University of Glasgow

Ascher, Simon (2024) *Environmental and techno-economic analysis of biomass and waste gasification facilitated by machine learning.*
PhD thesis.

https://theses.gla.ac.uk/84564/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

Environmental and Techno-Economic Analysis of Biomass and Waste Gasification Facilitated by Machine Learning

Simon Ascher

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Engineering
College of Science and Engineering
University of Glasgow



April 2024

Abstract

Bioenergy is essential for limiting global warming to 1.5°C, the goal set out by the Intergovernmental Panel on Climate Change (IPCC) in 2015. Gasification is one type of technology that can deliver bioenergy by converting biomass into fuels and chemicals via thermal conversion. When compared to other conversion technologies, gasification stands out due to its potential to efficiently turn biomass into syngas and biochar, which have a plethora of applications. Therefore, this work sets out to improve our knowledge of biomass and waste gasification (and its process performance and efficiency), thereby encouraging its widespread commercialisation and industrial-scale implementation. This was done by developing a comprehensive modelling framework to holistically assess the environmental and techno-economic performance of a biomass or waste gasification scheme, as particularly the trade-offs between environmental and techno-economic factors remain poorly understood.

As a first step, a comprehensive data set of 312 samples was collected from the existing literature, to develop machine learning (ML) models to predict ten key biomass and waste gasification outputs (e.g. syngas yield and composition), which are vital for environmental and economic analysis. Gradient boosting regression was found to be highly accurate with a coefficient of determination of (R^2) of 0.90 when averaged across all model outputs. The developed models also boast a better generalisation capability than existing models, meaning they can predict gasification performance over a wide range of input parameters, such as feedstocks, reactor types, and gasifying agents. Crucially, investors' and policymakers' trust in the models was ensured by incorporating interpretability methods. These methods highlighted the feedstock's particle size and the gasifying agent as parameters of particular importance.

The developed ML models were then used to establish an environmental and techno-economic modelling framework, which utilises Monte Carlo methodology to account for sources of uncertainty within the framework. By implementing the framework through an easy-to-use graphical user interface, non-expert stakeholders can easily use the tool to rapidly assess the performance of a range of different gasification feedstocks and operational choices.

Ultimately, the developed modelling framework was used to conduct a case study analysing three promising Scottish feedstocks (forestry residues, barley straw, and distillers' malt draff). Through multi-objective optimisation and sensitivity analysis, the best feedstock and process conditions were identified. When considering carbon capture and storage (CCS), a maximum carbon avoidance of -2,948 $kg\ CO_2$-eq/tonne of feedstock at a benefit-cost ratio (BCR) of 1.1 was found for the feedstock forestry residues. Other solutions considering CCS showed an improved economic performance of BCR = 1.41, despite a slightly diminished environmental performance of approximately -2,400 $kg\ CO_2$-eq/tonne of feedstock. Optimal solutions not considering CCS still led to negative emissions of up to -950 $kg\ CO_2$-eq/tonne and an improved economic performance (BCRs ranging from 1.72 to 1.85). The case study effectively illustrates how the developed modelling framework can be used to rapidly and easily compare a range of feedstocks and gasification system choices. It also lays a foundation for more expansive modelling frameworks in the future which could consider a range of other bioenergy options, aiding with decisions on how to best utilise our limited biomass resources.

Contents

Abstract **i**

List of Publications **xii**

Acknowledgements **xiii**

Declaration **xiv**

Nomenclature **xv**

1 Introduction **1**

1.1 Global Warming and the Climate Crisis 1
1.2 Bioenergy and its Role in Avoiding the Climate Crisis 3
1.3 Communicating Research to Policymakers through the Use of Models . . . 6
1.4 Knowledge Gaps 8
1.5 Research Aims and Objectives and the Contribution of this Thesis 9
1.6 Thesis Structure 10
1.7 Published Work as Part of this Thesis 11

2 Literature Review **13**

2.1 Chapter Overview 13
2.2 Bioenergy 13
2.2.1 Thermochemical Methods 15
2.2.1.1 Pyrolysis 17
2.2.1.2 Gasification 17
2.2.2 Biochemical Methods 19
2.3 Machine Learning Methods to Model Thermochemical Processes 21
2.3.1 Section Summary 21
2.3.2 Introduction and Background 21
2.3.3 Machine Learning Techniques 25
2.3.3.1 Regression Methods 25
2.3.3.2 Artificial Neural Networks 26

2.3.3.3 Decision Trees and Related Methods . . . 28
2.3.3.4 Support Vector Machine . . . 30
2.3.4 Machine Learning Applications . . . 31
2.3.4.1 Artificial Neural Networks for Process Modelling . . . 34
2.3.4.1.1 Hybridised Use of ANNs with Existing Thermodynamic Equilibrium Models . . . 37
2.3.4.1.2 Time-series Approaches . . . 37
2.3.4.2 Hybrid Machine Learning Models . . . 39
2.3.4.2.1 Use of Particle Swarm Optimisation for Network Learning . . . 39
2.3.4.2.2 Dimensionality Reduction Methods . . . 40
2.3.4.3 Use of Other Machine Learning Methods . . . 41
2.3.4.3.1 Regression Methods . . . 41
2.3.4.3.2 Decision Tree-based Methods . . . 42
2.3.4.3.3 Support Vector Machines . . . 43
2.3.4.4 Use of Machine Learning in Bioenergy-related Fields . . . 43
2.3.5 Challenges and Prospects . . . 45
2.3.5.1 Challenges . . . 45
2.3.5.2 Prospects . . . 46
2.3.6 Conclusions . . . 47
2.4 Environmental and Techno-Economic Analysis of Thermochemical Processes 47
2.4.1 Environmental Analysis of Thermochemical Processes . . . 47
2.4.2 Techno-Economic Analysis of Thermochemical Processes . . . 50
2.4.3 Combined or Holistic System Analysis . . . 51
2.4.4 Monte Carlo Methodology to Account for Uncertainty within Models 54
2.5 Chapter Summary and Conclusions . . . 55

3 A Comprehensive Artificial Neural Network Model for Gasification Process Prediction 57
3.1 Chapter Overview . . . 57
3.2 Introduction . . . 58
3.3 Materials and Methods . . . 60
3.3.1 Data Collection and Description of the Data Set . . . 60
3.3.2 Data Pretreatment . . . 61
3.3.2.1 Encoding of Categorical Variables . . . 61
3.3.3 Data Normalisation . . . 64
3.3.3.1 Mean Substitution of Missing Values . . . 64
3.3.4 Preliminary Data Analysis . . . 65
3.3.5 Artificial Neural Network Development . . . 65

3.3.5.1 Artificial Neural Network Background 65
3.3.5.2 Artificial Neural Network Optimisation and Performance Assessment 67
3.3.5.3 Workflow Summary 68
3.4 Results and Discussion 69
3.4.1 Preliminary Data Analysis 69
3.4.2 Artificial Neural Network Model Optimisation 70
3.5 Chapter Summary and Conclusions 77

4 Interpretable Machine Learning to Model Biomass and Waste Gasification 78
4.1 Chapter Overview 78
4.2 Introduction 79
4.3 Materials and Methods 81
4.3.1 Data Collection, Preliminary Analysis, and Predictor Selection . . . 81
4.3.2 Data Preparation 81
4.3.3 Model Development 84
4.3.3.1 Model Types 84
4.3.3.2 Model Optimisation and Comparison 84
4.3.4 Interpretability Analysis 86
4.4 Results and Discussion 87
4.4.1 Selection of Optimal Prepossessing Steps 87
4.4.2 Model Optimisation and Comparison 88
4.4.3 Interpretability Analysis 92
4.5 Chapter Summary and Conclusions 103

5 Environmental and Economic Modelling of Biomass and Waste Gasification 104
5.1 Chapter Overview 104
5.2 Design of Modelling Framework 105
5.2.1 Object-Oriented Programming Structure of the Program 105
5.2.2 Uncertainty Accounting through Monte Carlo Methodology 106
5.2.2.1 Distribution Types Used to Model Uncertainty 108
5.2.2.2 Accounting for the Prediction Model's Uncertainty within the Monte Carlo Simulation 109
5.3 Process Modelling Methodology 111
5.3.1 Environmental Modelling through Life Cycle Assessment 112
5.3.1.1 The Allocation Problem 115
5.3.2 Economic Modelling 116

5.3.2.1 Capital Expenditure (CAPEX) Estimation . . . 117
5.3.2.2 Operation and Maintenance (O&M) Cost Estimation . . . 119
5.3.2.3 Cost Estimation Methods . . . 119
5.3.2.3.1 Cost Indices . . . 119
5.3.2.3.2 The Power-Sizing Method . . . 119
5.3.2.3.3 The Cost Estimating Relationship (CER) Method 120
5.3.2.3.4 Conversion between Different Countries and Currencies . . . 121
5.4 Process Models . . . 121
5.4.1 Gasification and Gas Cleaning . . . 121
5.4.1.1 Agent Provision . . . 124
5.4.1.1.1 Oxygen . . . 124
5.4.1.1.2 Steam . . . 124
5.4.2 Combined Heat and Power (CHP) Generation . . . 125
5.4.2.1 Syngas Combustion . . . 127
5.4.3 Feedstock Pretreatment . . . 128
5.4.3.1 Bale Breaking and Chopping . . . 128
5.4.3.2 Drying . . . 129
5.4.3.3 Milling/Grinding . . . 133
5.4.3.4 Pelleting . . . 136
5.4.4 Biochar Use . . . 138
5.4.4.1 Soil Carbon Sequestration as a Result of Biochar Application to Soil . . . 139
5.4.4.2 Reductions in Soil N_2O Emissions as a Result of Biochar Application to Soil . . . 141
5.4.5 Carbon Capture and Storage . . . 141
5.4.5.1 CO_2 Transport and Storage Costs . . . 144
5.4.5.2 Costs due to Monoethanolamine (MEA) Degradation for Amine-based Carbon Capture . . . 145
5.5 Complete Gasification System Model . . . 146
5.5.1 System Optimisation . . . 148
5.5.2 Decision Support Tool with a Graphical User Interface (GUI) Based on the Developed Framework . . . 149
5.6 Chapter Summary and Conclusions . . . 152

6 A Case Study Using the Developed Modelling Framework: Gasification of Barley Straw, Forestry Residues, and Distillers' Malt Draff in Scotland 153
6.1 Chapter Overview . . . 153

6.2 Introduction and Background . . . 154
6.3 Methodology . . . 155
6.3.1 General Process Data . . . 155
6.3.2 Feedstock Availability and Composition . . . 156
6.3.3 Economic Factors . . . 157
6.3.3.1 Interest/Discount Rate . . . 157
6.3.3.2 Energy prices . . . 158
6.3.3.3 Feedstock Prices . . . 158
6.3.3.4 Biochar Price . . . 159
6.3.4 Carbon Tax . . . 160
6.3.5 System Optimisation . . . 160
6.3.5.1 Sensitivity Analysis . . . 161
6.4 Results and Discussion . . . 162
6.4.1 Baseline Results . . . 162
6.4.2 Optimisation Results . . . 174
6.4.3 Sensitivity Analysis Results . . . 179
6.4.4 Updated Baseline Results for the Feedstock Forestry Residues . . . 181
6.5 Chapter Summary and Conclusions . . . 184

7 Conclusions, Limitations, and Future Work **187**
7.1 Summary and Conclusions . . . 187
7.2 Limitations and Future Work . . . 188

References **192**

Appendix **231**
A Detailed Screenshots of the Developed Decision Support Tool with a Graphical User Interface (GUI) . . . 231
B Baseline Energy Generation Results for Forestry Residues, Barley Straw, and Draff . . . 242
C Optimisation Results for Barley Straw and Draff . . . 246

List of Tables

2.1 Summary of ML models used for the modelling of gasification. 32
2.2 Summary of ML models used for the modelling of pyrolysis. 33

3.1 Summary of predictor variables. 62
3.2 Summary of target variables. 62
3.3 Predictor combinations used for model training. 63
3.4 Example of one-hot encoding for the variable gasifying agent. 63
3.5 Optimised and fixed hyperparameters used during ANN training. 68
3.6 Summary of performance indicators and hyperparameter choices for best performing models. 71
3.7 Model performance indicators by output. 75

4.1 Comparison of different predictor and target preprocessing methods. 88
4.2 5-fold cross-validated R^2 and RMSE scores for all model outputs and model types. 89

5.1 Performance summary of CAPEX models fitted to gasifier data. 124
5.2 Conversion efficiencies and parasitic loads of various CHP units. 125
5.3 Performance summary of CAPEX models fitted to CHP data. 127
5.4 Performance summary of CAPEX models fitted to feedstock dryer data. . . 133

6.1 Annual operating hours of gasification plants. 155
6.2 Ultimate composition of forestry residues, barley straw, and draff. 157
6.3 Proximate composition and lower heating value (LHV) of forestry residues, barley straw, and draff. 157
6.4 Optimised parameters and their respective values. 161

List of Figures

1.1 Overview of different bioenergy conversion pathways. 4
1.2 Illustration of the different components making up the developed biomass and waste gasification modelling framework and their interactions. 11

2.1 Prevalence of different reactor types at different gasifier sizes. 19
2.2 Research interest in pyrolysis and gasification modelling. 23
2.3 Prevalence of machine learning modelling for gasification and pyrolysis. . . 25
2.4 Illustration of single layer perceptron (single hidden layer back-propagation network) architecture. 27
2.5 Illustration of decision tree and RF methodology. 29
2.6 SVM classification illustrated using a simple example of two linearly separable classes. 31

3.1 Spearman correlation coefficient matrix of continuous input features. . . . 66
3.2 Flowchart of the procedure employed for the model development. 69
3.3 Distribution of model predictions against target values. 73
3.4 Network structure of best performing ANN trained on set OPT. 74
3.5 Distribution of model predictions. 76

4.1 Data pretreatment methods to handle missing data. 83
4.2 Flowchart of the methodology and workflow of this study, illustrating the two stages model development and interpretability analysis 85
4.3 Scatter plots of target values vs. predictions for all 10 optimised GBR submodels. 90
4.4 Scatter plot of scaled target values vs. predictions for GBR models trained for all ten outputs. 91
4.5 Stacked feature importance graphs 95
4.6 Normalised permutation and SHAP feature importance analysis results. . . 97
4.7 SHAP feature importance analysis for GBR model predicting the syngas yield. 99
4.8 SHAP explanations of some selected individual predictions. 101

5.1 The five main object types used within the program and their interactions. 107
5.2 Finding prediction boundaries using the kernel density method. 110
5.3 Comparison of local and global error method. 111
5.4 The four key stages of LCA. 113
5.5 Simplified system boundary. 115
5.6 CAPEX levels and their components. 118
5.7 Scatter plot of normalised fixed bed and fluidised bed gasifier CAPEX. . . 123
5.8 Regression fits to the normalised CAPEX of gasifiers. 123
5.9 Scatter plots of the normalised CHP CAPEX. 126
5.10 Regression fits to the normalised CAPEX of CHP units. 127
5.11 Scatter plot and regression fit to the normalised CAPEX of feedstock grinders. 130
5.12 Scatter plot of normalised feedstock dryer CAPEX by reference. 131
5.13 Regression fits to the normalised CAPEX of feedstock dryers. 132
5.14 Energy consumption for milling in relation to the employed mill screen size. 134
5.15 Scatter plot and regression fit for the normalised CAPEX of hammer mills. 135
5.16 Energy consumption for pelleting in relation to feedstock's particle size. . . 137
5.17 Scatter plot and regression fits to the normalised CAPEX of pellet mills and coolers. 138
5.18 Energy consumption for amine-based carbon capture process. 143
5.19 (a) Energy consumption and (b) carbon recovery rate for RVPSA carbon capture process. 144
5.20 Illustration of the design of the developed decision support tool with a GUI. Dashed arrows indicate optional steps. 150
5.21 Page layout of the developed decision support tool. 151

6.1 Baseline environmental results for forestry residues as the feedstock. 163
6.2 Baseline environmental results for barley straw as the feedstock. 164
6.3 Baseline environmental results for draff as the feedstock. 165
6.4 Net baseline economic results for forestry residues as the feedstock. 167
6.5 Net baseline economic results for barley straw as the feedstock. 168
6.6 Net baseline economic results for draff as the feedstock. 169
6.7 Baseline economic results shown by process for forestry residues as the feedstock. 170
6.8 Baseline economic results shown by process for barley straw as the feedstock. 171
6.9 Baseline economic results shown by process for draff as the feedstock. . . . 172
6.10 Optimisation results differentiated by feedstock. 174
6.11 Optimisation results for the feedstock forestry residues shown for the model parameter gasification temperature [°C]. 175

6.12 Optimisation results for the feedstock forestry residues shown for the model parameter ER. 176
6.13 Optimisation results for the feedstock forestry residues shown for the model parameter gasifying agent. 177
6.14 Optimisation results for the feedstock forestry residues shown for the model parameter reactor type. 177
6.15 Optimisation results for the feedstock forestry residues shown for the model parameter interest rate [%]. 178
6.16 Optimisation results for the feedstock forestry residues shown for the model parameter CCS (included or not included). 178
6.17 Optimisation results for the feedstock forestry residues shown for the model parameter carbon tax [£/tonne CO_2-eq]. 179
6.18 Sensitivity analysis results for the feedstock forestry residues shown for the model parameter gasification temperature [°C]. 180
6.19 Sensitivity analysis results for the feedstock forestry residues shown for the model parameter electricity price [£/kWh]. 181
6.20 Sensitivity analysis results for the feedstock forestry residues shown for the model parameter system size [tonnes/hour]. 182
6.21 Environmental results for the baseline scenario, updated according to the optimisation and sensitivity analysis findings for the feedstock forestry residues. 183
6.22 Economic results for the baseline scenario, updated according to the optimisation and sensitivity analysis findings for the feedstock forestry residues. 185

List of Publications

Journal Articles

- Simon Ascher, William Sloan, Ian Watson, and Siming You. "A comprehensive artificial neural network model for gasification process prediction". In: *Appl. Energy* 320 (2022), p. 119289. DOI: 10.1016/j.apenergy.2022.119289
- Simon Ascher, Xiaonan Wang, Ian Watson, William Sloan, and Siming You. "Interpretable machine learning to model biomass and waste gasification". In: *Bioresour. Technol.* 364 (2022), p. 128062. DOI: 10.1016/j.biortech.2022.128062
- Simon Ascher, Ian Watson, and Siming You. "Machine learning methods for modelling the gasification and pyrolysis of biomass and waste". In: *Renew. Sustain. Energy Rev.* 155 (2022), p. 111902. DOI: 10.1016/J.RSER.2021.111902

Acknowledgements

I would like to start by thanking all my family and friends who have supported me along the way. It was often the small things, and things completely unrelated to my PhD, that kept me in a good mental space and able to complete this work, so thanks for that. In particular, I would like to thank my family for their belief in me and their support while being away from home all this time. Next, I would like to thank various flatmates over the past three and a half years. Be it productive or silly discussions, endless cups of tea, or anything else – thank you Sean, Amir, Jan, Chris, and Michael. The same goes for all my other friends who have made university life and life in general as much fun as it has been. Finally, a special thank you to Scotland for providing endless distractions and adventures, allowing me to refill my mental energy tank, and to my partner Iona for being the best companion on many of these adventures.

On the academic side, I would like to express my sincere gratitude to my first supervisor, Siming You, for his support and guidance throughout my research journey. Similarly, I would like to thank my second supervisor, Ian Watson, especially for the bigger-picture discussions and many more light-hearted chats. They always made me feel more relaxed and helped me focus on the important things. A big thank you also goes out to the rest of my group, namely Jade Lui, Yi Fang, Yize Li, Luís Rodrigues, Bauyrzhan Biakhmetov, Asam Ahmed, and Rohit Gupta, for their support and advice.

Declaration

I hereby declare that the contents of this thesis are original and my own work. All references made to the work of others are duly cited and included in the bibliography. This thesis has not been submitted, as a whole or in part, for any other degree at the University of Glasgow or any other institution.

All research data is openly available within the relevant chapters of this work.

Simon Ascher
April 2024

Nomenclature

AD	anaerobic digestion
ANFIS	adaptive neuro-fuzzy inference system
ANN	artificial neural network
ar	as received
ASU	air separation unit
AV	annual value
BCR	benefit-cost ratio
BEC	bare erected cost
BECCS	bioenergy with carbon capture and storage
BR	Bayesian regularisation backpropagation
C_2H_n	higher order hydrocarbons
CO_2-eq	carbon dioxide-equivalents
CAPEX	capital expenditures
CER	cost estimating relationship
CEPCI	Chemical Engineering Plant Cost Index
CHP	combined heat and power
CCS	carbon capture and storage
CCUS	case of carbon capture utilisation and storage
CFD	computational fluid dynamics
CNN	convolutional neural networks
DACCS	direct air carbon capture and storage
daf	dry ash-free basis
db	dry basis
DT	decision tree
ETS	emissions trading scheme
ER	equivalence ratio
FC	fixed carbon
FU	functional unit
FV	future value

GBR	gradient boosting regression model
GUI	graphical user interface
HHV	higher heating value
i	interest/discount rate
IEA	International Energy Agency
IPCC	Intergovernmental Panel on Climate Change
LCA	life cycle assessment
LCI	life cycle inventory
LCIA	life cycle impact assessment
LCSA	life cycle sustainability assessment
LHV	lower heating value
LIME	local surrogate models
LM	Levenberg-Marquardt backpropagation
LOGSIG	logistic sigmoid function
MAPE	mean absolute percentage error
MEA	monoethanolamine
MIMO	multiple-input multiple-output
ML	machine learning
MLP	multilayer perceptron
MLR	multiple linear regression
MSE	mean squared error
MSW	municipal solid waste
N	system/component life span
NARX	autoregressive exogenous neural networks
NET	negative emissions technology
NPV	net present value
O&M	operation and maintenance
PCA	principal component analysis
PDP	partial dependence plot
PV	present value
PR	polynomial regression
PSO	particle swarm optimisation
R^2	coefficient of determination
RF	random forest model
RMSE	root mean square error
RMSPE	root mean squared percentage error

RVPSA	rapid vacuum pressure swing adsorption
R^2_{adj}	adjusted coefficient of determination
SD	standard deviation
SCC	Spearman's correlation coefficients
SE	standard error
SHAP	SHapley Additive exPlanations
SL	super learner model
SOFC	solid oxide fuel cell
SVM	support vector machine
TANSIG	tangent sigmoid function
TOC	total overnight cost
TEA	techno-economic analysis
TLR	technology readiness level
UN	United Nations
UNFCCC	United Nations Framework Convention on Climate Change
VM	volatile matter
wb	wet basis
XGB	extreme gradient boosting

Chapter 1

Introduction

1.1 Global Warming and the Climate Crisis

In the 21^{st} century, limiting global warming to avoid a climate crisis is one of the greatest challenges humanity faces. Climate change has been shown to be largely caused by the anthropogenic emissions of greenhouse gases (GHGs). In particular carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O) have been key drivers responsible for climate change since the pre-industrial period. Especially the use of fossil fuels, across a plethora of sectors and industries, has led to an increase in the atmospheric concentrations of these gases **[4, 5]**.

Climate change poses many risks to both natural and human systems, encompassing extreme weather events such as heatwaves, storms, floods, and droughts, which threaten lives, infrastructure, and ecosystems. Additionally, rising sea levels from melting glaciers and ice caps exacerbate coastal flooding and erosion, endangering coastal communities and ecosystems. Climate change's repercussions extend to food and water security, economic stability, and public health, with vulnerable populations facing heightened risks. Moreover, the economic impacts are substantial, encompassing infrastructure damage, agricultural losses, and increased healthcare expenses, exacerbating social and economic inequalities. This can lead to social and political instability through resource scarcity and conflict over natural resources, amplifying geopolitical tensions. All these factors mean that mitigating these risks necessitates urgent and joint efforts to reduce GHG emissions and safeguard the well-being of present and future generations **[5–9]**.

The Intergovernmental Panel on Climate Change (IPCC), a scientific body founded by the United Nations (UN), has been at the forefront of informing the public about the scientific basis of climate change, its impacts, and possible adaptation and mitigation strategies. This is done by providing regular assessments based on scientific literature on climate change and its impacts.

In the recent past, the *Paris Agreement* was a landmark international treaty adopted

in December 2015 at the 21st Conference of the Parties (COP21) hosted by the United Nations Framework Convention on Climate Change (UNFCCC), which is a part of the UN aimed at guiding the global response to climate change. It set the chief goal to prevent "the increase in the global average temperature to well below 2°C above pre-industrial levels" and to pursue efforts "to limit the temperature increase to 1.5°C above pre-industrial levels" **[9]**. In 2018 the *Katowice Climate Package* reaffirmed the *Paris Agreement* and set provisions into place to implement it **[10]**.

The *Paris Agreement* stated that GHG emissions must peak before 2025 at the latest and decline by around 45% (when compared to 2010 levels) by 2030 to meet the 1.5°C target **[9]**. According to the most recent IPCC assessment report (AR6), released in 2023, a mean increase in global surface temperature of around 1.1°C above 1850 to 1900 levels was seen from 2011 to 2020 **[8]**. Some sources even suggest that global warming has already breached the 1.5 °C limit with a warming of 1.52°C in the 2023 calendar year **[11]**. Now, in 2024, the growth of GHG emissions has slowed but is likely yet to reach its peak **[12]**. In 2021, global GHG emissions were estimated at 54.59 billion tonnes of carbon dioxide-equivalents (CO_2-eq). Whilst the GHG emissions of much of the Western world, such as the EU, UK, and USA, peaked as early as around 1980, countries like China and India are still rapidly growing and so are their GHG emissions **[12]**. However, it is essential to highlight, that to this date most of the net anthropogenic CO_2 emissions have come from North America and Europe which are responsible for 23% and 16% of all emissions between 1850 and 2019 **[8]**.

The unit CO_2-eq, as introduced in the previous paragraph, is the standard unit used to quantify GHG emissions. It is used by the UNFCCC and IPCC for the target-setting of countries' emissions reduction targets and in their GHG reporting. Furthermore, it has been widely adopted within the scientific literature, in particular for life cycle assessment (LCA), which is a method for the comprehensive environmental assessment of products, services, and systems. In essence, the metric of CO_2-eq aims to consolidate the warming impacts of all GHGs into a singular metric. This calculation factors in both the potency of various gases contributing to global warming and their respective lifespans in the atmosphere.

In the early 1990s, the IPCC released its first set of emission scenarios which are a set of potential future pathways for greenhouse gas emissions, developed to explore different socioeconomic, technological, and policy drivers that could shape future climate outcomes **[13]**. These scenarios are often used as inputs for climate models to assess the potential impacts of different mitigation and adaptation strategies and have since been periodically updated. Bioenergy, and in particular bioenergy with carbon capture and storage (BECCS), play a critical role in many of the IPCC's emission scenarios for staying under the 1.5°C or 2°C global warming limit **[14, 15]**. However, bioenergy and BECCS

come with their own set of problems which need to be carefully analysed to ensure their positive impact on limiting global warming. This is further explored in the next section (Section 1.2).

1.2 Bioenergy and its Role in Avoiding the Climate Crisis

Bioenergy refers to energy derived from the conversion of organic materials known as biomass. Biomass contains carbon which has been absorbed by plants during their growth through photosynthesis. When producing energy from biomass, the stored carbon is released into the atmosphere. If biomass is responsibly managed and the same amount of biomass is replanted, the new biomass can absorb carbon and reduce bioenergy emissions to near-zero levels. Globally, bioenergy accounts for 55% of all renewable energy and 6% of the global energy supply, making it the largest renewable energy source **[16]**.

Biofuels are classified into three main generations based on the feedstock used and the production processes involved. First-generation biofuels (1G) are typically produced from food crops or feedstocks that are also used for food production. Examples include ethanol from corn (maize), sugarcane, or wheat, and biodiesel from vegetable oils like soybean oil or palm oil. First-generation biofuels have faced criticism for their potential to compete with food production, land use change, and limited greenhouse gas emission reductions **[17]**. Second-generation biofuels (2G) have widely been seen as a response to the controversy surrounding the sustainability of 1G biofuels **[17]**. They are produced from non-food crops, agricultural residues, woody biomass, or other lignocellulosic materials. Examples include cellulosic ethanol, produced from crop residues, grasses, or wood, and biodiesel from non-edible oil seeds or algae. Second-generation biofuels have the advantage of using non-food feedstocks, potentially reducing competition with food production and avoiding indirect land use change. They also offer the potential for greater greenhouse gas emission reductions compared to first-generation biofuels **[17–19]**. Third-generation biofuels (3G) are typically derived from algae. They offer significant potential for high yields and superior environmental performance when compared to 1G and 2G biofuels, as well as low land and resource use. However, the scale-up of the technology has proven difficult thus far. Finally, advanced biofuels, sometimes classified as fourth and fifth-generation biofuels, have been in their early stages of development, utilising technologies such as genetic engineering to increase the desired traits of organisms used in biofuel production **[20]**.

Numerous technologies and processes can convert biomass into bioenergy. The two main pathways are thermochemical and biochemical conversion pathways. During the thermochemical conversion, the organic components in the biomass are decomposed us-

ing heat. Conversely, biochemical conversion processes use microorganisms or enzymes to convert biomass into a usable form of energy **[18]**. Transesterification, a chemical conversion pathway, forms an exception. During this process, triglycerides are converted into fatty acid methyl esters (biodiesel) in the presence of alcohol (methanol or ethanol) **[21]**. Thermochemical conversion pathways include pyrolysis, gasification, liquefaction, and combustion, whereas anaerobic digestion, alcoholic fermentation, and photobiological reaction are three biochemical conversion pathways **[18]**. Figure 1.1 summarises these conversion pathways and their potential feedstocks and products.

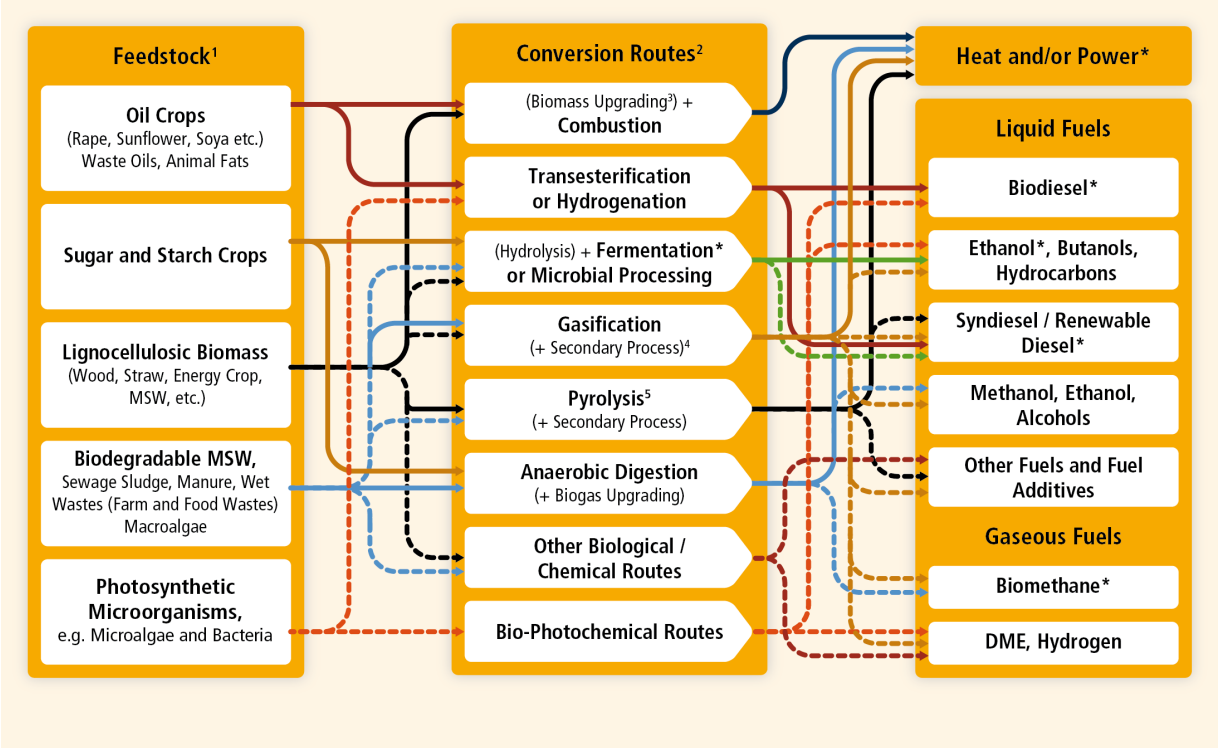


Figure 1.1: Overview of different bioenergy conversion pathways (Note: DME = dimethyl ether; MSW = municipal solid waste). Notes: 1. Parts of each feedstock, for example, crop residues, could also be used in other routes. 2. Each route also gives co-products. 3. Biomass upgrading includes any one of the densification processes (pelletisation, pyrolysis, torrefaction, etc.). 4. Anaerobic digestion processes release methane and CO_2 and the removal of CO_2 provides essentially methane, the major component of natural gas; the upgraded gas is called biomethane. 5. Could be other thermal processing routes such as hydrothermal, liquefaction, etc. DME=dimethyl ether. Used under the permission of the IPCC **[22]**.

One can distinguish between direct and indirect thermochemical conversion approaches. During combustion, the energy contained in the biomass is directly converted into heat and electricity. Indirect processes, such as pyrolysis and gasification, produce three main products at different rates: bio-oil, a gaseous product (pyrolysis gas in the case of pyrolysis and syngas (also referred to as synthesis gas or product gas) in the case of gasification), and biochar **[19, 23]**. Bio-oils and syngas can be further converted to biofuels like bioethanol,

whereas biochar represents a stable form of carbon with varying properties depending on factors such as the exact feedstock used, the chosen conversion pathway, and the specific conditions of the conversion process.

Exploring pyrolysis and gasification further, pyrolysis occurs in the absence of oxygen. In contrast, during gasification, the conversion occurs in a controlled low-oxygen environment facilitated by an external agent called the gasifying agent. The gasifying agent is typically air, steam, oxygen, or a combination of these. Pyrolysis is typically carried out at lower temperatures, ranging between 300 and 650°C, whereas gasification typically occurs at a higher temperature ranging from around 700 to 1000°C **[3, 24]**. Generally, pyrolysis produces mainly bio-oil and biochar, whereas gasification is primarily associated with the production of syngas. However, more recently this line has been blurred and gasification has been considered as an alternative biochar production pathway due to the biochar's altered properties **[25–27]**.

As previously mentioned, bioenergy plays a critical role in meeting global warming targets. The use of modern bioenergy has seen an increase of about 3% per year between 2010 and 2022 and this upward trend continues. However, to meet net zero emissions targets, its use needs to increase by a further 8% per year between 2022 and 2030. At the same time, social and environmental consequences need to be monitored to ensure a sustainable increase in bioenergy deployment. Particularly, biofuels could play a role in hard-to-electrify sectors such as aviation, shipping, heavy transport, and manufacturing **[16]**. When used for electricity and heat generation bioenergy has the potential to replace fossil fuels by providing dispatchable power. Unlike some other renewable energy sources like solar and wind, which are intermittent and dependent on weather conditions, bioenergy can generate electricity on demand, providing a stable and predictable baseline supply to the grid.

Bioenergy's potential for negative emissions is highlighted when considering the climate mitigation goals outlined in the IPCC's *Paris Agreement*, which considers negative emissions technologies (NETs) as an essential tool for limiting global warming to 2°C **[10, 28]**. A negative emissions technology (NET) is a technology or approach that removes CO_2 or other GHGs from the atmosphere, thereby contributing to climate change mitigation by creating a negative emissions scenario. These technologies aim to reduce the concentration of GHGs in the atmosphere, effectively offsetting emissions from other sources. Importantly, NETs should not be seen as a substitute for reducing GHG emissions at their source (e.g. reducing the use of fossil fuels), but rather as complementary strategies to help achieve net-zero emissions. Most likely, a range of NETs will be required to cost-effectively and sustainably meet the rates of carbon uptake required in the IPCC pathways. BECCS and biochar are two key options with estimated carbon saving potentials of 0.5 to 5 $Gt\ CO_2$/year and 0.5 to 2 $Gt\ CO_2$/year, respectively **[28]**. In comparison,

the potential of (i) afforestation and reforestation and (ii) direct air carbon capture and storage (DACCS) are considered to be 0.5 to 3.6 $Gt\ CO_2$/year and 0.5 to 5 $Gt\ CO_2$/year, respectively **[28]**.

The cost-effectiveness of NETs is essential in realising their wide-scale implementation. In addition to BECCS's significant potential for negative emissions, it is a cost-competitive NET option. BECCS's costs are estimated to be $US 100 to 200 per tonne of CO_2 by 2050. DACCS on the other hand has high initial costs of $US 600 to 1,000 per tonne of CO_2, even though as the technology matures nth plants are estimated to have costs of $US 100 to 300 per tonne of CO_2. The cost of mitigation through biochar from pyrolysis is inherently uncertain, due to the lack of commercially available biochar from large-scale plants, but has been estimated at $US 90 to 120 per tonne of CO_2 **[28]**. Gasification represents an interesting opportunity due to its potential to contribute to negative emissions through two NETs, namely BECCS and biochar. Additionally, it may be more economical than pyrolysis due to the production of syngas as a primary high-value product **[27]**.

BECCS is subject to several side effects (both positive and negative) which need to be assessed before its large-scale uptake. They fall into the following broad categories: (i) climate effects induced by biomass provision, (ii) resource needs, and (iii) broader environmental and sustainability effects transmitted via the coupled land-energy system **[28–30]**.

The effect of biochar application as a soil amendment, which is its typical application, is also subject to significant uncertainty. It has been shown to have beneficial effects on crop productivity and yield and may decrease emissions of other GHGs like N_2O and CH_4 **[28, 31]**. Water retention has also been improved in some studies, which is of particular importance to regions of the world at increased risk of aridification due to climate change **[32]**. However, many of these potential benefits are highly dependent on factors such as the soil type, incorporation method, and biochar composition **[26, 27]**.

1.3 Communicating Research to Policymakers through the Use of Models

The implementation of new technologies is heavily dependent on the decisions made by policymakers. Regulatory frameworks govern the adoption and deployment of new technologies, and incentives, such as subsidies, tax incentives, or grants, encourage investment in these new technologies. Furthermore, funding for research and development can accelerate the timelines for the development and deployment of innovative technologies. Finally, policymakers play a role in shaping public perception and acceptance of new technologies through communication, education, and public engagement initiatives. The public's attitude towards new technologies is crucial to facilitate the uptake of these technologies. All

these factors mean that the effective communication of research findings to policymakers is of great importance.

India serves as an example of how research can inform policy-making. With its large forest cover (24% of the country) and agriculture-based economy, India has a significant amount of biomass available for the use of bioenergy **[33]**. In particular 2G biomass, in the form of deadwood, forestry residues, and agricultural waste and residues, is abundant and not well utilised at the moment, with many of these types of biomass being currently burned **[33]**. Additionally, fuelwood and other types of biomass are still frequently used as cooking fuels in rural areas, leading to numerous health concerns **[34]**. Utilising this waste biomass for gasification could contribute to the country's rising energy demands while reducing environmental emissions resulting from the direct burning of biomass. It can also address health risks associated with the use of unclean cooking fuels by upgrading gasification products to clean cooking fuel alternatives. The Indian government has acted on this huge opportunity by introducing, amongst other measures, the "biomass gasification program" and "central financial assistance biomass gasifier program", which effectively promote and incentivise the use of biomass gasifiers to improve off-grid power in rural areas and generate power for industries **[33, 35]**.

One effective way of communicating research to policymakers is through the use of models. They can be used to provide structured and evidence-based insights into complex systems. By finding variables which influence the outputs of a system and exploring variations in these variables, a range of scenarios can be compared. By comparing the outcomes of different scenarios, policymakers can better understand the potential impacts of their decisions and make informed choices. Models can also be used to provide visual representations of engineering concepts, data, and results, making complex information more accessible and understandable. In turn, this facilitates communication between policymakers and engineers or scientists, improving collaboration between the two groups. Ultimately, this makes models a powerful decision-support tool by providing quantitative insights and evidence-based recommendations.

The complexity of modern engineering projects demands policymakers' attention to a multitude of factors, with environmental, economic, and social considerations standing out as paramount. Consequently, there's a growing necessity for holistic modelling approaches that account for these diverse factors and their interconnections, often necessitating interdisciplinary approaches or the integration of extensive data sets. Machine learning (ML) has emerged as a pivotal tool in crafting such models, owing to its capacity to process vast amounts of data. Its widespread adoption across science, technology, and all fields of life has significantly reshaped these domains and society at large **[36]**.

Interpretable ML models are essential for fostering trust among policy and decision-makers in developed models. Interpretable models provide transparency into the decision-

making process, enabling stakeholders to understand the factors influencing predictions or decisions. This transparency helps build confidence in the model's outputs and facilitates informed decision-making. In addition, they allow the developer to ensure that the model is (i) fair and predictions are unbiased, (ii) reliable and robust, and (iii) causal (i.e. only causal relationships are picked up by the model) **[37, 38]**. In addition, uncertainty around bioenergy technology's performance has been another major factor hindering investment **[26, 27]**. Hence, robust uncertainty accounting frameworks allow policymakers to assess the reliability and limitations of the model's predictions, providing a clear understanding of the level of confidence associated with each outcome. By integrating interpretable models with uncertainty quantification techniques, policymakers can gain insights into both the model's reasoning and the level of uncertainty, thus enhancing trust and enabling more confident adoption of the model's recommendations in policy and decision-making processes.

1.4 Knowledge Gaps

The first and arguably most crucial step in training any ML model is collecting high-quality training data. Existing data sets on the biomass and waste gasification process, have been limited in several ways: (i) Due to data availability restrictions, data sets have often been too small to effectively train a ML model. A large enough data set is important to ensure that ML models can effectively generalise, capture complex patterns, and minimise the risk of overfitting. (ii) Existing data sets were often too narrow in scope, leading to models that only generalise to a narrow range of scenarios, such as only singular feedstock or reactor types. A data set more narrow in focus can result in more consistent data, which may improve performance in certain applications. However, in other applications, the ability to generalise and make predictions for a wide range of feedstock and processing choices is more crucial. (iii) Many data sets lacked transparency about the way the data was collected and converted to the same basis if necessary. This may lead to a lack of trust in the data and introduce unexpected uncertainty or errors. In addition, this makes it challenging to compare different data sets. (iv) Frequently, only a select number of model outputs were collected in data sets, leading to models that can only predict some gasification outputs. Especially when considering further gasification process modelling (e.g. in the form of LCA or techno-economic analysis (TEA)) other relevant outputs should also be collected in data sets.

As developed ML models are highly dependent on the data sets used to train them, existing models faced all the same issues as the ones mentioned above regarding existing data sets. However, an additional issue with existing models is their lack of interpretability. The black-box nature of ML algorithms can limit their usefulness, as they often do

not provide transparent insight into the model's underlying decision-making process. Improving models' interpretability (a model is considered interpretable when the reasoning behind its predictions can be readily understood by humans) is crucial for two reasons. Firstly, it allows researchers to guard against bias in their models and extract maximum knowledge from their data and models, improving the efficiency of data and model utilisation. Secondly, interpretable models' transparency makes them easier to communicate to policy and decision-makers because they provide insights into the factors influencing predictions or decisions, facilitating better understanding and trust in the model's outputs.

LCA and TEA are labour and time-intensive due to their comprehensive nature, involving extensive data collection, analysis, and modelling across various stages of a system's life cycle. This means estimating the environmental and economic performance of biomass and waste gasification schemes has typically been slow. Previous studies have often considered only one of these aspects, but modern engineering modelling needs to take a more holistic approach, taking into account more relevant factors to modern engineering developments **[39–41]**. However, to this date holistic gasification models combining LCA and TEA do not exist to the best of the author's knowledge, meaning that existing findings are often incomplete. Here, the recent advances in computational power and ML would allow for new types of models which can rapidly model complex engineering systems with greater confidence and predict parameters relevant for further environmental and economic analysis. By implementing methods to account for uncertainties within such a modelling framework, a model's results could be judged much more accurately. Considering these factors, when turned into decision support tools for policy and decision-makers, these types of holistic models could aid with accelerating the development of new technologies, such as biomass and waste gasification, and encourage their widespread commercialisation and industrial-scale implementation.

1.5 Research Aims and Objectives and the Contribution of this Thesis

The overarching goal of this thesis is to establish a comprehensive biomass and waste gasification modelling framework to holistically assess the environmental and economic performance of a gasification scheme. As a result, our knowledge of gasification for bioenergy from biomass and waste (and its process performance and efficiency) can be improved, encouraging its widespread commercialisation and industrial-scale implementation. This is accomplished by addressing the current gaps in the scientific literature discussed in Section 1.4 and may be broken down into several sub-goals.

The initial aim of this thesis was to create a comprehensive and cohesive data set focusing on the biomass and waste gasification process, specifically tailored to generalise

across a diverse array of feedstock types and gasification conditions. This data set can then be used to develop ML models which can predict gasification outputs across a wide range of factors, such as feedstock types (woody biomass, herbaceous biomass, plastics, municipal solid waste, and sewage sludge), gasifying agents (air, steam, oxygen, and combinations of these), and reactor options (fixed bed and fluidised bed).

The developed ML models must meet three key criteria: (i) accuracy – to ensure the models are reliable and their predictions are trustworthy; (ii) comprehensiveness – to ensure the models generalise to a wide range of gasification systems and feedstocks, and to make them useful for further system modelling (e.g. LCA and TEA); (iii) interpretable – to ensure the models' predictions can be understood and accurately communicated to stakeholders. Robust methods for finding the best hyperparameters are necessary to ensure the models' optimal performance with regard to the three above-mentioned criteria.

