langeveld, F. Kapteijn, J. Moulijn, *Catalysis Today* **2003**, *35*–41.
- [16] E. Crezee, B. W. Hoffer, R. J. Berger, M. Makkee, F. Kapteijn, J. A. Moulijn, *Applied Catalysis A: General* **2003**, *251*, 1–17.
- [17] L. Ye, X. Duan, H. Lin, Y. Yuan, *Catalysis Today* **2012**, *183*, 65–71.
- [18] M. Banu, S. Sivasanker, T. Sankaranarayanan, P. Venuvanalingam, *Catalysis Communications* **2011**, *12*, 673–677.
- [19] US Department of Energy, Top Value Added Chemicals From Biomass, **2004**, <http://www.nrel.gov/docs/fy04osti/35523.pdf> (visited on 08/07/2014).
- [20] P. Gallezot, P. J. Cerino, B. Blanc, G. Fleche, P. Fuertesl, *Journal Of Catalysis* **1994**, *146*, 93–102.

- [21] G. Li, H. Cai, Y. Liu, T. Su, *Hebei J. Ind. Sci. & Technol.* **2006**, *26*, 497–499.
- [22] A. M. Ruppert, K. Weinberg, R. Palkovits, *Angew. Chem. Int. Ed.* **2012**, *51*, 2564–2601.
- [23] *Biorefineries - Industrial Processes and Products: Status Quo and Future Directions*, (Eds.: B. Kamm, P. R. Gruber, M. Kamm), WILEY, **2010**.
- [24] P. F. H. Harmsen, M. M. Hackmann, H. L. Bos, *Biofuels Bioprod. Bioref.* **2014**, 306–334.
- [25] International Sugar Journal, *Vol. 115, Vol. 115*, **2013**, pp. 229–288.
- [26] A. Bridgwater, R. Chinthapalli, P. Smith, Identification and market analysis of most promising added-value products to be co-produced with the fuels, Project Report, Aston University, **2010**.
- [27] Roquette Freres, Roquette Group, **2014**, <http://www.roquette.com/>.
- [28] International Starch Institute, Sweetner analysis - Sorbitol, **2013**, <http://www.starch.dk/isi/glucose/sorbitol.asp>.
- [29] T. Kilpiö, A. Aho, D. Murzin, T. Salmi, *Ind. Eng. Chem. Res.* **2013**, *52*, 7690–7703.
- [30] M. A. Zulkifley, M. M. Mustafa, A. Hussain, A. Mustapha, S. Ramli, *PLoS ONE* **2014**, *9*.
- [31] Research and Markets, Polyethylene Terephthalate (PET) - A Global Market Watch, 2011 - 2016, **2012**, <http://www.researchandmarkets.com/research/trmp3c/polyethylene>.
- [32] Jiangsu Sanfangxiang, Jiangsu Sanfangxiang Group, **2014**, <http://www.sfxjt.com/>.
- [33] Indorama, Indorama Ventures, **2014**, <http://www.indoramaventures.com/EN/Home/index.php>.
- [34] Grand View Research - Market Research & Consulting, Bio-Based Polyethylene Terephthalate (PET) Market By Application and Segment forecasts to 2020, **2014**, <http://www.grandviewresearch.com/press-release/global-bio-based-polyethylene-terephthalate-pet>.
- [35] Coca-Cola, The Coca-Cola Company, **2014**, <http://www.coca-cola.com/global/glpl.html>.
- [36] P & G, Procter and Gamble Company, **2014**, http://www.pg.com/fr_DE/index.shtml.
- [37] Y. Banat, Z. A. El-Rub, A Technical and Economic Feasibility Study of: Production of Polyethylene Terephthalate by Direct Esterification Using Pervaporation, Research Report, University of Twente, Faculty of Chemical Technology, **2001**.
- [38] American Chemistry Council, Ed., CPI Polyurethane 2014, PU players set to up the pace, CPI Polyurethane Technical ConferenceDallas,Texas,US, Sept. 22, 2014, **2014**.
- [39] UTECH-Polyurethane, Shifting production in CASE market highlighted at conference, **2014**, <http://utech-polyurethane.com/information/shifting-production-in-case-market-highlighted-at-conferencecase/>.