After the development of suitable ML models, the aim was to facilitate further environmental and economic system modelling through the use of these models. The aim was to establish LCA and TEA models within a larger uncertainty accounting framework. By accounting for all uncertainty within a modelling framework, the reliability and robustness of predictions can be maximised. Additionally, this provides insight into the limitations of the models and allows for more informed decision-making under uncertainty. To enhance the usability of the developed modelling framework a further aim was the development of a graphical user interface (GUI), enabling non-technical decision-makers to utilise the modelling tool seamlessly. Additionally, by incorporating a simple multi-objective optimisation routine, the tool can effectively be used to facilitate case studies and the comparison of various parameter choices in the face of potential trade-offs.

The various components (which are informed by the aims of the thesis) making up the complete modelling framework and their interactions are illustrated by Figure 1.2. A final objective of this work is to demonstrate how similar tools could be developed for the simple and fast assessment of other types of bioenergy or renewable energy in general.

1.6 Thesis Structure

The structure of this thesis is as follows: This chapter (Chapter 1) introduces central concepts and themes related to the thesis. It then introduces the main aims and objectives of the research and explains their importance. Chapter 2 reviews and critically evaluates existing research related to the thesis. Chapter 3 presents an artificial neural network (ANN) to predict biomass and waste gasification outputs. The following chapter (Chapter 4) then explores alternative ML methods to improve upon the ANN model developed in the previous chapter. Interpretability methods are utilised to better understand the inner workings of the developed ML models and improve decision-makers' trust in the models.

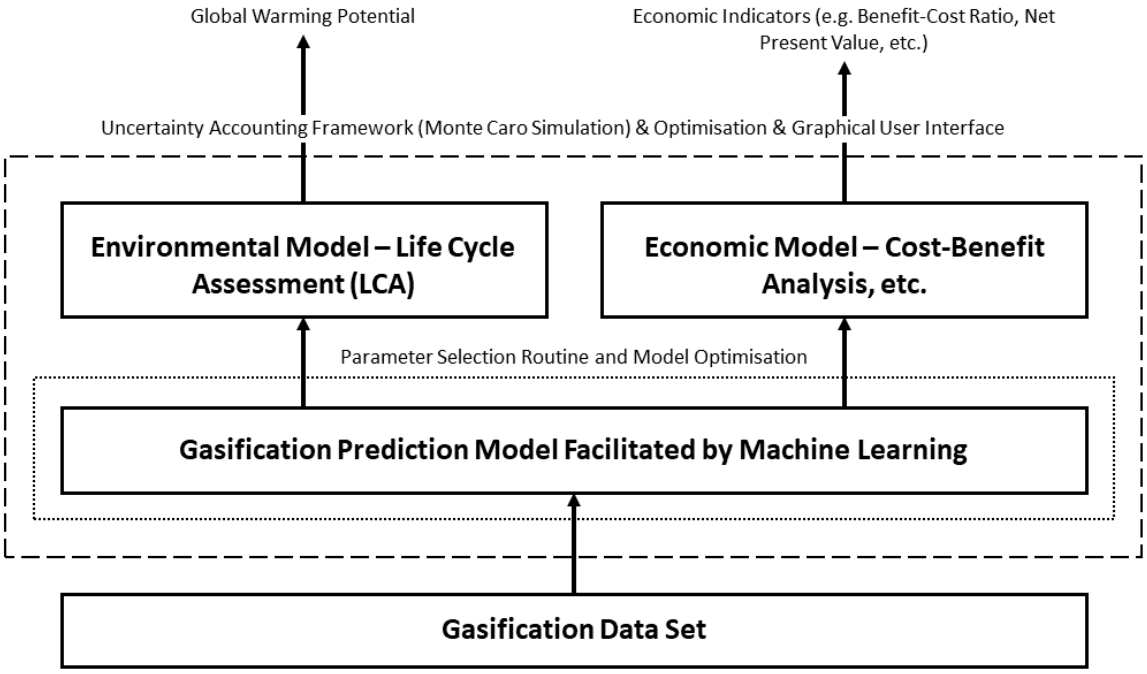


Figure 1.2: Illustration of the different components making up the developed biomass and waste gasification modelling framework and their interactions.

Chapter 5 explores a holistic biomass and waste gasification modelling framework built upon the previously developed ML models. It incorporates LCA and TEA to assess a gasification scheme's environmental and economic feasibility, respectively. A robust uncertainty accounting framework founded in Monte Carlo simulation and an optimisation method is incorporated into the greater modelling framework. Chapter 6 illustrates how the developed modelling framework can be used to construct a GUI-based decision support tool and exemplifies its usefulness through a case study. Finally, Chapter 7 synthesises key findings, acknowledges constraints, and outlines potential avenues for further research.

1.7 Published Work as Part of this Thesis

A portion of this work has previously been published in academic journals, as shown in the List of Publications. All technical work related to the published papers has been done by the author and the writing is their own, except for aid in the form of proofreading, reviewing, and editing. Chapters 3 and 4 both appeared as papers in similar formats, with Chapter 4 incorporating some extensions/corrections to the original paper. Parts of Chapter 2 have appeared as a paper, with the chapter expanding upon the contents of the paper. The specific contributions of papers to chapters are listed below:

- Chapter 2:
  Simon Ascher, Ian Watson, and Siming You. "Machine learning methods for modelling the gasification and pyrolysis of biomass and waste". In: *Renew. Sustain. Energy Rev.* 155 (2022), p. 111902. DOI: 10.1016/J.RSER.2021.111902

- Chapter 3:
  Simon Ascher, William Sloan, Ian Watson, and Siming You. "A comprehensive artificial neural network model for gasification process prediction". In: *Appl. Energy* 320 (2022), p. 119289. DOI: 10.1016/j.apenergy.2022.119289

- Chapter 4:
  Simon Ascher, Xiaonan Wang, Ian Watson, William Sloan, and Siming You. "Interpretable machine learning to model biomass and waste gasification". In: *Bioresour. Technol.* 364 (2022), p. 128062. DOI: 10.1016/j.biortech.2022.128062

Chapter 2

Literature Review

2.1 Chapter Overview

This chapter lays the groundwork for the research conducted in this thesis. Its purpose is to introduce key ideas and concepts that are relevant to this work and to identify gaps in knowledge that this research aims to address. Some topics critical to the understanding of the thesis may be well-established and have previously been reviewed by other researchers. As such, these have been reviewed in less detail than Section 2.3, which is of particular importance to this work. This section focuses on the use of machine learning (ML) methods to model thermochemical processes – a crucial aspect of this work and a primary source of its novelty.

In addition to the extensive review of the use of ML to model thermochemical processes, Section 2.2 explores the various forms of bioenergy, with a particular focus on thermochemical processes, although biochemical processes are also discussed. Section 2.4 chiefly discusses the environmental and techno-economic analysis of thermochemical processes. It also delves into the development of more holistic and comprehensive models, particularly models including other aspects such as social factors and ones incorporating uncertainty accounting approaches such as Monte Carlo methodology. Finally, Section 2.5 summarises the chapter and gives some concluding remarks.

2.2 Bioenergy

Bioenergy refers to energy derived from biomass (i.e. organic material). Biomass can include various organic materials such as wood, agricultural residues, organic waste, and dedicated energy crops. As previously shown in Figure 1.1, a large range of bioenergy conversion pathways exist. The two main pathways, thermochemical and biochemical, are discussed in more detail in Section 2.2.2 and 2.2.1, respectively.

As previously mentioned in Chapter 1, bioenergy is expected to play a critical role in

achieving the climate goals set out by the Intergovernmental Panel on Climate Change (IPCC). Whilst its role is expected to be critical, it will likely play a transient role in our path to a net-zero future, leading to the role of bioenergy declining after around 2050. This is mainly due to three main drivers outlined by Reid *et al.* (2020). Firstly, the cost of competing technologies is expected to reach the same levels or lower than their bioenergy alternatives. Secondly, particularly land-intensive bioenergy will become more costly in the future with land becoming a more limited resource. Thirdly, bioenergy will lose its position as a relatively well-established and cheap energy option, due to the cost of other alternatives declining as they scale. This makes it important to critically evaluate the impacts of land-intensive bioenergy to avoid potential problems such as a decrease in food security and loss of habitat and biodiversity **[42]**.

As land-intensive bioenergy directly scales with the use of land, which is fundamentally limited in supply, other types of biomass, such as byproducts, residues, and waste materials have been seen as a promising alternative. These types of biomass avoid land use changes and can improve the efficiency of existing systems and supply chains whilst effectively promoting more circular economies. However, waste and byproducts come with several challenges. For one, they can lead to more complexity, as they are inherently related to other systems. This also means that they may not be as scalable in the short term as more dedicated energy crops. Additionally, waste materials and byproducts can be contaminated or less well suited for bioenergy developments due to other reasons, leading to the need for more advanced feedstock pretreatment **[18, 42–44]**.

Although the electricity sector has rapidly transitioned towards renewable alternatives, the transport and heating sectors have proven to be more challenging when it comes to reducing their environmental impact. Some of the transport sector (e.g. trains and cars) is predicted to be mostly electric in the near future. However, particularly aviation and shipping are expected to be more difficult to electrify. For aircraft, the energy density of their fuels is essential, with current jet fuels having an energy density nearly 50 times larger than that of modern batteries. This makes advanced biofuels derived from biomass a promising option to minimise the carbon impact of aviation. In fact, according to the International Energy Agency's (IEA's) predictions, advanced biofuels will comprise 70% of aviation fuel by 2060. For shipping advanced biofuels are projected to make up 50% of the fuel mix by the same year **[42]**.

As previously mentioned in Section 1.2, particularly bioenergy with carbon capture and storage (BECCS) is heavily relied upon to limit global warming to the desired 1.5°C or 2°C goal **[14, 15]**. This is the case despite BECCS largely remaining an unproven technology at scale **[45]**. However, researchers have pointed out that BECCS needs to be deployed early to ensure it can effectively aid in limiting global warming. For instance, Hanssen *et al.* (2020) estimated the global CO_2 sequestration potential of BECCS to be

2.5 $Gt\ CO_2$ annually over a 30-year time horizon **[46]**. This number increases to 40 $Gt\ CO_2$ annually over an 80-year time horizon, indicating the importance of rapid research and development, allowing for the industrial scale deployment of BECCS. Additionally, it is important to consider factors such as geospatial conditions, original land use, and the exact bioenergy technology to ensure that BECCS effectively aids with mitigating climate change.

There are three main pathways for carbon capture: post-combustion capture, pre-combustion capture, and oxy-fuel combustion capture. During post-combustion capture, CO_2 is captured from the flue gas produced during the direct combustion of biomass, or the combustion of intermediate energy-carriers like syngas. Typically, a chemical solvent such as amine is used to capture CO_2 from the flue gas stream. Once captured, the CO_2 is separated from the solvent, compressed, and transported for storage. A key benefit of post-combustion capture is that it can be retrofitted to existing plants. Pre-combustion capture involves the separation of CO_2 from syngas before its combustion and often utilises pressure swing adsorption for the capture of the CO_2. Finally, oxy-fuel combustion capture uses a pure oxygen environment which leads to the production of a pure CO_2 stream. It also avoids the production of nitrogen oxides (NO_x) which are harmful to human health and can contribute to the formation of acid rain and smog **[47]**.

2.2.1 Thermochemical Methods

One can distinguish between direct and indirect thermochemical conversion approaches. During combustion, the energy contained in the biomass is directly converted into heat and electricity. Indirect processes, such as pyrolysis and gasification, produce three main products at different rates: bio-oil, syngas (also referred to as synthesis gas or product gas), and biochar **[19, 23]**. Bio-oils and syngas can be further converted to biofuels like bioethanol, whereas biochar represents a stable form of carbon, with varying properties depending on the exact feedstock, and conversion technology and conditions.

Other thermochemical methods include co-firing with other fuels (typically fossil fuels), hydrothermal liquefaction, hydrothermal carbonisation, and torrefaction **[48]**. However, this review focuses on gasification and pyrolysis as they are most relevant to the work presented here.

Biochar is a highly porous, carbon-rich material produced during the pyrolysis or gasification of organic materials. Depending on the exact feedstock, and treatment type and conditions, biochar with a variety of properties and potential applications can be produced. The primary use case for biochar is for agriculture and soil management. When incorporated into soil, biochar can improve soil fertility, water retention, and nutrient availability. It also enhances soil structure and microbial activity, which can lead to increased crop yields and improved resilience to drought and other environmental stresses. Additionally,

biochar can sequester carbon from the atmosphere, helping to mitigate climate change by locking carbon in the soil for potentially hundreds to thousands of years **[49–52]**.

To ensure the long-term environmental benefits of biochar for soil application, the biochar's stability must be ensured. One key factor affecting biochar's stability is the treatment temperature (i.e. gasification or pyrolysis temperature), with higher temperatures typically improving the biochar's stability **[53, 54]**. Matustik *et al.* (2022) pointed out that whilst environmental studies analysing the effects of applying biochar to soil have generally found that biochar leads to climate change mitigation, methodological decisions can have a significant effect on the results. For this reason, it is essential for researchers to clearly state the choices and assumptions made during environmental analysis, such as life cycle assessment (LCA), to ensure results remain accurate and can be compared on a fair basis **[54]**.

Beyond agriculture, biochar has also been explored for applications such as water filtration, environmental remediation, fuel cells, and as a catalyst. Its porous structure and high surface area make it effective at adsorbing pollutants from water and air. In particular, biochar engineered for maximum surface area could be turned into a high-value product for these types of applications and make it comparable to activated carbon products **[49–52, 55]**.

One of the main reasons why biochar is yet to be commonly used on a large scale is the lack of information on its economic feasibility. Campion *et al.* (2023) found that the profitability of biochar production is highly uncertain, mainly due to factors such as feedstock, scale, location, process choice and conditions, etc. In particular, externalities have not been satisfactorily addressed in the literature so far. Better addressing these factors and making case studies case-specific will improve the trust in the technology and lay a foundation to develop policies encouraging the uptake of biochar technologies **[26]**.

Whilst biochar has typically been associated with the pyrolysis process, more recently gasification has been considered as another promising biochar production pathway. Gasification may be often optimised for the production of syngas, but this does not detract from the fact that biochar has the potential to significantly improve the feasibility of gasification schemes. The primary benefit of producing biochar from gasification is the increased process efficiency, as in comparison to e.g. traditional charcoal-making the produced gas is fully utilised leading to a high overall process efficiency. Furthermore, the higher temperature during gasification, in comparison to pyrolysis, can lead to a more stable product, improving its climate change mitigation potential. In the future, thermochemical processes will likely have to balance economic and environmental considerations. By focusing on either syngas (economic focus) or biochar production (environmental focus), gasification is a promising technology to offer that flexibility, depending on factors such as market demand, stakeholder priorities, feedstock availability, and local conditions **[51, 55–58]**.

2.2.1.1 Pyrolysis

During the pyrolysis process, the feedstock is converted into bio-oil, biochar, and gases by breaking down larger hydrocarbon molecules into smaller ones **[24]**. The process temperature, heating rate, and residence time are key parameters determining product yields. Generally, higher temperatures (> 600°C) favour gas production due to the dominance of secondary cracking reactions **[59]**. Bio-oil production generally peaks at temperatures in the 400 to 550°C range **[60]**.

Another essential parameter is the process' heating rate. In combination with the residence time, it defines the type of pyrolysis, i.e. flash, fast, or slow pyrolysis. Gases and bio-oil are predominantly produced at high heating rates and low residence times. Slow pyrolysis, on the other hand, characterised by an increase in residence time and a decrease in heating rate leads to an increase in the yield of biochar. The influence of the reaction conditions has previously been reviewed in greater detail by Kan *et al.* (2016) **[60]**.

Typical product yields for the pyrolysis and gasification of wood have previously been summarised by Bridgwater. Product yield splits for liquids, solids, and gases of 75/12/13, 30/35/35, and 5/10/85 have been found for fast pyrolysis, slow pyrolysis, and gasification, respectively. Here, fast pyrolysis occurred at around 500°C with a hot vapour residence time of approximately 1 s. Slow pyrolysis on the other hand took place at around 400°C with a residence time in the range of days. Finally, a gasification temperature of 700 to 900°C was considered **[61]**.

Recent research efforts in the field of biomass and waste pyrolysis focused on improved process and reactor designs as well as product upgrading and applications. For instance, work on the efficient and cost-effective upgrading of bio-oil is essential in making the process commercially viable **[62]**. Additionally, turning biochar into a high-value product may significantly help with its economic feasibility. One such product may be activated biochar with properties engineered for adsorption, catalysis, and wastewater treatment applications **[63]**. Furthermore, work on improved pyrolysis methods and reactor designs continues and novel approaches include microwave-assisted and solar pyrolysis **[63–65]**. In particular, microwave-assisted pyrolysis might be a valid option to resolve the drawbacks of conventional pyrolysis, making it an environmentally friendly method to produce biofuels from different municipal waste streams **[65]**.

2.2.1.2 Gasification

In general, gasification aims to convert a carbonaceous feedstock into a gas product, often called syngas or producer gas, of a usable heating value. Additionally, gasification produces biochar (typically in smaller quantities than pyrolysis) as a byproduct. Syngas can either be used for energy generation via an internal combustion engine, turbine, or CHP plant or be upgraded to value-added chemicals **[24]**. Gasification typically involves the following

four processes: (i) drying, (ii) pyrolysis, (iii) partial combustion and tar cracking, and (iv) reduction.

Air, steam, oxygen, or a combination of the three have successfully been used as the gasifying agent, which facilitates the process. Air is commonly employed in small or medium-sized plants due to its inexpensiveness. Due to the high nitrogen concentration in air, the produced syngas is diluted with nitrogen, resulting in a relatively low heating value syngas of 4 to 7 MJ/m^3. The use of steam or oxygen as a gasifying agent results in greater capital and operating costs but increases the heating value of the syngas to 10 to 18 MJ/m^3 and 12 to 28 MJ/m^3, respectively **[24]**. Steam improves the heating value by promoting the formation of H_2, resulting in a final gas with a higher H/C ratio. Using oxygen yields a gas high in CO and CO_2 with a much smaller proportion of H_2.

Gasifiers come in several different designs, depending on the intended feedstock, scale of production, and desired product properties. Fixed bed gasifiers are arguably the simplest type of gasifier. They are relatively forgiving when it comes to feedstock quality, as they can tolerate high ash and moisture content feedstocks of relatively large particle sizes. Tar accumulation is typically no issue for fixed bed gasifiers, but their design inherently limits their maximum size. The gasifiers' complexity increases for fluidised bed and entrained flow gasifiers. Their design makes them suitable for large-scale applications (which typically treat tens to hundreds of tonnes of feedstock per day) and they typically have a relatively high process efficiency. However, high tar production tends to be a problem for fluidised bed gasifiers, necessitating additional gas cleaning equipment. Entrained flow gasifiers have long been used by the coal industry, but adapting them to biomass has been challenging, as inhomogeneous and high moisture content feedstocks can cause problems during operation. More recently, plasma gasification has received significant attention. It entirely eliminates the tar generation problem and can utilise a vast array of feedstocks to produce very energy-dense syngas. However, the technology is still young and hence relatively costly and complex **[18, 66, 67]**.

Figure 2.1 shows the prevalence of different gasifier designs in the data set collected for the development of ML models, as described in Chapters 3 and 4. From the figure, it can be seen that fluidised bed gasifiers were the most common type of gasifier studied in the academic literature. However, fixed bed gasifiers have also received significant attention, especially at lab scale. Most gasifiers were of pilot scale, highlighting the pre-commercial status of the technology.

Significant research efforts in the field of gasification have been focused on the production of biofuels **[68]**. Hydrogen production by gasification has also been of interest to researchers **[69]**. To ensure industrial-scale deployment, efforts have been made to identify the optimal operating conditions which allow for efficient and economic production of hydrogen **[70]**. Other ongoing research is closely linked to aiding industrial-scale deploy-

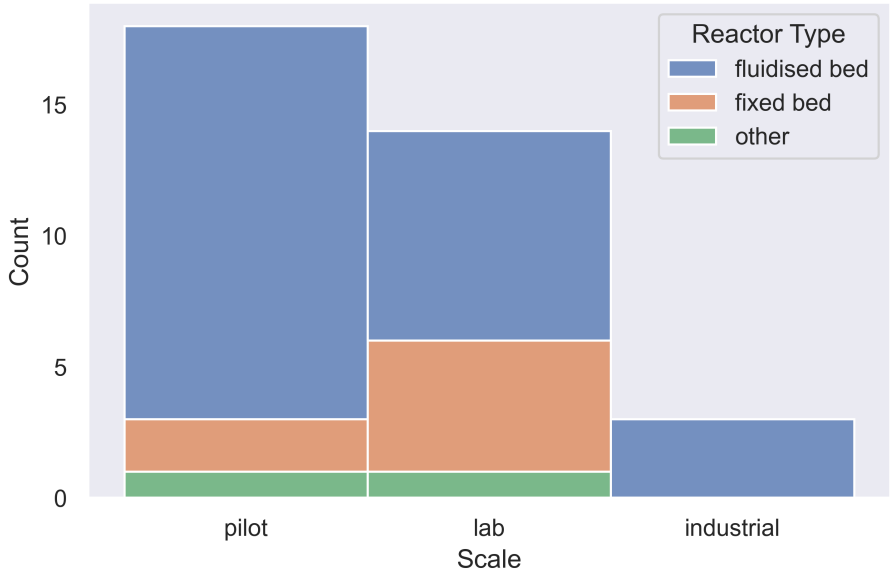


Figure 2.1: Prevalence of different reactor types at different gasifier sizes, based on the data set used in Chapters 3 and 4 (and as published in **[1, 2]**). Note: lab $<$ 100 kW_e (150 kW_{th}); 100 kW_e (150 kW_{th}) $\leq$ pilot $<$ 1 MW_e (1.5 MW_{th}); industrial $\leq$ 1 MW_e (1.5 MW_{th}). Data from literature sources published between 1992 and 2018.

ment. For instance, tar reduction and its removal are major technical issues, as tar can cause significant problems to downstream equipment **[71, 72]**. In-situ or post-gasification tar removal is often expensive. Therefore, the use of catalysts, which are still potentially costly, to produce cleaner syngas more suitable for industrial-scale deployment has been explored **[71]**.

2.2.2 Biochemical Methods

Biochemical conversion processes use microorganisms or enzymes to convert biomass into a usable form of energy. Three main processes have been the focus of research, namely anaerobic digestion (AD), fermentation, and photobiological methods which typically use microalgae to produce hydrogen **[18]**. AD has established itself as an effective technology for the management of agricultural and urban waste, however, more recently it has also been getting attention in a biorefinery context **[73]**. Fermentation has been used at scale to produce ethanol, which is commonly used as an additive to vehicle fuels **[48]**. Finally, photobiological methods have the potential to efficiently produce biohydrogen **[18, 74]**.

AD is a biological process where microorganisms break down organic matter, such as food waste or animal manure, in the absence of oxygen. This process produces biogas, which is primarily composed of methane (CH_4) and carbon dioxide (CO_2), and digestate. The generated biogas can either be directly combusted for heating purposes or electricity generation by using an engine, turbine, or combined heat and power (CHP) plant. It can

also be upgraded to biomethane, which can be used as a natural gas substitute, vehicle fuel, or platform chemical. Digestate, a byproduct of the AD process, describes the nutrient-rich residue produced during the process. It is the solid and liquid materials left behind in the reactor vessel and is rich in organic matter, nitrogen (N), phosphorus (P), and other essential nutrients, making it a valuable fertiliser **[73, 75]**.

Despite the technology's maturity, AD still faces several challenges limiting its widespread application. In theory, any biodegradable material can be utilised for AD, but in practice feedstock variability causes significant issues. Feedstocks vary drastically in terms of physical and chemical properties, making it challenging to select the most suitable digester technology. This can make it challenging to maintain optimum process conditions, leading to reduced process performance. In fact, low process efficiency can be a problem in general. During digestion, the accumulation of several intermediate products, such as Ammonia (NH_3) and volatile fatty acids, can inhibit the process and in the worst case lead to digester failure. Finally, the relatively low quality of the produced biogas (biogas typically has a CO_2 content of around 40%) limits its application for direct utilisation **[73]**. Current research focuses on tackling these challenges through improved pretreatment methods, co-digestion, new biorefinery concepts, and the use of machine learning (ML) to control the process and maintain process stability. **[73, 75–77]**.

Fermentation utilises microorganisms (typically yeasts) to convert simple sugars to alcohol and CO_2. Sugar crops like sugar cane and sugar beet, or starch crops like maize, are common feedstocks utilised at large scale. The produced ethanol is purified through distillation which adds a significant energy expenditure and contributes to the process' relatively low net efficiency. Brazil might be the most notable example of countries successfully producing fuel from biomass, with many cars running on 100% bioethanol. The sugar cane utilised in Brazilian biorefineries is the most efficient feedstock due to its high sugar content and fuel yield per hectare **[48]**. However, this example effectively illustrates the "food vs fuel" problem within bioenergy, which refers to the ethical and practical dilemma of diverting agricultural resources, such as crops or arable land, from food production to fuel production, potentially exacerbating food scarcity and raising food prices. The current research focuses on the use of fermentation within a biorefinery context, the fermentation of lignocellulosic feedstocks, and the use of fermentation for biohydrogen production **[48, 78]**.

Photobiological bioenergy methods harness the ability of photosynthetic organisms (typically microalgae) to convert sunlight and carbon dioxide into hydrogen. Some types of biomass can produce H_2 during photosynthesis by splitting water molecules into their elemental components. For this process to continuously occur, anaerobic conditions need to be present. The two main pathways to achieve this are to either simultaneously produce H_2 and O_2, or to first culture microalgae and then promote the production of H_2

under sulphur-deprived conditions. The main challenge of this technology is its scalability. Hence, recent research efforts are mainly focused on scaling the technology and getting it to a commercial level **[18, 74]**.

2.3 Machine Learning Methods to Model Thermochemical Processes

2.3.1 Section Summary

Over the past two decades, the use of machine learning (ML) methods to model biomass and waste gasification/pyrolysis has increased rapidly. Only 70 papers were published in the 2000s compared to a total of 549 publications in the 2010s (according to a keyword search of "Web of Science"). However, the approaches and findings have yet to be systematically reviewed. In this work, the machine learning methods most commonly employed for modelling gasification and pyrolysis processes are discussed with reference to their applications, merits, and limitations. Whilst coefficients of determination (R^2) can be difficult to compare directly, due to some studies having vastly different approaches and aims, it has generally been found that most studies consistently achieved a high prediction accuracy with $R^2 > 0.90$. Artificial neural networks have been most widely used due to their potential to learn highly non-linear input-output relationships. However, a variety of methods (e.g. regression methods, tree-based methods, and support vector machines) are appropriate depending on the application, data availability, model speed, etc. It is concluded that ML has great potential for the development of models with greater accuracy. Some advantages of machine learning models over existing models are their ability to incorporate relevant non-numerical parameters and the power to generate a multitude of solutions for a wide range of input parameters. More emphasis should be placed on model interpretability to better understand the processes being studied.

2.3.2 Introduction and Background

Thermochemical methods have been extensively investigated for their great potential to convert biomass and waste into bioenergy and serve as a valid waste treatment option. Generally, one may differentiate between the following five conversion pathways: combustion, gasification, pyrolysis, carbonisation/torrefaction, and liquefaction **[24]**. This work focuses on gasification and pyrolysis due to their widespread applications and significant environmental benefits.

Gasification and pyrolysis represent cost-effective ways to recover energy from biomass and waste feedstocks **[79]**. Whilst combustion is well suited for the direct generation

of electricity and heat, pyrolysis and gasification can produce a wider range of products such as liquid transport fuels, bulk chemicals, and biochar. These products may serve to effectively reduce the carbon footprint of relevant development and offer greater flexibility towards demand management. Transport fuels, such as biodiesel, may reduce the environmental impact of the transport sector when appropriate production strategies are employed **[80–82]**. Whereas biochar represents a solid form of carbon which, along with carbon capture and storage (CCS), has the potential to turn bioenergy into a carbon-negative industry (e.g. upon soil application) **[83–85]**.

Dependent on factors such as capacity and feedstock, gasification-based systems have been shown to outperform combustion-based systems when considering their energy recovery, conversion efficiency, and potential to generate revenue **[86]**. For instance, one study found that municipal solid waste (MSW) gasification resulted in a net energy production to the grid of 685 kWh/tonne MSW as compared to 544 kWh/tonne MSW for waste incineration **[86]**. However, this potential increase in efficiency can be associated with greater process complexity **[87]**.

Importantly, gasification and pyrolysis also have the potential to environmentally outperform the technologies based on waste combustion processes **[88, 89]**. One study comparing four commercial plants found that the environmental impact of gasification was lower than incineration, due to improved energy efficiency and syngas cleaning technology **[88]**. Environmental performance is expected to improve even more as these technologies develop and further improvements to, for instance, syngas cleaning are made.

The great complexity of thermochemical processes calls for the development of mathematical models for process design, optimisation, and intensification. Models can be used to identify optimal operating conditions of a system and study the effects of different parameters on the outputs. Models can also be used to achieve real-time process control and optimisation and aid with the scale-up of a system or initial system sizing **[24]**.

Figure 2.2 shows how the research interest in modelling techniques for the gasification and pyrolysis process has increased over the last two decades. A rapid increase in publications can especially be observed for the second half of the 2010s. Data was obtained from the "Web of Science" database using the keywords model/modelling/modeling/simulation, waste/biomass, and the respective thermochemical process (i.e. gasification or pyrolysis) **[90]**.

A range of approaches have been used to model the complex chemical reactions, as well as mass and heat transfer inside the reactor, during gasification and pyrolysis. Conventionally, thermodynamic equilibrium, kinetic, and computational fluid dynamics (CFD) models have been the main focus. However, more recently machine learning (ML) techniques, such as artificial neural networks (ANNs), have received increasing attention due to their potential to efficiently model these highly complex processes.

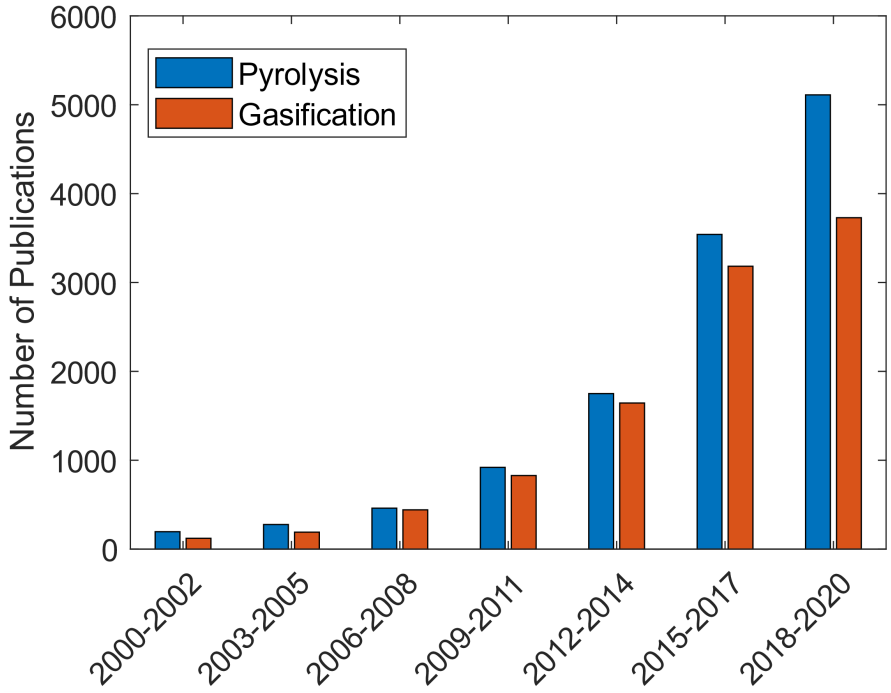


Figure 2.2: Research interest in pyrolysis and gasification modelling throughout the years.

Thermodynamic equilibrium models are effective at determining the theoretical limits of a process **[91]**. The main drawbacks of equilibrium models stem from the high number of assumptions made and include a lack of accuracy when it comes to modelling the effects of specific reactor designs. Furthermore, they are not well suited for conditions which are dissimilar to equilibrium conditions (e.g. low-temperature gasification) **[92]**. Kinetic models can improve the performance of equilibrium models by considering the reactor design. Meanwhile, this also limits them to a specific reactor design **[91, 92]**. Similar limitations apply to CFD models. Additionally, the high complexity of CFD models results in high computational demands.

ML models represent a different approach. Where conventional models try to accurately represent the physical and chemical processes occurring within a reactor, ML models are generally not concerned with the actual processes. Instead, an attempt is made to map inputs to the corresponding outputs. It is left up to the employed algorithm to understand the underlying mathematical relationships resulting from a vast number of highly complex processes. In this field, it has the potential to more accurately model thermochemical processes and better predict their outputs than conventional models **[93, 94]**.

Most gasification review papers have a strong focus on thermodynamic equilibrium models and kinetic rate models **[91, 92, 95–97]**, with Aspen Plus models **[95]** and CFD models **[96, 97]** being reviewed to a lesser extent. ANN methodology was the only ML technique receiving limited attention in past review papers **[91, 95–97]**. Moreover, ANN modelling was generally mentioned as a side note and not reviewed in greater detail. For

example, Baruah and Baruah reviewed only two studies using ANNs to model gasification **[96]**. One of the reviewed studies showed that an ANN was able to better predict gasification products than a real-gas equilibrium model which incorporated tar, char, and permanent gas reactions **[98]**. It is important to highlight that other relevant ML options, such as support vector machine (SVM) and random forest (RF) approaches, have not been considered by any of the mentioned reviews **[91, 92, 95–97]**.

A recent 2020 review by Kostetskyy and Broadbelt (2020) discusses the latest advances in biomass fast pyrolysis modelling **[99]**. A large proportion of this review focused on conventional approaches, such as atomistic and mechanistic techniques, to model the pyrolysis process. The authors included a section discussing recent advances in ML techniques and their potential applications within the wider field of chemical engineering. The successful implementation of other ML methods, such as SVM and RF approaches, as well as advanced neural networks, such as adaptive neuro-fuzzy inference systems (ANFIS) and convolutional neural networks (CNN), has been acknowledged **[99]**.

On top of Kostetskyy and Broadbelt's review, several other closely related review papers have been published throughout the years. For example, Hameed *et al.* (2019) compared kinetic, mechanistic, and network/structural models **[100]**. Kaczor *et al.* (2020) paid special attention to how chemical reactions were implemented in different approaches and focused on solutions appropriate for subsequent CFD implementation **[101]**. In an earlier work, Sharma *et al.* (2015) focused on kinetic, particle, and reactor-based models, as well as the effects of process parameters and catalyst use on the pyrolysis process **[102]**. Finally, Babu's 2008 review focused on mathematical techniques for the estimation of kinetic parameters to model biomass pyrolysis **[103]**. From this, it becomes clear that ML-based techniques for pyrolysis modelling have received little attention, apart from Kostetskyy and Broadbelt's review.

In conclusion, a lack of a systematic review on the application of ML to model biomass and waste conversion via thermochemical methods, such as gasification or pyrolysis, is evident. Hence, this work aims to provide a clear picture of the current state of the art of these methods within the field. To the author's knowledge, this work is the first review considering a wide range of ML methods for modelling gasification and pyrolysis.

This review is structured as follows: Section 2.3.3 represents a brief introduction to numerous popular ML techniques which have been used in the field, and their merits and relative advantages and disadvantages are compared. Section 2.3.4 takes a closer look at how researchers have used ML to model thermochemical processes. Section 2.3.5 outlines the current challenges in the successful deployment of ML and how they may be overcome before Section 2.3.6 concludes the findings of this work.

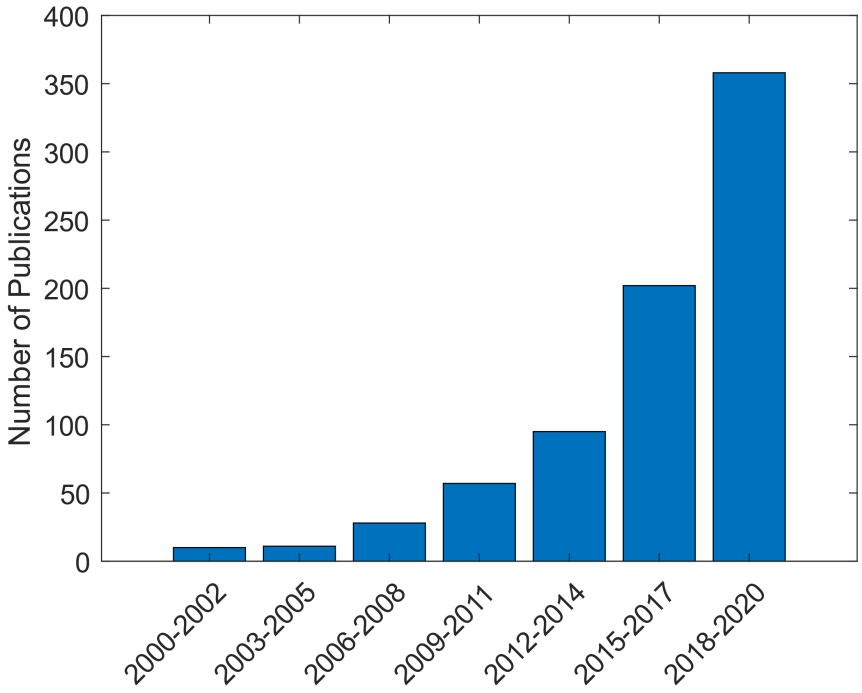


Figure 2.3: Prevalence of machine learning use for gasification and pyrolysis modelling applications.

2.3.3 Machine Learning Techniques

Four ML methods have been commonly employed to model pyrolysis and gasification processes: regression analysis, ANNs, tree-based methods such as RF, and SVMs. Sections 2.3.3.1, 2.3.3.2, 2.3.3.3, and 2.3.3.4 provide a brief introduction to these methods.

Figure 2.3 shows the rapid increase in publications applying ML methods to model gasification and pyrolysis processes. The Web of Science search keywords used to obtain the data illustrated by Figure 2.3 were the following: TS = (gasification OR pyrolysis) AND TS = (waste OR biomass) AND TS = ("machine learning" OR "regression" OR "deep learning" OR "neural network" OR "perceptron" OR "ANFIS" OR "autoencoder" OR "restricted Boltzmann machine" OR "support vector machine" OR "support vector regression" OR "k-nearest neighbor" OR "k-nearest neighbour" OR "genetic algorithm" OR "gene expression programming" OR "decision tree" OR "regression tree" OR "random forest" OR "gradient boosting"). Here "TS", "OR", and "AND" are keywords used in the search engine, where "TS" stands for topic search and "OR" and "AND" have their usual meaning.

2.3.3.1 Regression Methods

Regression analysis can be used to assess the relationship between dependent (target) variables and independent variables (inputs). The simplest form of regression analysis is

linear regression for which a straight-line equation is fitted to describe the relationship between one dependent and one independent variable. This is generally done using the least squares method **[104]**.

In addition to linear regression, many other types of regression exist with various applications. For example, logistic regression is well suited for discrete problems whereas shrinkage methods such as ridge regression can make sense of highly correlated independent variables. Polynomial regression may be used to fit a nth-degree polynomial to data that does not follow linear relationships. Kernel smoothing methods represent another option which splits the domain into segments and locally fits several simple models. Combining these simpler models will in turn yield a fit to the overall data set **[104]**.

Whilst other ML methods, such as ANNs, are often considered a black box, regression analysis generally allows for easy model interpretation. For instance, linear regression analysis clearly indicates how an independent variable affects a dependent variable. However, the method's simplicity is also its downfall as poor performance must be expected for data following non-linear relationships. More advanced regression techniques, such as polynomial regression, can better fit non-linear data while retaining strong interpretability. However, in this case, model tuning becomes more challenging.

2.3.3.2 Artificial Neural Networks

ANNs are a programming paradigm inspired by the workings of the biological neural network within the human brain. The term neural network describes a larger number of different models and methods. Figure 2.4 shows the simple and widely employed single layer perceptron or single hidden layer back-propagation network.

An ANN is made up of nodes called artificial neurons. Each neuron is connected to other neurons to which it can pass on received signals. Neurons are arranged into different layers, namely an input and output layer, as well as any number of hidden layers (one for a single layer perceptron). Each neuron has an associated weight, as well as a bias for hidden and output layer neurons. In short, the network learns by feeding back its predictions, comparing them to the corresponding inputs, and adjusting weights accordingly **[104]**.

ANNs are not a new idea as the first functional networks were published in the 1960s and the idea itself existed for even longer **[105]**. However, more recently with the increase of processing power ANNs have experienced a renaissance. Their potential to model highly non-linear problems makes them appealing for modelling thermochemical processes. In fact, ANNs can be considered a universal function approximator. This means, in theory, a neural network can be created to model any given function **[106]**.

ANNs have widely been accepted as a powerful tool for the modelling of highly non-linear processes. However, their black box nature and lack of interpretability have been a drawback highlighted by many authors **[93, 107–109]**. Whilst this issue cannot be

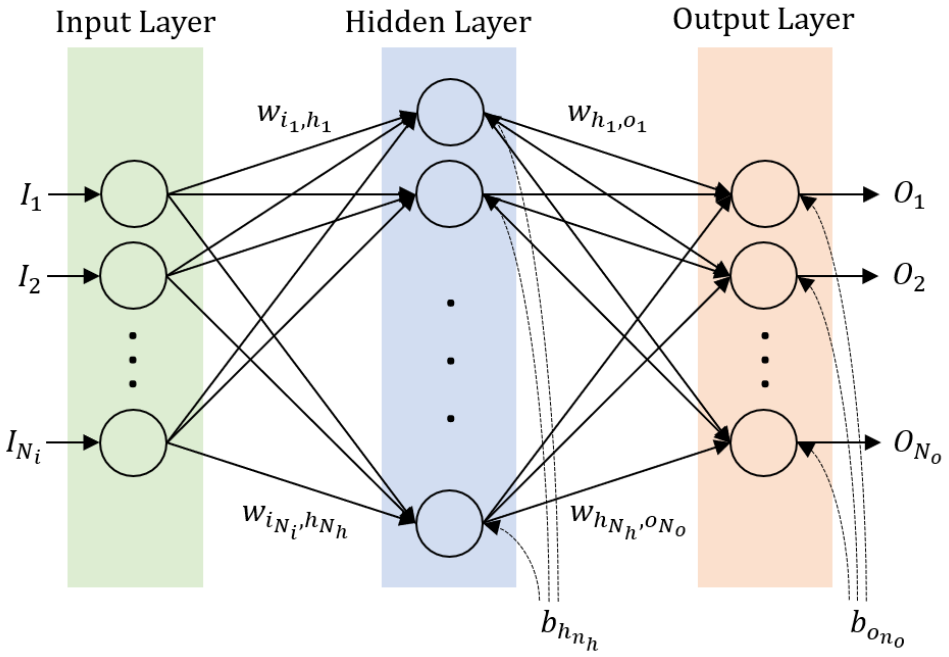


Figure 2.4: Illustration of single layer perceptron (single hidden layer back-propagation network) architecture. I and O represent the perceptron's inputs and outputs, respectively. N_i, N_h, and No indicate the number of neurons in the ANN's input, hidden, and output layer, respectively. Weights between two neurons are denoted by w, where the two subscripts denote the neurons which are connected by the weight (e.g. w_{i_1,h_1} is the weight connecting the first input layer neuron with the first hidden layer neuron). Finally, b represents bias terms feeding into hidden and output layer neurons with the subscripts h_{n_h} and o_{n_o} representing each hidden and output layer neuron, respectively.

neglected, significant efforts have been made to create robust methods which improve their interpretability **[110]**. Olden *et al.* (2004) found that the connection weights and partial derivatives methods predicted the true parameter importance better than for instance Garson's algorithm **[110]**. Model-agnostic methods are a flexible way to interpret the output of any ML model. Feature permutation or the use of global surrogate models are two examples of model-agnostic methods **[37]**.

2.3.3.3 Decision Trees and Related Methods

Decision trees are a powerful methodology suitable for regression and classification problems. While growing a decision tree the algorithm splits the feature space into several regions. A split point is determined for each region and the splitting process is repeated for each resulting region. Initially, a large tree structure is grown this way which likely results in overfitting. Here a strategy called pruning comes into play which removes excess branches. This is often done by introducing a cost-complexity criterion which captures the trade-off between the tree's generalisation capability (i.e. the model's ability to adapt properly to new, previously unseen data) and goodness of fit **[105]**.

Many algorithms attempt to improve on individual decision trees by utilising methodologies such as bagging, boosting, or ensemble methods. One such method is called RF. RF methodology creates an ensemble of decision trees. Importantly, each tree is trained on only a fraction of the data set. The data used to train individual trees is randomly chosen. Ultimately, for regression, the outputs of all decision trees are averaged to compute the RF model's final prediction. RF for classification uses majority voting instead for which each tree casts a vote indicating its predicted class. The class with the most votes wins and is returned as the RF model's predicted class **[104, 111]**. A simplified illustration of the decision tree and RF methodology is shown in Figure 2.5

Whilst simple decision trees can be trained rapidly, they are also prone to overfitting. Additionally, the method's hierarchical nature leads to high variance. This means small changes in the training data could result in a very different tree. Ensemble methods using a vast number of trees can help with overcoming this problem. Furthermore, these methods result in models robust to overfitting as many ensemble algorithms have in-built regularisation mechanisms. However, the large number of trees also significantly increases computational demands **[104]**.

Two key ensemble methods are RF and gradient boosting. The fundamental difference between the two methods is that RF builds many large trees, whereas gradient boosting uses many weak learners (i.e. shallow trees which are small trees with a low depth). A well-tuned gradient boosting model tends to outperform an RF model **[104]**. However, tuning can be challenging. RF on the other hand has fewer hyperparameters and generally performs well even for sub-optimal hyperparameter choices.

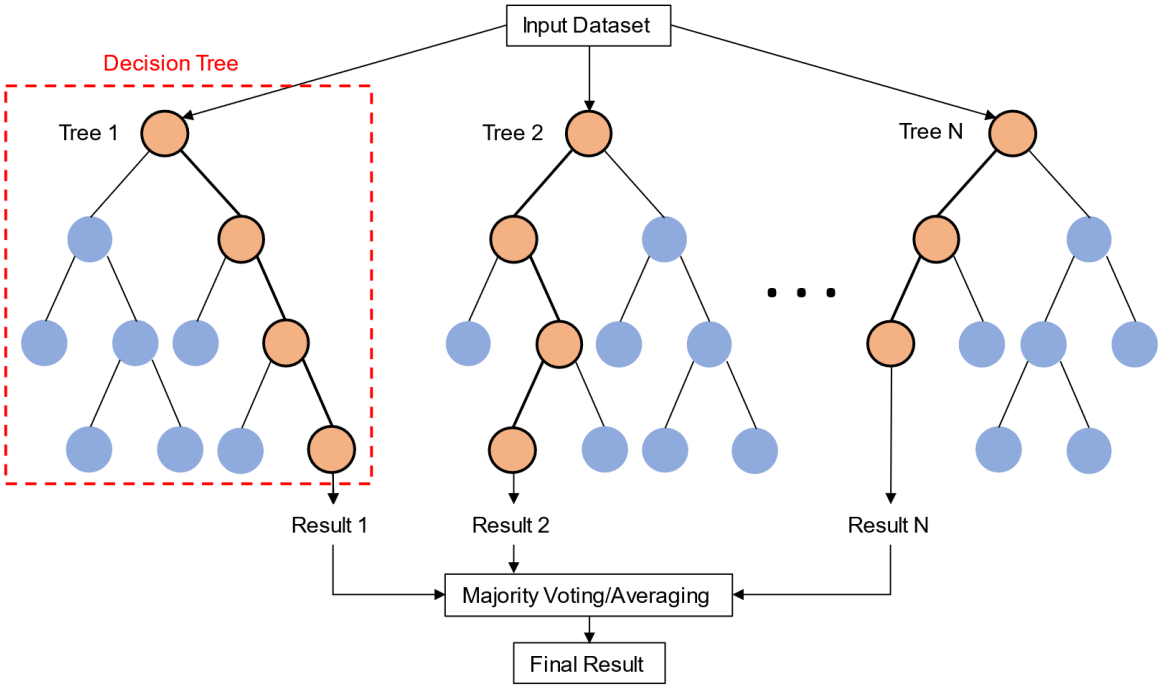


Figure 2.5: Illustration of decision tree and RF methodology. The red box shows an individual decision tree, whereas the complete figure illustrates RF methodology. Branches highlighted in orange show the paths yielding a tree's prediction for a given input. Furthermore, the figure shows how RF averages (majority voting for classification) the results of all trees to calculate the model's final prediction.

Other issues with basic decision trees are the lack of smoothness of their underlying functions and difficulties in capturing additive structures (e.g. noise in the data set). These issues have been addressed by the development of e.g. multivariate adaptive regression splines **[104]**.

In practice, tree-based methodologies are often simpler to implement than other ML methods. Data generally requires minimal pre-processing and especially missing predictors are easily handled due to the nature of the algorithms. Furthermore, the implementation of categorical predictors is straightforward, as these do not need to be turned into numerical values **[104]**. Finally, while ANNs are often considered a black box, tree-based methods allow the user to extract feature importance with relative ease. For example, one can determine how a feature affects splits in the tree structure to determine the relevance of the corresponding feature. This can help with troubleshooting the model, improve trust in the model, and increase understanding of the problem at hand which in turn may allow for model improvements.

2.3.3.4 Support Vector Machine

SVMs are a type of algorithm most commonly used for binary classification tasks, but the methodology is also well-suited for other classification and regression problems. Several applications are illustrated in Section 2.3.4.3.3. Figure 2.6 exemplifies the functional principle of SVMs using a simple example of two linearly separable classes. The training algorithm finds the maximum-margin hyperplane which divides two classes in such a way that the distance between the hyperplane and the nearest data points from either class is maximised. Construction of the hyperplane is achieved by using support vectors – data points from either class closest to the hyperplane.

For applications where data are not linearly separable, soft-margin methodology can be employed by using for instance the hinge loss function. For the hinge loss function, a high cost is associated with incorrectly classified instances. Correctly identified instances incur a low or no cost for instances with a small or large margin from the hyperplane, respectively **[104]**. To extend SVM methodology to non-linear problems, so-called kernel functions can be used to map inputs into a higher dimensional feature space. Kernel functions are a set of mathematical functions that transform the input into the required form. Common choices for kernel functions are nth-degree polynomial or radial basis functions **[104]**. Finally, the concepts illustrated for classification tasks can be extended to regression problems as first illustrated by Drucker *et al.* (1996) **[112]**.

Empirical evidence suggests that SVMs perform well for many real-world applications **[104]**. They have been shown to yield good results in terms of prediction accuracy for data sets of a limited size. Especially high-dimensional data sets make SVMs a good option **[113, 114]**. Another advantage of SVMs is their low sensitivity to outliers which

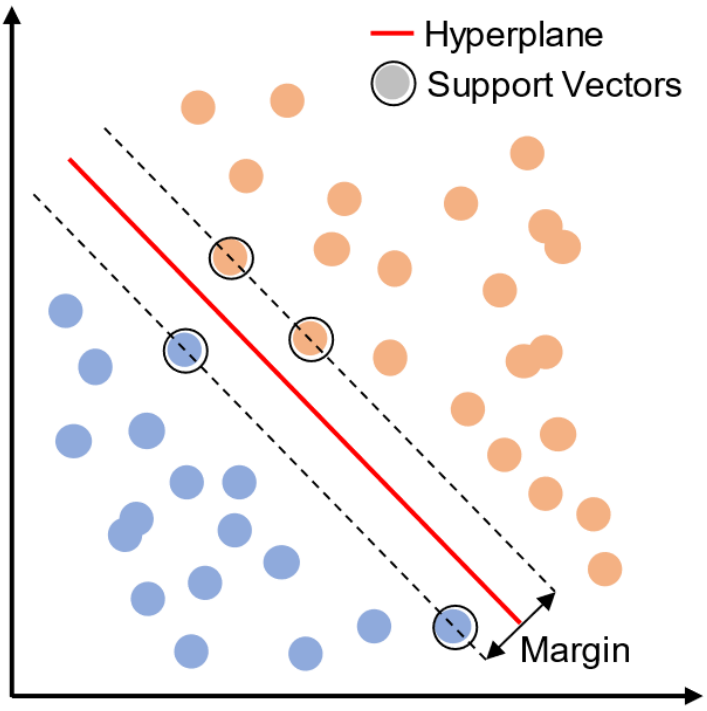


Figure 2.6: SVM classification illustrated using a simple example of two linearly separable classes.

results from the role the support vectors play in constructing the optimal hyperplane **[104]**. Additionally, it is often thought that SVMs may have an edge over the curse of dimensionality. However, Hastie *et al.* (2009) showed that this is not necessarily the case **[104]**. Finally, the selection of an appropriate kernel function is essential in building a well-functioning model **[104]**.

2.3.4 Machine Learning Applications

Tables 2.1 and 2.2 give an overview of publications employing ML methods to model gasification and pyrolysis, respectively. Only publications published in 2016 or later were considered. For studies which developed a range of models, performance parameters are given as a range. If a study considered more than one ML method, performance parameters are quoted for all considered methods. Generally, models achieved an acceptable prediction performance. The high predictive accuracy of ANNs stands out with R^2 values close to 1 and generally >0.9. Multi-output and single-output models have successfully been employed without either method outperforming the other.

Table 2.1: Summary of ML models used for the modelling of gasification.

ML Model	Year	Feedstock Type	Predicted Parameters	Size Data set	Number of Inputs/ Outputs	R^2	RMSE	Comments	Ref.
ANN	2016	MSW	Syngas tars, char LHV, and syngas yield	67	9/(1-3)	0.899-0.983	-	Double hidden layer better than single layer	**[108]**
ANN	2017	Woody and herbaceous biomass	CO, CO_2, H_2, and CH_4 yields	63	6/1	0.986-0.993	0.052-0.092	-	**[115]**
ANN	2018	Woody and herbaceous biomass	CO, CO_2, H_2, and CH_4 yields	70	7/4	0.914	-	-	**[116]**
SVM, RF	2018	Woody biomass	CO, CO_2, H_2, CH_4 yields, and syngas HHV	5237[a]	8/1	-	0.348-1.520, 1.796-5.144	Problem turned into classification problem	**[117]**
ANNs, NARX	2019	Woody biomass	CO, CO_2, H_2, CH_4 yields, and syngas HHV	3831[a]	8/1 or 9/1	0.930-0.999	0.020-0.853	NARX showed better accuracy than other ANNs	**[118]**
PR, SVM, DT, ANN	2020	Woody biomass	CO, CO_2, H_2, CH_4 yields, and syngas HHV	4826[a]	3/1	0.541-0.865, 0.713-0.915, 0.807-0.944, 0.823-0.944	0.101-0.582, 0.097-0.278, 0.058-0.180, 0.045-0.198	PCA used to reduce 16 inputs features to 3	**[94]**
NARX, PR	2020	Woody biomass	Reactor temperature, syngas composition, and syngas HHV	3500[a]	15/6	0.984-0.994, 0.832-0.874	0.016-0.034 ,0.054-0.094	PR suitable for real-time process control	**[119]**
ANN	2020	Woody biomass	Syngas tar content	120	8/1	> 0.970	-	No significant differences between single and two hidden layer ANN	[48]
ANN	2020	Woody and herbaceous biomass, sewage sludge, MSW, plastics	Syngas composition (CO, CO_2, H_2, CH_4) and yield	203	9/1	> 0.940	-	Considered effect of bed material	**[120]**
ANN	2020	Simulated (Woody, animal, and herbaceous biomass)	Power production of downdraft gasifier	1032	11/1	0.999	0.496	Training data obtained from thermodynamic equilibrium model	**[121]**
ANNs	2021	Woody biomass, Plastics	H_2 production	30	4/1	0.990	-	-	**[122]**
ANN	2021	Food waste	CO, CO_2, H_2, and CH_4 yields	40	3/1	> 0.980	< 0.300	-	**[123]**
Ensemble	2021	MSW	Syngas yield and LHV and LHV of other products	67	9/3	0.972-0.990	0.065-0.152	Ensemble method combines DT, XGB, RF, ANN, and SVM	**[124]**

DT – decision tree; HHV – higher heating value; LCA – life cycle assessment; LHV – lower heating value; MLR – multiple linear regression; NARX – autoregressive exogenous neural networks; PR – polynomial regression; PCA – principal component analysis; R^2 – coefficient of determination; RMSE – root mean squared error; XGB – extreme gradient boosting

a – refers to individual data samples; b – different sized data sets were used for different model outputs; c – given in g instead of normalised form

Table 2.2: Summary of ML models used for the modelling of pyrolysis.

ML Model	Year	Feedstock Type	Predicted Parameters	Size Data set	Number of Inputs/ Outputs	R^2	RMSE	Comments	Ref.
ANN, SVM	2016	Animal biomass	Biochar mass yield	33	5/1	0.804, 0.963	0.835[c], 0.365[c]	SVM performed better than ANN	**[125]**
ANN	2017	Woody, animal, and herbaceous biomass	Kinetic constants	150	3/3	0.930-0.940	-	-	**[126]**
ANN, SVM	2018	Biomass (unspecified)	Char, gas, bio-oil yields, and bio-oil LHV	71-175[b]	9/1	0.800-0.932, 0.895-0.938	1.030-4.500, 0.980-3.340	SVM performed better than ANN	**[127]**
ANN	2018	Herbaceous biomass	Temperature dependent mass loss	1000	2/1	0.970-0.990	-	-	**[128]**
ANFIS	2019	Herbaceous biomass	Kinetic constants	192	4/3	> 0.970	< 5.006	-	**[129]**
RF	2019	Woody and herbaceous biomass	Biochar yield and C content	128-245[b]	(8-12)/2	0.755-0.864	3.403-6.939	Input features were varied	**[130]**
MLR, RF	2020	Woody and herbaceous biomass, Algae	Bio-oil yield and H content	137-264[b]	(7-8)/2	0.016-0.352, 0.790-0.920	1.410-7.960, 0.540-3.050	MLR unable to predict outputs	**[131]**
RF	2020	Woody and herbaceous biomass, Algae	Biochar yield and characteristics	276-538[b]	7/1	0.780–0.870	-	ML model integrated with LCA and economic analysis	**[132]**
ANN	2021	Herbaceous biomass, sewage sludge	Char, tar, and gas yield	-	13/3	0.984	2.427	-	**[133]**

DT – decision tree; HHV – higher heating value; LCA – life cycle assessment; LHV – lower heating value; MLR – multiple linear regression; NARX – autoregressive exogenous neural networks; PR – polynomial regression; PCA – principal component analysis; R^2 – coefficient of determination; RMSE – root mean squared error; XGB – extreme gradient boosting

a – refers to individual data samples; b – different sized data sets were used for different model outputs; c – given in g instead of normalised form

2.3.4.1 Artificial Neural Networks for Process Modelling

ANNs have arguably been the most popular type of ML approach used for the process modelling of gasification and pyrolysis. These thermochemical processes are extremely complex which calls for advanced models **[134]**. ANNs lend themselves to the modelling of these processes due to their capability to learn highly non-linear relationships between inputs and outputs.

In the early 2010s, various researchers observed that ANNs had scarcely been employed to predict gasification product yields and properties **[91, 134]**. Puig-Arnavat *et al.* (2013) explored the potential of ANNs to model the thermochemical conversion of biomass. Models for a circulating fluidised bed and bubbling fluidised bed gasifier were presented. For the training of the two models, only wood gasification at atmospheric pressure was considered to ensure high data homogeneity. However, the collected data sets were small containing data from only 18 and 36 experimental runs. Despite the proposed models achieving a high predictive accuracy ($R^2 > 0.97$), a major drawback is their limitation to a single type of feedstock and the specific range of gasifier conditions used in model training **[134]**.

Baruah *et al.* (2017) addressed the issue of other studies (e.g. **[134, 135]**) associated with the consideration of only one type of feedstock by using experimental data for different types of woody biomass for model development. Furthermore, the study identified that fixed bed downdraft gasifiers in particular had received little attention when it comes to ANN-based modelling despite being one of the most commonly employed types of gasifiers for small-scale applications. It is stated that expanding the used data set is desirable to improve the range of cases to which the model is applicable **[115]**.

This lack of generalisation capability in earlier ANN studies in the field of biomass gasification was also identified by George *et al.* (2018). For this reason, the authors chose to train their model on five different feedstocks to ensure generalisation. The model considered the feedstock's carbon, hydrogen, oxygen, moisture, and ash content, as well as operating temperature and equivalence ratio as the key input parameters. However, it was acknowledged that a large number of parameters could be relevant and further exploration of factors such as the feedstock's polymeric composition or reactor geometry was essential. Additionally, it is stated that predicting other critical factors, which are essential to our understanding of the gasification process, such as tar and char yields might be an interesting direction for future research **[116]**.

An attempt at including additional input parameters which are relevant to the process was recently made by including the reactor's bed material as a parameter for model training. The effect of different bed materials has widely been studied but was never considered in an ANN before. Serrano *et al.* (2020) simply implemented this using ordinal encoding **[120]**. This technique assigns an integer value to each unique bed material option.

This simplistic approach may result in acceptable results in certain cases, but for many cases, more advanced methods to encode categorical variables are called for. Especially, parameters which do not have a clear order or ranking (e.g. bed material, gasifier type, or gasifying agent) need to be treated with care. For example, one-hot encoding is a commonly used method, which has been employed in other fields, to handle these types of variables **[136]**. One-hot encoding transforms a variable, which can take on n distinct categorical values, to n binary variables. A value of 1 indicates the presence of the variable and a value of 0 the absence.

Many early studies used trial and error to decide network structure and layout **[98, 125, 129, 137, 138]**. However, choices on network structure, training algorithms, threshold functions, etc. can all have a significant effect on the ANN's performance. Serrano et al. compared the effects of different activation functions and network topologies on the predictive power of ANNs. They compared two ANN types (multi-layer feed forward and cascade forward network) and found that no single combination resulted in the best performance for all parameters **[120]**. In another work, the performance of a single hidden layer model and a two hidden layer model was compared **[107]**. It was found that the potential performance improvement obtained from a two-hidden layer model was not justifiable considering the increase in computational power required **[107]**. Results such as this are case-specific and it is important to consider and compare a range of different options.

Pandey *et al.* (2016) chose Levenberg-Marquardt back-propagation as the training algorithm for their ANN **[108]**. This training algorithm has been used in many comparable studies which also found that it performs well **[93, 98, 107, 108, 118, 137]**. Nonetheless, a lack of comparison between training algorithms becomes apparent and future research should aim to improve this. However, the study attempted to improve previous works which often used trial and error or empirical equations (e.g. **[116]**) to decide on the optimum ANN configuration (e.g. number of hidden layers and nodes in hidden layers) by approaching network hyperparameter tuning in a more systematic fashion. For this, a procedure employing 100 Monte Carlo runs with different initial guesses, data segmentations, etc. was implemented to decide on the best network structure **[108]**.

As previously mentioned, the prediction of other factors than syngas composition and yield is important to fully map the gasification process and extend the capacity of model-based process design. As tar generation can be one of the limiting factors hindering industrial scale deployment of gasification by affecting downstream applications, some authors attempted to predict tar generation using ANNs **[107, 108]**. Serrano and Castelló identified the heterogeneous nature of tar generation data, as well as its scarcity, as key reasons for the limited development of tar prediction models **[107]**. In their study three different models were trained for the different tar sampling methods, as well as a fourth

trained on the entire data set. A data set of 120 data patterns was used for the model development. In general, the proposed ANN models were found to predict the tar generation well, with coefficients of determination > 0.97 for test and validation and relative errors below 20% for most samples. The models behaved well in most of the operating regions, with exceptions in the boundary regions where insufficient data was available for the model to learn the data structure well **[107]**.

A point frequently made when talking about ANNs is their "black box" nature which limits interpretability **[107, 108]**. This means the process linking inputs to outputs is opaque. This makes it challenging to draw conclusions about the fundamental mechanisms of a considered process from the network's prediction. Many factors, such as heat and mass transfer effects, which are considered by e.g. kinetic models are not, and often cannot easily be, incorporated in ANNs **[107]**. There have been attempts to improve interpretability by evaluating the relative importance of impact factors. Garson's equation has frequently been used for this purpose **[107, 115, 118, 134, 139]**. It returns the percentage contribution of any given independent variable on the considered dependent variable. For this, the absolute weights between the hidden-to-output layer nodes are partitioned into components linked to each input neuron. This means the equation does not provide information about the direction of the relationship **[139]**. The equation is given by

$$f_I = \frac{\sum_j^{n_H} \left(\frac{I_{V_j}}{\sum_k^{n_V} I_{V_j} O_j} \right)}{\sum_i^{n_V} \left(\sum_j^{n_H} \left(\frac{I_{V_j}}{\sum_k^{n_V} I_{V_j} O_j} \right) \right)} \tag{2.1}$$

where f_I is the relative importance of input neuron I. The numerator is the sum of all contributions of one input variable V on the corresponding output variable. The denominator contains the sum of all such summations for all input variables. I and O are the absolute weights of the i^{th} and k^{th} input and output nodes respectively. These are associated with the j^{th} hidden node. Finally, n_V and n_H are the number of input variables/nodes and the number of hidden layer neurons, respectively.

The performance of Garson's equation for correctly identifying variable importance has previously been found to be lacking due to its use of absolute connection weight values **[110, 140]**. It has been shown that different methods for quantifying variable importance may well result in different importance rankings **[110]**. For this reason, researchers should consider using a range of methods to be sure of their findings. One method that was found to perform well is the "Connection Weights" method introduced by Olden and Jackson **[110, 141]**. Like Garson's equation, the input-hidden and hidden-output connection weights are used for the calculation. However, importantly, the raw connection weights are used instead of absolute values **[110, 141]**. Another approach to improve the inter-

pretability of results is the hybridisation of theory-based models, such as thermodynamic equilibrium models, with ANNs. This approach is discussed in more detail in the following section (Section 2.3.4.1.1).

2.3.4.1.1 Hybridised Use of ANNs with Existing Thermodynamic Equilibrium Models

To improve interpretability, some authors attempted to combine thermodynamic equilibrium models with ANN methodology. In the approach taken by Guo et al., an ANN was used to fill in the gaps of a partial first principles model – i.e. the ANN was used as an estimator for parameters which could not be measured and proved challenging to model from first principles **[142]**. Whilst this approach is promising, this early study leaves open numerous questions. For example, inputs are matched with the corresponding outputs for most ANNs. However, it was stated that the desired outputs were not measurable, making it unclear which parameters the inputs were matched with instead. Additionally, overfitting behaviour was identified but not further addressed. Finally, discrepancies between model and experimental results were identified but their size was not statistically quantified **[142]**.

In a recent work **[121]**, the role of ANN and thermodynamic equilibrium model was reversed as compared to the earlier work by Guo *et al.* (2001) **[142]**. The study identified that proper ANN model training required large amounts of data. To obtain the data, large-scale experiments are required which are both expensive and time-consuming. Hence, it was suggested that the data obtained from a thermodynamic equilibrium model should be used for ANN training **[121, 142]**.

The main benefits of this combined approach are simplicity and convenience. The approach allows for fast computation of many different input data sets once the model is sufficiently trained. This offers the user a convenient way to assess the effects of variations in for example feedstock composition due to different growing conditions or climatic differences on the system's net output power. Thus, intermediate calculation steps become obsolete. A very high $R^2 > 0.999$ was achieved for the training and test stage of the ANN. This is not surprising as the training data comes from the equilibrium model which uses the same underlying equations throughout the data set. It is unclear why a validation stage was not included, as the data set appears to be sufficiently large to do so. Finally, it needs to be stressed that the underlying model is still a thermodynamic equilibrium model. As such the hybrid model is constrained by the shortcomings and assumptions made within the equilibrium model **[121]**.

2.3.4.1.2 Time-series Approaches

Many early ML-based models adopted a time-independent approach **[119]**. Whilst this may predict the final outputs of the process well, a time-dependent model is required

for on-line process control and optimisation. Kinetic and CFD models are generally not suitable for on-line process control or implementation as model predictive controllers due to their computational intensiveness. Equilibrium models, on the other hand, are far less computationally intensive and represent a valid option for on-line control purposes. However, they only capture a snapshot of the process due to their time-independent approach. Thus, they are better suited for the analysis and optimisation of a static process at set operating conditions **[119, 135]**.

Early studies used conventional ANNs for time-dependent modelling. More recent studies utilised autoregressive exogenous neural networks (NARX) which are more catered towards time-series data **[118, 119]**. NARX models contain a feedback mechanism which allows them to learn a time series by using input/output pairs of previous time steps as inputs to the current time step. In addition, this type of model uses an exogenous time series.

In general, researchers found that time-series approaches are better suited for on-line control and model predictive control than equilibrium models which have conventionally been used. This is especially true for NARX which were found to learn a system's dynamics well **[118, 119, 135]**. However, to train these types of models, all training data is obtained from experiments based on one specific reactor. The high homogeneity of the training data allows for greater accuracy but also means that the model is only applicable to the specific reactor from which the training data has been obtained.

One strength of the NARX approach is its ability to easily implement real-time readings from sensors (e.g. thermocouples) installed throughout the reactor. For example, Yucel *et al.* (2019) **[118]** explored the effects of the temperature distribution throughout a fixed-bed downdraft gasifier on the syngas composition and HHV. It is generally accepted that the gasification temperature is one of the key parameters dictating the process. It was found that the temperature distribution within the gasifier accounted for 34.53% of the variation in the outputs and especially the temperatures of the reduction, combustion, and drying zones were found to be most influential **[118]**.

Elmaz and Yücel compared a NARX with a simpler polynomial regression model **[119]**. It was found that the conventional validation methods used for ML applications, such as k-fold validation, might be suboptimal for a time-series approach. Other methods, such as rolling window analysis, may be better suited for time-series approaches to evaluate the stability and performance of the model **[119]**.

In addition to choosing appropriate validation techniques future research considering time-series approaches needs to improve on previous researchers' works by (i) using formal statistical descriptors (e.g. R^2 and RMSE) to evaluate model accuracy, (ii) providing a sufficient description of training data and how it has been obtained, and (iii) describing the network/model structure to an appropriate level of detail. Improving on these points is

essential for allowing other researchers to understand the methods employed and reproduce findings if necessary.

2.3.4.2 Hybrid Machine Learning Models

The hybridisation of ML models with other ML techniques or advanced statistical methods has shown great potential in the wider energy sector **[143]**. More recently, researchers have also looked towards hybrid models to model thermochemical processes **[132, 144, 145]**. Compared to ordinary models, hybrid models generally have higher accuracy and generalisation ability at the cost of an increase in computational load. For instance, ensemble methods, such as gradient boosting, can turn many weak and poorly performing models into one effective one **[146]**. A recent study by Kardani *et al.* (2021) **[124]** nicely illustrated how an ensemble model could improve model robustness despite only limited data being available for model training. Finally, incorporating statistical methods, such as Monte Carlo simulation, can greatly improve the credibility of a model's results by accounting for the related uncertainties **[132]**. Two frequently considered methods – particle swarm optimisation (PSO) for ANN training and dimensionality reduction algorithms – are reviewed in greater detail in Sections 2.3.4.2.1 and 2.3.4.2.2.

2.3.4.2.1 Use of Particle Swarm Optimisation for Network Learning

As identified in Section 2.3.4.1, existing research often paid little attention to identifying the optimal network structure. As part of this, the effect of choosing the most suitable training algorithm has often been neglected. More recently, PSO has been considered as an advanced algorithm for model training **[147]**. PSO is a type of metaheuristic inspired by the movement of a fish school or flock of birds. As a metaheuristic, few or no assumptions are made about the problem which is to be optimised **[148]**.

Chiñas-Palacios (2021) **[144]** explored the potential of a hybrid model combining PSO and ANN methodology to model a biomass gasification plant. The hybrid PSO neural network was compared to two more conventional networks: cascade forward propagation and feed-forward back propagation. The optimal ANN configurations, for the various considered networks, were found after performing 215 simulation runs. The number of hidden neurons varied between 3, 10, and 100 and the particle population for the PSO varied between 10, 100, 600, and 1000. Intermediate values were not considered which might result in significant performance penalties. It was found that using PSO resulted in poor performance when using suboptimal particle populations. However, when using the optimal configuration, a mean squared error (MSE) reduction ranging from 11-37% was achieved for a range of parameters when compared to the two more conventional ANN configurations. This highlights the potential performance improvements with PSO, as well as the importance of appropriate parameter choices **[144]**.

In an earlier work, Aghbashlo *et al.* (2019) used a PSO algorithm for the tuning of a different type of neural network **[129]**. ANFIS methodology was used which combines the learning ability of neural networks with fuzzy logic principles. Upon using previously unseen biomass samples, their model was shown to be robust and adequately accurate with a R^2 value of > 0.91 **[129]**. Obafemi *et al.* (2019) identified that there was a lack of comparison between different training algorithms **[138]**. In the past, many researchers simply chose the Levenberg-Marquardt algorithm for model training **[108, 116, 120, 149–151]**. Future research needs to consider and compare different training options (e.g. Levenberg-Marquard algorithm, Bayesian regularisation, PSO, etc.) to find the most suitable one. Finally, PSO may also have the potential to automatically design an ANN by determining the most suitable hyperparameters **[151]**. This type of methodology may make choosing the optimal network structure and hyperparameters by trial and error obsolete.

2.3.4.2.2 Dimensionality Reduction Methods

Generally, the purpose of dimensionality reduction is the transformation of training data from a high- to low-dimensional space whilst retaining the data's variation. This is often desirable for predictive models with a large number of observations. In practice, principal component analysis (PCA) is the most commonly employed method, but other methods such as kernel-based methods or autoencoder methodology do exist **[152]**. For a more in-depth, general review of the topic, the reader is referred to Van Der Maaten *et al.* (2009) **[152]**.

Recently, Shahbeig and Nosrati (2020) used PCA to identify the principal reactions occurring during the pyrolysis of biological waste. The authors showed that PCA allowed for the use of high dimensional experimental data, such as differential thermogravimetric data, for ML model training **[145]**. Elmaz *et al.* (2020) trained four different ML models to predict biomass gasification products. The training data set consisted of an initial 16 features which were reduced to three using principal component analysis. High R^2 values of up to 0.944 were achieved for the best-performing ML methods, ANN and RF **[94]**.

Hosseinpour *et al.* (2017) successfully combined an ANN-based model with PCA to predict biomass HHV based on the feedstock's proximate composition (i.e. fixed carbon, volatile matter, and ash). Clearly, the initial data set already has low dimensionality. Thus, in this case, PCA was used to eliminate the linear correlation between predictors before the data was used for ANN training **[153]**.

It was found that linear techniques such as PCA do not always perform well with complex non-linear data. Some researchers showed that non-linear techniques outperformed linear techniques on selected artificial data sets. However, these performance improvements of non-linear techniques could not always be replicated on real-world data sets

[152]. Nonetheless, non-linear techniques for dimensionality reduction may be a worthwhile direction to explore in the future for thermochemical process data sets.

2.3.4.3 Use of Other Machine Learning Methods

ML algorithms are commonly split into supervised and unsupervised learning. For this review, a further split by learning task (i.e. regression, classification, and clustering) is desirable. Identifying the learning task at hand allows the ML practitioner to select an appropriate ML algorithm. For gasification and pyrolysis modelling, most learning tasks are regression problems. A lot of the early research within the field considered solely ANNs (as reviewed in Section 2.3.4.1). However, more recently, other ML techniques have received increasing attention. The three most frequently employed methods beyond ANNs are discussed below in Sections 2.3.4.3.1, 2.3.4.3.2, and 2.3.4.3.3.

2.3.4.3.1 Regression Methods

A range of different types of regression analysis exist. Linear regression has been used for many years in statistics and is one of the simplest forms of analysis to assess the relationship between dependent and independent variables **[104]**. Like many other real-world phenomena, thermochemical processes are known to follow complex highly non-linear relationships. It is no surprise, therefore, that linear regression has been found to be inadequate for the modelling of thermochemical processes such as the pyrolysis of biomass **[131]**.

However, regression methods are not limited to linear regression and other methods (e.g., polynomial regression, stepwise regression, and lasso regression) exist that are better suited for higher dimensionality data and the modelling of non-linear relationships **[131]**. One of the key benefits of regression-based methods compared to other ML methods is their lower computational demands. Elmaz and Yücel (2020) compared a NARX model to a simpler polynomial regression model. The polynomial regression model did not perform as well achieving $R^2 > 0.80$ for all outputs as compared to $R^2 > 0.98$ for the NARX model outputs **[119]**. In a separate study, even larger performance differences were identified upon comparing quadratic and cubic polynomial regression with support vector regression, decision tree regression and an ANN **[94]**. However, a less accurate model may still have its merits especially when it displays a much lower complexity and computational cost. Importantly, the authors noted that the polynomial regression model was still able to predict trends in the data. This makes this type of model well-suited for real-time model predictive control applications implemented with a conventional microcontroller. Ultimately, required model accuracy and computational demands need to be weighed against each other to find the most appropriate approach in the face of trade-offs.

2.3.4.3.2 Decision Tree-based Methods

In practice, tree-based methods, such as regression trees or more advanced ensemble methods, have seen less frequent use than ANNs for the prediction of gasification or pyrolysis product yields and properties. However, in a recent comparison between the ANN and regression tree methodology for the prediction of syngas composition and heating value, both methods were found to achieve similar prediction accuracy **[94]**. This highlights the importance of comparing different ML algorithms on a case-by-case basis. When comparing various ML techniques, it is key to initially select techniques which are suitable for the problem at hand. For example, a recent study by Tang *et al.* (2020) **[131]** compared linear regression to RF methodology for the prediction of the bio-oil yield and bio-oil hydrogen content from biomass pyrolysis. The well-known non-linear nature of biomass pyrolysis made linear regression a poorly suited approach to the problem and a comparison with ANN or SVM methodology seemed more worthwhile.

Whilst a simple decision tree is easy to follow and understand, ensemble methods suffer from the same black box nature as ANNs **[104]**. However, the interpretability of tree-based methods can still exceed that of ANNs, as the feature importance is generally more easily extracted. For example, Tang *et al.* (2020) **[131]** found that the biomass composition generally has a larger effect than the pyrolysis conditions on the process outputs. On the other hand, studies by Cheng *et al.* (2020) **[132]** and Zhu *et al.* (2019) **[130]** found that pyrolysis temperature had a strong effect on the production of biochar from a range of biomass feedstocks. Interestingly, the findings by Zhu *et al.* illustrated that the feature importance obtained from RF methodology could significantly differ from the rankings obtained using Pearson correlation coefficients **[130]**.

Cheng *et al.* (2020) **[132]** proposed the novel integration of an RF-based predictive model for pyrolysis with life cycle assessment and economic analysis. Their study aimed to achieve a comprehensive evaluation of a range of different pyrolysis feedstocks on a more holistic basis by considering environmental and economic factors. This type of approach shows the power of ML techniques which allow for the rapid estimation of a range of different scenarios without the need for expensive experiments. This type of approach is not limited to RF, and other ML techniques and thermochemical processes may also be considered for similar implementations in the future.

Some other advantages of tree-based methods are their capability to handle missing predictor values, as well as the easy implementation of categorical values. Many authors used existing experimental studies to collect a process database for model training **[120, 126, 127, 130]**. However, due to low homogeneity in the published literature, it can be challenging to amass a sufficiently large database without any missing values. Many ML methods struggle with handling missing predictor values. Tree-based methods, on the other hand, are well suited for handling missing predictor values without having to

discard observations which could reduce the size of the training set **[104]**. Furthermore, categorical variables which may have a significant effect on the model outputs are easily implemented in the tree structure of such models.

2.3.4.3.3 Support Vector Machines

SVMs have been commonly used for classification tasks. However, recently several researchers have looked towards SVM as an alternative to ANN for the modelling of thermochemical processes **[94, 125, 127, 145]**. Upon the comparison between SVM and ANN, both Cao *et al.* (2016) **[125]** and Chen *et al.* (2018) **[127]** found that SVM showed better performance, in terms of RMSE and R^2, at predicting pyrolysis product distributions and properties. However, it is important to note that these results cannot necessarily be generalised. The two studies used small data sets of 33 and 71/166/175 (for the prediction of char and gas yield, bio-oil yield, and bio-oil HHV, respectively) data samples, respectively. SVMs are known to generally perform well in a high dimensional space with small to medium-sized data sets **[113, 114]**. Correspondingly, the larger data sets used in Chen *et al.*'s 2018 **[127]** study reduced the performance difference between SVM and ANN. In contrast, Elmaz *et al.* (2020) **[94]** gave the edge to ANN. The authors found that a multilayer perceptron performed better than an SVM at predicting the syngas composition and HHV from biomass gasification. The key difference between this study and **[125, 127]** is the significantly larger data set that consists of 4,826 time-dependent data samples.

For the case of small data sets, SVM can show better robustness to overfitting than ANN as suggested by Cao *et al.* (2016) **[125]** and Chen *et al.*'s (2018) **[127]** studies where the ANN models had higher coefficients of determination and lower RMSE during model training than SVM. However, the subsequent decrease in performance during testing indicated overfitting.

Selecting an appropriate kernel function is essential in creating a well-performing model. Choosing a linear function results in similar results to linear regression, making more advanced non-linear functions generally more attractive. The radial basis function has widely been accepted as a good option for the modelling of non-linear processes **[154]**. All studies reviewed here selected the radial basis function as the kernel function **[94, 125, 127, 145]**. In conclusion, SVMs represent an interesting alternative to ANNs or other ML techniques, especially for small, high-dimensional data sets.

2.3.4.4 Use of Machine Learning in Bioenergy-related Fields

Researchers in a vast number of different fields have been looking towards ML for simulation and modelling tasks. Bioenergy systems tend to be complex systems spanning a range of fields. Hence, developing comprehensive ML models for e.g. sustainable waste

management schemes will likely require knowledge across different disciplines.

A thorough review of how ML has been employed in related fields is beyond the scope of this review. However, several authors published dedicated reviews on these topics. For example, Guo *et al.* (2021) published a review on the use of ML methods in the field of recycling and organic solid waste treatment **[155]**. Various treatment methods were discussed, among them incineration, gasification, and pyrolysis are highlighted. The authors noticed that ANN were the most employed method accounting for more than half of all considered studies. Furthermore, it was noted that combining different ML techniques in integrated models or ML techniques with other innovative methods is seen as promising for improved performance. Finally, the authors stressed the importance of demystifying the "black box" of many ML techniques by, for example, quantifying the contributions of input parameters **[155]**.

Sewsynker-Sukai *et al.* (2017) reviewed the use of ANNs for the modelling and optimisation of biofuels production through biological methods **[156]**. Coupling genetic algorithms with ANNs was seen as very attractive due to their power to solve a vast array of optimisation problems. Additionally, the authors discussed several papers comparing ANNs to response surface methodology for the optimisation of biofuel production processes. ANNs were found to considerably outperform models based on response surface methodology due to their capability to represent highly non-linear processes without being limited to quadratic equations. Finally, ANN applications in real-time process monitoring and control were suggested as a potential direction for further research **[156]**.

Mosavi *et al.* (2019) shifted the focus from process modelling/optimisation applications to an energy systems approach. In their thorough review, a vast range of ML techniques for numerous research areas was considered. Reviewed ML applications ranged from wind speed forecasting and building energy demand estimations to district heating system optimisation. However, thermochemical biomass and waste conversion technologies were not discussed in the review. Two methods that stand out are ensemble methods and hybrid ML models due to their improved accuracy and generalisation ability. As a closing remark, the study pointed out that the case-specific development of ML models for a particular application resulted in the best performance despite requiring more time and effort **[157]**.

One of the earliest uses of ANN has been for the prediction of feedstocks' heating values. Dong *et al.* (2003) predicted the LHV of MSW from its physical composition (i.e. weight fraction of different waste components). Upon comparing the proposed model to various conventional models based on the waste's physical, ultimate, and proximate compositions, the ANN model was found to predict the LHV best. In this early work, the distinction between the training, validation, and test stages is not clear **[158]**.

A similar approach was used for the prediction of the gross calorific value of Indian coal. Proximate and ultimate data were used, and the model was found to have an improved

performance when compared to conventional linear models. This is likely due to the fact that some parameters exhibited a linear relationship with the gross calorific value while others did not **[159]**.

Hosseinpour *et al.* (2017) considered various soft computing techniques for the estimation of biomass' HHV using its proximate composition information **[153]**. All models performed well and had lower mean absolute percentage error scores than various empirical equations in the literature. Upon comparing the various proposed models, the authors concluded that the advanced model combining ANN and partial least squares methodology performed best on the test data set ($R^2 = 0.9732$).

2.3.5 Challenges and Prospects

2.3.5.1 Challenges

Data availability for model development remains a significant challenge. Running experiments can be expensive and time-consuming while usually a large amount of training data is required for developing ML-based models. Many studies collected experimental data from the literature, but inconsistencies in how data is quoted can make this process challenging and subsequent comparison difficult. Furthermore, there is a trade-off between a model's prediction accuracy and its generalisation capability related to the type of training data used. Using highly homogeneous experimental data from e.g. only one type of gasifier generally results in a model with high predictive accuracy for modelling the process of the particular type of gasifier. However, this means the model is only applicable to the narrow type of training data used. It will be valuable to build more diverse models with a lower, but still satisfactory prediction accuracy. Again, this would increase training data requirements as the data needs to span a wider range of parameters.

A problem previously pointed out is the black box nature of some ML algorithms which limits data and model interpretability. Whilst simple methods, such as linear regression, do not suffer from this problem they have also been shown to have non-satisfactory prediction accuracy **[131]**. ANNs which have been the most employed algorithm can make it challenging to draw valuable mechanistic conclusions from their findings. Efforts have been made at illuminating ANNs and other ML methods **[110, 141, 160]**. Section 2.3.4.1 described the popularity of Garson's equation to determine the variable importance of input features. As this method has been found to be lacking, more recent approaches need to be considered to improve interpretability **[110, 140]**. One such option is the use of a global surrogate model for which a simpler, more interpretable model is trained to approximate the behaviour of a more complex black box model **[37]**. Another option is feature permutation to assess the importance of different input parameters. This method shuffles the values of one feature at a time and computes the resulting increase in the

prediction error.

A lack of clarity in published literature is another factor holding back the successful deployment of more ML models and the improvement of existing ones. For instance, a lack of clarity in how data is split for model training, validation, and testing has been identified. Additionally, some studies have been unclear on which stage their results refer to. The use of various statistical parameters to assess a model's performance can make comparisons between models challenging. Generally, R^2 and RMSE have been used, but some studies used, for instance, R **[116]** and MSE **[108, 116]**. Not all studies quoted error values in the normalised form **[125]**. Future studies should consider using R^2 (or adjusted R^2) and RMSE in normalised form to facilitate model comparison. Quoting error values as percentage errors (e.g. root mean squared percentage errors (RMSPE)) is another option which is generalisable and allows for an intuitive assessment of the error's size. Finally, vast differences in training data make model comparison challenging.

2.3.5.2 Prospects

Further work on illuminating the black box of ML algorithms to enhance explainability is considered essential. This way new interconnections between a system's inputs and outputs might be learned. The development of more comprehensive prediction models, capable of predicting a larger range of relevant factors, is essential for the creation of more holistic models. Understanding how the gasification/pyrolysis process ties into a larger scheme or system analysis will encourage the deployment of these technologies by minimising uncertainties for stakeholders. For example, Cheng et al. used the results from a pyrolysis prediction model in a larger framework incorporating LCA and economic analysis **[132]**.

Existing models may be improved by considering a wider range of training algorithms, approaching hyperparameter tuning in a more systematic fashion, and incorporating additional relevant input parameters such as categorical data into the model structure. Another promising avenue of research has been the combination of various ML methods or the combination of ML methods with other advanced methods. One such example is the use of PSO for model training and optimisation (Section 2.3.4.2). Since data sets for model training could often be limited in size, ensemble methods might be another good option to improve model robustness as illustrated by Kardani *et al.* (2021) **[124]**.

For control purposes models with a low computational demand (e.g. regression models) and acceptable accuracy may allow for better real-time process control as compared to conventional models. Conversely, more demanding and accurate models such as NARX can also be used for process optimisation. Combining these could allow for system scale-up and incentivise a more widespread deployment of the studied technologies. Importantly, future models need to find a balance between computational demands and accuracy to

allow for their implementation through conventional microcontrollers.

2.3.6 Conclusions

ML methods have shown great potential towards modelling the gasification and pyrolysis processes and predicting the processes' product yields and properties. Their predictive performance is already on a similar or even superior level to that of conventional models. Thus, it should not come as a surprise that the number of publications increased drastically over the past two decades. In the period from 2000 to 2009, a total of 70 papers have been published. On the other hand, 549 publications were found between 2010 and 2019 which represents an increase of 684%.

Whilst R^2 values can be difficult to compare, due to some studies having vastly different approaches and aims, it has generally been found that most studies achieved a high prediction accuracy with $R^2 > 0.90$. Despite their high accuracy, ML models do not lend themselves to interpretation, as e.g. thermodynamic equilibrium models do, and it has thus been challenging to interpret modelling results at times. Developing more interpretable models is essential in improving the underlying understanding and design capacity of the processes.

Utilising the strong predictive performance of ML models by incorporating them as part of more comprehensive system models can encourage industrial-scale deployment by minimising uncertainties for investors and policymakers. For this, the availability of sufficient training data is essential. Considering the nature and availability of training data is key in deciding on the most appropriate ML method for any given use case. Finally, as data availability and computational power increase, new avenues for research become available. For instance, more complex models may be used for real-time process control and optimisation applications.

2.4 Environmental and Techno-Economic Analysis of Thermochemical Processes

2.4.1 Environmental Analysis of Thermochemical Processes

LCA has established itself as the arguably most popular method to assess the sustainability and environmental impacts of a product, service, or system throughout its entire life cycle. It can be used to assess various factors such as resource use, energy consumption, emissions, and waste generation to provide insights for sustainable decision-making. LCA methodology has been guided by ISO (International Organization for Standardization) 14040 and 14044, which aim to provide a standardised and consistent approach to

environmental assessment **[161, 162]**.

Most researchers have focused on two phases of thermochemical biomass conversion systems. The first stage deals with the cultivation, collection, and transportation of the biomass to the plant. During this stage, factors such as the application of fertilisers and pesticides, and the effect of transporting the biomass to the plant are typically studied. Land use changes have often been ignored but can play a critical role in a system's life cycle. Biomass' relatively low density and low volume generally make it infeasible to transport it for long distances. Typically, distances of $> 200\ km$ have been considered in the literature **[163]**. However, transportation has been found to generally have a minor effect on the net environmental performance of a system **[164]**.

The second phase describes what happens at the plant, like biomass pretreatment, thermochemical conversion, and cleaning and upgrading of products. This stage varies widely depending on the exact thermochemical conversion method considered, the utilised feedstock, and the desired end products. Particularly factors such as feedstock drying are susceptible to significant variation, with the drying of low-moisture feedstocks having a negligible effect while high-moisture feedstocks have a drastic effect. The third phase, which has received little attention in the current LCA literature, is the end-of-life phase. It is concerned with decommissioning the plant and looks at the demolition or recycling of the facility **[163]**.

These phases are often discussed in terms of an LCA's system boundary and are generally defined as cradle-to-grave, cradle-to-consumption, cradle-to-bioenergy, and wheel-to-bioenergy. Most past research considered a cradle-to-bioenergy or wheel-to-bioenergy system boundary. A cradle-to-bioenergy system includes all steps from cultivation to the production of bioenergy but does not consider its use or the disposal stage. Finally, wheel-to-bioenergy comprises the transportation of biomass to the plant and its subsequent conversion **[165]**.

A key consideration in any LCA is the employed functional unit (FU). The FU represents the quantified performance or function of the product, service, or system that is being analysed in the LCA. It serves as a reference point for objective comparison between different alternatives that serve the same final function. Alhazmi *et al.* (2021) reviewed the literature on the thermochemical conversion of agricultural waste to bioenergy and found that 70% of the studies considered in their review defined their FU in terms of mass (e.g. tonnes or kg of feedstock utilised or end product produced). The other 30% of reviewed works defined their FU in terms of energy (e.g. kWh or MJ) **[165]**.