- [40] K. Kersh, Global Bio-based Chemical Capacity Springs to Scale, State of the Market report, Lux Research, **2011**.
- [41] L. Hojabri, X. Kong, S. S. Narine, *Journal of Polymer Science Part A: Polymer Chemistry* **2010**, *48*, 3302–3310.
- [42] A. S. More, T. Lebarb  , L. Maisonneuve, B. Gadenne, C. Alfos, H. Cramail, *European Polymer Journal* **2013**, *49*, 823–833.
- [43] M. D. Zenner, Y. Xia, J. S. Chen, M. R. Kessler, *ChemSusChem* **2013**, *6*, 1182–1185.
- [44] J.-C. Choque, G. Fleche (R. Freres), US 6,297,409 B1 (US), **Feb. 2000**.
- [45] A. B. Bindwal, P. D. Vaidya, *Energy & Fuels* **2014**, *28*, 3357–3362.
- [46] P. Brahme, L. Doraiswamy, *Ind. Eng. Chem. Process Des. Dev.* **1976**, *15*, 130–137.

- [47] J. Wisniak, M. Hershkowitz, R. Leibowitz, S. Stein, *J. Chem. Eng. Data* **1974**, *19*, 247–249.
- [48] J. Zhang, B. Hou, A. Wang, Z. Li, H. Wang, T. Zhang, *AICHE Journal* **2015**, *61*, 224–238.
- [49] S. M. Payne, F. M. Kerton, *Green Chem.* **2010**, *12*, 1648–1653.
- [50] J. Witte, Master Thesis, ETH Zurich, **2013**.
- [51] L. Berg, 4,966,658 (US), **July 5**.
- [52] J. Besnoin, K. Choi, *Journal of Macromolecular Science Part C* **1989**, *29*, 55–81.
- [53] V. Bhaskar, S. K. Gupta, A. K. Ray, *Polymer Reaction Engineering* **2001**, *9*, 71–99.
- [54] C. Laubriet, B. LeCorre, K. Y. Choi, *Ind. Eng. Chem. Res.* **1991**, *30*, 2–12.
- [55] D. van Krevelen, W. P. J. Nijenhuis, *Properties of Polymers*, 4th Ed., Elsevier Science, **2009**.
- [56] K. C. Seavey, Y. A. Liu in *Step-Growth Polymerization Process Modeling and Product Design*, Wiley, **2009**, pp. 557–588.
- [57] M. Johnson, L. Moore, D. Ylvisaker, *Journal of Statistical Planning and Inference* **1990**, *26*, 131–148.
- [58] M. D. Morris, T. J. Mitchell, *Journal of Statistical Planning and Inference* **1995**, *43*, 381–402.
- [59] G. E. P. Box, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **6**, 81–101.
- [60] D. Stoyan, *Biometrical Journal* **1996**, *38*, 454–454.
- [61] D. R. Jones, *Journal of Global Optimization* **2001**, *21*, 341–383.
- [62] A. Forreser, A. Sobester, A. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*, **2008**.
- [63] J. E. Gentle, *Numerical Linear Algebra for Applications in Statistics*, Springer, **1998**.
- [64] H. Althaus, M. Chudacoff, R. Hischier, N. Jungbluth, M. Osses, A. Primas, *Final report ecoinvent data v2.0* **2007**, *8*.
- [65] M. Morales, P. Y. Dapsens, I. Giovinazzo, J. Witte, C. Mondelli, S. Papadokonstantakis, K. Hungerbuhler, J. Perez-Ramirez, *Energy Environ. Sci.* **2015**, *8*, 558–567.
- [66] Ecoinvent Centre, Ecoinvent Database, **2010**, <http://www.ecoinvent.ch/>.
- [67] H. Baumann, A.-M. Tillman, *The Hitch hiker's Guide to LCA, An orientation in life cycle assessment methodology and application*, **2004**.
- [68] R. V. Siemons, *A Development Perspective Foir Biomass-Fuelled Electricity Generation Technologies: Economic Technology Assessment in View of Sustainability*, **2002**.
- [69] Alibaba Group, Alibaba Enterprises, **2014**, <http://www.alibaba.com/>.