An environmental impact category in LCA represents a specific aspect of the environment, such as climate change or resource depletion, which is assessed to understand the potential environmental impacts of a product or system. Global warming potential (GWP) was found to be the most commonly studied impact category, with other envi-

ronmental or ecosystem-related impact categories such as acidification and eutrophication potential also being popular. Ecosystem-related impact categories such as human health and resource depletion have generally received less attention **[163, 165, 166]**.

A potential issue with the use of dedicated energy crops for bioenergy generation is that they are often genetically modified to maximise their yield to allow them to compete with fossil fuels. This typically comes with the use of herbicides and pesticides which may lead to resistant weeds and insects which may in turn be dangerous to non-genetically modified crops **[167]**. Agricultural residues have also been found to have a better environmental performance than dedicated energy crops, even though the use of energy crops can be simpler and lead to lower maintenance **[168]**.

The utilisation of heat, particularly low-temperature heat, was found to be subject to significant barriers, making it a source of uncertainty in LCA and other kinds of sustainability analysis. Storing or transporting heat can be inefficient. If direct utilisation at the plant is not possible district heating may be a relevant option. However, district heating is typically only economical if sufficient heat demand is present within a 1.5 km radius **[166]**.

Uncertainties in LCA studies are unavoidable and by systematically addressing these uncertainties, the quality and usefulness of LCA studies can be improved. This can be done through uncertainty or sensitivity analysis which provides additional information for decision-makers in addition to the regular LCA results. A sensitivity analysis can be conducted by simply varying influential parameters, or probabilistic approaches like Monte Carlo simulation can be used to map the uncertainty in certain model parameters **[165]**.

LCA studies are often limited by a lack of available data and the high complexity of systems amongst other factors. In practice, this means that LCA practitioners need to employ simplifications when required. According to Beemsterboer *et al.* (2020) simplifying logics can be categorised into the following five categories: exclusion, inventory data substitution, qualitative expert judgement, standardisation, and automation **[169]**. Future research must ensure that appropriate simplifying logics are employed as required and discussed transparently. The effect of the made assumptions can be quantified by incorporating a sensitivity analysis.

Studies typically focus on one thermochemical conversion pathway and often on a particular feedstock. This, alongside widely different assumptions and local conditions, makes it challenging to compare various thermochemical methods or feedstocks. Additionally, researchers use different software and FUs for their analysis, compounding this effect. To address this effect, frameworks to assess a range of feedstocks or compare various thermochemical (or biochemical) bioenergy alternatives are required **[163]**.

In conclusion, future research on the environmental effects of thermochemical biomass and waste conversion should focus on the potential of feedstocks such as agricultural

residues, waste products, and by-products. It should also embrace simplified models when necessary to facilitate thorough uncertainty accounting, which is key to effectively communicating research to decision-makers. Finally, as highlighted by Alzhami *et al.* (2021), future LCA research must consider geographical factors so that the best feedstock and thermochemical process combinations for a given location can be identified **[165]**.

2.4.2 Techno-Economic Analysis of Thermochemical Processes

TEA is essential in providing important insights into the economic feasibility of engineering developments. It can effectively guide decision-making and ensure investment in the most suitable technological alternative, thus avoiding misspent efforts. Limited process data is a major obstacle, particularly for novel technologies, but this can be overcome by using relevant data from commercially available systems or components. To conduct TEA, three broad stages are typically completed. Initially, the system and system boundary are selected. Next, the process models are created to represent the components that make up the overall system. Finally, economic analysis is conducted by estimating capital and operational costs, as well as sources of income. **[170]**

In 2016, Patel *et al.* reviewed techno-economic and life cycle assessment studies concerned with various thermochemical conversion technologies. For gasification, they found that pressurised systems were generally more expensive than atmospheric systems, which is largely due to feeding being substantially more complex. However, this factor was less important for large capacity plants, where the increase in complexity can be justified by the improvement in performance. The authors also found that the co-gasification of biomass with coal may be an economical option. Importantly, this can only be an interim solution while fossil fuels are phased out. Particularly integrated gasification combined cycle plants with carbon capture showed potential for successful and economical co-gasification. For pyrolysis, the importance of biochar as a product was highlighted, particularly due to its higher yield in comparison to gasification. However, the revenue from biochar is highly uncertain and depends on a vast range of factors such as the processing conditions and feedstock, making it an important factor to consider during sensitivity analysis. Ultimately, when comparing various thermochemical processes, no one technology was found to be better than another. A plethora of factors, such as the desired end-product and local conditions have large effects on the economic feasibility of different thermochemical processes. For instance, gasification may generally be better suited to produce hydrogen than pyrolysis, whereas pyrolysis is typically more effective at producing liquid transport fuels **[163]**.

Recently, Campion *et al.* (2023) conducted a thorough review of the benefits and costs of biochar. Biochar has shown great potential but is yet to be used on a large scale, with particularly economic factors being a major roadblock. The profitability of

biochar developments is highly uncertain due to factors such as location, feedstock, and market conditions varying greatly on a case-by-case basis, ultimately hindering investment. External costs and benefits, such as increased water retention due to biochar application, are challenging to incorporate into models but will be key to consider in future research. Furthermore, the economic potential of other use cases than biochar for soil application will need to be considered **[26]**.

The potential of various thermochemical biomass conversion methods and feedstocks for negative CO_2 emissions was studied by Cheng *et al.* (2021). A combination of ML, LCA, and techno-economic analysis was used by the authors to find the break-even cost of each negative emission technology (NET) for American conditions. The slow pyrolysis of woody and crop residues was found to be economical in the current carbon market, whereas other conversion technologies rely on an increase in carbon credit prices due to for example new policies. Likely fluctuations in electricity and feedstock prices in the future are major limitations in the work **[171]**.

Technology maturity, markets for products, and policies incentivising pioneer projects are some of the major barriers limiting the development of thermochemical processes **[172]**. To accelerate the commercialisation of bioenergy technologies, further techno-economic analysis and subsequent optimisation are required **[173]**. Especially, integrated analysis considering economic, environmental, and technological factors is required to effectively communicate trade-offs between these factors to decision-makers and policymakers **[174]**. By developing models that can handle uncertainty, decision-makers' and policymakers' confidence in the results can be improved **[170]**.

2.4.3 Combined or Holistic System Analysis

Often overlooked, a third aspect of engineering system analysis is social considerations. Social factors are important to take into account when assessing the effectiveness and sustainability of a system. However, they can be hard to quantify and have received less attention in the literature than environmental and economic factors. In the future, more holistic approaches covering multiple aspects are required for decision-makers to make informed decisions. Life cycle sustainability assessment (LCSA) has been suggested as a response, which integrates economic and social aspects into a system's or product's LCA **[166, 175]**.

This is illustrated by Padilla-Rivera *et al.*'s 2019 study that analysed the literature on the sustainability assessment of bioenergy. They found that of the 8,542 initially screened works, only 32 studies considered sustainability from an economic, environmental, and social viewpoint. Most works focus on only one aspect. The investment cost was the most frequently considered economic indicator, whereas job creation potential and GWP were the most frequently considered economic and environmental indicators, respectively.

Public perception and socioeconomic repercussions were other social indicators commonly analysed. Land use changes, biodiversity, and water use were found to be other frequently considered environmental factors **[166]**.

Shackley *et al.* (2012) proposed the production of biochar from Cambodian rice husk via gasification instead of pyrolysis. By utilising gasification instead of pyrolysis, the system can produce electricity at reasonable efficiency while still producing significant amounts of biochar. The authors holistically assessed the system by studying the biochar's properties and subsequent effects on soil, as well as the economic and carbon abatement potential of the system. Even though a carbon abatement potential of up to 860 kg CO_2-eq per tonne of feedstock was shown, the authors noted that further studies are required to ensure there are no contamination and pollution issues related to biochar production from rice husk. Furthermore, social factors such as effects on local farmers and the access to resources were touched upon by the authors but require further analysis **[176, 177]**.

Jin *et al.* (2018) presented an integrated sustainability model for analysing a bioenergy system using forestry residues for electricity generation. It takes into account sustainability across three dimensions – social, economic, and environmental – through an integrated approach. The model considers various exogenous factors such as population growth, land-use changes, and policy alterations. It considered biomass cultivation, harvest, and transport, as well as its conversion and the use of the generated products. The studied biomass conversion technology was biomass co-firing with coal. Through the use of a causal loop diagram, the paper establishes interrelationships between system elements and maps their environmental, economic, and social impacts. However, the authors acknowledged that the model provides only a basic estimate of sustainability due to the inherent complexity of real-world phenomena, particularly social aspects that are challenging to quantify and accurately model. Nonetheless, the modular nature of the model allows for future improvements with more complex and accurate submodels replacing current ones, enabling better analysis and decision-making in the realm of bioenergy sustainability **[178]**.

Cheng *et al.* (2020) studied the potential of slow pyrolysis as a NET. Their work uses ML to model the bioenergy process and combines it with economic and environmental analysis. This way, the authors argued they could address the knowledge gap of previous studies focusing on either individual feedstocks under varying pyrolysis conditions or multiple feedstocks under the same conditions. A trade-off between economic and environmental factors was observed, with certain feedstock favouring one or the other. The authors solely focused on the LCA impact category GWP, as the main purpose of the study was to address slow pyrolysis' potential as a NET. A basic Monte Carlo simulation, which leaves room for improvement, was implemented, with the ML model's predictions and certain financial inputs being varied. For the Monte Carlo simulation, the ML model's predictions were described by a triangular distribution defined by the model's minimum,

mean, and maximum predictions across all feedstocks, without taking into account the fact that a given feedstock may perform differently than the average of all feedstocks. Similarly, many economic parameters were simply defined by a base price and a minimum and maximum value of the base price ± 30%. The utilised ML model was unable to predict detailed biochar properties which may influence factors such as the biochar's selling price. More training data will likely be necessary to develop ML models capable of predicting such factors. Ultimately, the authors found that biochar yield was the key parameter influencing the techno-economic and environmental analysis. The biochar yield itself was most significantly influenced by the pyrolysis temperature, with lower temperatures favouring higher biochar yields, despite lower temperatures also being associated with some adverse effects like a decrease in biochar HHV and carbon mass fraction **[132]**.

Baruah and Baruah (2021) developed a decision support system to analyse decentralised biomass gasification systems. The decision support system is implemented in Microsoft Excel and is made up of a data manipulation and database interface, where the user can impute relevant process data and set up the model, and an interface allowing the user to explore different scenarios. For this, the study incorporated biomass availability data and energy demand data for a representative rural village in India. Energy requirements were based on a previously conducted survey, whereas biomass production was estimated through a geospatial data model. Ultimately, by considering the specific feedstock supply and energy demand scenarios, Dhaincha was found to be the most suitable feedstock for the studied village **[179]**. Future research should consider incorporating LCA to also inform the user about the system's environmental implications.

Fernanda Rojas Michaga *et al.* (2022) assessed the techno-economic and environmental feasibility of jet fuel production from forestry residues via gasification and the subsequent Fischer-Tropsch process. By incorporating CCS, the authors showed that the proposed process could produce carbon-negative sustainable aviation fuel. The required selling price to ensure break even was only increased from 3.03 £/kg without CCS to 3.27 £/kg with CCS. CCS significantly improved the process' environmental performance. The authors highlighted the importance of policy to encourage the uptake of bioenergy technologies such as this one. In the British conditions studied by the authors, the jet fuel's minimum selling price decreased to 1.49 £/kg under the UK's "Renewable Transport Fuel Obligation" policy. Plant size was another factor that could drastically improve the system's economic performance due to economies of scale. However, one would have to ensure that sufficient feedstock is available locally for this to be a valid option **[180]**.

2.4.4 Monte Carlo Methodology to Account for Uncertainty within Models

Uncertainty analysis aims to quantify the effects of uncertainties in input data and methodological choices on a system's outputs. It is of particular importance for bioenergy systems due to their complexity and relative novelty. Especially, LCA and TEA typically require large amounts of data from various sources. Much of this data may require estimation and can be affected by unknown factors at the time of the analysis. Thus, understanding uncertainties is essential to ensure the effectiveness of models and avoid misleading conclusions **[170]**.

Sensitivity analysis is a simple way to address uncertainties in a model. It is a technique that assesses how the variation in the output(s) of a model can be attributed to variation in the inputs. This way, the effect of uncertainty or variability in the input parameters on the model's outputs can be understood. For TEA, some researchers suggested that variables that would lead to a 5% variation in net present value (NPV) or a 1% variation in internal rate of return for a 1% variation in the original variable should be considered **[181]**. Whilst sensitivity analysis is a useful tool for identifying the effects of the most influential parameters, it does not capture interactions between model inputs. For instance, it may be useful to assess worst- and best-case scenarios, where all uncertain input variables take on extreme values affecting the outputs in the same direction.

Monte Carlo simulation is a more comprehensive way to account for uncertainty than a simpler sensitivity analysis. It is a computational method to understand the impact of uncertainty and variability in a model. This is done by running a large number of randomised simulations where input variables are defined by probability distributions. Monte Carlo simulation results in a range of possible outcomes instead of a singular one. It is particularly useful when dealing with complex systems or models with multiple uncertain parameters. Through Monte Carlo simulation the likelihood of different outcomes can be quantified, allowing for more informed decisions under uncertainty.

Looking at gasification in particular, Barahmand *et al.* (2022) found that about half of the studies analysing the environmental, economic, and social impacts of gasification utilised a Monte Carlo or sensitivity analysis approach to account for uncertainties **[41]**. In general, Monte Carlo simulation and sensitivity analysis have been the most popular methods to account for uncertainties, but other methods have been proposed to address uncertainties within LCA and TEA. For instance, methods which are more discernible and have a reduced computational demand have been developed **[182]**. However, alternative methods may require more expertise from the practitioner and may not be as widely supported through existing LCA and TEA software.

In practice, researchers have started to integrate LCA and TEA with Monte Carlo methodology. For instance, in a previous work, the author used the Monte Carlo method-

ology in combination with TEA and LCA to assess the feasibility of MSW treatment through AD and gasification in Glasgow. Using Monte Carlo methodology, the likelihood of break-even was quantified across different scenarios. Additional sensitivity analysis was used to highlight the importance of particular input variables **[164]**. Similarly, Campbell *et al.* (2018) quantified the effects of volatile markets on the financial viability of biochar and biofuel production from forest biomass. Using probability distributions to describe key economic parameters, the authors found that the market prices of products most strongly affected the system's economic feasibility **[183]**. Fang *et al.* (2022) used a Monte Carlo approach to generate data for a kinetic gasification model linked to an RF model. For this, suitable probability distributions for key process parameters were explored, based on the realistic ranges of experimental data **[184]**.

Some authors used simple percentage increases/decreases from a baseline value to define process parameter distributions **[185, 186]**. This is useful to highlight trends in the data, but it does not accurately reflect reality. Other researchers have improved on this approach by collecting experimental or other empirical data to fit various probability distributions to process parameters **[137, 184, 187]**. When using empirical data, it is important to assess the quality of the distributions in isolation at the fitting stage and not at model evaluation. Tests like the Bayesian information criterion can be used **[187]** to compare distribution fits. Future case studies should make an effort to identify the boundary conditions that most accurately define the real-world probability distribution of a given parameter.

In summary, Monte Carlo methodology or other forms of uncertainty accounting are an essential tool for the comprehensive engineering analysis of bioenergy systems, and LCA and TEA in particular. As uncertainties during the environmental and techno-economic modelling of bioenergy systems are typically unavoidable, it is of the utmost importance to incorporate means to address and quantify these uncertainties. Many modern LCA and TEA frameworks and software incorporate Monte Carlo approaches, allowing for easy implementation by practitioners. Ultimately, the Monte Carlo methodology is an effective way to make models more robust and improve decision-makers' trust in model outputs.

2.5 Chapter Summary and Conclusions

This chapter reviewed the current academic literature relevant to this research, with a particular focus on the use of ML methods to model thermochemical processes. In general, ML models were found to be an effective means to model thermochemical processes, with the potential to improve on existing model types. However, data availability concerns and their lack of interpretability need to be addressed to maximise their usefulness. Data availability was also an issue for environmental and techno-economic modelling, highlight-

ing the importance of robust uncertainty accounting frameworks to address uncertainties due to data quality. The thorough accounting of model uncertainties also helps with improving decision-makers' trust in developed models, which in turn aids with accelerating commercialisation and industrial-scale implementation of technologies. Finally, gasification was found to be a promising technology to convert biomass and waste as it can balance environmental and economic requirements by adjusting its processing conditions. By developing more comprehensive and holistic modelling frameworks, trade-offs between these factors can be more accurately addressed.

Chapter 3

A Comprehensive Artificial Neural Network Model for Gasification Process Prediction

3.1 Chapter Overview

The viability and the relative merits of competing biomass and waste gasification schemes depend on a complex mix of interacting factors. Conventional analytical methods that are used to aid decision-making rely on a plethora of poorly defined parameters. In this chapter, a method is utilised that eschews the uncertainty in process representation by using a machine learning, data-driven, approach to predicting a set of ten key measures of gasification technology's performance. An artificial neural network (ANN) that is novel in its use of both categorical and continuous data inputs is developed, which makes it flexible and broadly applicable in assessing gasification process designs. It is the first model applicable to a wide range of feedstock types, gasifying agents, and reactor options. A strong predictive performance, quantified by a coefficient of determination (R^2) of 0.9310, was confirmed. The approach has the potential to simplify further system analysis, such as life cycle assessment (LCA) and techno-economic analysis (TEA) and thus allow for more transparency in the decisions made by policymakers and investors.

Section 3.2 introduces the chapter's background and discusses the most relevant previous research and how the work done as part of this chapter improves on this research. Section 3.3 details how the ANN model was developed. Particular attention is paid to data collection and data pretreatment, before the actual model development stage. The developed ANN model is discussed in Section 3.4 and its performance and optimal hyperparameter choices are put into the context of the existing literature. Finally, Section 3.5 summarises the most important findings of the chapter and the conclusions that can be drawn from them.

3.2 Introduction

The provision of sustainable energy and the transition to more sustainable waste treatment methods are two major challenges of the 21^{st} century. In 2018 The World Bank estimated an annual production of 2.01 billion tonnes of municipal solid waste (MSW), a third of which is not managed in environmentally safe ways **[188]**. Due to climate change concerns, non-fossil fuel alternatives for the generation of energy are of greater importance than ever. Biomass as an abundant resource has shown great potential to aid with this transition in both developed and developing countries **[189]**.

Thermochemical processes, such as gasification, represent a suitable way to tackle both problems by recovering energy from biomass and waste feedstocks **[79]**. Whilst the gasification of various biomass and waste feedstocks has been studied extensively, the technology's wider environmental, economic, and social impacts need to be better understood to allow for scale-up and efficient implementation. For this, accurate process models are essential. In the past thermodynamic equilibrium, kinetic, and computational fluid dynamics (CFD) models have widely and successfully been used for this purpose. A more recent alternative has been the use of machine learning (ML) models, such as artificial neural networks (ANNs), which can effectively map highly non-linear input-output relationships and thus have the potential to predict gasification outputs more accurately than conventional models **[3]**.

The use of ML within the field of bioenergy has just started gaining momentum. ML methods have been shown to be an effective tool to optimise thermochemical processes **[190]**. Prediction models can also ensure the generation of products with desired properties by identifying suitable operating conditions **[191, 192]**. Other opportunities for innovation through ML are, among others, on-line process control and waste-to-energy network design. For instance, image data has successfully been used to quantify the major gas species concentrations in a gasifier in real-time **[193]**. Robust data-driven optimisation models for waste-to-energy network designs can aid with finding the most optimal waste-to-energy networks in light of uncertainties such as varying feedstock prices **[194]**.

An ANN can refer to many different models. In this field a simple multilayer perceptron structure has generally been employed, which is made up of neurons arranged in an input and output layer and one or more hidden layers **[3, 107, 115, 116, 120, 122, 123]**. These existing studies showed that ANNs can predict gasification outputs with a high prediction accuracy. For example, George *et al.* (2018) trained an ANN model on lab scale data from a bubbling fluidised bed gasifier **[116]**. Their model was shown to predict the syngas production from local biomass with a high accuracy with a regression coefficient of R = 0.987 and mean squared error (MSE) of 0.71. Baruah *et al.* (2017) developed an ANN model to predict the H_2, CO, CO_2, and CH_4 yields from a downdraft gasifier **[115]**. The model predicted the outputs excellently, with coefficients of determination of $R^2 > 0.98$

and an average relative error of 2.65% when compared to experimental data.

However, existing studies were generally narrow in scope and used small data sets for model training. This means the developed models are often only applicable to a few select feedstocks **[116]**. Additionally, existing models are unable to predict the influences of different operational choices, such as the reactor design and gasifying agent, on the gasification process. Combined with the fact that many models do not predict all important outputs, this leads to limitations regarding the applicability of existing models towards optimal system and process design. For example, Serrano and Castelló modelled the tar production from the gasification of woody biomass with good prediction accuracy ($R^2 > 0.97$) using an ANN **[107]**. They achieved this by collating a data set of 120 samples from the literature. They reasoned that only samples with the same metadata (e.g. bubbling fluidised bed reactor, reactor size between 50 and 300 mm, and operation at atmospheric pressure) could be added to the model. Shenbagaraj *et al.* (2021) developed a model to predict a wider range of outputs (syngas composition and yield) with high accuracy of $R^2 > 0.97$ **[123]**. However, the developed model is specific to supercritical water gasification for a single feedstock type (food waste). The narrow scope of their work meant that only a small data set of 40 samples could be collected from the literature. Li *et al.* (2021) trained a gradient boosting model on a larger data set of 295 data points, but their model is only suitable for hydrothermal gasification of three different waste feedstocks **[195]**. Similarly, Kargbo *et al.* (2021) developed a model for a specific reactor design and feedstock combination. Their model was trained on a small data set of 40 samples obtained from a two-stage gasifier using waste wood **[196]**.

It has previously been identified that none of the existing ML-based gasification models are applicable across a wide range of feedstocks, gasifying agents, and reactor designs **[3]**. Hence, this work aims to significantly improve on existing studies by training a feedforward ANN on a large data set covering a broad range of feedstocks and operational conditions. For this, a highly generalisable model was developed. To the best of our knowledge, it is the first model applicable to a wide range of feedstock types (woody biomass, herbaceous biomass, plastics, municipal solid waste, and sewage sludge), gasifying agents (air, steam, oxygen, and combinations of these), and reactor options (fixed bed and fluidised bed). The model's generalisation capability has been maximised by the novel use of categorical data so that the model is highly practical for general gasification process design. Categorical variables (e.g. gasifying agent, reactor type, and bed material) allow factors that are critical to gasification, but which have been omitted by previous models, to be captured. To further improve the model's practicality for general process design, it was formulated to predict a wide array of outputs that are essential for further modelling such as life cycle sustainability assessment (LCSA), which studies a system's environmental, economic, and social impacts. This in turn allows policy makers and investors to make optimal

decisions by knowing the wider impacts of the system. The choice of ANN input features has been optimised. Furthermore, a novel optimisation superstructure to systematically compare and identify the best-performing network structures and hyperparameter choices was employed to improve on previous studies which employed trial and error methods.

3.3 Materials and Methods

3.3.1 Data Collection and Description of the Data Set

To train the proposed ANN model, a data set was collected from the literature. A broad and extensive search was conducted, drawing upon a large array of sources. Ultimately, 28 suitable sources were collated to create a data set of 312 samples **[197–225]**. To ensure a model with high generalisation capability, a wide range of feedstocks, gasifier types, and operating conditions were considered. The procedures for collecting data from literature were: (i) parameters were converted to the basis used in the data set (e.g. ultimate composition data given as wet basis (wb) would be converted to dry ash-free basis (daf) to fit in with the rest of the data set); (ii) the feedstock's lower heating value (LHV) was calculated from the feedstock's higher heating value (HHV) where necessary **[226]**; (iii) the quoted particle size refers to the lowest dimension of the particle (e.g. a pellet of $2{\times}2{\times}10$ mm would be quoted as 2 mm); (iv) all higher order hydrocarbons (C_2H_n) in the syngas were treated as C_2H_4 (as C_2H_4 is generally the dominant species **[218, 227]**); (v) cold gas efficiency (CGE) and carbon conversion efficiency (CCE) were calculated where necessary. Due to this, some efficiencies of $> 100\%$ were obtained. The complete data set may be obtained from the online version of the paper this Chapter is based upon (https://doi.org/10.1016/j.apenergy.2022.119289).

The process' cold gas efficiency (CGE) is given by

$$CGE = \frac{Y_G \mathrm{LHV}_G}{\mathrm{LHV}_F} \times 100\% \tag{3.1}$$

where Y_G is the syngas yield in Nm^3/kg of the feedstock on a wet basis. LHV_G and LHV_F represent the LHV of the produced syngas and feedstock, respectively. It indicates how much of the feedstock's LHV is converted to the syngas' LHV.

The process' carbon conversion efficiency is defined as the percentage of carbon in the feedstock converted to syngas and is given by

$$CCE = \frac{Y_G \left(\frac{12}{28} V_{CO}\rho_{CO} + \frac{12}{44} V_{CO_2}\rho_{CO_2} + \frac{12}{16} V_{CH_4}\rho_{CH_4} + \frac{24}{28} V_{C_2H_4}\rho_{C_2H_4}\right)}{C_F} \times 100\% \tag{3.2}$$

where Y_G is the syngas yield in Nm^3/kg of feedstock on a wet basis. V represents the

vol.% yields of different gas species and ρ is their respective density in kg/Nm^3. Finally, C_F is the feedstock's carbon percentage on a wet basis.

Tables 3.1 and 3.2 summarise the predictor and target variables collected for the database. In total 18 continuous and 8 categorical predictors were collected as part of the database. Three sets of predictors (i.e. ALL, OPT, and -CAT) were considered in this work to assess the effect of different parameter combinations on model training. Predictor set ALL uses all predictors for model training, except for predictors with a significant number of missing values (i.e. the feedstock's lignocellulosic composition (cellulose, hemicellulose, and lignin) and residence time). For predictor set OPT a range of additional parameters were excluded from the data set used for model training: (i) feedstock type and feedstock shape were excluded as it was assumed that the same information was already captured by the continuous feedstock composition data and feedstock particle size data, respectively; (ii) volatile matter content and fixed carbon content were excluded as they are considered dependent variables which are related to other compositional parameters **[115]**; (iii) feedstock LHV, feedstock oxygen content, and feedstock nitrogen content were excluded due to their strong correlations with other input parameters as identified in Section 3.3.4 Predictor set -CAT is used to assess the effect of removing categorical variables from the model. Hence set -CAT is a copy of set OPT but excludes all categorical variables and the remaining set only contains continuous variables. The three different predictor combinations considered in this work are summarised in Table 3.3.

The target variables used in the model are all variables listed in Table 3.2, except for the process' CGE and CCE. These were not considered as target variables during modelling, as they often could be estimated based on other predictor and target variables. In other words, they were considered functions of some of the other predicted parameters.

3.3.2 Data Pretreatment

3.3.2.1 Encoding of Categorical Variables

As previously mentioned, the use of a range of categorical variables to improve the ANN's prediction accuracy is a major point of novelty of this work. Unlike other ML methods, such as random forest, categorical variables cannot be directly fed into an ANN **[104]**. Thus, it is necessary to develop methods through which the categorical data can be represented by numerical values.

To the best of the authors' knowledge, only Serrano *et al.* (2020) previously made use of categorical variables to improve a gasification model **[120]**. In their study, the bed material of a fluidised bed gasifier was considered a categorical variable. The authors encoded this as an ordinal variable. However, the values of the variable follow no intrinsic order or scale and, therefore, using an encoding method such as one-hot encoding would

Table 3.1: Summary of predictor variables.

	Min	Max	Mean	SD*	No. of Samples
Feedstock Information					
Type	Woody biomass, Herbaceous biomass, Sewage sludge, Municipal solid waste, Plastics, Other				312
Shape	Chips, Dust, Fibres, Particles, Pellets, Other				312
Particle Size [mm]	0.25	70.00	4.82	6.71	270
LHV [$MJ/kg\ wb$]	11.50	42.90	18.44	6.55	312
Ultimate Analysis					
C [% daf]	40.07	86.03	51.42	9.23	312
H [% daf]	3.79	14.23	6.86	2.00	312
N [% daf]	0	7.32	1.07	1.87	312
S [% daf]	0	1.61	0.32	0.45	279
O [% daf]	0	53.40	39.89	10.88	312
Proximate Analysis					
Ash [% db]	0.27	44.00	7.43	11.90	312
Moisture [% db]	0	27.00	8.37	5.65	312
Volatile Matter [% db]	56.00	89.11	77.56	8.98	264
Fixed Carbon [% db]	9.07	23.82	15.60	3.18	222
Lignocellulosic Composition					
Cellulose [% db]	29.60	46.20	39.42	5.08	28
Hemicellulose [% db]	14.00	29.60	20.91	4.67	28
Lignin [% db]	14.00	33.00	24.99	5.39	28
Gasifier Conditions and Information					
Temperature [°C]	553.00	1050.00	799.13	80.88	312
Operation Mode	Batch, Continuous				312
Residence Time [min]	6.00	403.00	98.34	97.30	124
Steam/Biomass [wt/wt]	0	4.04	1.02	0.73	133
ER	0.09	0.87	0.30	0.11	235
Gasifying Agent	Air, Air + Steam, Air + Oxygen, Steam, Oxygen, Other				312
Reactor Type	Fixed-bed, Fluidised bed, Other				312
Bed Material	Silica, Alumina, Calcium Oxide, Olivine, Dolomite				234
Catalyst	Present, Not present				312
Scale	Lab, Pilot				312

* Standard deviation

Table 3.2: Summary of target variables.

	Min	Max	Mean	SD*	No. of Samples
Syngas Composition					
N_2 [$vol.\%\ db$]	0	74.90	38.16	25.76	286
H_2 [$vol.\%\ db$]	3.10	73.85	22.49	17.28	312
CO [$vol.\%\ db$]	2.20	50.00	19.43	9.59	309
CO_2 [$vol.\%\ db$]	0	38.25	16.10	6.45	309
CH_4 [$vol.\%\ db$]	0.25	16.00	5.17	3.30	309
C_2H_n [$vol.\%\ db$]	0	9.50	2.06	1.45	229
Syngas LHV [$MJ/Nm^3\ db$]	1.30	15.56	7.19	3.47	312
Tar content [$g/Nm^3\ db$]	0	364.00	29.89	47.85	193
Yield and Process Efficiency					
Syngas yield [$Nm^3/kg\ wb$]	0.49	6.19	1.76	0.96	278
Char yield [$g/kg\ wb$]	3.60	301.52	94.56	76.49	87
CGE [%]	16.70	122.87	61.27	20.76	278
CCE [%]	20.61	126.89	70.98	17.35	275

* Standard deviation

Table 3.3: Predictor combinations used for model training.

	Included Predictor Variables
Set ALL	Carbon, Hydrogen, Sulphur, Nitrogen, Oxygen, Ash, Moisture, Volatile Matter, Fixed Carbon, Particle Size, LHV, Temperature, ER, Steam to Biomass Ratio, Operation Mode, Catalyst Use, System Scale, Feedstock Type, Feedstock Shape, Gasifying Agent, Reactor Type, Bed Material
Set OPT	Carbon, Hydrogen, Sulphur, Ash, Moisture, Particle Size, Temperature, ER, Steam to Biomass Ratio, Operation Mode, Catalyst Use, System Scale, Gasifying Agent, Reactor Type, Bed Material
Set -CAT	Carbon, Hydrogen, Sulphur, Ash, Moisture, Particle Size, Temperature, ER, Steam to Biomass Ratio

seem more appropriate. Thus, here we used one-hot encoding for our nominal variables. A categorical variable which can take on n distinct categorical values is transformed into n binary variables, where values of 1 and 0 indicate the presence and absence of the variable, respectively. This means a categorical input variable with five categories will require five input neurons. An example of this is shown by Table 3.4 for the variable gasifying agent. The variable gasifying agent can take on any of the following six categories: *Air*, *Air + Steam*, *Air + Oxygen*, *Steam*, *Oxygen*, and *Other*. When using one-hot encoding the variable is converted into six new binary variables (v1-v6). The following variables were encoded using this method: feedstock type, feedstock shape, gasifying agent, reactor type, and bed material.

Table 3.4: Example of one-hot encoding for the variable gasifying agent.

	Gasifying Agent					
Variable	Air	Air + Steam	Air + Oxygen	Steam	Oxygen	Other
v_1	1	0	0	0	0	0
v_2	0	1	0	0	0	0
v_3	0	0	1	0	0	0
v_4	0	0	0	1	0	0
v_5	0	0	0	0	1	0
v_6	0	0	0	0	0	1

All categorical variables that can only take on two values (e.g. the process' operation mode (batch/continuous), system scale (lab/pilot), and catalyst use (present/not present)) were encoded as binary values. This means the variable can take on a value of 0 or 1 (-1 and 1 after normalisation). In the current database all data came from lab or pilot

scale systems and, therefore, the system scale was encoded using this method. One-hot encoding or ordinal encoding could be employed if data from e.g. industrial-scale plants were to be added. Ordinal encoding simply encodes each label as an integer, retaining the feature's order. Many papers which were used to compile the presented database simply stated whether the studied system was of lab/bench or pilot scale. However, some papers quoted the system's power output instead. In this case, a system was considered of lab scale for power ratings of $< 100\ kW_e$ or $< 150\ kW_{th}$. A pilot scale system was defined as $< 1\ MW_e$ or $< 1.5\ MW_{th}$.

It is obvious that the addition of a catalyst can strongly affect the gasification process. Various types of catalysts have been shown to improve factors such as process efficiency, tar removal, or hydrogen yield **[199, 220, 228, 229]**. However, it can be hard to quantify the exact effect of a catalyst numerically. Hence the authors chose a simplified model which encodes the use of a catalyst as a binary as previously explained (i.e. 0 for no catalyst present and 1 for catalyst present). This way the network can draw useful information and learn from whether a catalyst is present during the process.

3.3.3 Data Normalisation

Data normalisation is an essential prerequisite for ANN modelling because it speeds up learning and ensures that all input features are within a common range. Equation 3.3 has been used to normalise the data set between -1 and 1:

$$x_{\text{norm}} = l + \left(\frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \right) (u - l) \tag{3.3}$$

where x is the observed value and x_{norm} is the normalised value. x_{max} and x_{min} represent the maximum and minimum values of the parameters, respectively. l and u are the lower and upper bound values between which the data is to be normalised. In this work, a lower and upper bound of -1 and 1 were selected.

3.3.3.1 Mean Substitution of Missing Values

As the database used for model training was collected from many studies (29), it is unavoidable that there are inconsistencies in the format and types of data reported and some studies do not quote all parameters collected as part of the database. However, missing values can lead to issues during model training. For instance, the authors noticed a large loss of target-prediction pairs when a significant number of missing values was apparent in the training data set. This means missing predictor data can result in the ANN being unable to make a prediction. In turn this led to fewer comparisons between the network's predicted value and the experimental target value which caused issues regarding the reproducibility of results. For this reason, the mean substitution method was employed for all

predictors before model training, where a variable's missing values are replaced with the variable's mean value **[230]**. As the testing stage is the key stage of interest to evaluate the model's performance, this method was not employed for target variables and thus only affected model training.

3.3.4 Preliminary Data Analysis

Some initial analysis of the raw data was conducted by calculating the mean and standard deviation (SD) of variables as reported in Tables 3.1 and 3.2. Then Spearman's correlation coefficients (SCC) were calculated for the continuous predictor data to measure the monotonic relationship between data pairs. SCC was employed instead of the more commonly used Pearson correlation coefficient as the data was not known to follow a Gaussian distribution and SCC is less affected by outliers. Equation 3.4 was used to calculate the SCC:

$$\mathrm{SCC} = \frac{\sum_{i=1}^{N}(R(x_i) - \bar{R}(x))(R(y_i) - \bar{R}(y))}{\sqrt{\sum_{i=1}^{N}(R(x_i) - \bar{R}(x))^2}\sqrt{\sum_{i=1}^{N}(R(y_i) - \bar{R}(y))^2}} \tag{3.4}$$

where N is the sample size, ($R(x_i)$ and $R(y_i)$ are the ranks of individual samples of the two tested variables, and $\bar{R}(x)$ and $\bar{R}(x)$ indicate the mean ranks of the two variables.

A SCC of 0 indicates that two parameters are not correlated. The closer to ± 1 the stronger the monotonic relationship. In this work, a SCC of $\geq|\ 0.6\ |$ was used to indicate a strong relationship which would lead to the removal of one of the parameters before model training. Figure 3.1 shows the SCC matrix of all continuous predictors, except for parameters with a significant number of samples missing. The analysis of the SCC led to the subsequent removal of the feedstock's nitrogen content due to a significant correlation with the feedstock's sulphur and ash contents. Additionally, the feedstock's LHV and oxygen content were removed due to their correlations with the feedstock's carbon content. Highly correlated features are removed before training as they contain similar information and retaining both generally reduces model performance by adding additional noise to the model **[231]**.

3.3.5 Artificial Neural Network Development

3.3.5.1 Artificial Neural Network Background

Conceptually, an ANN is a non-linear model which can fit any given function making it a universal function approximator **[232]**. It can be used to predict an arbitrary number of response variables given one or more explanatory variables. The type of ANN utilised in this work is often referred to as a feed-forward multilayer perceptron (MLP). An MLP is

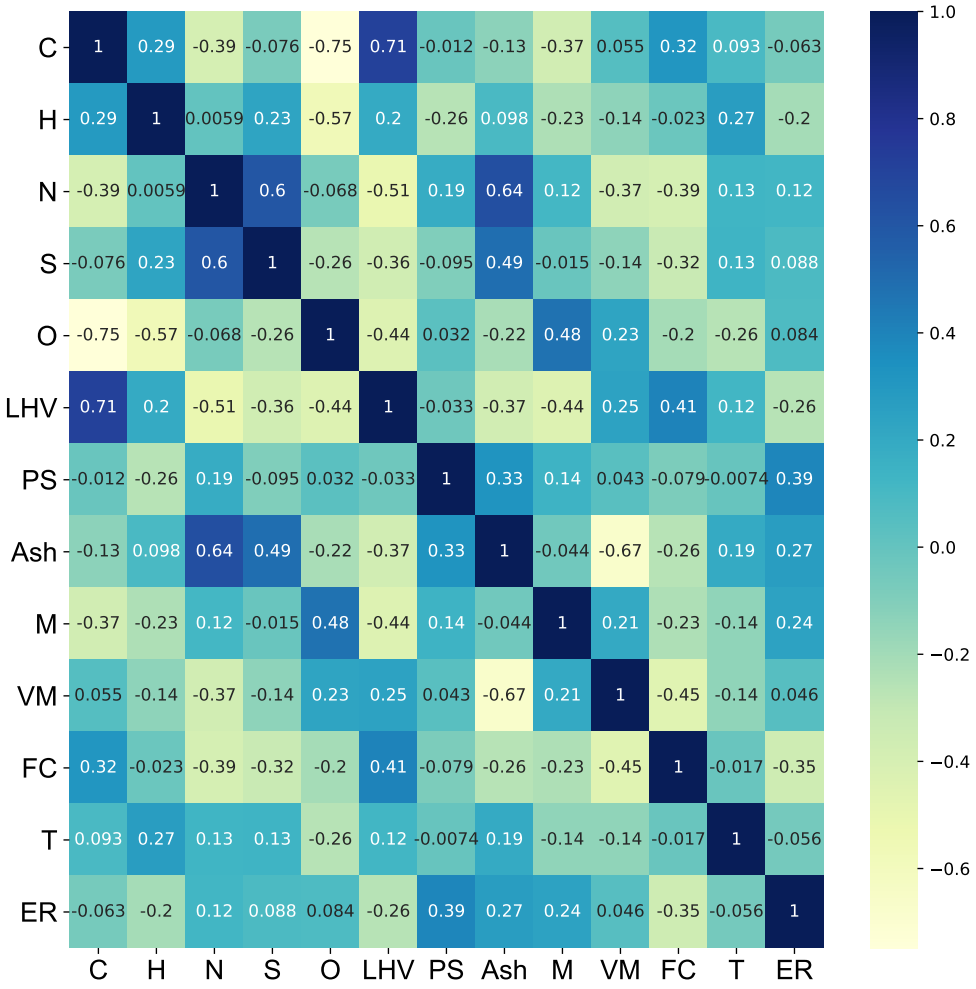


Figure 3.1: Spearman correlation coefficient matrix of the following continuous input features: C- Carbon, H –Hydrogen, N –Nitrogen, S –Sulphur, O –Oxygen, LHV –LHV, PS – Particle Size, M – Moisture, VM – Volatile Matter, FC – Fixed Carbon, T – Temperature, SB – Steam to Biomass Ratio, and ER – Equivalence Ratio.

made up of an input and output layer and one or more hidden layers. Each layer contains nodes with weighted connections which allow for the transfer of information from one layer to the next.

Training is generally achieved through the backpropagation of errors **[233]**. This means, through an iterative process the optimum weights which minimise the prediction error for an independent test data set are found. Over the years various training algorithms have been developed. In this work, Levenberg-Marquardt backpropagation (LM) and Bayesian regularisation backpropagation (BR), which are implemented as part of the MATLAB deep learning toolbox, were considered **[234–236]**. The output of a fitted ANN is given by Equation 3.5

$$y_k = f_o\left(\sum_h w_{hk} f_h\left(\sum_i w_{ih} x_i\right)\right) \tag{3.5}$$

where y_k is the network's estimation of the response variable **[237]**. This is calculated from the sum of products of the weights w for i input variables x for the hidden nodes h. These are transferred by the activation functions f_o and f_h for the output and hidden nodes, respectively **[238]**.

3.3.5.2 Artificial Neural Network Optimisation and Performance Assessment

To find the best possible network configuration, five essential hyperparameters affecting the performance of the ANN were varied. The optimised parameters were: (i) number of neurons in 1^{st} hidden layer, (ii) number of neurons in 2^{nd} hidden layer, (iii) data split for test set, (iv) training function, (v) hidden layer transfer function. Initial tests resulted in the use of the parameter ranges and options indicated in Table 3.5. A test data split of 15% has frequently been employed in literature, hence a range from 10-20% was selected for optimisation purposes **[3]**. By initially considering various training functions, LM and BR were found to be the two preferred options which were further considered in the optimisation routine. Furthermore, initial tests allowed for the elimination of all but the hyperbolic tangent sigmoid (TANSIG) and logistic sigmoid (LOGSIG) functions as hidden layer transfer function options. The step sizes (given in brackets in Table 3.5) for the numerical parameters (i)-(iii) were selected by weighing the required accuracy against computational demands. Other parameters, such as the output layer's transfer function and the network's performance function were kept constant throughout and are also given in Table 3.5. Finally, 5-fold cross-validation was used to minimise the risk of overfitting and improve confidence in the results.

The model's prediction performance was measured by considering the root mean square error (RMSE), R^2, and adjusted coefficient of determination (R^2_{adj}). R^2_{adj} is a modified version of the R^2 which considers the number of predictors in the model. Whilst R^2

Table 3.5: Optimised and fixed hyperparameters used during ANN training.

Parameters	Type/Value
Optimised parameters	
No. neurons 1^{st} hidden layer	2-24 (2)
No. neurons 2^{nd} hidden layer	0-25 (5)
Split test set	10-20% (2.5%)
Training function	LM, BR
Hidden layer transfer function	TANSIG, LOGSIG
Fixed parameters	
Performance function	MSE
Learning rate	0.1
Output layer transfer function	Linear function
Maximum number of learning epochs	1000

generally increases with the addition of new predictors, despite potentially not adding any explanatory power, this is not the case for R^2_{adj}. Thus, by using R^2_{adj} an increase in R^2 due to the addition of unnecessary predictors is avoided, allowing for a fairer comparison between models trained with different input features. Equations 3.6, 3.7, and 3.8 describe the three performance indicators:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i^o - y_i^p)^2} \tag{3.6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i^o - y_i^p)^2}{\sum_{i=1}^{N}(y_i^o - \bar{y}^o)^2} \tag{3.7}$$

$$R^2_{\text{adj}} = 1 - \frac{(1-R^2)(N-1)}{N-k-1} \tag{3.8}$$

where N is the number of samples and y_i^o and y_i^p are the observed and predicted values, respectively. $\bar{y}^o$ and k represent the mean of all predicted values and the number of explanatory variables in the model (i.e. number of predictor variables used for model training), respectively.

3.3.5.3 Workflow Summary

This study's workflow is illustrated by Figure 3.2. The figure shows how the various steps and procedures introduced throughout Section 3.3 are linked to each other. It is shown how continuous and categorical variables had initially been pretreated separately before they were combined again after the normalisation step. At this point, the data was ready for model training. Data was fed into the optimisation and cross-validation stage of this work which is shown by the dashed box in the bottom half of the figure. Here the

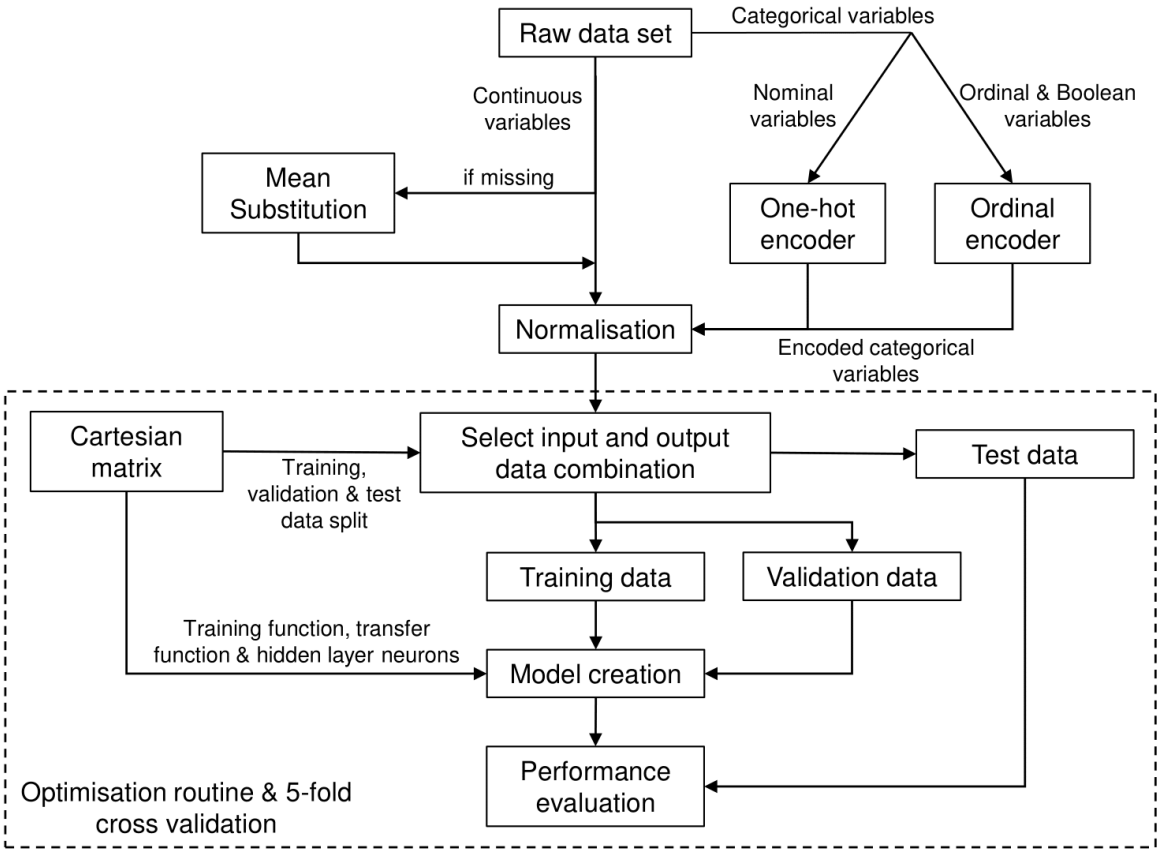


Figure 3.2: Flowchart of the procedure employed for the model development.

Cartesian matrix contains all possible hyperparameter combinations introduced in Section 3.3.5.2. Models were trained and cross-validated for each of these combinations until the code had iterated through all combinations. Finally, the trained models' performance was evaluated by considering their RMSE, R^2, and R^2_{adj} to find the best-performing model. All data analysis and modelling work has been implemented in the MATLAB programming environment **[239]**.

3.4 Results and Discussion

3.4.1 Preliminary Data Analysis

The raw data's range, mean, and SD were calculated and are shown by Tables 3.1 and 3.2. Some of the key findings resulting from this preliminary analysis shown in the Tables are discussed here. When considering the syngas composition, it was found that the nitrogen content showed the highest variation of all variables with an SD of 25.76 vol.% db. This can be explained by the use of various gasifying agents which result in syngas with vastly different nitrogen contents. The syngas LHV is similarly varied for the same reason. For

instance, using air as a gasifying agent produces syngas of a lower LHV, as compared to steam or oxygen, due to a strong dilution by nitrogen.

Studying the gasifier's operational conditions reveals a high variation in temperatures with an SD of 80.88°C. This was expected as the temperature is one of the key factors widely varied during many experimental studies. The ER on the other hand showed very little variation with an SD of 0.11. Many studies considered an ER of 0.30 which was also found to be the parameter's mean value.

Looking at the feedstocks' compositional data the particle size showed some outliers (70 mm); however, most of the samples were >10 mm which led to a small SD of 6.71 mm. When studying the feedstocks' elemental composition, the carbon contents were found to differ heavily with an SD of 9.23% daf. The high SD can be explained by the large range of different feedstocks being collected as part of the database. Whilst plastic waste was found to have carbon contents of up to 86.03% daf, most herbaceous biomass feedstocks have carbon contents $< 50\%$ daf.

The number of samples collected for a given variable is another important factor to consider. ML relies on the availability of large amounts of data and models tend to have a harder time learning the relationships affecting a variable when fewer samples are available. For instance, gasification is mainly considered a technology to produce syngas. Hence, many authors did not report on the process' char yields as shown by the relatively small number of samples (87 total) collected for this variable. This in turn was found to diminish the model's prediction accuracy for this variable as further discussed in Section 3.4.2.

3.4.2 Artificial Neural Network Model Optimisation

Figure 3.3 shows the five-fold cross-validated results of the best performing model structures trained on the three different data sets described in Section 3.5. Results are shown in normalised form as they are shown across all output variables. Considering the best-performing model trained with all predictors collected as part of the database (set ALL) resulted in a well-performing model with RMSE = 0.1283, $R^2 = 0.9329$, and $R^2_{adj} = 0.9199$. This is clearly illustrated by the fact that the data points in Figure 3.3a are tightly distributed around the dotted perfect fit line with a unity slope. Using an optimised selection of predictors (set OPT) yielded a model with a similarly high prediction accuracy and low error (RMSE = 0.1307, $R^2 = 0.9310$, and $R^2_{adj} = 0.9254$). Generally, a model with more predictors will spuriously achieve a higher R^2 than a model with fewer predictors. However, when considering R^2_{adj} instead, where excessive predictors are penalised, the model trained on set OPT was found to be the best-performing model. The removal of all categorical variables (set -CAT), as shown in Figure 3.3c, led to a significant drop in model accuracy as illustrated by the wider dispersion of data points from the perfect fit line and

increased RMSE = 0.2126 and decreased R^2 = 0.8016 and R^2_{adj} = 0. 7959. This highlights the importance of incorporating key categorical variables to improve the model's accuracy. In conclusion, these findings indicate that the use of a well-optimised combination of predictors, including meaningful categorical variables, can significantly increase the prediction accuracy of an ANN model. A summary of these findings is presented in Table 3.6.

Table 3.6: Summary of performance indicators and hyperparameter choices for best performing models.

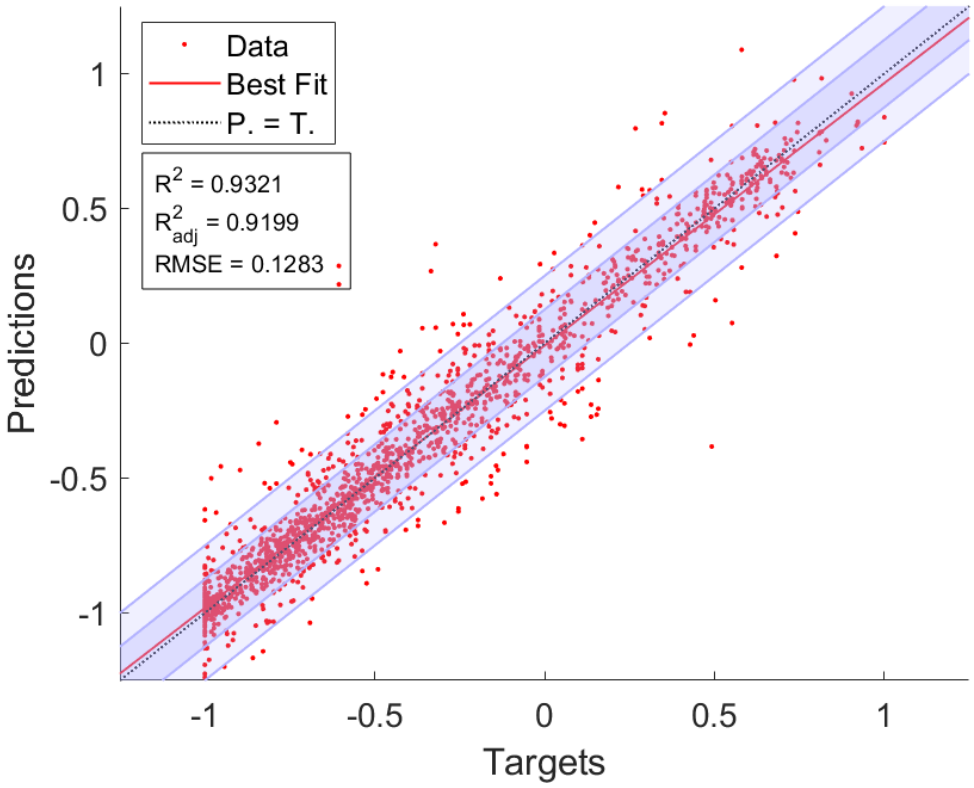
	Performance Indicators[a]				Hyperparameters			
	RMSE	R^2	R^2_{adj}	Training time [s]	Hidden layer neurons	Training function	Hidden layer transfer function	Test split [%]
Set ALL	0.1283 (0.0145)	0.9321 (0.0153)	0.9199 (0.0175)	29.53 (9.64)	8, 10	BR	TANSIG	12.5
Set OPT	0.1307 (0.0194)	0.9310 (0.0289)	0.9254 (0.0305)	43.40 (16.15)	18	BR	TANSIG	12.5
Set - CAT	0.2126 (0.0280)	0.8016 (0.0540)	0.7957 (0.0557)	4.35 (1.44)	20, 25	LM	LOGSIG	12.5

a – values in brackets indicate the SD over all cross validation runs

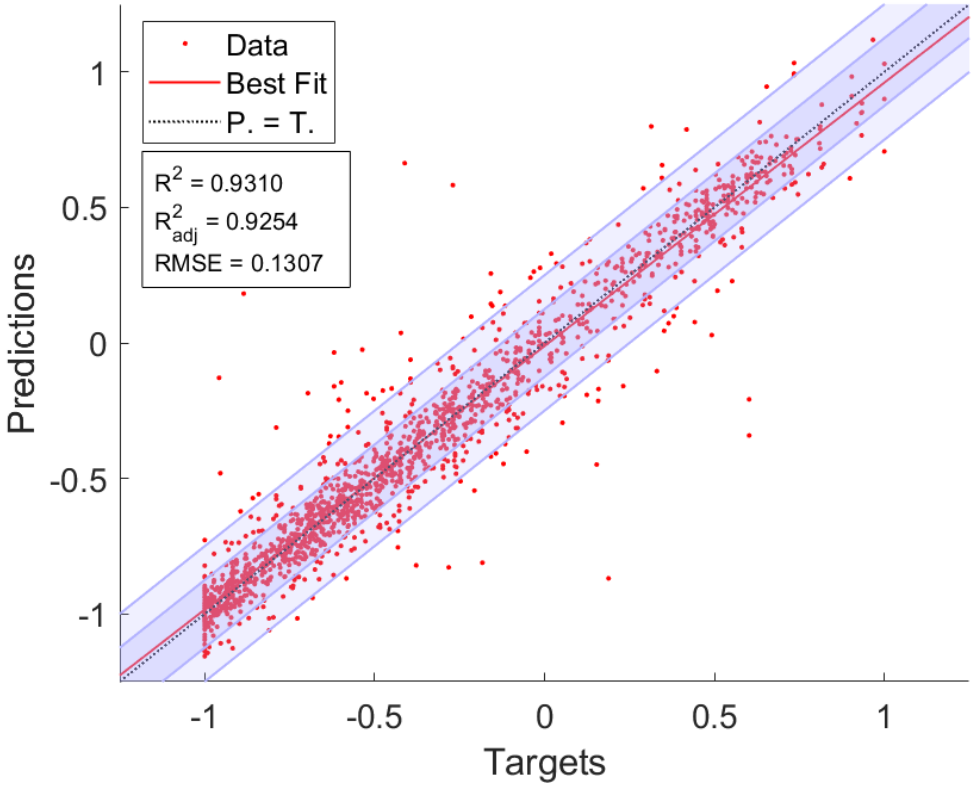
Based on these findings, set OPT was considered the most suitable combination of predictors for model training and the model's optimised structure is illustrated by Figure 3.4. Hence models trained on this data set were further evaluated. When considering the next best-performing network structures, it becomes clear that various network structures and hyperparameter choices can result in a well-performing model. For example, a model trained with a double hidden layer structure of 24 and 15 neurons utilising the LM training function and a model trained with the BR training function and a single hidden layer with 16 neurons resulted in a similar predictive performance to the best-performing model with $R^2 > 0.92$ and RMSE < 0.14. This agrees with existing literature where both BR and LM were found to be suitable training functions for the ANN modelling of various thermochemical processes **[95, 98, 120, 240]**. Furthermore, various data splits for testing and validation ranging from 12.5-17.5% were found to produce acceptable results. Finally, the TANSIG hidden layer transfer function was found to generally be the best choice for the trained models.

A limitation of this work is the use of lab and pilot scale data only for model development. At the time of development, no industrial scale data has been available to train the model on. It remains uncertain how well the model performs on unseen industrial-scale data. However, through the novel use of categorical variables, the presented approach could easily be adapted were industrial scale data to become available as highlighted in Section 3.3.2.1.

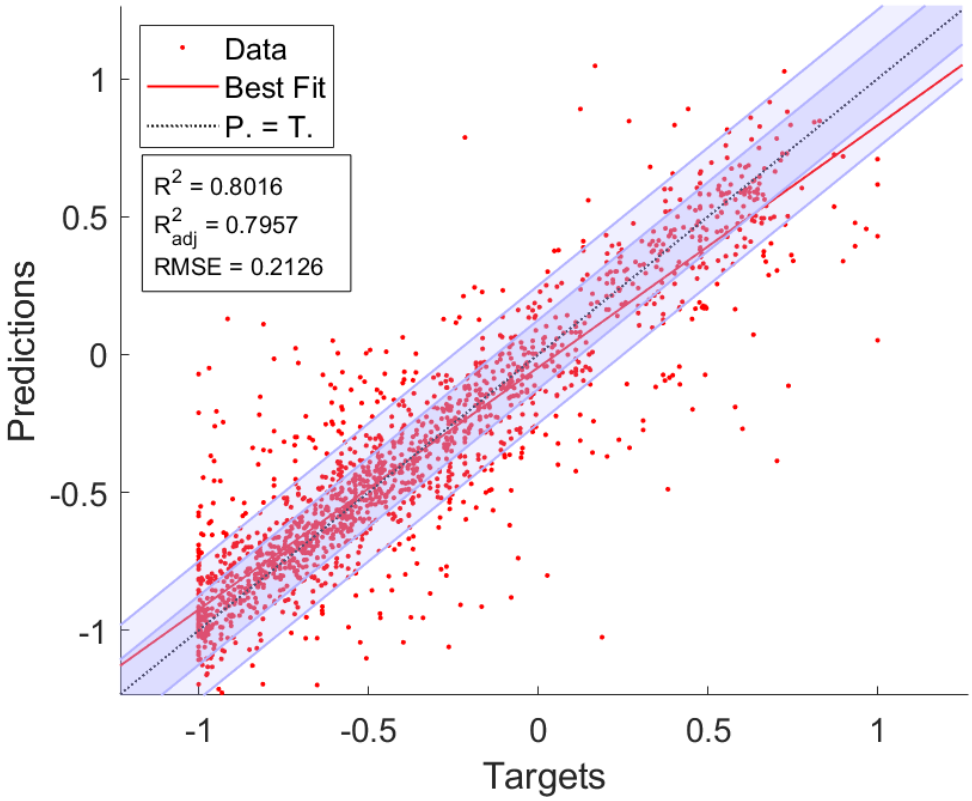
In general, adding a 2^{nd} hidden layer was not shown to improve the model's prediction


Data
Best Fit
P. = T.
R^2 = 0.9321
R^2_{adj} = 0.9199
RMSE = 0.1283
Predictions
Targets
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1

(a)


Data
Best Fit
P. = T.
R^2 = 0.9310
R^2_{adj} = 0.9254
RMSE = 0.1307
Predictions
Targets
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1

(b)



(c)

Figure 3.3: Distribution of normalised model predictions against target values for all 5 cross-validation runs of the best-performing model. (a), (b), and (c) show the results for predictor combinations set ALL, set OPT, and set -CAT respectively. The black dotted line represents the perfect fit line for the case where all predictions would equal their target values, and the solid red line indicates the best fit to the data points. The inner blue region shows the 10% error range, whereas the lighter outer region shows the 20% error range.

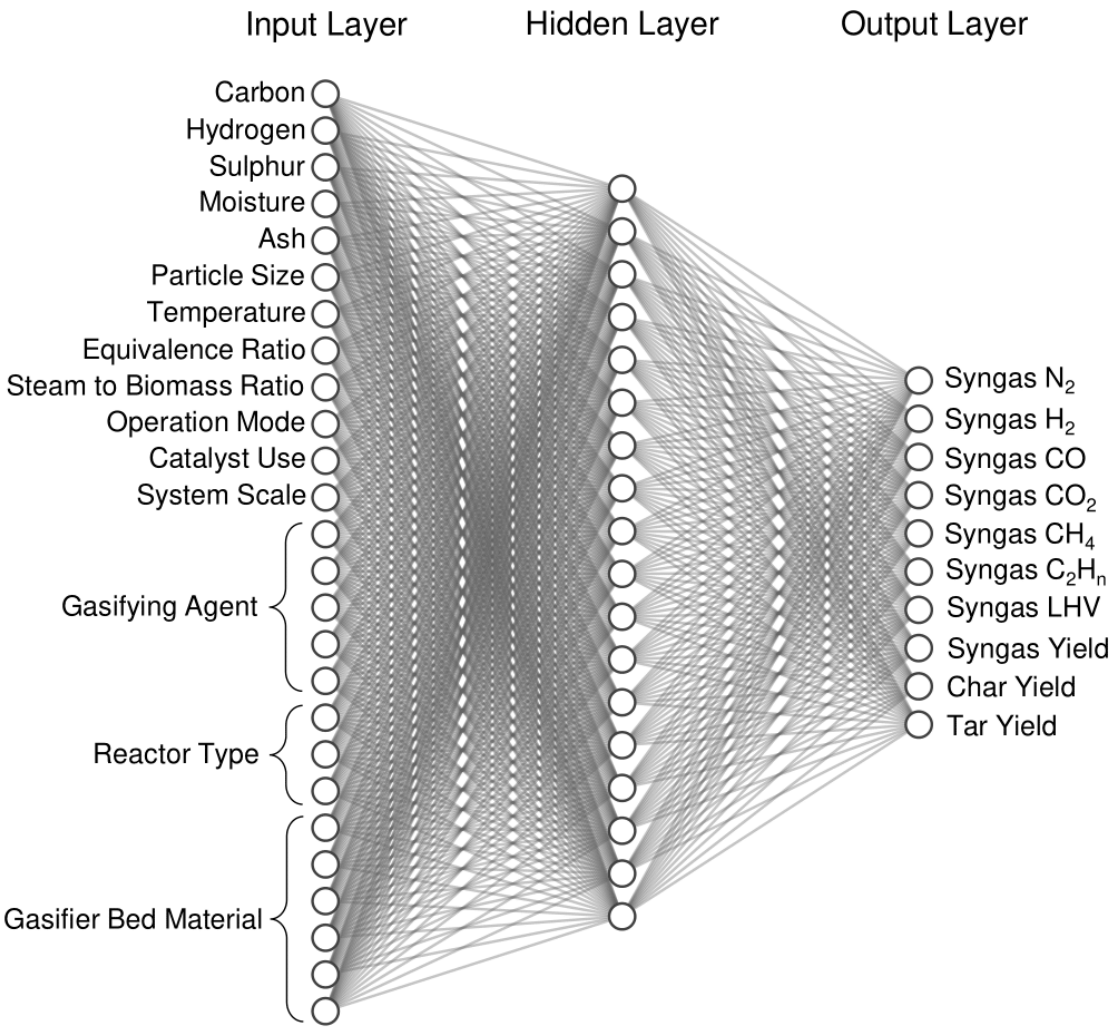


Figure 3.4: Network structure of best performing ANN trained on set OPT.

accuracy. This agrees with theory and the universal approximator theorem which states that a single hidden layer ANN is sufficient to represent any given function arbitrarily well **[232]**. Thus, the increased complexity from adding a 2^{nd} hidden layer cannot be justified. When it comes to the choice of training function, it was found that the BR function generally resulted in a better-performing model than LM. This was in particular the case for single hidden layer network structures, albeit at the cost of an increased computational demand.

Figure 3.5 shows the prediction accuracy for all ten individual target outputs for the best-performing model structure trained with set OPT. The model was found to predict some outputs with greater accuracy than others, as highlighted by the wider dispersion of data points for some parameters. The performance indicators for the individual model outputs are summarised in Table 3.7. The findings shown in the table highlight that some factors (e.g. the process' syngas yield and H_2 content) can be predicted with higher accuracy than others (e.g. char yield and syngas C_2H_n content). The highest R^2 of 0.9791 and lowest RMSE of 0.0731 were obtained for the syngas N_2 content and tar yield, respectively. Conversely, the lowest R^2 of 0.5202 and highest RMSE of 0.3020 were obtained for the syngas C_2H_n content and char yield, respectively. Understanding and accounting for the uncertainty in certain predictions is essential for further system analysis such as LCSA.

Table 3.7: Model performance indicators by output.

	N_2	H_2	CO	CO_2	CH_4	C_2H_n	Gas LHV	Gas yield	Char yield	Tar yield
RMSE	0.1047	0.1094	0.1084	0.1078	0.1796	0.1883	0.0989	0.0821	0.3020	0.0731
R^2	0.9791	0.9515	0.9261	0.8954	0.7990	0.5202	0.9571	0.9558	0.7106	0.9190
R^2_{adj}	0.9770	0.9471	0.9193	0.8858	0.7805	0.4584	0.9532	0.9512	0.5852	0.9063

When considering R^2_{adj}, only a minor decrease from R^2 was found for the outputs which the model predicted well. However, the two outputs which the model predicted poorly, the syngas C_2H_n content and char yield, showed a significant decrease. The former decreased by 13% and the latter by 21%. This decrease and poor performance can be explained by the significantly smaller sample size available for the training of these two features compared to other features. For instance, only 87 samples could be collected for the process' char yield, whereas the data set contained 312 samples for the excellently predicted features syngas LHV and H_2 content. Adding additional samples for these outputs would likely allow the model to better learn the relationships affecting the outputs, leading to an increase in prediction accuracy.

As previously mentioned, one of the key aims of this study was to create a highly generalisable model which is suitable across a wide range of feedstocks and gasifier design choices. For this reason, a like-for-like comparison between existing studies and this work is not always possible or sensible due to vastly different scopes and aims. Existing studies

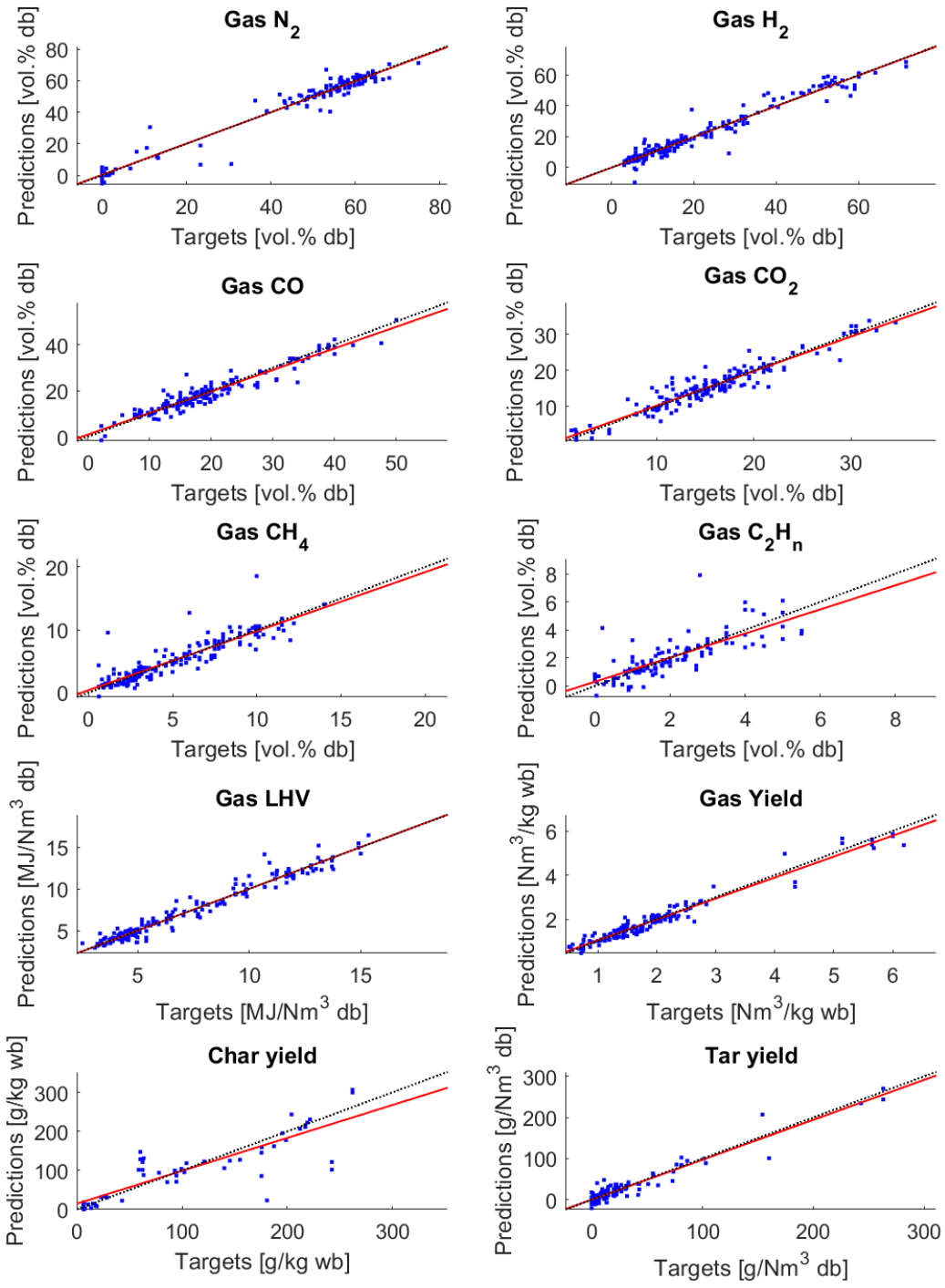


Figure 3.5: Distribution of model predictions against target values by output for all 5 cross-validation runs for the model trained using set OPT.

often developed gasifier-specific models or models specific to a select few feedstocks **[134, 135, 241]**. For instance, Kargbo *et al.* (2021) optimised a two-stage gasification process using an ANN **[196]**. Their model was found to generally predict most gasification outputs excellently with $R^2 = 0.99$. However, the tar production was predicted less well with R^2 of 0.74–0.82. Whilst their model generally performs better than the model presented in this study, their model is only applicable to the specific two-stage gasifier design used in the study. Furthermore, their model was trained on a small data set of 40 samples containing only a single feedstock (woody biomass). This limits the model's use to these specific conditions. Elmaz *et al.* (2020) applied four regression techniques to data obtained from a 10kW downdraft fixed-bed gasifier **[94]**. Decision tree regression and MLP were found to predict the gasification outputs best with $R^2 > 0.9$. Serrano *et al.* (2020) trained ANN models for different gasifier bed material options with a prediction accuracy of $R^2 = 0.90$-0.97 **[120]**. Whilst the mean absolute percentage error (MAPE) for a model considering only silica sand as bed material was acceptable (MAPE = 2.25-24.89%), larger errors (MAPE = 8.18-38.91%) were obtained for a model designed for a range of bed material options. In comparison, a MAPE of 11.14% was obtained in this work when averaged across all outputs for a model suitable across a range of bed material options.

The developed model could be used for a variety of applications. Its power to make predictions over a wide range of feedstock types, gasifying agents, and reactor options makes it particularly suitable for modelling and designing optimal gasification systems and process conditions. Since the model can predict more gasification outputs than existing models, it has the potential to simplify further modelling and analysis relying on these outputs. For instance, techno-economic analysis or LCSA could be streamlined by using the predictions made by the model, which can be further combined with multi-objective optimisation to decide on optimal gasification process conditions.

3.5 Chapter Summary and Conclusions

A highly generalisable artificial neural network (ANN) model of excellent prediction accuracy ($R^2 = 0.9310$ and RMSE = 0.1307) has been developed to model the gasification of biomass and waste. The developed model is the first model able to predict gasification outputs over a wide range of feedstock, reactor, and gasifying agent options. Thus, investors and policymakers can quickly and conveniently compare a multitude of system options and their performance. This in turn reduces the need for costly and time-consuming experimental studies. Since the model predicts many key gasification outputs, the suggested approach has the potential to facilitate the development of integrated gasification design models by combining the developed model with e.g. life cycle sustainability assessment (LCSA), cost-benefit analysis, and multi-objective optimisation.

Chapter 4

Interpretable Machine Learning to Model Biomass and Waste Gasification

4.1 Chapter Overview

Machine learning (ML) has been regarded as a promising method to better model thermochemical processes such as gasification. However, their black-box nature can limit how much one can trust and learn from the developed models. Here seven different ML methods have been adopted to model the gasification of biomass and waste across a wide range of operating conditions. Gradient boosting regression has been found to outperform the other model types with a coefficient of determination (R^2) of 0.90 when averaged across ten key gasification outputs. Global and local model interpretability methods have been used to illuminate the developed black-box models. The studied models were most strongly influenced by the feedstock's particle size and the type of gasifying agent employed. By combining global and local interpretability methods, the understanding of black box models has been improved. This allows policymakers and investors to make more educated decisions about gasification process design.

Section 4.2 introduces the chapter's background and discusses how interpretability methods are needed to improve upon previously developed ML models. The chapter's general workflow is introduced in Section 4.3. Furthermore, the considered ML methods and interpretability methods are presented. Section 4.4 compares the performance of the various ML model types and elucidates the interpretability analysis results. Particular attention is paid to illustrating how global and local interpretability methods can explain the parameters a model generally considers for a prediction and how individual predictions are being made, respectively. Finally, Section 4.5 summarises the most important findings of the chapter and the conclusions that can be drawn from them.

4.2 Introduction

Bioenergy will likely play a critical role in achieving the sustainable development goals set out by the 2015 Paris Agreement. To date, only a small fraction of its potential has been tapped **[242]**. Thermochemical methods, such as gasification, are key technologies highlighted as part of the UK government's Net Zero Strategy. Amongst other reasons, gasification stands out to be an attractive renewable energy technology, as it has, unlike wind and solar, the potential to provide a stable baseload. When combined with carbon capture and storage (i.e. bioenergy with carbon capture and storage (BECCS)), gasification could become a negative emission technology that is highly demanded to fulfil the 1.5°C climate goal. Additionally, gasification can aid with reducing the emissions of the hardest-to-decarbonise sectors, such as transport and heating **[243]**.

Whilst the potential of gasification is clear, its widespread commercialisation and industrial-scale implementation are still subject to limited process performance and efficiency **[243]**. It has become an important topic to search for new methods and tools towards improved process design. Machine learning (ML) represents one such method. ML is a type of artificial intelligence, which uses a data-driven approach to develop methods which "learn" from provided data. This way models can be developed without any predetermined equations. For this work, ML is a promising method to achieve high-accuracy gasification production predictions, which can simplify model-based process design and optimisation. However, existing models often focused on a narrow range of feedstock types and gasifier setups **[3]**. For instance, Kardani *et al.* (2021) compared five soft computing methods, but focused on municipal solid waste (MSW) gasification only **[124]**.

In addition to low data availability and the narrow scope of existing models, limited interpretability has been another major problem. When first developed, many ML methods have been considered a black box, but more recent efforts to illuminate the inner workings of these methods have led to improved process understanding and analysis **[37]**. Indeed, interpretability is crucial for allowing researchers to guard against bias in their models and extract maximum knowledge from their data and models, improving the efficiency of data and model utilisation. A variety of methods such as correlation coefficients, hypothesis tests, and variance-based methods have been developed to assess the importance of a model's variables to aid with interpretability **[244]**. A model is considered interpretable when the reasoning behind its predictions can be readily understood by humans.

In the field of bioenergy, conventionally, model-specific techniques predominantly aimed at explaining a model's global behaviour have been used. Global model interpretability is concerned with a holistic view of the model and how the distribution of model outcomes is affected by the model's features. For instance, Serrano and Castelló (2020) used Garson's algorithm to explain the variable importance of a tar prediction model for a bubbling fluidised bed gasifier **[107]**. Zhu *et al.* used Gini importance-based variable importance

assessment to analyse a random forest (RF) model for the prediction of biochar yield from biomass pyrolysis **[130]**. Zhang *et al.* (2022) used Pearson correlation coefficients, Gini feature importance assessment, and partial dependence plots (PDPs) to study the effect of biomass composition and pyrolysis conditions on bio-oil characteristics **[245]**. The developed RF was found to rely more heavily on the biomass composition than the pyrolysis conditions to make its predictions.

Model-agnostic methods, which aim to separate the explanations from the ML model, have been a more recent development **[246, 247]**. The key advantage of model-agnostic techniques lies in their flexibility. This way a researcher is free to use any model type and different model types can be compared by using the same interpretability methods. SHAP (SHapley Additive exPlanations) **[246]** is one such method which has recently been used **[109, 231]**. SHAP can be used as a global method to explain the average effect of features on an output, but also as a local method to explain individual predictions. Whilst, in this field, these local methods have not been explored yet, they can powerfully communicate to stakeholders how a model has made a particular prediction. The merits of the interpretability methods selected for this work are discussed in more detail in Section 4.3.4.

In this work, a range of ML methods to model biomass and waste gasification have been developed and systematically compared. Instead of focusing on a narrow range of feedstocks and operating conditions like many existing studies **[3]**, an expansive data set including an array of different feedstock and gasifier types has been used. Model performance has been optimised by identifying the best model preprocessing steps for each model type and tuning their hyperparameters using a search algorithm.

As compared to existing interpretability-related studies, the consideration of three different feature importance assessment methods and the use of global and local methods allowed for the creation of a more complete picture of how the developed models made their predictions. Gini, permutation, and SHAP feature importance assessment have been compared, which mitigates limitations of individual methods and reduces uncertainty in the results **[140]**. The use of local explainability methods to guide the decisions of stakeholders and investors by providing an intuitive way to communicate an ML model's prediction process has been illustrated. By combining local with global methods, thermochemical process prediction black box models can be better understood. The developed models can predict syngas yield, syngas lower heating value (LHV), and syngas tar content, as well as the char yield and syngas composition in terms of N_2, H_2, CO, CO_2, CH_4, and C_2H_n. These models and their predictions can simplify further gasification system design and analysis such as life cycle sustainability assessment (LCSA) during which a system's environmental, economic, and social impacts are studied.

4.3 Materials and Methods

4.3.1 Data Collection, Preliminary Analysis, and Predictor Selection

The data set used for model development has been presented in greater detail in Chapter 3. It contains information on the feedstock's composition and its preparation, as well as data describing the gasification system and how it was operated. In total, up to 312 samples were collected from the current gasification literature for a range of different gasifier types, operating conditions, and feedstocks. The preliminary analysis followed a similar procedure to the one described in Section 3.3.4. In short, Pearson's and Spearman's correlation coefficients were calculated to measure the linear/monotonic relationship between predictors. It is important to identify highly correlated features before model training, as retaining them can limit a model's performance and hinder feature importance assessment **[231]**.

Predictor variables were excluded for several reasons. The feedstock's N and O content and LHV were excluded due to their high correlation with other predictors. Volatile matter (VM) and fixed carbon (FC) were removed as literature considers them dependent variables **[115]**. Feedstock type and shape were deemed redundant as the same information was already captured by the feedstock's ultimate and proximate composition and particle size. Finally, several predictors were dropped as they contained too many missing samples. Namely, the feedstock's cellulose, hemicellulose, and lignin contents, as well as the operating pressure, residence time, and steam-to-biomass ratio were dropped for this reason.

The following predictors were ultimately used for model development: feedstock C, H, S, ash, and moisture content, feedstock particle size, gasifier temperature, gasifier operation mode (batch/continuous), gasifier scale, equivalence ratio (ER), catalyst usage, gasifying agent, reactor type, and bed material.

All analysis was done in the Python programming language (version 3.8.13).

4.3.2 Data Preparation

Careful data preparation is an essential step before model training. It ensures maximum performance and trust in the developed models. Many ML algorithms require data to be in a uniform and specific format. For instance, missing or invalid entries cannot be processed by most algorithms and thus need to be treated beforehand. In this work, all data was cleaned, and categorical predictors were encoded using ordinal or one-hot encoding. Multilayer perceptron neural networks and support vector machines (SVMs) prefer features which are normally distributed. Thus, each feature was standardised by

removing the mean and scaling samples to unit variance.

Regarding categorical variables, the following variables were ordinally encoded: operation mode (batch or continuous), system scale (lab or pilot), and catalyst use (present or not present). The gasifying agent, bed material, and reactor type were one-hot encoded. Both encoding techniques have been discussed in more detail in Section 3.3.2.1.

To study the effects of various data preparation and preprocessing options a range of scenarios was defined. For this, predictor and target data were treated independently. Predictor data was prepared based on one of the following four procedures: (i) Continuous predictors only and drop row if missing value present (CONT+DROP); (ii) Continuous predictors only and mean impute missing values (CONT+MEAN); (iii) Add categorical predictors and drop row if missing value present (ALL+DROP), (iv) Add categorical predictors and mean impute missing values (ALL+MEAN).

The aim of comparing these four options was twofold. Firstly, the effect of adding categorical predictors to the training data set was studied by comparing CONT+DROP and CONT+MEAN to ALL+DROP and ALL+MEAN. Secondly, the effect of mean imputing missing values was studied. This is one of the simplest and most employed methods to deal with missing values. One main benefit of mean imputing missing values, compared to dropping the row is that it does not lead to a reduction in the size of the data set used for model development. During mean imputation, all instances where a sample is missing for a variable are filled with the mean of all samples that are present for that variable.

Target data was only processed for the training set. If missing values were present in the test set a prediction was not made. This way test data remained independent and was not affected by any data preprocessing. The training data was prepared based on one of the following three procedures: (i) Drop row if missing value present (TAR-DROP); (ii) Mean impute missing values (TAR-MEAN); (iii) Impute missing values using RF submodel (TAR-RF).

Methods TAR-DROP and TAR-MEAN follow similar procedures as the methods described for the preparation of the predictor data. Method TAR-RF employs RF submodels to fill in missing values. One RF submodel was fitted for each target variable with missing values present. Each submodel was trained using the complete data set and then used to make predictions for the missing target values. Using this method, a size reduction in the data set can be avoided (similar to the mean imputation method) whilst filling the missing instances with more meaningful values than simply the variable's mean. RF was chosen for the submodels due to its strong out-of-the-box performance **[248]**. A visual explanation of the employed methods is shown in Figure 4.1.

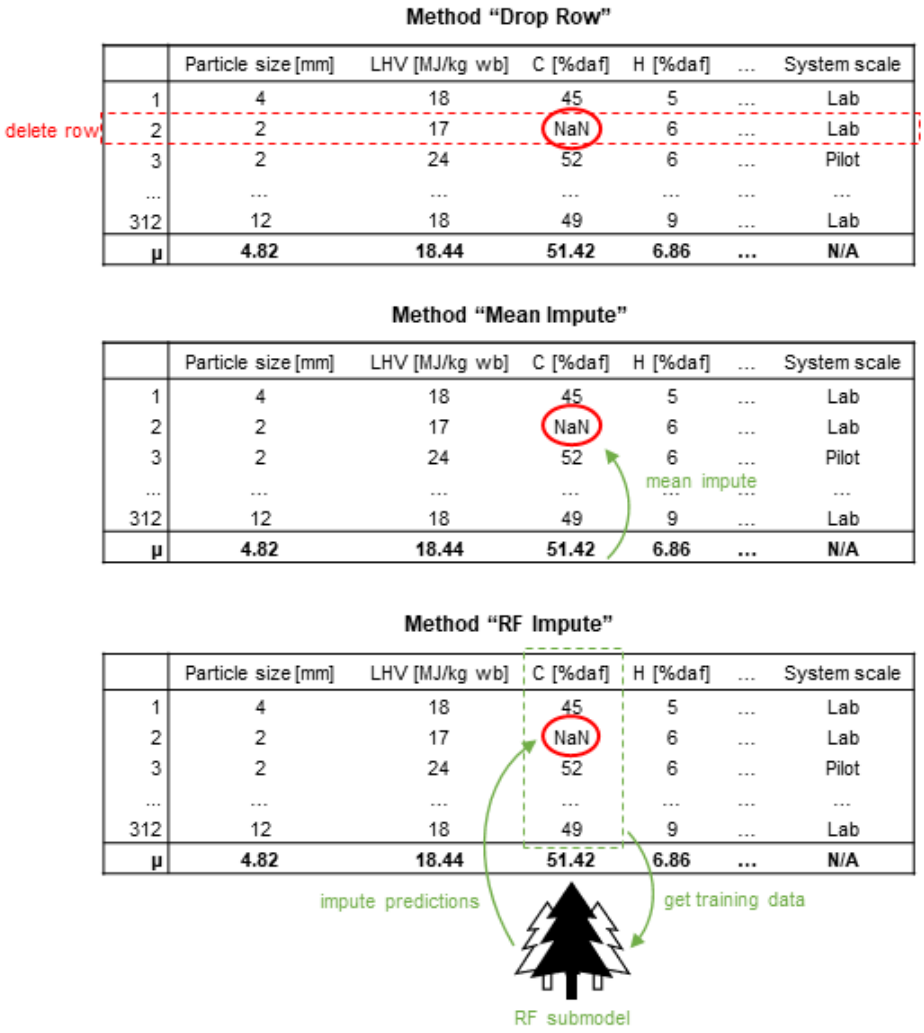


Figure 4.1: Illustration of the different data pretreatment methods employed to handle missing data in the data set.

4.3.3 Model Development

4.3.3.1 Model Types

A range of different model types have been compared in this work. One important family of models is tree-based models. On a fundamental level, tree-based models are algorithms which infer simple decision rules from the feature space to predict a target variable **[104]**. In practice, many models employ a collection of trees to reduce the variance associated with using a singular decision tree. RF is one such method introduced by Breiman (2001) **[111]**. It is a bootstrap aggregating, also called bagging, algorithm which averages the results of many independent learners. By averaging the results of many decision trees, the noise problems of individual trees can be circumvented **[104]**. Whilst bagging algorithms average the results of many independent models, boosting algorithms iteratively train and adjust the weights of many weak learners to create a powerful ensemble model. In this work, three different algorithms falling under this category were considered, namely gradient boosting for regression (GBR), XGBoost, and AdaBoost.

SVMs are another class of well-performing algorithms and were also considered in this work. Whilst modern deep neural networks often outperform SVMs or some of the other mentioned model types when vast amounts of data are available, for the size of data set available for this work, the presented algorithms could be among the best-performing algorithms in terms of prediction accuracy **[249]**. For this reason, instead of a deep neural network, a simpler multilayer perceptron neural network (ANN) was considered as another alternative. It is worth noting that artificial neural networks have been the most popular ML method to model biomass and waste gasification **[3]**. The final tested model type is the super learner (SL) concept which combines all previously developed models into one ensemble model **[250]**.

4.3.3.2 Model Optimisation and Comparison

Whilst some algorithms, such as RF and AdaBoost, are well known for their good out-of-the-box performance, other algorithms, such as ANN and SVM, require careful hyper-parameter tuning to maximise their performance **[248]**. In the past, trial and error has often been manually implemented to tune ML models; however, more advanced methods have become available to find the best hyperparameter combinations **[3]**. In this work, a search algorithm has been used to optimise each type of model. Specifically, a parameter grid was defined for each model type. These grids contained the hyperparameters which were to be optimised and possible options/values each parameter could take. The algorithm then searched the parameter space for the best-performing combination of hyperparameters. Model performance was assessed by determining the 5-fold cross-validated coefficient of determination (R^2) which is given by Equation 3.7. The highest R^2 score

indicates a model's best-performing hyperparameter combination. The root mean square error (RMSE) is another utilised performance measure. It is given by Equation 3.6 and measures the mean difference between observed values and an estimator's predictions.

R^2 and RMSE were used in conjunction as performance measures to quantitatively compare all the model types. They were calculated for test sets and the cross-validated models. For the model development, the data set was split into a training and test set using an 85% to 15% split. Another approach for accurately judging a model's generalisation capability is using cross-validation. This work employed k-fold cross-validation, during which the dataset was split into k parts. The model was then trained on k-1 folds and tested on the last fold. This process was then repeated k times so that each fold was used for testing once. Cross-validation allows one to obtain an independent measure of model performance, whilst using all available data for model development. Here 5-fold cross-validation has been used.

The overall methodology and workflow of this study are illustrated in Figure 4.2, where the left-hand side of the figure focuses on the ML model development described in this section.

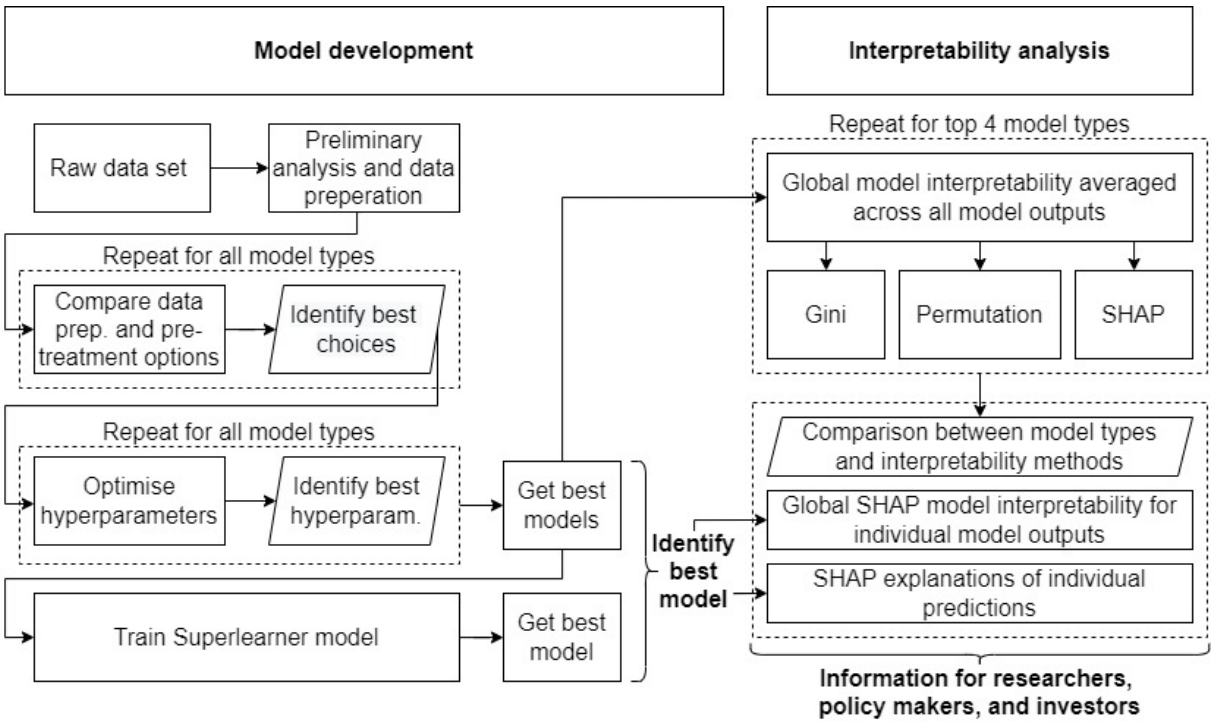


Figure 4.2: Flowchart of the methodology and workflow of this study, illustrating the two stages model development and interpretability analysis

4.3.4 Interpretability Analysis

Model interpretability and explainability are concerned with extracting knowledge about relationships learned by a model or the patterns of the underlying data. Interpretability may be understood as a human's ability to understand the cause of an ML model's prediction. A highly interpretable model is easy to comprehend, and the model's results can be consistently predicted by a human. In contrast, humans struggle to understand the predictions of a model with low interpretability **[37]**.

Many ML models have been considered a black box, as their inner workings are often poorly understood. One approach to make ML more interpretable is the use of models such as linear regression and decision trees which are inherently interpretable **[37]**. However, these simple models generally do not offer the same prediction performance as more complex models **[131]**. A more recent approach is model-agnostic methods which aim to make any black box model interpretable. One intuitive approach is called permutation feature importance during which a feature's values are randomly shuffled **[111]**. The resulting increase in the prediction error yields a measure of the feature's importance. This provides a global insight into the model's behaviour and automatically accounts for interaction effects of features **[37]**.

Another global method uses the Gini index, also known as Gini impurity, which is used by tree-based models to determine where nodes should be split. This provides a straightforward measure to interpret tree-based methods such as RF and GBR. One disadvantage of Gini-based variable importance assessment is that it is biased towards inputs with more categories. This is less of an issue for mostly continuous and largely uncorrelated features. Furthermore, this method does not apply to other model types **[244]**.

Whilst global methods explain the workings of the overall model, they remain limited in explaining individual predictions. Here methods such as local surrogate models (LIME) **[247]** and SHAP **[246]** come into play. LIME fits an interpretable surrogate model locally. The local models only explain individual predictions and are not required to be a good global approximation of the model. Whilst LIME has been considered a promising method, it still has several problems, especially for tabular data. For instance, hyperparameter tuning can be challenging with opposite findings being possible depending on the chosen smoothing kernel width. Another issue is the instability of LIME's explanations, which means that repeating the same explanation can lead to significantly different results **[37]**.

SHAP is a flexible technique which can be used for global interpretation and to explain individual predictions. It has a strong theoretical foundation in game theory and uses the concept of allocating optimal credits based on Shapley values to estimate the importance of features. SHAP force plots provide an intuitive visualisation of how different features affect an individual prediction. One advantage of SHAP for global interpretation, over the Gini and permutation feature importance assessment methods, is that SHAP not only informs

about the importance of features but also their relationship with the output. Additionally, SHAP's predictions are fairly distributed among feature values. These factors are critical in guaranteeing trust in the method **[37]**.

In this work, Gini, permutation, and SHAP feature importance assessment were used for global interpretation of the developed models. SHAP was also used to give examples of how individual predictions can be explained locally. The workflow outlined here is summarised by the right-hand side of Figure 4.2, which illustrates the workflow of the interpretability analysis and how it interacts with the model development stage described in Sections 4.3.1, 4.3.2, and 4.3.3.

For interpretation, the results of the analysis were illustrated in graphical form. Global methods were illustrated using bar charts by showing the importance scores across all ten model outputs/targets in the same figure which maximises the information contained per figure. Results were then systematically compared between model types and feature importance assessment methods. As previously mentioned, SHAP also lends itself to the explanation of individual predictions and can globally illustrate the relationship between features and model outputs. Hence additional figures were created to illustrate these aspects.

4.4 Results and Discussion

4.4.1 Selection of Optimal Prepossessing Steps

As data preparation and preprocessing are key stages in ML model development, various methods were compared as outlined in Section 4.3.2. The predictor and target preprocessing methods ALL+DROP and TAR-RF performed best for all but one model type. Predictor methods CONT+DROP and CONT+MEAN which did not include categorical variables performed significantly worse than their counterparts which included categorical variables. This suggests that the use of meaningful categorical variables can noticeably increase the models' prediction accuracy. Mean imputation has frequently been employed due to its simplicity **[104]**. However, comparing mean imputation (CONT+MEAN and ALL+MEAN) to simply dropping the row if missing data was present (CONT+DROP and ALL+DROP) showed that dropping the row generally led to a higher R^2. These findings indicate that other methods, such as dropping rows with missing data or imputation by a submodel, also need to be considered to ensure optimal model performance.

The target preprocessing method TAR-RF outperformed its alternatives by a considerable margin, which means the added complexity from developing the RF submodels was well justified. By leaving the test data untouched, an unbiased evaluation of the model was possible whilst maximising the models' performance by avoiding missing samples in the training set. The performance of all methods is shown in Table 4.1.

Table 4.1: Comparison of different predictor and target preprocessing methods (as described in Section 4.3.2) and their effect on the average 5-fold cross-validated R^2 scores for all model types. The best-performing preprocessing combination is shown in bold for each model type.

Predictor method	Target method	RF	GBR	XGBoost	AdaBoost	SVM	ANN
CONT+DROP	TAR-DROP	0.73	0.73	0.66	0.67	0.33	0.35
CONT+MEAN	TAR-DROP	0.75	0.73	0.73	0.64	0.29	0.27
ALL+DROP	TAR-DROP	0.81	0.84	0.77	0.79	0.52	0.56
ALL+MEAN	TAR-DROP	0.82	0.81	0.8	0.74	0.47	0.51
CONT+DROP	TAR-MEAN	0.75	0.72	0.7	0.69	0.35	0.05
CONT+MEAN	TAR-MEAN	0.74	0.7	0.71	0.61	0.3	0.20
ALL+DROP	TAR-MEAN	0.83	0.84	0.79	0.78	0.55	-0.49
ALL+MEAN	TAR-MEAN	0.81	0.81	0.79	0.7	0.48	-0.37
CONT+DROP	TAR-RF	0.78	0.74	0.72	0.71	0.37	0.50
CONT+MEAN	TAR-RF	0.79	0.77	0.77	0.67	0.32	0.35
ALL+DROP	TAR-RF	**0.86**	**0.87**	0.83	**0.82**	**0.56**	**0.63**
ALL+MEAN	TAR-RF	0.84	0.84	**0.84**	0.75	0.49	0.42

4.4.2 Model Optimisation and Comparison

Initially, the performance between the out-of-the-box models with optimal preprocessing choices and optimised models was compared by considering the average 5-fold cross-validated R^2 scores shown in Table 4.2. It was found that RF and XGBoost did not benefit from hyperparameter optimisation, as the out-of-the-box models were found to perform best. The performance of all other model types improved after hyperparameter optimisation as compared to the out-of-the-box models. For instance, the R^2 of GBR slightly increased from 0.87 to 0.90, whereas hyperparameter optimisation had a more substantial effect on the ANN model which showed an increase from 0.63 to 0.81. Generally, optimisation was shown to have a relatively minor effect on ensemble and tree-based methods. However, wherever computational demands during model training are of no concern, hyperparameter optimisation can be an effective means to obtain the best possible model structure.

In addition to the average 5-fold cross-validated R^2 scores, Table 4.2 also shows the prediction performance of individual submodels predicting the ten outputs modelled as part of this work. Similarly, the prediction performance in terms of the 5-fold cross-validated RMSE values is shown in brackets. All optimisation results shown in Table 4.2 used the best preprocessing steps determined during the previous model development stage, as described in Section 4.4.1. From these results, it becomes clear that, overall, GBR was the best-performing model type. It achieved the highest performance for seven and eight out of ten outputs in terms of R^2 and RMSE respectively. RF, XGBoost, AdaBoost, ANN, and the SL model were found to perform at an acceptable level with mean R^2

Table 4.2: 5-fold cross-validated R^2 and RMSE scores (shown in brackets) for all model outputs and model types. The highest-scoring model types are shown in bold for each output.

	RF	GBR	XGBoost	AdaBoost	SVM	ANN	Super Learner
N_2 $[vol.\%\ db]$	0.96 (4.47)	**0.97** (**4.08**)	**0.97** (4.19)	0.96 (4.19)	0.24 (19.7)	0.96 (4.33)	**0.97** (4.09)
H_2 $[vol.\%\ db]$	0.86 (5.13)	0.89 (**2.53**)	**0.91** (2.89)	0.88 (2.68)	0.58 (5.52)	0.83 (3.36)	0.88 (2.61)
CO $[vol.\%\ db]$	0.94 (3.90)	0.94 (2.36)	0.82 (2.35)	**0.95** (2.14)	0.53 (6.47)	0.93 (2.53)	0.94 (**2.10**)
CO_2 $[vol.\%\ db]$	**0.94** (2.82)	**0.94** (**1.57**)	0.79 (1.61)	**0.94** (**1.57**)	0.74 (3.41)	0.88 (2.25)	0.93 (1.59)
CH_4 $[vol.\%\ db]$	0.83 (1.58)	**0.90** (**0.85**)	0.77 (1.06)	0.88 (0.93)	0.76 (1.30)	0.84 (1.02)	0.89 (0.87)
C_2H_n $[vol.\%\ db]$	0.77 (0.82)	**0.87** (**0.40**)	0.64 (0.51)	0.78 (0.50)	0.69 (0.61)	0.64 (0.61)	0.85 (0.43)
LHV $[MJ/Nm^3]$	0.93 (0.79)	**0.94** (**0.60**)	**0.94** (0.63)	0.92 (0.69)	0.90 (0.78)	0.88 (0.78)	**0.94** (**0.60**)
Tar $[g/Nm^3]$	0.59 (15.94)	0.76 (**10.29**)	**0.85** (13.73)	0.41 (14.49)	0.09 (20.47)	0.62 (12.2)	0.75 (9.98)
Gas yield $[Nm^3/kg\ wb]$	0.93 (0.26)	**0.96** (**0.20**)	0.91 (0.25)	0.92 (0.29)	0.90 (0.31)	0.72 (0.45)	0.95 (0.22)
Char yield $[g/kg\ wb]$	**0.83** (25.11)	**0.83** (21.69)	0.76 (**21.27**)	0.82 (21.92)	0.14 (49.57)	0.76 (26.47)	0.82 (24.84)
Mean	0.86 (6.08)	**0.90** (**4.46**)	0.84 (4.85)	0.85 (4.94)	0.56 (10.81)	0.81 (5.40)	0.89 (4.73)

scores > 0.80. Particularly, the SL model had a similar performance to GBR. Whilst ANN models performed exceptionally well in previous works (e.g. see Section 2.3.4.1, they were outperformed here. As ANNs require large amounts of data, a potential reason for this is the relatively limited amount of data available to train the model, despite the data set used in this study being still larger than the ones used in previous works **[115, 124, 251]**.

The SVM model performed well for a few outputs, such as the syngas yield and LHV, but very poorly for other outputs such as the char yield and syngas' N_2 content. Interestingly, the syngas' N_2 content was an output which was learned extremely well by all other model types with R^2 scores > 0.96. In contrast, SVM only achieved a R^2 of 0.24. This suggests that, without extensive further optimisation, SVM is not a suitable method for modelling some of the studied gasification outputs based on the data set provided in this study. When considering the strong out-of-the-box performance of tree-based models such as GBR and RF, it becomes hard to justify the optimisation of SVM models which requires the careful selection of an appropriate kernel function and other hyperparameters.

Figure 4.3 shows scatter plots of known target values against model predictions for all optimised GBR submodels. Data points in yellow represent the training set, whereas the test set is shown in blue. The outputs which were modelled well with a high R^2 and low RMSE, as shown in Table 4.2, showed little dispersion. Syngas yield and LHV are two examples of outputs with a high prediction accuracy. Other outputs, such as the tar content in the produced syngas, showed a much larger dispersion. When considering test data across all outputs, most predictions have an error of $< 10\%$ and very few predictions have an error $> 20\%$, with most outliers being constrained to the syngas tar concentration and char yield submodels. This has been illustrated graphically in Figure 4.4.

A like-for-like comparison of this work to existing studies is not always possible, due to vastly different aims and assumptions. However, as the ANN model developed in Chapter 3 was developed on the same data set a more direct comparison can be made.

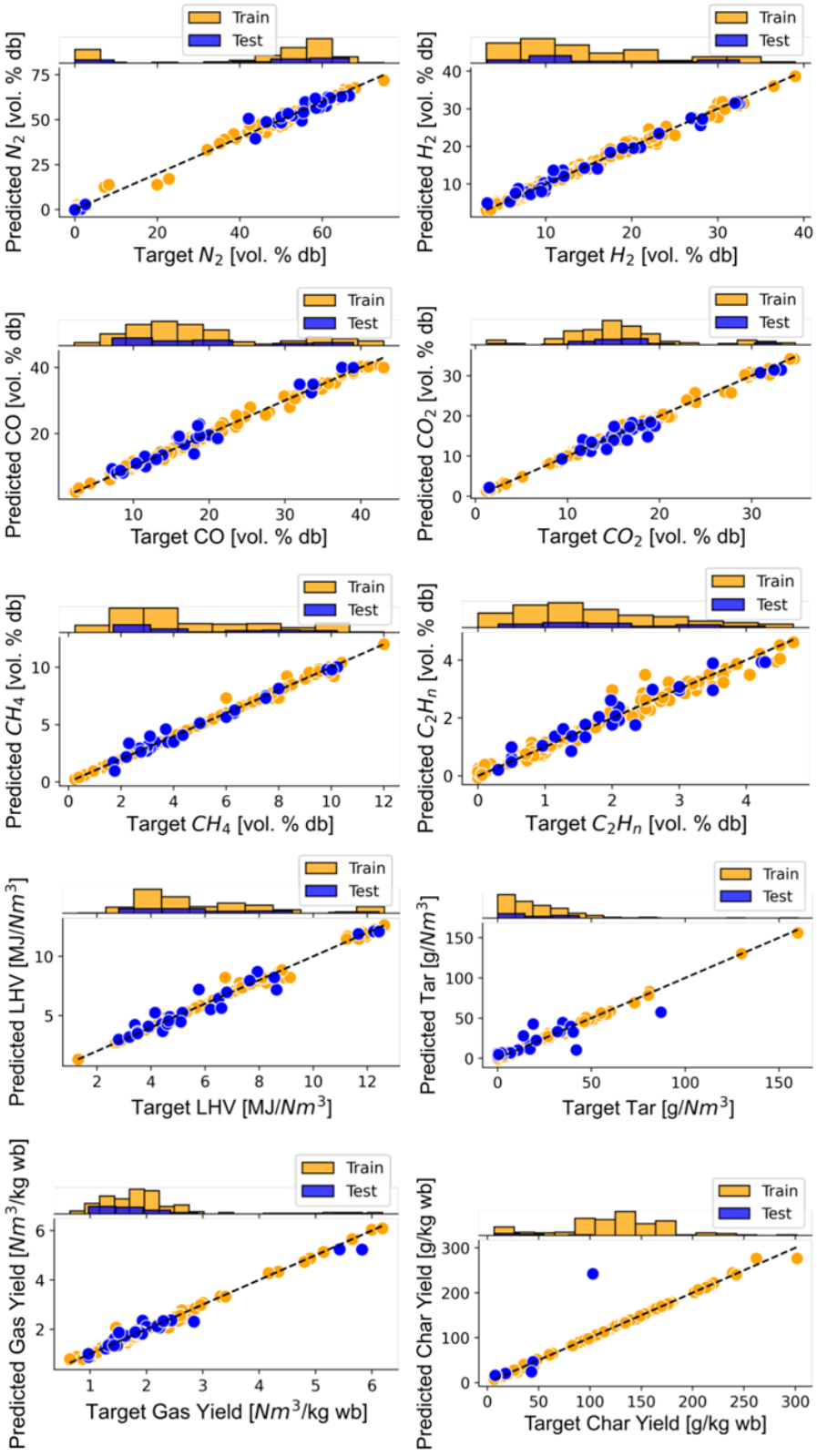


Figure 4.3: Scatter plots of target values vs. predictions for all 10 optimised GBR sub-models. The figures show both the training (yellow) and test (blue) performance of each model. The black dashed line indicates perfect predictions

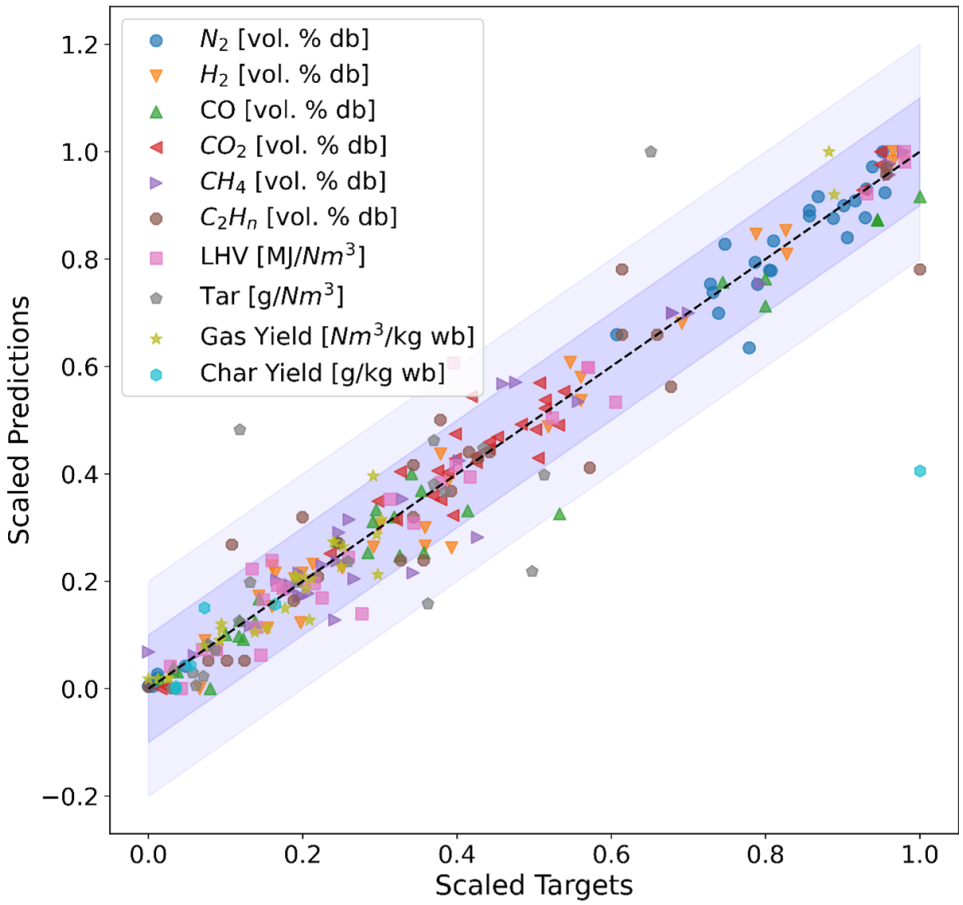


Figure 4.4: Scatter plot of scaled target values vs. predictions for GBR models trained for all ten outputs. The black dashed line indicates perfect predictions. The 10% and 20% error regions are illustrated by the darker and lighter shaded blue regions.

By comparing the averaged R^2 score (averaged across all model outputs) of this previous model, an improvement from 0.86 to 0.90 was achieved by the best-performing GBR model presented in this chapter. Looking at the work of other researchers, Serrano *et al.* (2020) developed an ANN model suitable for a range of different fluidised bed gasifier bed materials **[107]**. Their model achieved R^2 scores ranging from 0.57 to 0.98 for a range of different outputs (e.g. gas composition in terms of CO_2, CO, CH_4, H_2 and gas yield). When looking at the mean absolute percentage error (MAPE) achieved in their work, there were significant errors ranging from 9.18% to 38.91%. The best-performing model developed in this work compares favourably with most MAPEs being around or lower than 10%. Furthermore, this model is also suitable for a wider range of gasifiers, such as fixed-bed gasifiers.

Whilst GBR was found to be the best-performing model type in this work, a range of model types were found to be suitable in literature and no single model type appears to be dominant **[3]**. Elmaz *et al.* (2020) trained polynomial regression, SVM, ANN, and decision tree models on data from an in-house gasifier fed with pine cones and wood pellets **[94]**. They found decision trees and ANN to be the preferable model types. Despite their data set being more homogeneous than the one used in this study, their decision tree models achieved a lower test performance than the ones studied in this work with $R^2 = 0.81$-0.94.

Sun *et al.* (2022) took an alternative approach for model optimisation than the one proposed in this work **[252]**. By employing particle swarm optimisation, they developed an ANN model for the prediction of syngas yield, gas species concentrations, and char yield with an excellent test performance of $R^2 = 0.97$. However, the authors highlighted that the model had only been trained on data from pine wood gasification and increasing the size of the data set by incorporating a wider range of feedstocks and gasification conditions was deemed important to improve the model's applicability.

The model developed in this work has the potential to simplify general gasification process design. By using a varied data set for model training a large range of different gasification systems and feedstocks could be optimised and compared. Furthermore, the model's predictions could be used in a more holistic system modelling context. For instance, predictions could be directly used for LCA or TEA.

4.4.3 Interpretability Analysis

Initially, the global interpretability of models was studied. The GBR model was taken as the baseline model as it was found to be the best-performing model. It was then compared to the three next best-performing model types, namely AdaBoost, RF, and XGBoost. It must be noted that at the time of writing the SHAP method was not supported for AdaBoost, hence only Gini and permutation importance were computed for this model type.

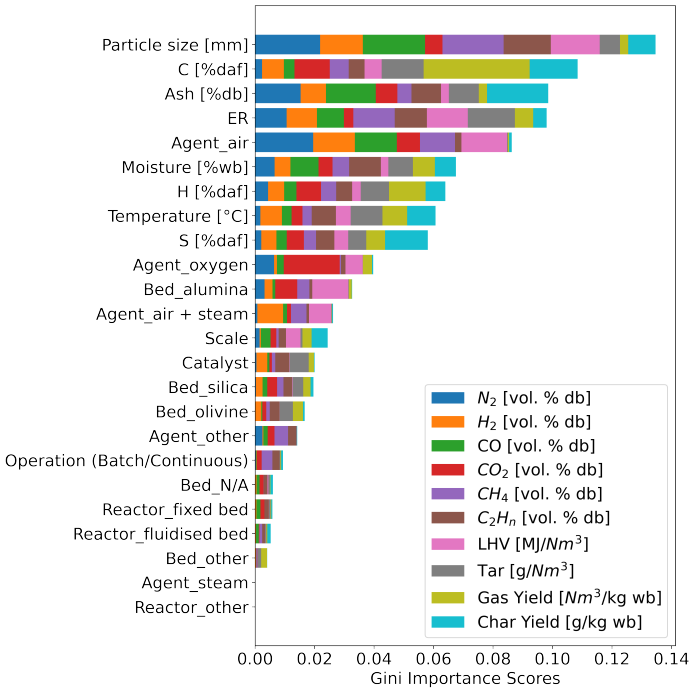
Figures 4.5a and 4.5b show the Gini feature importance of the GBR and RF models across all ten outputs. The x-axis shows the importance scores, whereas the y-axis shows the predictors used to train the models. Submodels predicting the different outputs can be differentiated by their colours. RF is shown for comparison purposes, because as a bagging model, it operates differently from the GBR model. However, other model types are also discussed. Additional figures for other model types and feature importance assessment methods are available from the GitHub page which was shared as part of the paper (https://github.com/SimonAscher/Interpretable-Machine-Learning-to-Model-Biomass-and-Waste-Gasification), as these additional figures would exceed the scope of an appendix. The feedstock's particle size and choice of gasifying agent were found to be the key predictors across all four studied model types, with the particle size being the top predictor for GBR, RF, and AdaBoost. The top ten predictors account for most of the variation in the GBR and RF models' predictions. This is illustrated by a drop of over 50% in the combined Gini importance scores from the 10^{th} to 11^{th} most important predictor.

Looking at the ten most influential predictors, it can be seen that they are identical for GBR and RF, with predictors such as temperature, ER, and proximate and ultimate composition strongly affecting the models' predictions. Some minor differences for AdaBoost were apparent, whereas XGBoost showed some more major deviations. Where most submodels generally substantially contributed to a feature's importance for GBR, RF, and AdaBoost, this is not the case for XGBoost. For this model type, a feature's combined importance was often heavily dominated by a few submodels. For instance, the score of XGBoost's top predictor gasifying agent oxygen was largely made up of the three submodels predicting the syngas yield, syngas CO_2 content, and syngas LHV.

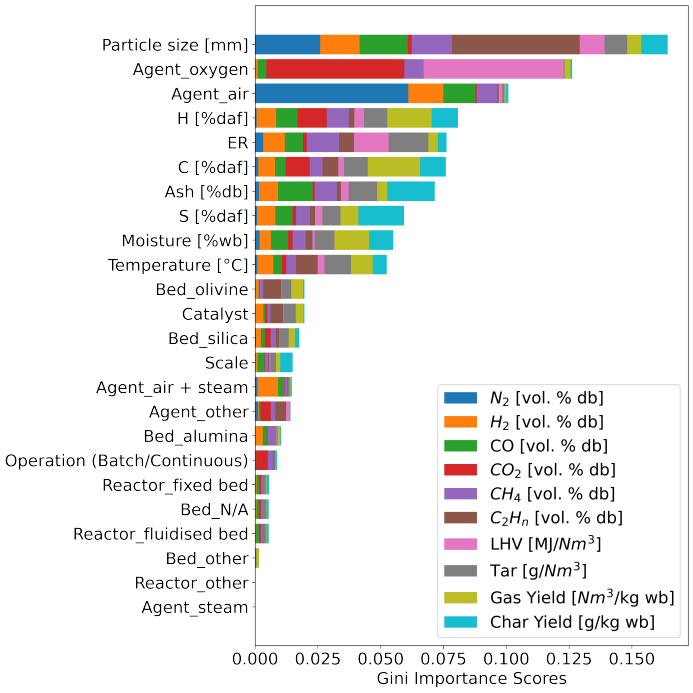
Comparing Gini to permutation importance for RF showed few differences, with the top three predictors remaining the same. Similarly, the top predictors for GBR remained nearly unchanged when comparing Gini to permutation importance, with only catalyst usage displacing oxygen as a gasifying agent in the top ten as shown in Figure 4.5c. The use of a catalyst was found to impact the prediction of the syngas' tar content most strongly. This is in good agreement with the literature, as many studies used a catalyst to reduce the tar content in the produced syngas **[200, 220]**.

Whilst the char yield submodel only contributed marginally to the Gini importance findings, it dominated other submodels based on permutation importance assessment. For permutation importance, the gasification temperature was found to be the most important predictor for GBR, with more than two-thirds of the combined score coming from the char yield prediction submodel.

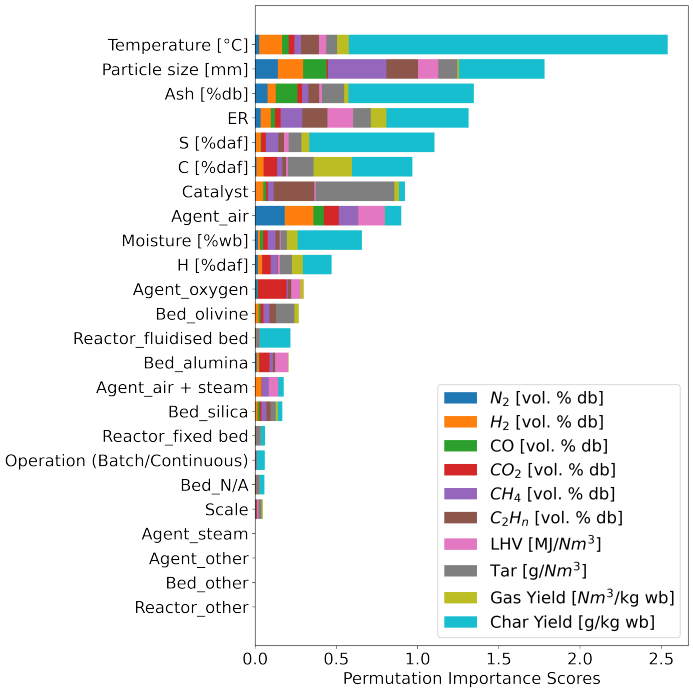
In general, when looking at the combined feature importance of all submodels, all studied model types were in good agreement for permutation-based feature importance.


Particle size [mm]
C [%daf]
Ash [%db]
ER
Agent_air
Moisture [%wb]
H [%daf]
Temperature [°C]
S [%daf]
Agent_oxygen
Bed_alumina
Agent_air + steam
Scale
Catalyst
Bed_silica
Bed_olivine
Agent_other
Operation (Batch/Continuous)
Bed_N/A
Reactor_fixed bed
Reactor_fluidised bed
Bed_other
Agent_steam
Reactor_other
N_2 [vol. % db]
H_2 [vol. % db]
CO [vol. % db]
CO_2 [vol. % db]
CH_4 [vol. % db]
C_2H_n [vol. % db]
LHV [MJ/Nm^3]
Tar [g/Nm^3]
Gas Yield [Nm^3/kg wb]
Char Yield [g/kg wb]
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14
Gini Importance Scores

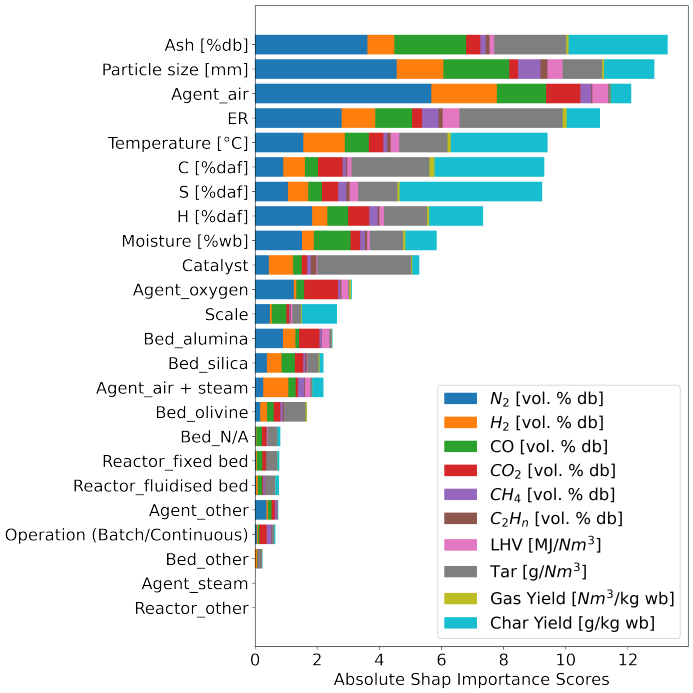
(a)


Particle size [mm]
Agent_oxygen
Agent_air
H [%daf]
ER
C [%daf]
Ash [%db]
S [%daf]
Moisture [%wb]
Temperature [°C]
Bed_olivine
Catalyst
Bed_silica
Scale
Agent_air + steam
Agent_other
Bed_alumina
Operation (Batch/Continuous)
Reactor_fixed bed
Bed_N/A
Reactor_fluidised bed
Bed_other
Reactor_other
Agent_steam
N_2 [vol. % db]
H_2 [vol. % db]
CO [vol. % db]
CO_2 [vol. % db]
CH_4 [vol. % db]
C_2H_n [vol. % db]
LHV [MJ/Nm^3]
Tar [g/Nm^3]
Gas Yield [Nm^3/kg wb]
Char Yield [g/kg wb]
0.000 0.025 0.050 0.075 0.100 0.125 0.150
Gini Importance Scores

(b)



(c)



(d)

Figure 4.5: Gini feature importance of (a) gradient boosting models and (b) random forest models, as well as (c) permutation and (d) SHAP feature importance analysis of gradient boosting models. Importance scores are shown on the x-axis. The y-axis shows the predictors used to train models for the 10 considered outputs. A predictor's total score is made up of the importances from the different submodels (i.e. model outputs) which are illustrated by the different colours shown in the legend.

However, the overall importance of individual predictors could result from very different submodels. For instance, the particle size, which was found to be the most important predictor for three out of four model types, had significant contributions to its importance from most outputs for GBR and XGBoost. In contrast, for AdaBoost, its importance was heavily dominated by the outputs CO and C_2H_n. When looking at RF, N_2, C_2H_n, and char yield were the outputs which mostly contributed to the particle size's importance as a predictor.

The feature importance results yielded by the SHAP method are shown in Figure 4.5d for GBR. Top predictors were found to be similar to Gini and permutation feature importance. Where permutation importance was heavily dominated by individual outputs, SHAP-based feature importance scores showed more balanced contributions. A finding worth highlighting is that whilst the N_2 content submodel contributed to a similar level as other submodels to the importance of air as a gasifying agent for Gini and permutation feature importance, this contribution was much larger for SHAP feature importance. RF and XGBoost already showed a much larger contribution from this submodel to the importance of air as a gasifying agent for Gini and permutation feature importance. For SHAP feature importance, this dominance was more obvious and more than half of the importance score of air as a gasifying agent resulted from the N_2 content submodel. These findings are in good agreement with the literature, where the use of air as an agent has been linked to diluted syngas with a high N_2 concentration **[78, 253]**.

Even though all three methods generally agreed well with each other, there was some variation in the importance of predictor variables. The most striking difference was that the combined permutation-based feature importance scores could be heavily dominated by a single output, whilst this was less obvious for the Gini and SHAP-based scores. Gini-based feature importance is generally prone to favour high cardinality features, such as numerical inputs **[111]**. However, in this work, it was found that some categorical variables (e.g. type of gasifying agent used) were also important.

So far, Gini importances were shown in normalised form, but the permutation-based feature importance scores (Figure 4.5c) and SHAP-based feature importance scores (Figure 4.5d) were shown as absolute values. Typically, outputs with higher feature values obtain higher absolute importance scores, which makes these outputs contribute more strongly to the global importance of any given predictor variable (i.e. shown on the y-axis). This led to outputs with higher feature values (e.g. char yield and syngas N_2 content) dominating outputs with lower feature values (syngas yield and LHV). This potential problem is clearly illustrated by Figure 4.5c, where the char yield contributes significantly more to the overall importance scores than any other output, especially the previously discussed outputs with a low feature value.

Alternative versions of the permutation and SHAP-based feature importance figures,

with normalised importance scores, are shown in Figure 4.6. To normalise the vectors of predictor importances, two main vector regularisation/normalisation methods were considered, namely L1 and L2 normalisation. L1 normalisation uses the sum of the absolute values of the vector, whereas L2 normalisation uses the square root of the sum of the squared vector values. Whilst, L2 normalisation is typically preferred, as it generally outperforms L1 normalisation from an accuracy point of view, L1 normalisation offers some unique benefits. Mainly, L1 normalisation creates sparse solutions, which are highly relevant from an interpretability point of view **[254]**. Furthermore, the sum of all values in the normalised vector is equal to 1. In this application, this guarantees that all ten considered model outputs contribute equally to the overall importance scores of the predictors. It is to be noted that, certain model outputs may be considered more important than others, however, assigning weights to certain outputs would be subjective. For this reason, the approach of assigning the same weight to each output was preferred. Additionally, feature importance graphs for individual outputs can be studied if required.

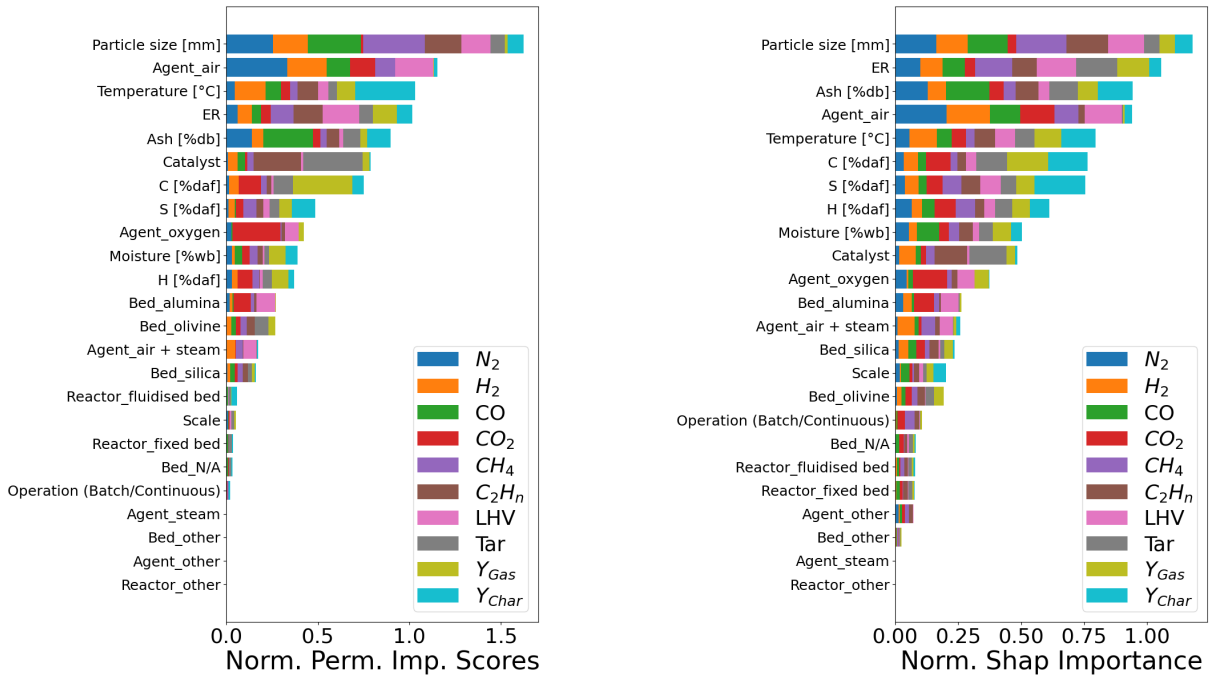


Figure 4.6: Normalised permutation (a) and SHAP (b) feature importance analysis results for gradient boosting models. Importance scores are shown on the x-axis. The y-axis shows the predictors used to train models for the 10 considered outputs. A predictor's total score is made up of the importances of the different submodels (i.e. model outputs) which are illustrated by the different colours shown in the legend.

When comparing these updated results to the original results presented in Figure 4.5, a few differences can be observed. Most importantly, the char yield submodel no longer dominates the permutation importance results, and the syngas N_2 content and char yield

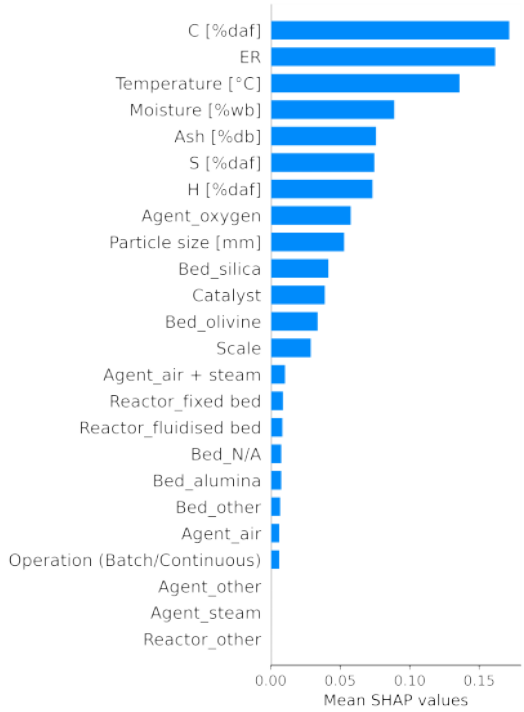
submodels no longer have as strong of an impact on the SHAP importance results. Now, outputs such as the syngas yield, which played a minimal role in the original SHAP results, play a much more significant role as they are weighted equally to all other outputs. Thus, the updated results may be considered a fairer representation of which predictors the developed models most heavily rely on across the ten predicted variables.

Using this normalised method, the order of top predictors slightly changed for both the permutation and SHAP importance methods. Most notably the particle size is now also the top predictor for the permutation method, making it the top predictor for all interpretability methods. However, in general, the same predictors remained important for all interpretability method types. Thus, the general findings and resulting conclusions are the same for either method.

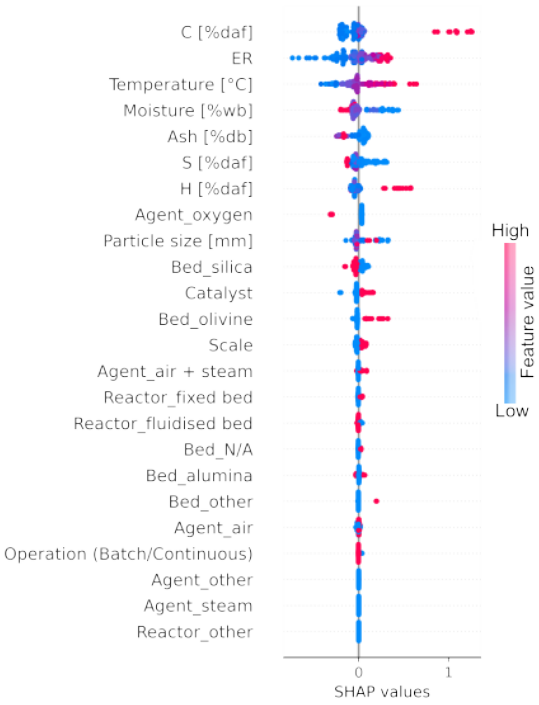
Whilst Figure 4.5 shows the absolute importance of features across all outputs, the nature of their relationships with the outputs remains unexplained. The SHAP method can be used to not only study the importance of features but also their relationships with individual output parameters. Figure 4.7a shows the absolute importance of features in a similar fashion to Figure 4.5 but this time for a single output. In this instance, the syngas yield is shown. Figure 4.7a not only illustrates the feature importance but also the feature effects. Red represents a high feature value, whereas blue represents a low feature value. The further away a point is from the baseline SHAP value of zero, the stronger it affects the output. This way a feature's relationship with the SHAP value (and in turn the predicted output) can be better understood.

The feedstock's carbon content most strongly affected the GBR model's prediction for the syngas yield. From Figure 4.8b it becomes clear that a low feedstock carbon content led to a minor decrease in syngas yield. A high carbon content led to a large increase in syngas yield. Most data points show a low feature value and are bulked in the negative or near zero SHAP value region. The data points with a high feature value (SHAP value close to 1) correspond to plastic waste and other feedstocks with a very high carbon content. The temperature and ER show a smoother distribution of data points. Again, low temperatures and ERs decreased the syngas yield, whereas high feature values increased the syngas yield. Feature values near the middle of the range (shown in purple) were shown to have little effect on the model's predictions. For the binary variable catalyst usage, it was shown that the use of a catalyst generally had a positive effect on the syngas yield. Figures for the remaining nine model outputs can again be found on the GitHub page which was shared as part of the original paper (https://github.com/SimonAscher/Interpretable-Machine-Learning-to-Model-Biomass-and-Waste-Gasification), as these additional figures would exceed the scope of an appendix.

Figure 4.8 illustrates how the SHAP method can be used to explain individual model predictions. Two example model outputs and feedstocks are shown in the figure. The



(a)



(b)

Figure 4.7: Feature importance analysis by SHAP method for GBR model predicting the syngas yield. Absolute importance scores are shown in (a), whereas (b) shows the influence of individual predictions on the overall importance scores. Red represents a high feature value (in this case syngas yield), whereas blue represents a low feature value. The further away a point is from the baseline SHAP value of zero, the stronger it effects the output.

figures represent an intuitive way to guide the decisions of stakeholders and investors and improve their understanding of how the developed model makes a particular prediction. Figures 4.8a and 4.8b show the syngas yield from barley straw gasification for a base case at 800°C (a) and a high-temperature case at 1,000°C (b). The gasification process underwent continuous operation without a catalyst in a pilot-scale fluidised bed gasifier using silica as a bed material and air as an agent. A temperature of 800°C, particle size of 4 mm, and an ER of 0.3 were selected for the base case (Figure 4.8a), as these values are close to the mean of all samples for these variables. The force plots show how individual feature values perturb the model's prediction from a baseline value. For the base case, as shown in Figure 4.8a, most feature values lowered the predicted syngas yield (shown in blue). Most notably, the barley straw's low carbon and relatively high moisture contents led to a decrease in the model's predicted value. When raising the temperature to 1,000°C as shown in Figure 4.8b, a significant increase in the predicted syngas yield can be noticed (shown in red). The model's predicted syngas yield rose from 1.63 to 2.21 Nm^3/kg wb. The high temperature value solely balanced the effects of all feature values lowering the predicted yield.

For Figure 4.8c, the gasifier setup and conditions remained unchanged from the ones previously described (i.e. shown in Figure 4.8a and Figure 4.8b), however, the feedstock has been changed to MSW. Two variables have been altered for Figure 4.8d, namely, steam was used as a gasifying agent instead of air and the feedstock's particle size was reduced from 4 mm to 1 mm. The syngas LHV has been selected as the output of interest in this example. MSW's high feedstock carbon content led to an increase in the model's prediction. In contrast, a high feedstock ash content of 16.82% db and a particle size of 4 mm led to a reduction in the predicted LHV. Figure 4.8d reveals that smaller particle sizes can increase the predicted LHV. Similarly, using steam as a gasifying agent was found to significantly increase the model's predicted LHV. Interestingly, this was illustrated by the one-hot encoded variable representing the gasifying agent air equalling zero. This could be interpreted as the model understanding that air as a gasifying agent reduced the LHV. The opposite applied to the variable gasifying agent oxygen equalling zero – here the model learned that the gasifying agent not being oxygen adversely affected the syngas' LHV.

In summary, Figure 4.8 shows how the SHAP method could be used to communicate an ML model's prediction process to a non-expert audience. This way the ML practitioner can offer a simple explanation of how certain system set-ups affect the outputs which stakeholders or policymakers may be concerned with. Different system or feedstock choices may be studied by creating similar figures for an array of operating conditions and feedstock combinations.

Studies assessing the interpretability of bioenergy ML models are still scarce. Researchers have often looked towards PDPs and Pearson's or Spearman's correlation coeffi-

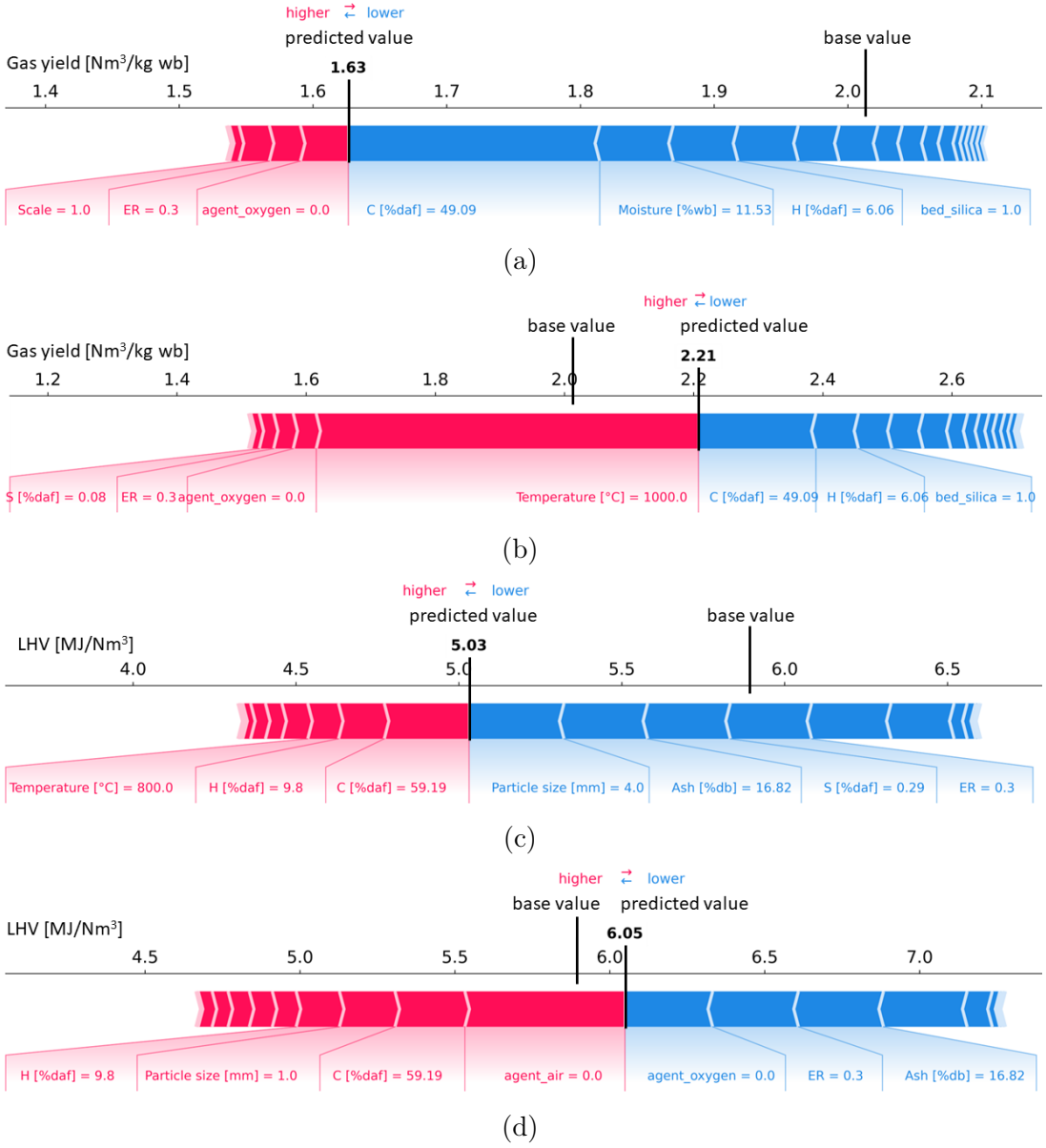


Figure 4.8: SHAP explanations of some selected individual predictions. The syngas yield from barley straw gasification is shown for a base case at 800°C (a) and a high-temperature case at 1,000°C (b). The resulting syngas LHV from MSW gasification is shown for a base case using air as a gasifying agent (c) and an alternative case using steam as a gasifying agent (d). The force plots start at the base value (the average of all predictions). Each predictor (and its corresponding Shapley value) is represented by an arrow which either increases (shown in red) or decreases (shown in blue) the model's predicted value with respect to the base value. A predictor's importance is shown by the size of its arrow, where a larger arrow represents a more important predictor. Ultimately, the model's predicted value is illustrated by the point where the red and blue arrows meet.

cients **[130, 231, 251, 255]**. PDP are an effective and straightforward means to illustrate the relationship between a predictor variable and model output. However, as each input and output combination requires an individual plot this method can become infeasible, especially for multiple-input multiple-output (MIMO) models like the one presented in this work. For this reason, identifying the importance of features through global methods, to then study the relationship in more detail using PDP, represents a good alternative. Furthermore, confounding factors cannot be captured by PDP.

Another popular interpretability method has been permutation feature importance **[251, 255, 256]**. A key benefit of permutation importance is its easy implementation and the fact that it can be implemented for any model type. However, by randomly shuffling a feature, unrealistic data instances can be created (e.g. feedstock ultimate composition no longer summing to 100%) which may limit interpretability.

Zhao *et al.* (2021) studied the supercritical water gasification of biomass for hydrogen production **[251]**. By using feature permutation on an RF model, the authors identified the biomass concentration and temperature to be the two most influential factors affecting hydrogen production. This work found gasifier temperature to be of medium importance with features such as the gasifying agent and particle size being more important in predicting hydrogen production. However, these factors were not considered in Zhao *et al.*'s work.

Li *et al.* (2021) developed a GBR model to predict the syngas yield from the hydrothermal gasification of wet waste **[195]**. By assessing the feature importance, the authors showed that the model heavily relied on the gasifier temperature to make its predictions. Other factors such as the feedstock composition had a lesser effect. In comparison, in this work the gasifier temperature was found to be important but not multiple times more important than other factors as in Li *et al.*'s study. However, a direct comparison is difficult as the studied gasification process is significantly different to the ones studied in this work.

A challenge for all types of feature importance assessment is correlated features. As they contain largely the same information, their importance can be split across two or more features, making the correlated features appear less important than they are. This highlights the importance of removing strongly correlated features to ensure that important features do not get lost. Hence, a rigorous predictor selection process is required when feature importance assessment is one of the aims of model development. In this work, this issue has been addressed by computing Pearson's and Spearman's correlation coefficients to subsequently remove strongly correlated features. Finally, the difference between causation and correlation must be highlighted. Feature importance assessment can only bring to light which features the model uses to make its predictions. Although the behaviour of the model does not necessarily correspond to the real-world process, it may provide one with a good starting point on which variables may significantly affect some of the outputs

one is interested in. Additionally, it allows sensibility checks to confirm whether inputs deemed important by the theory are also deemed important by the model.

By combining both global and local methods, a more complete picture of a model's workings can be established. Global models powerfully describe the average behaviour of an ML Model and highlight trends in the data. Local explanations on the other hand are useful for understanding which factors might influence a given gasification system. This provides researchers with an idea of which parameters could be changed to optimise a given output. For instance, the developed method could be used to screen and test for promising process conditions before moving on to more expensive and time-consuming physical experiments. Furthermore, SHAP makes it easy to communicate the entire thought process to stakeholders. However, care needs to be taken when interpreting SHAP results as interpretability methods can be fooled to hide biases **[257]**. As such, it is the researcher's responsibility to avoid creating misleading explanations.

4.5 Chapter Summary and Conclusions

Gradient boosting regression (GBR) has been found to outperform other model types with a coefficient of determination (R^2) of 0.90 when averaged across all ten model outputs. Global and local interpretability methods were combined to extract new information from the developed models. This information can be used to guide the decisions of investors and policymakers by increasing their confidence in the models' results. This reduction in uncertainty can in turn promote the uptake of a new technology. The feedstock's particle size and the gasifying agent were found to be among the top predictor variables influencing the models.

Chapter 5

Environmental and Economic Modelling of Biomass and Waste Gasification

5.1 Chapter Overview

This chapter mostly focuses on the methods used to establish a comprehensive biomass and waste gasification modelling framework. In contrast, the following chapter (Chapter 6) discusses the results which can be obtained through the developed modelling framework in the form of a case study.

The chapter commences with Section 5.2, providing a detailed account of the programming methodology underlying the modelling framework and detailing how Monte Carlo methodology was used to account for uncertainties throughout the modelling framework. In Section 5.3, various approaches for modelling the subprocesses constituting the entire gasification scheme are introduced. Life cycle assessment (LCA) is presented as a method for modelling the environmental impacts of a system, while the benefit-cost ratio (BCR) and net present value (NPV) methods are introduced to assess a system's economic performance. Section 5.4 offers an exhaustive description of each process contributing to the complete gasification scheme, while Section 5.5 explores how these distinct components come together to form the overall modelling framework. This final section also explains how system parameters can be optimised for the system's two objectives: its environmental performance and its economic performance. Additionally, it explains how the developed modelling framework was turned into an accessible decision support tool through the development of a graphical user interface (GUI).

All model code may be accessed through the author's GitHub page: https://github.com/SimonAscher/Gasification_LCA_TEA.

5.2 Design of Modelling Framework

This chapter introduces a comprehensive modelling framework, built upon the Python programming language, with the overarching goal of precisely capturing the environmental and economic dynamics of biomass and waste gasification schemes. To achieve this life cycle assessment (LCA) and techno-economic analysis (TEA) methodologies were incorporated into a Monte Carlo simulation. The framework further enhances accessibility and expedites analysis through a user-friendly graphical interface, facilitating the evaluation of diverse system configurations, as further outlined in Chapter 6. Additionally, it leverages the machine learning (ML) models developed in Chapters 3 and 4 to predict gasification outputs affecting the LCA and TEA.

5.2.1 Object-Oriented Programming Structure of the Program

Object-oriented programming (OOP) is a programming paradigm that structures code into objects, instances of classes representing certain concepts. Each object encapsulates data and behaviour, allowing for the modelling of intricate systems in a modular and reusable manner. OOP is founded on four main principles: encapsulation, inheritance, polymorphism, and abstraction. Encapsulation involves bundling data and methods within a class to restrict direct access from external code, promoting information hiding and modularity. Inheritance allows classes to inherit properties and behaviours from other classes, fostering code reuse and hierarchy. Polymorphism enables objects to take on multiple forms, allowing different classes to be treated interchangeably. Abstraction focuses on simplifying complex systems by emphasising essential features and hiding unnecessary details. The merits of OOP include enhanced code organisation, maintainability, scalability, and improved reusability through modular and extensible code structures **[258]**.

In the context of this work, five main types of objects (denoted by italics) were used to structure the program:

- *Results* – These objects are used to store the overall environmental, economic, and energy generation results of a simulation.
- *Process* – These objects represent the real-life processes making up the overall gasification scheme (e.g. gasifier, combined heat and power (CHP) unit, pretreatment, etc.) and contain all relevant information related to a given process.
- *Requirement* – These objects represent energy, resource, or monetary requirements of some sort (e.g. electricity, heat, oxygen, the present value of a cash flow, etc.).
- *Cost/Benefit* – These objects are used to represent the costs/benefits of a given process.

- *Global Warming Potential* – These objects are used to represent the global warming potential (GWP) of a given process.

These five object types interact hierarchically. The *Results* object is composed of a collection of *Process* objects which represent the different components of a given gasification scheme. In turn, *Process* objects are made up of a range of different *Requirement* objects. For instance, the *Process* feedstock drying may require heat and electricity for the drying, initial capital investment to purchase and set up the dryer (i.e. CAPEX), and continuous operational and maintenance expenditures to operate it (i.e. O&M cost). Thus, the feedstock drying *Process* object would contain the following four *Requirement* objects: heat, electricity, the present value of a cash flow (due to CAPEX), and the annual value of a cash flow (due to O&M cost). Finally, a *Process* object's *Requirements* are used to calculate their resulting environmental and economic effects, which are stored as *GlobalWarmingPotential* and *Cost/Benefit* objects, respectively. For instance, an electricity *Requirement* would lead to the consumption of grid electricity, which in turn would incur a cost and an environmental impact related to the carbon intensity of the grid. Other *Requirement* objects may only occur an environmental or economic effect (e.g. the present value of a cash flow only occurs an economic effect). Figure 5.1 illustrates a condensed version of these interactions.

5.2.2 Uncertainty Accounting through Monte Carlo Methodology

The LCA and TEA of engineering systems require vast amounts of data at every step of the analysis. In the case of LCA, existing software such as *GaBi* or *SimaPro* can provide some of this data but other data may still be challenging to obtain. This is where the general understanding of the practitioner about the involved processes and systems is key. However, there are bound to be instances where data is lacking and practitioners have to make assumptions in the realm of large uncertainty **[259]**. In these cases, a robust uncertainty accounting framework can effectively map the uncertainty in a given model.

A robust uncertainty accounting framework is especially important when communicating results to stakeholders, such as investors, policymakers, and the public. Clear communication about uncertainties in LCA and TEA builds trust and transparency, helping stakeholders make informed decisions and understand potential risks. This is of particular importance for emerging technologies, such as gasification, and can guide their research and development efforts to address these challenges and improve their overall performance.

In this work, the Monte Carlo simulation methodology is employed to account for uncertainties within the modelling framework. Monte Carlo simulation is a commonly employed methodology for modelling uncertainties in engineering analyses, including LCA and TEA. It can effectively handle complex systems with many variables and interde-

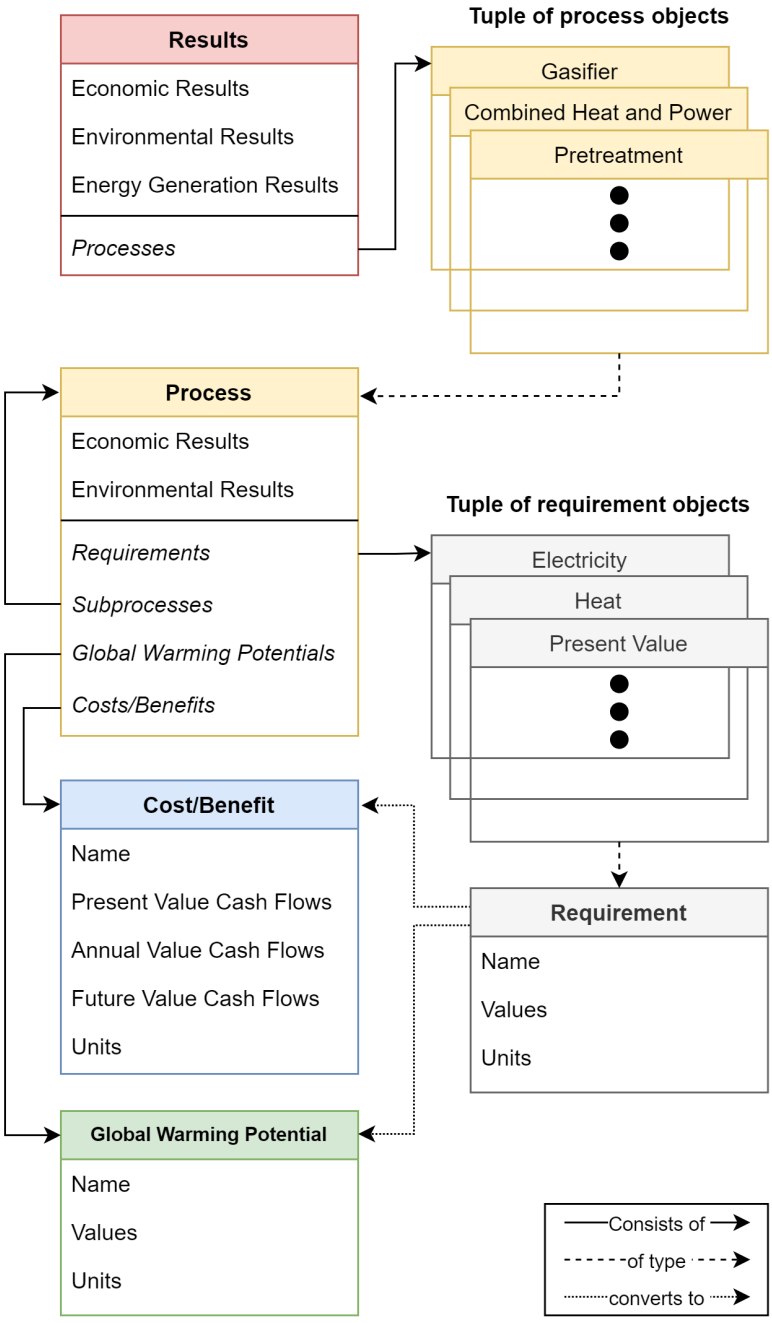


Figure 5.1: The five main object types (Results (red), Process (yellow), Requirement (grey), Cost/Benefit (blue), and Global Warming Potential (green)) used within the program and their interactions.

pendencies. By assigning probability distributions to uncertain parameters rather than relying on single-point estimates, the uncertainty in these parameters can be more realistically captured. This adaptability to situations where data is incomplete or uncertain is especially important for real-world engineering scenarios. Finally, Monte Carlo simulation produces probabilistic outputs, allowing engineers to obtain a probability distribution of the results. These can be effectively communicated to decision-makers, who can assess the likelihood of different outcomes and make informed decisions based on risk tolerance and project objectives.

5.2.2.1 Distribution Types Used to Model Uncertainty

For the Monte Carlo Simulation, certain parameters within the greater modelling framework were allowed to vary according to the following three distinct distribution types: normal distribution, triangular distribution, and uniform distribution. In addition, a fixed-value method was utilised, during which the parameter value was kept constant rather than subject to variation.

The normal or Gaussian distribution is defined by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)} \tag{5.1}$$

where *f(x)* is the probability density function (PDF), σ is the standard deviation, and μ is the mean. As a result of the central limit theorem, the normal distribution is often suitable for modelling naturally occurring phenomena. It tends to be a good default option to model uncertainty where a distribution's σ and μ are known.

The triangular distribution is defined by

$$f(x) = \begin{cases} \frac{2(x-l)}{(m-l)(u-l)} & \text{for } l \leq x \leq m \\ \frac{2(u-x)}{(u-m)(u-l)} & \text{for } l \leq x \leq u \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

where *f(x)* is the PDF, l is the lower limit of the distribution, m is the mode of the distribution (i.e. the peak of the distribution), and u is the upper limit of the distribution. The values must fulfil the condition $l <= m <= u$.

The triangular distribution is prevalent in ill-defined problems, where the underlying distribution is unknown, yet partial information regarding limits and mode is available. It is also versatile, allowing for the representation of skewed or asymmetrical uncertainties with a peak around the mode. Its three variables can be interpreted as the optimistic, most likely, and pessimistic estimates of the modelled parameter. This simplicity makes it accessible to stakeholders and policymakers who may not have a strong statistical background. It can effectively model parameters where a certain value is expected (m), but

variations are anticipated and the values for the modelled parameter in extreme or fringe cases are known. All this makes the triangular distribution a popular choice in economic modelling.

The uniform distribution is defined by

$$f(x) = \frac{1}{u - a} \tag{5.3}$$

where *f(x)* is the PDF, and l and u are the lower and upper limits of the distribution, respectively. This means any value within the interval defined by l and u is equally likely to be drawn from this distribution. In this work, the uniform distribution was used in place of the triangular distribution when the most probable value was unknown.

Finally, during the above-mentioned fixed-value method parameter values are not varied. This may be most appropriate when the value of a parameter is not subject to change or known to a large degree of certainty. In practice, the "NumPy" library (version 1.22.4) for Python was used to draw random samples from the above-defined distributions.

5.2.2.2 Accounting for the Prediction Model's Uncertainty within the Monte Carlo Simulation

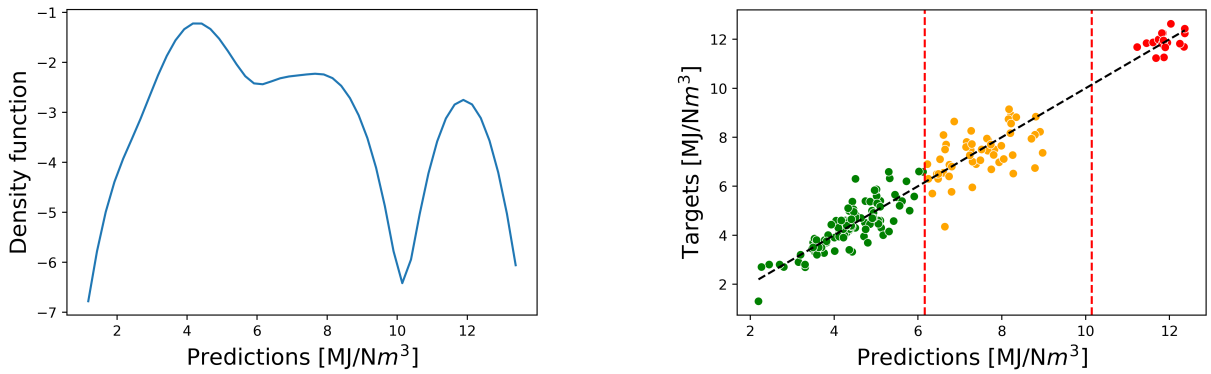
While the predictions of the ML prediction model, developed in Chapters 3 and 4, could directly be used in further analysis, this would neglect the inherent uncertainty in the prediction model. For this reason, a method to account for the uncertainty in the prediction model's predictions within the greater uncertainty accounting framework has been proposed.

Initially, it has been hypothesised that each model prediction is subject to some uncertainty. Globally this uncertainty can be quantified by the model's root mean square error (RMSE). Assuming the model's predictions vary according to a normal distribution of the form given by Equation 5.1, the uncertainty in a model's predictions can be accounted for by inserting the model's predicted value for the PDF's mean (μ) and the model's RMSE for the PDF's standard deviation (σ). In other words, the prediction model will now produce a distribution of predictions centred over the predicted value and varied according to the model's global RMSE, instead of simply providing a single prediction. This method will be referred to as the *global error method*. A major flaw of this method is that it cannot factor in whether certain predictions are being made more confidently than others. This flaw has been addressed by developing an improved method referred to as the *local error method*.

The *local error method* hypothesises that the model's predictive performance varies in certain regions. These regions can be identified by considering the errors of neighbouring predictions. Kernel density estimation has been used to cluster predictions into groups.

For this, Gaussian kernels have been used for each model output. The kernel's bandwidth was the key parameter which needed to be tuned, as the resulting kernel density estimate heavily depends on the used bandwidth. A too-low bandwidth leads to an undersmoothed curve picking up on spurious data and leading to too many minima in the resulting function. Using a too-high bandwidth leads to an over-smoothed curve, potentially missing some of the nuances of the underlying data. Hence, the bandwidth has been manually tuned for each output parameter to find density estimates deemed optimal for each output parameter. The local minima in the kernel density function, which indicate regions of low similarity between neighbouring predictions, were then translated to boundaries splitting predictions into groups. Finally, errors can be calculated for each resulting region.

The *local error method* first considers in which region a given prediction falls into. Then the error of this region is used to fit a PDF, which in turn can be used to get draws for a Monte Carlo simulation. Figure 5.2 gives an example of how the boundaries and regional errors of the *local error method* can be found for the prediction of one model output – the syngas' LHV. As shown in the figure, two boundaries and three regions were identified. The RMSEs for the first, second, and third regions (shown by the blue, orange, and green data points) were found to be 0.476, 0.829, and 0.371 respectively. In comparison, the overall RMSE for the syngas LHV prediction model was 0.600.



(a) (b)

Figure 5.2: (a) Illustration of the use of kernel density estimation for the determination of prediction boundaries, exemplified for the prediction of the syngas' LHV; (b) Scatter plot of predictions vs. targets and illustration of the resulting boundaries (shown by red dashed line) used to group predictions into error regions (illustrated by different colours). The black dashed line indicates the case where the predictions perfectly match their targets.

Figure 5.3 compares the *global error method* and *local error method*. Distributions and their corresponding scaled PDFs are shown for each error region. From the figure, it becomes clear that the distributions based on the *global error method* (shown by opaque bars) have a wider spread than the distributions based on the *local error method* in the first and third region. The distribution in the middle section was found to have a wider spread for the *local error method* than the *global error method*. In other words, the model's

performance in the first and third regions was better than the model's average performance and worse in the second region.

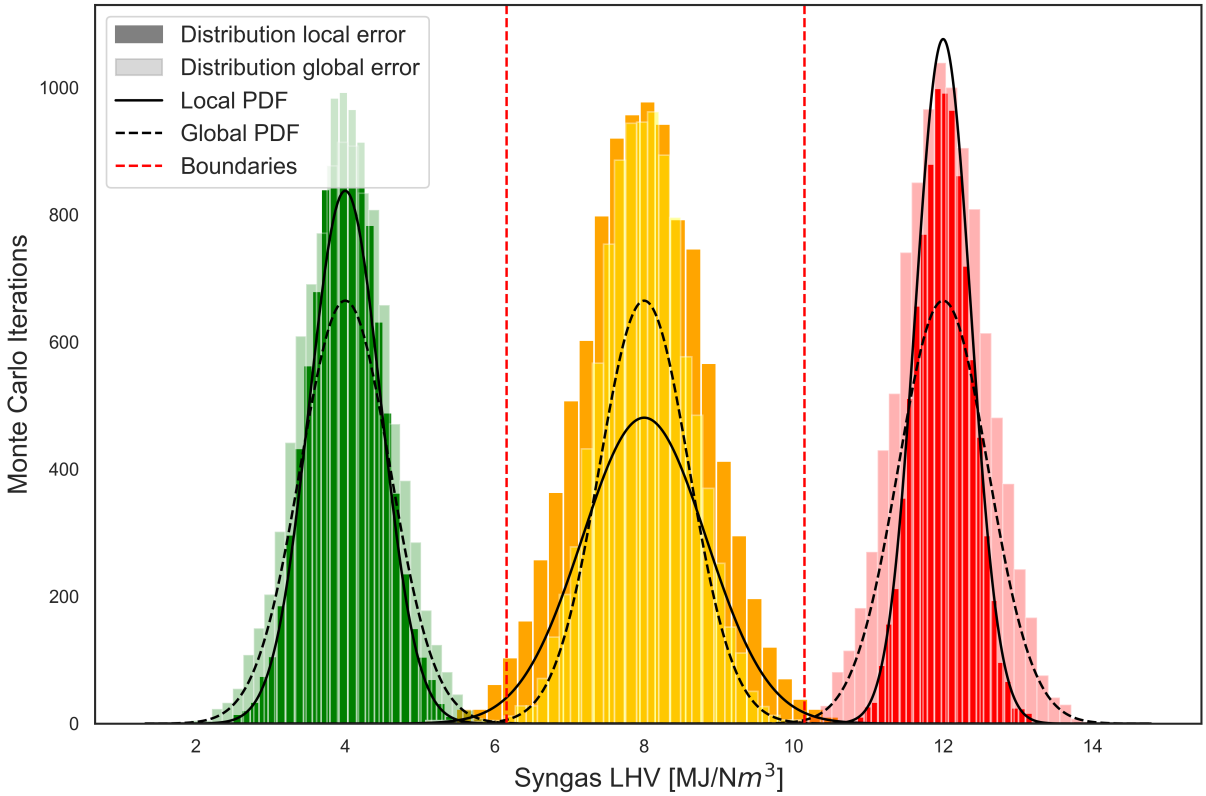


Figure 5.3: Comparison of the distributions resulting from the global and local error method for 10,000 Monte Carlo iterations, exemplified for the prediction of the syngas' LHV. Distributions are shown for the three exemplative predicted values 4 (green), 8 (yellow), and 12 (red) MJ/Nm^3, where each distribution falls into one of the 3 different error regions. Solid colours show the results for the local error method, whereas opaque colours indicate the use of the global error method. For plotting purposes, the fitted probability density functions (PDFs) were scaled by 1000.

5.3 Process Modelling Methodology

To model the entire biomass and waste gasification scheme, each process was modelled individually before being combined to make up the overall system. The development of general process models primarily relied on four key approaches (and a combination of them): (i) employing models based on physical and chemical reactions to describe a process, (ii) utilising empirical data for modelling purposes, (iii) incorporating the machine learning models developed in Chapters 3 and 4 to estimate process data, and (iv) treating a process as a black box.

When a process was treated as a black box, an approach similar to Skorek-Osikowska *et al.* (2017) was used **[260]**. Here, a process such as CCS is treated as a black box and

modelled using data from existing systems found in the literature. Whilst some processes could be modelled in isolation, other processes, such as feedstock drying, would affect other process models (e.g. by changing the feedstock moisture). Hence, it was important to have a feedback link between individual process models where required.

Section 5.4 details the exact approaches used for each individual process. The environmental modelling, in the form of LCA, is further described in Section 5.3.1, whereas the economic modelling is described in Section 5.3.2.

5.3.1 Environmental Modelling through Life Cycle Assessment

Environmental modelling of engineering systems is essential for identifying and mitigating potential environmental impacts, guiding sustainable design and decision-making, and ensuring responsible resource management throughout the life cycle of products and processes. LCA has established itself as the arguably most popular method for comprehensive environmental assessments, especially for evaluating the entire life cycle of products and services. The goal of LCA is to provide a holistic and scientifically sound assessment of the environmental performance of a product or service, enabling informed decision-making for sustainability. Its standardised framework, outlined in ISO 14040 and 14044, has contributed to its widespread adoption in environmental management and decision-making processes. ISO 14040 establishes the overarching principles and framework for conducting LCA, while ISO 14044 provides more detailed guidance and specific requirements for the practical implementation of LCA studies. Together, these standards ensure a standardised and consistent approach to environmental assessment, facilitating comparability and reliability of LCA results **[161, 162]**.

LCA typically involves four main stages, often referred to as phases, as illustrated by Figure 5.4. These stages provide a systematic framework for assessing the environmental aspects and potential impacts associated with a product, process, or service throughout its entire life cycle. The four stages of LCA are (i) goal and scope definition, (ii) life cycle inventory (LCI) analysis, (iii) life cycle impact assessment (LCIA), and (iv) life cycle interpretation.

The initial stage is goal and scope definition, where the study's objectives are clearly defined, system boundaries are outlined, and a functional unit (FU) is identified. The subsequent LCI stage entails the collection and compilation of data on all inputs and outputs at each life cycle stage, forming a comprehensive inventory of resources and environmental releases. Following this, the LCIA stage evaluates the potential environmental impacts using characterisation factors and quantifies relevant impact categories, such as global warming potential (GWP). The final interpretation stage involves summarising findings, assessing reliability, identifying areas for improvement, and providing recommendations to support informed decision-making and sustainable practices. It's important to note that

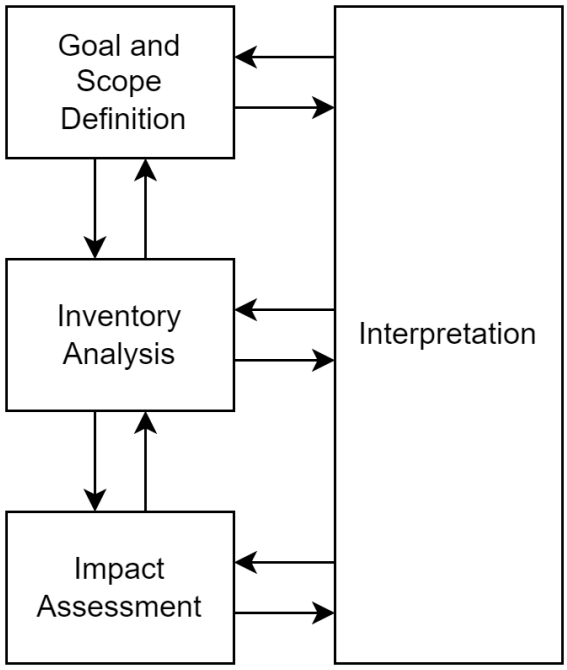


Figure 5.4: The four key stages of LCA. Adapted from ISO 14040 **[162]**.

these stages are iterative, and the process may involve revisiting earlier stages as more data becomes available or as the understanding of the system under analysis evolves. This means assumptions made during the analysis may need to be re-evaluated throughout the process. **[161, 162]**.

Impact categories are used to organise and quantify the environmental impacts associated with a product, process, or service. These impact categories help to assess the potential effects on the environment and provide a structured way to communicate the results of an LCA study. The choice of impact categories depends on the goals of the assessment and the system under consideration. Due to the urgency of climate change concerns, GWP is often considered one of the most important impact categories. However, other impact categories such as ozone depletion potential, acidification potential, and eutrophication potential, to name a few, inform us about other critical effects on our planet and ecosystems. More specifically, they inform us about the potential harm to the ozone layer, the contribution to acid rain, leading to soil and water acidification, and the potential for nutrient enrichment in water bodies, causing excessive growth of algae and other aquatic plants, respectively.

Due to data availability limitations, it is not always possible to assess all LCA impact categories. Since GWP was deemed the most pressing concern in our day and age, the analysis done in this work was limited to this impact category. The impact category GWP uses carbon dioxide equivalents (CO_2-eq) as its reference unit and the following equivalency factors, relative to the impact of fossil CO_2: fossil $CO_2 = 1$, biogenic CO_2

= 0, fossil CH_4 = 29.8, non-fossil CH_4 = 27.2, and fossil NO_2 = 273. These values are based on the IPCC Sixth Assessment Report (AR6) and are for a 100-year time horizon **[6]**. Biogenic carbon, which refers to carbon that is part of the natural carbon cycle, typically associated with living organisms and biological processes, is of particular interest as its release is considered to not incur a global warming response **[261]**. This makes it essential to rigorously account for the fossil or biogenic nature of carbon to accurately reflect its effect on global warming. In addition, the practitioner needs to be clear about the employed time horizon for their LCA study, as the dynamics of the biogenic carbon cycle may be challenging to accurately reflect on shorter time horizons **[261, 262]**.

The goal of this work is to create a comprehensive environmental model for biomass and waste gasification schemes. Any given gasification scheme is made up of several processes, which may have a range of alternatives to choose from (e.g. different CHP units or pretreatment options). All processes are described in Section 5.4. The developed LCA model is to be used as part of a greater tool to holistically assess the environmental and economic performance of biomass and waste gasification schemes, as further detailed in Section 5.5.

The FU of the system is the use or treatment of 1 tonne of feedstock and the selected allocation method, as well as the allocation problem in general, are discussed in more detail in Section 5.3.1.1. The system boundary includes the processes described in Section 5.4, with a number of them being optional (e.g. feedstock pretreatment). A simplified system boundary which shows the main processes and product flows is shown in Figure 5.5.

The following four types of processes are not considered in this work:

- **Infrastructure and construction:** Emissions resulting from the manufacture and construction of system components are generally insignificant and have hence been neglected **[263]**.
- **Transportation:** Emissions due to the transportation of the feedstock and the generated biochar have been shown to be negligible in the context of an overall gasification scheme and have hence been neglected **[164]**.
- **Feedstock cultivation:** Here the common assumption that net-zero carbon emissions are linked to the process from cultivation to the point of collection is followed **[161, 263]**.
- **Land use changes:** The feedstock generation/production is outside the system boundary. For this reason, it was assumed that the land use changes resulting from the scheme would be negligible **[263]**.

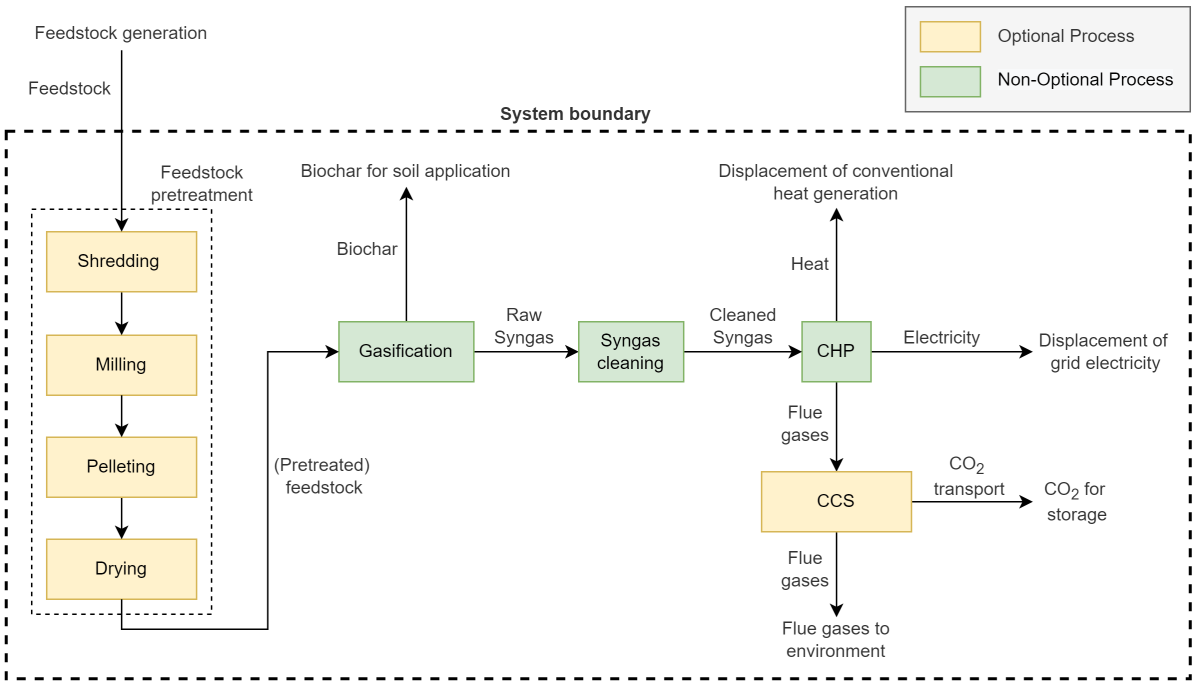


Figure 5.5: Simplified system boundary.

5.3.1.1 The Allocation Problem

The allocation problem is a well-established issue in LCA studies and it can significantly impact their results. During allocation, the LCA practitioner aims to assign the distribution of a system's or product's environmental burdens to the functions that caused these impacts in a representative fashion. This problem is most apparent when a system produces more than a single product. Since most bioenergy systems produce more than one product (e.g. electricity, heat, and biochar for the case of a gasification scheme), LCAs on bioenergy systems in particular are required to make sound allocation choices **[264]**. LCA practitioners typically look towards ISO 14044 and its two amendments in 2006 and 2020 for guidance on good LCA practices **[161]**. However, the latest amendment was not entirely successful in removing confusion about the hierarchy on how to best address the allocation problem **[265]**.

ISO 14044 states that allocation should typically be avoided. This should either be done by dividing processes into subprocesses so that each subprocess produces only a single product or through system expansion. System expansion involves expanding the system boundaries to include all relevant co-products and considering the environmental impacts of alternative uses for those co-products. This is done by comparing the system to an appropriate reference system, which helps to account for the environmental impacts of all products or processes without allocating burdens **[161]**. In the case of bioenergy LCAs, this might mean that the generated electricity is assumed to displace electricity from the grid mix of the study's country.

While system expansion is the recommended approach and is well-suited for case studies with a clearly established reference system, its applicability has certain limitations. The selection of the reference system plays a pivotal role in influencing the outcomes of the LCA, placing considerable weight on this decision. If the reference system is subject to uncertainties, an inappropriate choice of reference system may introduce errors in the LCA. For example, in the context of countries trying to rapidly decarbonise their electricity grids, the associated GWP of the grid mix can undergo significant changes within just a few years. Consequently, the LCA becomes highly location and time-specific. To mitigate this bias, practitioners may incorporate a sensitivity analysis or other approaches to address uncertainty in their LCAs. When implementing a sensitivity analysis, it might be prudent to present results for both the present scenario and likely future scenarios or consider outcomes for best, most likely, and worst-case scenarios.

Allocation should only be used if the system expansion method is not appropriate **[161]**. In this case, several allocation approaches may be considered. As a first choice, the physical relationships of the co-products should be used as a basis for allocation. For instance, several LCAs were found to allocate based on energy. The issue with this approach is the lack of differentiation between the quality of the energy (e.g. electricity and heat) **[264]**. When using the physical relationship of co-products as a basis is not possible, other options may be considered **[161]**. Economic value was used in some studies as an allocation basis **[264]**. This is typically a poor choice, as temporal and regional differences can significantly impact the market prices.

In cases where allocation is challenging, substitution can be considered. This involves replacing the co-product with an equivalent product or process that serves the same function but has known environmental impacts. Clearly, this approach requires expertise and judgement from the practitioner. Similar levels of expertise and judgement are required to establish cut-off rules. Establishing cut-off rules involves excluding certain co-products or processes from the analysis if they contribute to a small percentage of the overall system's environmental impacts. This can simplify the assessment while ensuring that significant contributions are not overlooked. However, the practitioner needs to be clear about these rules and be certain that the exclusion of certain co-products does not invalidate the LCA's results.

5.3.2 Economic Modelling

The scheme's economic feasibility was assessed using two main economic indicators, the benefit-cost ratio (BCR) and the net present value (NPV).

The NPV method is a financial technique used to evaluate the profitability of an investment or project. It involves discounting all expected future cash flows associated with the investment to their present value (PV) equivalent, including both inflows and

outflows and may be given by

$$NPV = \sum PV(Benefits) - \sum PV(Costs) \tag{5.4}$$

When using the NPV, a project is deemed economically justified for NPV ≥ 0 **[266]**.

The BCR is another financial metric used to evaluate the economic feasibility of a project or investment. It is calculated by dividing the PV of the project's expected benefits by the PV of its costs. Hence, it may be defined by

$$BCR = \frac{\sum PV(Benefits)}{\sum PV(Costs)} \tag{5.5}$$

In the context of project evaluation, a BCR > 1 indicates that the benefits of a project outweigh its costs, suggesting that the investment is potentially worthwhile. Conversely, a BCR < 1 suggests that the costs exceed the benefits, which may indicate a less economically viable project **[266]**.

Both methods can be used to assess the financial desirability of a project and provide decision-makers with a quantitative measure to evaluate projects and compare alternatives. By assessing the distance from the break-even point for each method (i.e. BCR $= 1$ and NPV $= 0$), one can effectively gauge the risk associated with the investment, providing insights into the extent to which the project's returns exceed or fall short of its costs. The greater the distance from break-even, the lower the risk, whereas a closer proximity indicates a higher level of financial risk.

Benefits typically result from the sale of a system's products or services, whereas costs include factors such as CAPEX for system components and operation and maintenance (O&M) costs. These are further discussed in Sections 5.3.2.1 and 5.3.2.2, respectively. The costs and benefits related to the system studied here, are detailed in Sections 5.4 and 5.5, respectively.

5.3.2.1 Capital Expenditure (CAPEX) Estimation

The CAPEX refers to the funds spent by a company to acquire, upgrade, or maintain physical assets such as property, industrial buildings, or equipment. A system's CAPEX can be defined at several levels, depending on whether factors such as engineering work, procurement, or financing costs are taken into account. This is illustrated by Figure 5.6, where the bare erected cost (BEC) is considered the lowest level. It comprises the cost of process equipment and supporting facilities, as well as the labour costs for their installation and/or construction. At the highest level, the total overnight cost (TOC) includes costs for engineering, procurement, and construction, as well as project and process contingencies and all overnight costs **[267]**.

CAPEX data is often uncertain and scarce for novel systems or components, meaning

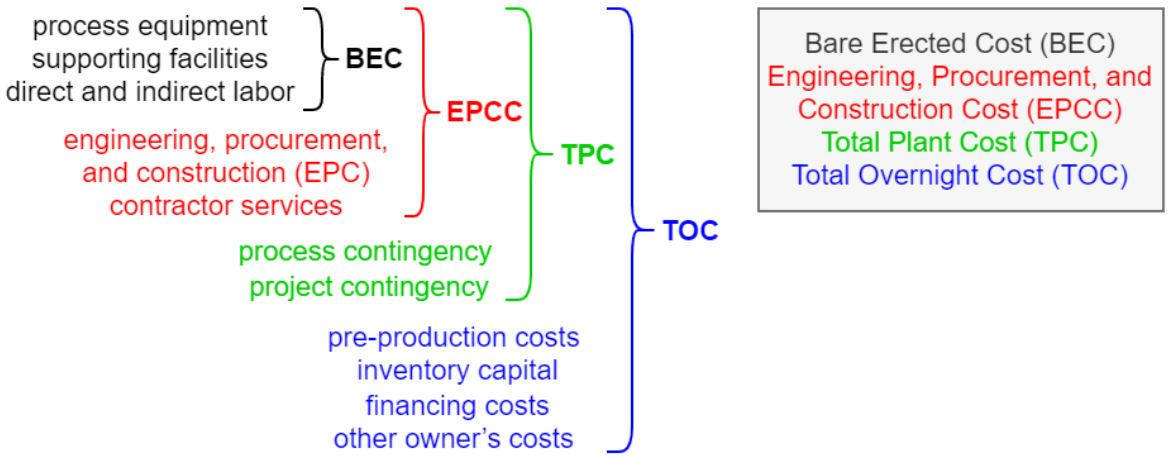


Figure 5.6: CAPEX levels and their components. Adapted from James *et al.* (2019) **[267]**.

that data typically needs to be adapted to a particular system **[268]**. When searching for CAPEX data in the existing literature, it has often been unclear which CAPEX level authors referred to which introduces a further degree of uncertainty. In addition, the underlying datasets to estimate the CAPEX of components or systems were often prepared by different engineers with different bases for engineering design and costing which adds further uncertainty. Similarly, the N^{th} plant assumption, which states that a given technology's cost decreases over multiple iterations due to research and development efforts, leads to further inherent uncertainties, in particular for pre-commercial technologies. In combination, these factors mean that uncertainty in capital costs of $\pm 30\%$ is not uncommon for many novel engineering systems **[269]**.

To minimise the uncertainty in a component's CAPEX, the CAPEX at the TOC level was considered throughout this work. Whenever literature data was quoted at a different level, values were standardised to TOC level using typical factors for CAPEX estimation based on the delivered equipment cost **[270, 271]**. For instance, the TOC is calculated by multiplying the delivered equipment cost by a factor of 3.8 for a solid processing plant or a factor of 4.8 for a fluid processing plant.

A range of cost estimation methods has been explored in Section 5.3.2.3 which can be used to calculate CAPEX values of components. Whenever possible a cost estimating relationship (CER) (described in Section 5.3.2.3.3) was established to model a system's or component's CAPEX. The CER method was deemed more reliable than other methods as it does not rely as heavily on individual reference systems. If there was not enough data available to establish a CER, the power-sizing method (described in Section 5.3.2.3.2) was defaulted to.

5.3.2.2 Operation and Maintenance (O&M) Cost Estimation

The operation and maintenance (O&M) cost of a process or system component is frequently modelled as a fraction of the component's CAPEX. By analysing a range of different O&M costs, utilised in a range of bioenergy studies **[272–275]**, it was found that the typical annual O&M cost ranges from 2% to 7% of the components CAPEX. Based on this a default O&M cost fraction defined by a triangular distribution with a lower limit, mode, and upper limit of 2%, 4%, and 7% was defined. Whenever no component-specific O&M cost data was available, the model reverted to this distribution. In cases where process/component-specific O&M cost data was available, this data was used instead. This is further described in the sections detailing the individual process models (i.e. subsections of Section 5.4). Finally, it was assumed that the cost for minor equipment refurbishments and replacements is included in the O&M cost **[274]**.

5.3.2.3 Cost Estimation Methods

Economic analysis for engineering projects involves a range of present and future economic considerations. To obtain reliable estimates, engineers rely on several techniques to calculate the costs and prices of engineering systems and their components.

5.3.2.3.1 Cost Indices

Indices, such as the Chemical Engineering Plant Cost Index (CEPCI) **[276]**, are dimensionless numbers that relate a cost (or price) from a reference/base year to the year of interest. They represent a convenient way to estimate a cost from historical data. They can be used at a lower level to estimate factors such as equipment, labour, and materials costs and at the top level to estimate the total system or project cost. Amongst other factors, costs typically vary with time due to factors such as inflation, technological advances, and changes in the availability of labour and materials. The cost of an item in year n can be estimated by

$$C_n = C_k \frac{I_n}{I_k} \tag{5.6}$$

where k is the reference/base year for which the item's cost is known, n is the year of interest for which the item's cost is to be estimated ($n > k$), C_n is the item's estimated cost in year n, and C_k is the item's known cost in reference/base year k **[266]**.

5.3.2.3.2 The Power-Sizing Method

Some costs, especially capital investments, are strongly dependent on the size (or capacity) of the plant/system. The power-sizing technique takes this into account by scaling costs according to a cost-capacity factor f which reflects economies of scale. Scaling

can be done in either direction, i.e. the costs of larger or smaller plants can be estimated. The cost of a designed plant C_d can be estimated by

$$C_d = C_r \left(\frac{S_d}{S_r} \right)^f \tag{5.7}$$

where C_d and C_r are the designed and reference cost of the system/plant respectively. S_d and S_r are the corresponding size/capacity of the designed and reference system/plant. f is the cost-capacity factor which determines the rate at which the cost varies with the plant's size. Note that C_d will be in the same currency and year as C_r, and S_d and S_r should be given in the same physical units **[266]**. In general, $f < 1$ which indicates decreasing economies of scale. Typical values for f can be found in the existing literature **[270]**. For bioenergy systems and their components, typical values for f are between 0.6 and 0.8, with a value of $f = 0.7$ being used as a default for this work **[277]**.

A key issue with the power-sizing technique is that entirely relies on a single reference value. When the available reference data is of high quality and matches the designed system well, this is a minor issue. However, when there is uncertainty in the reference value or the specifications of the designed and reference system vary, this may introduce potential errors and lower the trust in the estimate. To avoid the reliance on a single reference value, an alternative is the development of a cost estimating relationship (CER), i.e. fitting a regression model to a set of data points which are relating the system/plant cost and size/capacity. This is further explored in 5.3.2.3.3

5.3.2.3.3 The Cost Estimating Relationship (CER) Method

CERs are a form of mathematical models which correlate the cost of an engineering system to one or more design parameters (cost drivers). To develop a CER the collection of reliable data relevant to the designed system/plant is essential. As data often varies based on factors such as currency, base year, and geographical location, normalisation is required to account for differences due to these. Other methods, such as cost indexes (Section 5.3.2.3.1), or price inflation methods can be used to normalise data **[266]**.

Once data is collected and normalised, a regression model can be fitted to the collected data. In theory, any type of regression model could be fit to describe the data, but in practice, the relationship between cost driver(s) and system cost often follows a linear, power, logarithmic, or exponential relationship. Trends usually become apparent by plotting the data, after which an appropriate equation can be selected. The coefficients of the selected equation are typically determined by least squares fitting, although other methods exist **[266]**.

It is important to know how well the developed model can predict cost. For this goodness of fit measures, such as standard error (SE), RMSE, and R^2, can be used to determine the performance of the developed model and validate it. Judgement is required

when deciding between various fits. If several fits appear equally suitable, typically the one with the best goodness of fit should be chosen **[266]**.

5.3.2.3.4 Conversion between Different Countries and Currencies

Location can be an important driver of costs, with factors such as labour and equipment costs potentially varying to a large degree depending on the country and currency of interest. For instance, labour costs have been shown to vary up to 33% between the USA and Euro region for certain technologies **[278]**. Furthermore, currency fluctuations can be significant over the course of a year, which raises the question of which exchange rate to use.

Whenever possible country-specific data was used in this work. When this was unavailable, costs were converted by using the average exchange rate of a given cost item's reference year. In addition, this method was used with any of the other methods which have been described in this Section.

5.4 Process Models

5.4.1 Gasification and Gas Cleaning

The gasification process and the subsequent gas cleaning stage were mainly modelled using the ML models introduced in Chapters 3 and 4. This means the gasification process was treated as somewhat of a black box, where the developed ML models were used to establish a relationship between factors such as the feedstock properties and gasification conditions and the yield and composition of gasification products. Specific requirements depending on the used gasifying agent are further described in Section 5.4.1.1.

Auxiliary heat and electricity demands for both the gasification process and the subsequent gas cleaning stage were derived from the existing literature. When considering the auxiliary heat demands of the process, it is important to keep in mind that while the pyrolysis process is endothermic and the incineration process is exothermic, gasification sits somewhere in the middle and can be either **[88]**. In this study, a conservative approach was adopted, estimating that, on average, 20% of the heat generated in the CHP unit would be utilised to support the gasification and gas cleaning processes **[24, 164]**.

A range of literature was analysed to estimate the auxiliary electricity demands. It was found that electricity requirements for the gasification and gas cleaning stages were low, typically ranging from less than 2% to 5% of the energy contained in the feedstock input **[279–281]**. Hence, a triangular distribution defined by a lower limit, mode, and upper limit of 1.5%, 3%, and 5% was used in the model to characterise auxiliary electricity requirements for the gasification gas cleaning stage as a fraction of the energy contained in the feedstock input.

To estimate the CAPEX of a gasifier, 22 data points were collected from the literature **[269, 282–292]** to define a CER. As the developed model needs to be suitable for a range of gasifier designs and sizes, a regression model was considered more suitable than using the power-sizing technique based on individual reference values. To develop a CER, a cohesive data set of high quality is important. Unfortunately, limited gasifier cost data was available, so to avoid discarding data, several approaches were used to normalise all data points:

1. All data was converted to the same currency by using the average exchange rate of the year the data is from.
2. CEPCI indices were used to convert all data to the same year.
3. As the system size was quoted in a range of different units (e.g. tonnes feedstock /hour, MW_{el}, $MW_{feedstock-LHV}$, kWh feedstock-LHV/hour, etc.), all values were converted to the same units.
4. If balance of plant cost factors or overall installation factors were available, values were converted to their TOC equivalent. If no factors were given it was assumed that the cost was already given as TOC.
5. Data points including gas cleaning or power generation were converted to their gasifier-only equivalent.
6. Outliers were removed and regions of too high sparsity were identified.
7. Data was split into two data sets based on the reactor type (i.e. fluidised bed and fixed bed).

The cleaned and normalised data set for fixed and fluidised bed gasifiers is shown in Figure 5.7. A clear linear relationship between plant size and CAPEX can be seen for fixed bed gasifiers, which is a result of all the data originating from the same references **[260, 292]**. The resulting fit is shown in Figure 5.8a.

For fluidised bed gasifiers a non-linear relationship can be identified from Figure 5.7. Furthermore, the data can be split into two regions – small-scale gasifiers of size $< 5MW_{el}$ (approximately 2.1 tonnes feedstock/hour) and medium-scale gasifiers of $> 5MW_{el}$. After preliminary analysis and comparison of various regression fits, a linear model was fitted to the small-scale gasifier data and a power function was used to describe the medium-scale data. Figure 5.8b shows the fit of these models.

The performance of all models is summarised by Table 5.1. In general, cost appears to increase linearly for fluidised bed gasifiers of a size $< 5MW_{el}$. Beyond this, cost increases more slowly, which is likely a result of economies of scale. Across both regions, errors

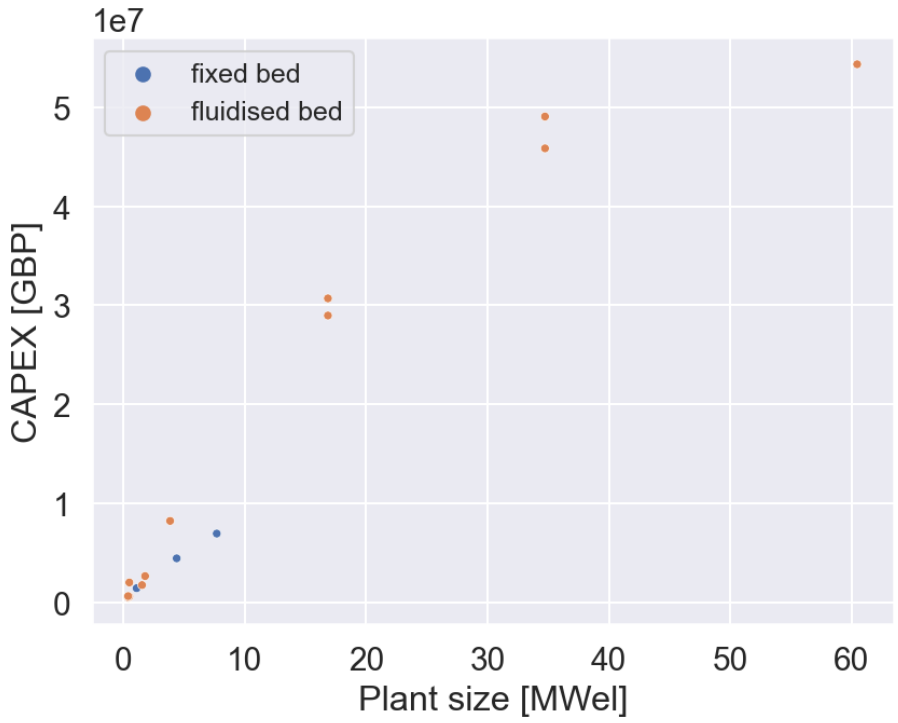
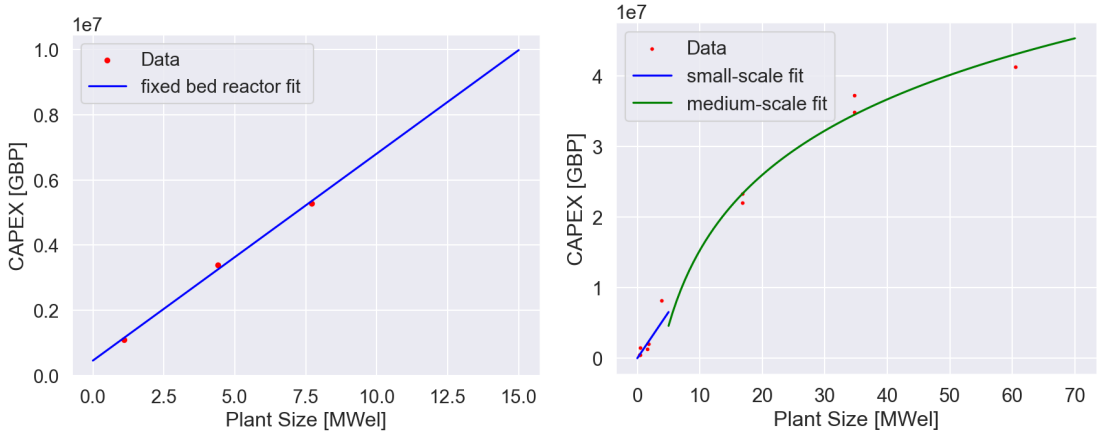


Figure 5.7: Scatter plot of normalised fixed bed and fluidised bed gasifier CAPEX.



(a) (b)

Figure 5.8: Regression fits to the normalised CAPEX of (a) fixed bed and (b) fluidised bed gasifiers.

are quite large, especially in the boundary region at a plant size of $5MW_{el}$ significant uncertainty exists, with a 30% difference between the CAPEX predicted by the small-scale and medium-scale model.

Table 5.1: Performance summary of CAPEX models fitted to fixed bed and fluidised bed gasifier data.

	RMSE [£]	MAPE [%]	R^2	Fitted function
Fixed bed	91,436	4	0.997	$y = 634,900x + 460,157$
Fluidised bed ($< 5MW_{el}$)	1,384,931	36	0.647	$y = 1,307,335x + 9,876$
Fluidised bed ($> 5MW_{el}$)	1,594,225	4	0.957	$y = 6.86 * 10^{10} x^{(2.23*10^{-4}} - 6.87 * 10^{10}$

The fitted CER models do not inherently consider uncertainty in their predictions. For the Monte Carlo simulation, a triangular distribution was defined from which random samples were drawn. The CER model's prediction for a given plant size was used as the triangular distribution's mode, whereas the lower and upper bounds were defined by the predicted value ± the predicted value multiplied by the model's MAPE.

Typically, the CAPEX for gas cleaning was quoted as a fraction of the total cost of a gasifier. Based on literature values, this fraction was defined by a triangular distribution with a lower bound, mode, and upper bound of 17%, 24%, and 33% **[269, 282, 287]**. During the Monte Carlo simulation, a random sample was drawn from this distribution to estimate the CAPEX of gas cleaning.

5.4.1.1 Agent Provision

5.4.1.1.1 Oxygen

ASUs are used to separate atmospheric air into its components — for instance nitrogen and oxygen. A range of different cryogenic and non-cryogenic technology options are available. Cryogenic distillation is currently the most cost-effective process and can produce oxygen at high purities. However, research on other technologies utilising adsorption, membranes, or chemical processes continues and may challenge cryogenic distillation in certain cases **[293]**.

To model the energy consumption for oxygen production by cryogenic distillation, data was compiled from several scientific papers **[294–297]**. After converting all data to a consistent unit, it was found that the specific power consumption for producing oxygen varied between 0.28 and 0.64 kWh/Nm^3 O_2 with a mean and standard deviation of 0.45 and 0.14 kWh/Nm^3 O_2 respectively. These values were used to define the mean and standard deviation of a normal distribution which is used during Monte Carlo simulation.

5.4.1.1.2 Steam

The production of steam comes at a significant energy expenditure. To determine the required amount of steam, the process' steam to biomass ratio was used. The theoretical thermal energy requirement for first raising the temperature of water to its boiling point and subsequently evaporating it Q_S was calculated using

$$Q_S = M_S(h_w + c_w(T_b - T_r)) \tag{5.8}$$

where M_S is the required mass of steam, h_w is the specific latent heat of vaporisation of water at the boiling temperature T_b, c_W is the specific heat of water, and T_r is the room temperature.

A penalty due to boiler inefficiencies was added to the theoretical energy thermal energy requirement. For this a range of literature on boiler performance, using various fuels, was reviewed **[298–302]**. In general, gaseous fuels, such as natural gas, were shown to have higher efficiencies than solid and liquid fuels, with efficiencies of up to 95% **[301]**. The lowest efficiencies found in the reviewed literature were 76.08% **[300]**. However, typical modern boilers achieve efficiencies of $> 90\%$ **[302]**. Hence, a triangular distribution with a lower limit, mode, and upper limit of 76%, 90%, and 95%, respectively.

5.4.2 Combined Heat and Power (CHP) Generation

The conversion of the cleaned syngas in a CHP unit to electricity and heat was modelled based on the conversion efficiencies and parasitic loads of existing CHP units. Table 5.2 summarises the relevant data for a few existing CHP units, based on a range of different technologies. During model execution, the most suitable CHP unit is selected based on the size of the system, unless a particular CHP unit is selected by the user.

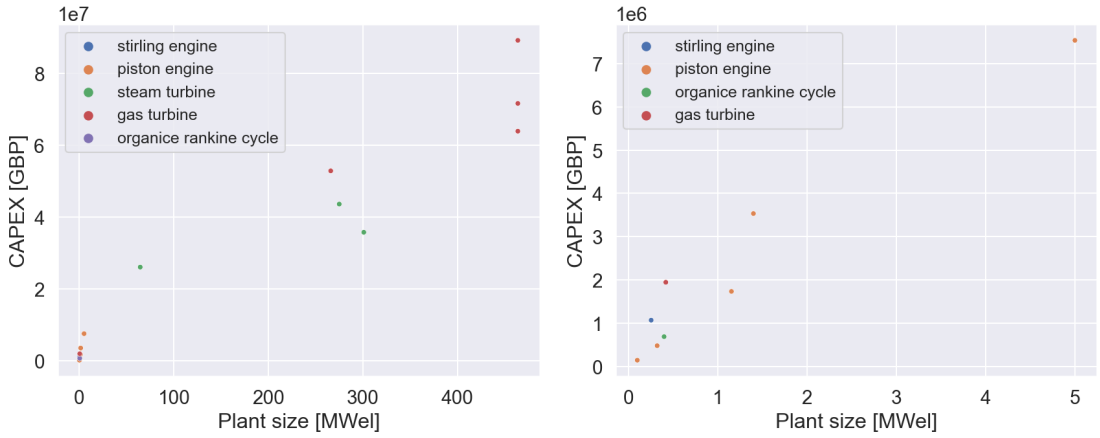
Table 5.2: Conversion efficiencies and parasitic loads of various CHP units.

	Stirling Engine 1.2 *kW*	SOFC 1.7 *kW*	Micro gas turbine 30 *kW*	Gas engine 1,634 *kW*
Manufacturer	Whispergen WhisperTech	ISM Sunfire	Capstone Turbine Corporation	Jenbacher
Electrical efficiency [%]	12	35	26	45
Thermal efficiency [%]	80	55	52	41.4
Parasitic loads [% of total electricity produced]	2	2	10	-
Reference	**[303, 304]**	**[304]**	**[303, 304]**	**[305]**

Emissions due to the combustion of the generated syngas in the CHP unit are modelled based on a simplified syngas combustion model, as described in Section 5.4.2.1.

To estimate the CAPEX of CHP units 17 data points were collected from the literature **[260, 267, 282, 283, 288, 291, 292, 306]** to define a CER. Like the gasifier's

CAPEX model, a regression model was considered more suitable than using the power-sizing method. Data was collected for a range of different generator types, as shown in Figure 5.9. No clear differences in CAPEX due to different generator types can be identified from the scatter plot. For this reason, and since the overall amount of available data is very limited, the developed models did not distinguish between different generator types.



(a) (b)

Figure 5.9: Scatter plots of the normalised CHP CAPEX by generator type. (a) Shows all data and (b) focuses on small-scale plants of a capacity $< 5MW_{el}$.

All data was normalised using the following steps:

1. All data was converted to the same currency by using the average exchange rate of the year the data is from.

2. CEPCI indices were used to convert all data to the same year.

3. If balance of plant cost factors or overall installation factors were available, values were converted to their TOC equivalent. If no factors were given it was assumed that the cost was already given as TOC.

4. Outliers were removed and regions of too high sparsity were identified.

Once all data was normalised, two regions were identified: small-scale CHP units of size $< 5MW_{el}$ and medium-to-large-scale CHP units of 5 to $500MW_{el}$. After preliminary analysis and comparison of various regression fits, a power function was fitted to the small-scale CHP unit data and a linear model was deemed most appropriate for the medium-to-large-scale CHP unit data. The fitted models are shown in Figure 5.10.

The performance of both models is summarised by Table 5.3. A good fit was achieved for small-scale plants, although the error is still significant. The model was fitted using data from piston engines, Stirling engines, gas turbines, and organic Rankine cycle generators.

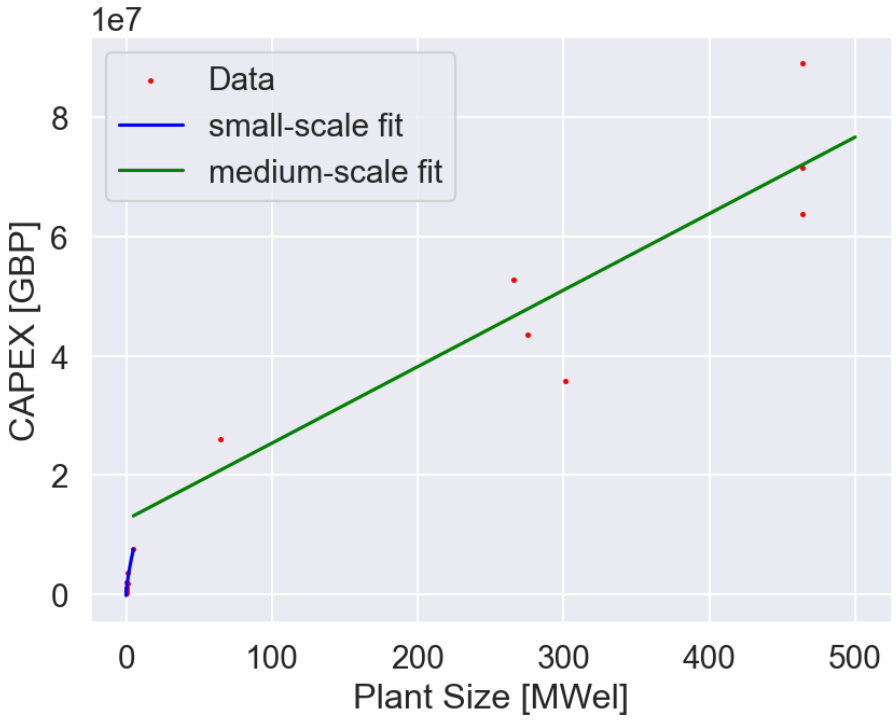


Figure 5.10: Regression fits to the normalised CAPEX of CHP units.

This means in this region the model approximates the cost of a range of different generators well. At a larger scale (i.e. ($> 5MW_{el}$)) the data was limited to gas turbines and steam generators. In this region, the CHP CAPEX is highly uncertain.

Table 5.3: Performance summary of CAPEX models fitted to CHP data.

	RMSE [£]	MAPE [%]	R^2	Fitted function
Small-scale ($< 5MW_{el}$)	515,339	46	0.948	$y = 2,485,589x^{0.707} - 211,467$
Medium-to-large-scale ($> 5MW_{el}$)	9,852,006	17	0.763	$y = 128,387x + 12,515,990$

To account for the uncertainty in the predictions, triangular distributions were used to draw samples from during the Monte Carlo simulation. Similar to the gasification and gas cleaning CAPEX model, the CER model's prediction for a given plant size was used as the triangular distribution's mode, whereas the lower and upper bounds were defined by the predicted value ± the predicted value multiplied by the model's MAPE.

5.4.2.1 Syngas Combustion

A simplified syngas combustion model has been proposed to calculate the GWP of syngas combustion. For this the following assumptions have been made: Firstly, CO, CH_4, and C_2H_n are completely converted to CO_2 during the combustion process, as described by the following balanced chemical reaction equations:

$$2CO + O_2 \rightarrow 2CO_2$$
$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$
$$C_2H_4 + 3O_2 \rightarrow 2CO_2 + 2H_2O$$

Secondly, all higher order hydrocarbons (i.e. C_2H_n) are treated as C_2H_4. Considering these assumptions, the GWP of syngas combustion has been calculated using

$$\mathrm{GWP_{syn}} = \left(f_{\mathrm{CO_2}} \rho_{\mathrm{CO_2}} + f_{\mathrm{CO}} \rho_{\mathrm{CO}} \left(\frac{2M_{\mathrm{CO_2}}}{2M_{\mathrm{CO}}} \right) + f_{\mathrm{CH_4}} \rho_{\mathrm{CH_4}} \frac{M_{\mathrm{CO_2}}}{M_{\mathrm{CH_4}}} + f_{\mathrm{C_2H_n}} \rho_{C_2H_4} \frac{2M_{\mathrm{CO_2}}}{M_{\mathrm{C_2H_4}}} \right) y_{\mathrm{syn}}\,\mathrm{FU} \tag{5.9}$$

where f are the volume fractions of different gas species in the produced syngas, M are the molar masses of the corresponding gas species ($M_{CO_2} = 44\ g/mol$, $M_{CO} = 28\ g/mol$, $M_{CH_4} = 16\ g/mol$, $M_{C_2H_4} = 28\ g/mol$), y_{syn} is the syngas yield in m^3/kg wb, ρ are the densities of the gas species ($\rho_{CO_2} = 1.98\ kg/m^3$, $\rho_{CO} = 1.14\ kg/m^3$, $\rho_{CH_4} = 0.657\ kg/m^3$, $\rho_{C_2H_4} = 1.18\ kg/m^3$), and FU is the functional unit in kg. f and y_{syn} are imputed by the ML prediction model, whereas the FU is user-defined and has been chosen to be 1,000 kg wb of feedstock treated.

Emissions were split into biogenic and non-biogenic emissions. For this, it has been estimated that emissions from the combustion of woody and herbaceous biomass are 98.78% and 98.33% biogenic respectively **[307]**. MSW was estimated to be 65% biogenic on average **[88, 308]**. However, the composition of MSW in particular may vary drastically from country to country due to other influencing factors. Consequently, users have the option to choose custom biogenic carbon fractions for their specific feedstock, provided that case-specific data is available.

5.4.3 Feedstock Pretreatment

5.4.3.1 Bale Breaking and Chopping

Some feedstock types may be received as bales. Typical feedstocks received in baled form include agricultural by-products such as corn stover, wheat straw, or sorghum stalks **[309]**. Alternatively, feedstocks may also be received in other forms that still need breaking up and size reduction. Bale breaking and feedstock particle reduction are typically done using a grinder **[309, 310]**.

Jannasch *et al.* (2001) quoted particle sizes of 25 to 100 mm after processing alfalfa and switchgrass **[310]**. The authors estimated an energy requirement of 14.9 $kWh/tonne$

but stated that considerable energy savings are possible based on the type of feedstock and the shape of the bales. For instance, an energy saving of 30% is suggested for processing switchgrass over alfalfa and breaking square bales takes 20% less time than round bales. In comparison, the tub grinder used in Wilson *et al.* (2014) reduced four types of cellulosic biomass (sorghum stalks, corn stover, wheat straw, and big bluestem) to a size of 178 to 229 mm and it was assumed that all feedstocks require the same energy **[309]**.

It is clear that there is significant uncertainty in the energy requirements for bale breaking and chopping. For this reason, these are modelled by using a triangular distribution defined by a mode of 14.9 $kWh/tonne$, and a lower and upper limit of 7.5 and 22.4 $kWh/tonne$ (i.e. a 50% decrease and increase from the mode), respectively.

Cost data for grinders which can be used for the initial particle reduction of feedstock bales was very limited, but eight data points were gathered by Sultana and colleagues **[311]**. They gathered a dataset of the CAPEX associated with primary grinders, which can be used for breaking bales and chopping biomass materials. This data set was used to model the CER between grinder capacity and CAPEX. The following steps were taken to normalise all data and prepare it for modelling:

1. All data was converted to the same currency by using the average exchange rate of the year the data is from.

2. CEPCI indices were used to convert all data to the same year.

3. Outliers were removed.

The cleaned and normalised data set for primary grinders is shown in Figure 5.11. Note that the grinder's capacity was converted to tonnes/hour to align with the rest of the model. In addition, the resulting linear fit is shown in the same figure.

In general, the fitted straight line predicts the grinder's CAPEX to a reasonable accuracy defined by the following performance metrics: $R^2 = 0.700$, RMSE = 4,152, and MAPE = 15%. The straight-line fit is given by $y = 2,107x + 10,351$. To account for the uncertainty in the model's predictions, triangular distributions were used to draw samples from during the Monte Carlo simulation. Like previously described CAPEX models, the CER model's prediction for a given plant size was used as the triangular distribution's mode, whereas the lower and upper bounds were defined by the predicted value ± the predicted value multiplied by the model's MAPE.

5.4.3.2 Drying

A feedstock's moisture content critically affects gasification and the moisture content of biomass feedstocks can vary greatly. For this reason, drying is a commonly included pre-treatment step. Several options are available to dry a feedstock. If solar drying or

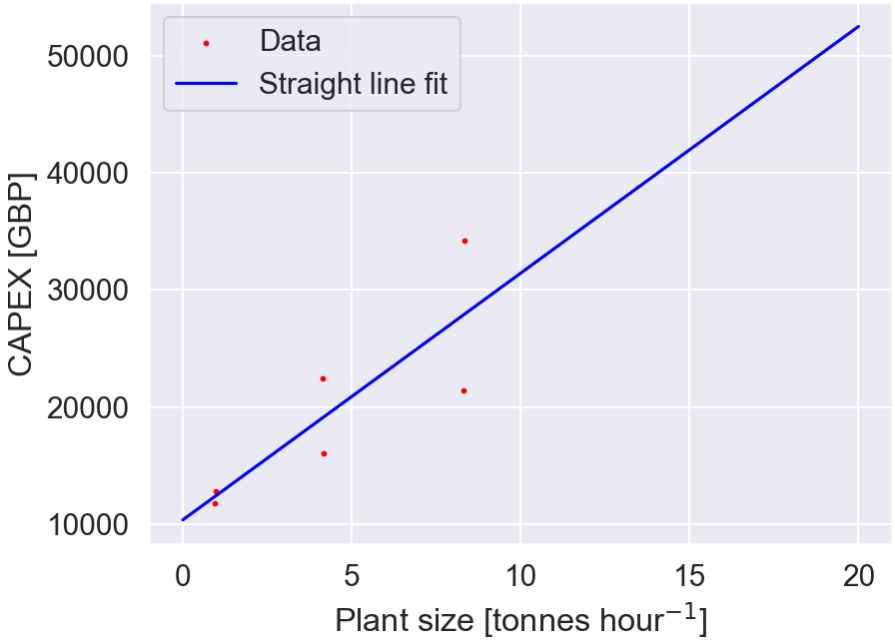


Figure 5.11: Scatter plot and regression fit to the normalised CAPEX of feedstock grinders.

recovered waste heat are used no or minimal additional energy is required for drying. Alternatively, dedicated dryers powered by gas (natural gas or syngas) or electricity can be used. Examples of these are indirect-heated convection dryers, contact dryers, direct-fired dryers, or microwave dryers.

The first stage of modelling the drying process was to calculate the mass of the evaporated water M_{ev} based on the initial feedstock mass M_i resulting from a change from the feedstock's initial moisture m_i content to its final moisture content m_f, as described by

$$M_{ev} = (M_i(1 - m_i)(m_i - m_f)/(1 - m_i) \tag{5.10}$$

In the second stage, the theoretical heat requirement for evaporation Q_{th} was calculated using

$$Q_{th} = M_{ev}\,(h_w + c_v(T_d - T_r)) \tag{5.11}$$

where h_w is the specific latent heat of vaporisation of water at the dryer temperature T_d, c_v is the specific heat of water vapour, and T_r is the room temperature.

Finally, a penalty due to dryer inefficiencies was added based on the dryer type. The final energy requirements in the form of electricity, syngas, or natural gas were calculated depending on the considered dryer type.

The CAPEX of a feedstock dryer was modelled based on literature data and data supplied from Vulcan Drying Systems **[312]**. In total 33 data points were collected based on seven sources **[270, 311–316]**. To establish a CER between water removal rate and

dryer CAPEX, the following steps were taken to normalise all data and prepare it for modelling:

1. All data was converted to the same currency by using the average exchange rate of
2. the year the data is from.
3. CEPCI indices were used to convert all data to the same year.
4. Outliers were removed.

The cleaned and normalised set is shown in Figure 5.12. It can be seen that feedstock dryers typically fall into one of the following three categories: (i) small-scale dryers with a water removal rate of $<$ 1,000 $kg\ H_2O$/hour, (ii) medium-scale dryers with a water removal rate between 1,000 and 10,000 $kg\ H_2O$/hour, and (iii) large-scale dryers with a water removal rate of $>$ 10,000 $kg\ H_2O$/hour. Furthermore, the figure clearly shows that the CAPEX obtained from Vulcan Drying Systems is significantly lower than that of older reference systems quoted in the literature. This may be due to the maturation of the technology or due to differences in the used technologies.

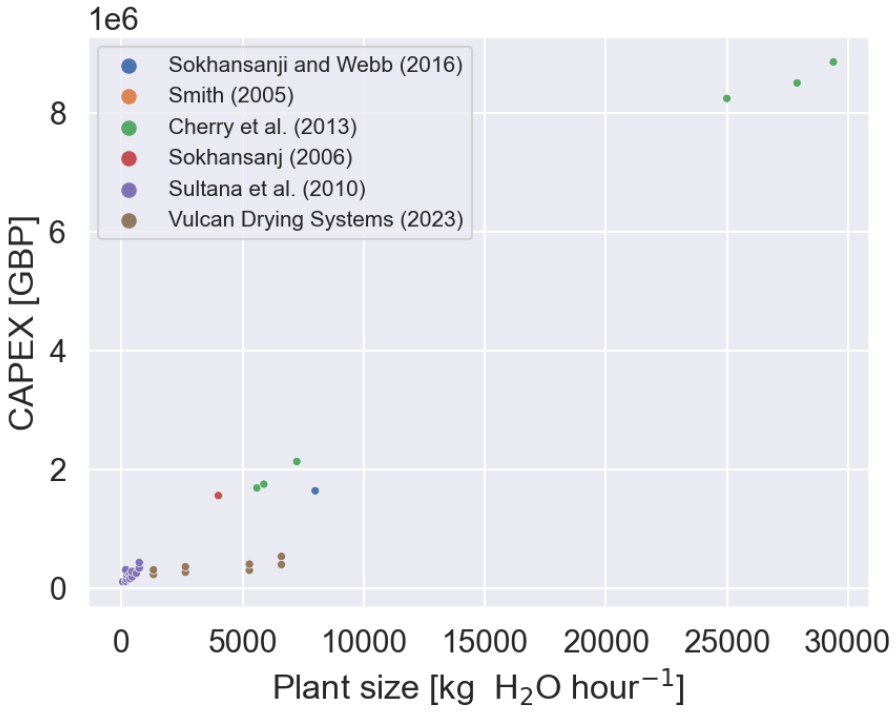


Figure 5.12: Scatter plot of normalised feedstock dryer CAPEX by reference.

The cleaned and normalised data set was used to establish CERs for the three previously mentioned sizes, as shown in Figure 5.13. Two different fits are shown for the central region, one using only data from Vulcan Drying Systems and one using all the available

data in that region. Based on communication with Vulcan Drying Systems, it was found that there is a relatively small increase in costs as the dryer's size increases. Whilst the increase in drum size leads to a cost increase, other cost factors such as labour, engineering works, and procurement only increase to a small degree if at all **[312]**.

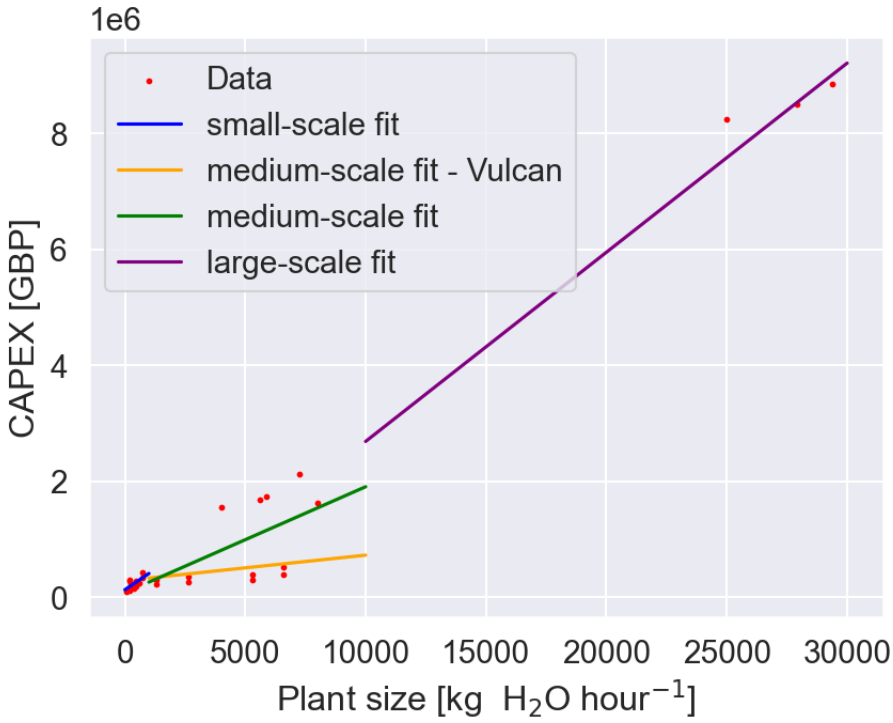


Figure 5.13: Regression fits to the normalised CAPEX of feedstock dryers.

The performance of all models is summarised by Table 5.4. The model for small-scale dryers achieved reasonable accuracy, aided by the fact that most data was available in this region (i.e. 13 data points). Unsurprisingly, the model fitted to the central region which used data from only one source (i.e. from Vulcan Drying Systems) performed significantly better than the model using data available from all sources. This, in combination with the fact that the data from Vulcan Drying Systems is more recent, ultimately led to the model only based on the data from Vulcan Drying Systems being used. Finally, the large-scale model showed a reasonable predictive performance. However, a severe lack of data in the 7,000 to 25,000 $kg\ H_2O$/hour dryer size region is apparent, leading to significant uncertainty in the CAPEX of dryers of that size. The 10,000 $kg\ H_2O$/hour region was particularly uncertain, as the medium-scale and large-scale models did not agree well in that region.

To address uncertainties within the models' predictions, triangular distributions were used to draw samples from during the Monte Carlo simulation. Similar to other CAPEX models, the CER models' predictions for a given plant size were used as the triangular distribution's mode, whereas the lower and upper limits were defined by the predicted

Table 5.4: Performance summary of CAPEX models fitted to feedstock dryer data.

	RMSE [£]	MAPE [%]	R^2	Fitted function
Small-scale (< 1,000 $kg\ H_2O$/hour)	67,057	28	0.443	$y = 281x + 131,386$
Medium-scale (Vulcan) (1,000 to 10,000 $kg\ H_2O$/hour)	58,300	15	0.578	$y = 33x + 283,123$
Medium-scale (Other) (1,000 to 10,000 $kg\ H_2O$/hour)	568,112	79	0.336	$y = 190x - 24,695$
Large-scale (> 10,000 $kg\ H_2O$/hour)	559,845	67	0.966	$y = 12.64x^{1.31} + 5991$

value ± the predicted value multiplied by the model's MAPE.

5.4.3.3 Milling/Grinding

Milling or grinding can be used to reduce the particle size of the feedstock to improve feedstock flow and heat transfer inside the reactor. The particle size post milling is most directly affected by the mill's screen size with typical screen sizes ranging between 2.78 and 6.5 mm **[309, 310]**. For instance, a 10 hp hammermill employed in Wilson *et al.* (2014) **[309]** resulted in particle sizes ranging from 20 to 31 mm and 15 to 18 mm when using a 6.5 and 3.2 mm screen, respectively. In their study, four types of cellulosic biomass (sorghum stalks, corn stover, wheat straw, and big bluestem) were compared. It appears that feedstock particle size post-grinding varies based on the feedstock type, with more fibrous materials resulting in a larger average particle size. This effect was more pronounced when a larger screen size was used. A feedstock's average bulk density on the other hand showed greater variation the smaller the selected screen size and resulting particle size. The authors argue that at smaller particle sizes the natural characteristics of a given feedstock (e.g. hollow stems of wheat straw) are compounded.

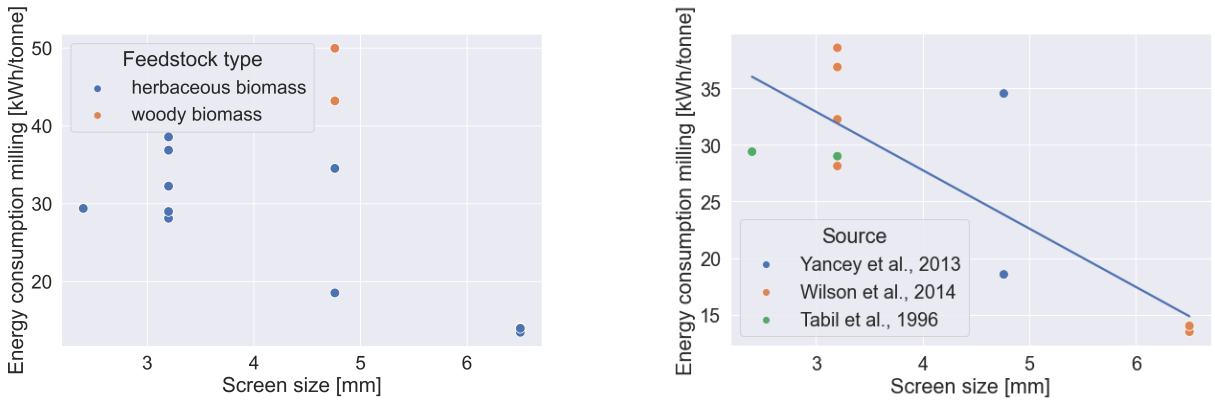
Based on Wilson *et al.* (2014) **[309]**, the average electricity consumption for milling using a 6.5 and 3.2 mm screen was 13.81 and 33.96 $kWh/tonne$, respectively. This represents a 41% difference in energy consumption between different screen sizes. In comparison, Jannasch *et al.* (2001) **[310]** estimated an electricity consumption of 55.9 $kWh/tonne$ for the grinding of switchgrass. No difference in energy consumption between the used 5.56 and 2.78 mm screen was considered. However, the authors noted that energy consumption could be halved to 28 $kWh/tonne$ if the mill throughput was to be optimised. An early study by Tabil and Sokhansanj (2016) **[317]** is in good agreement with the other studies quoting specific energy consumptions of 29.4 and 29.0 $kWh/tonne$ for milling alfalfa using screen sizes of 2.4 and 3.2 mm, respectively. Yancey *et al.* (2013) **[318]** compared the energy consumptions of grinding switchgrass, eucalyptus, lodgepole pine, and corn stover using a 150 hp hammermill with a 4.76 mm screen. The energy consumptions ranged from 18.55 to 49.97 $kWh/tonne$, with woody biomass showing higher energy consumptions (49.97 and 43.21 $kWh/tonne$) than herbaceous biomass (18.55 and 34.54 $kWh/tonne$). Whilst the energy consumptions are of a similar scale as the ones quoted in the other studied works, the particle size post-milling was significantly different. Mean

particle sizes ranging from 0.52 to 0.81 mm are manyfold smaller than the ones in Wilson *et al.* (2014) **[309]**.

To model the energy consumption from milling and the resulting particle size, data from the above-discussed studies was initially compiled and plotted for analysis. Figure 5.14 shows that an increase in the mill's screen size leads to a decrease in the mill's energy consumption. Notably, it appears that woody biomass incurs an increase in energy consumption. This is likely due to woody biomass being harder than herbaceous biomass. It must be noted, that very limited data on woody biomass was available. In practice a linear regression model described by

$$E_m = -5.17 s_s + 48.42 \tag{5.12}$$

was fitted to describe the relationship between screen size (s_s) and energy consumption (E_m) for herbaceous biomass. For woody biomass, an increase in energy consumption of 22.75 $kWh/tonne$ was estimated based on data from Yancey *et al.* (2013) for a 4.7 mm screen **[318]**. The linear model was then compared to the original data which resulted in a RMSE of 4.86 $kWh/tonne$. The linear regression model's prediction and the calculated RMSE were used as the mean and standard deviation of a normal distribution. Ultimately it is this normal distribution that samples are drawn from for the Monte Carlo simulation.



(a) (b)

Figure 5.14: Energy consumption for milling in relation to the employed mill screen size. Shown for (a) all feedstock types and (b) herbaceous biomass only, where the linear regression fit is shown by the blue line.

A similar approach was taken to model the resulting particle size after milling. After removing outliers, a linear regression model described by

$$S_p = -3.33 s_s + 5.08 \tag{5.13}$$

was fitted to describe the relationship between s_s and particle size (S_p). Upon comparing the model's predictions to the original data, a RMSE of 3.38 mm was calculated.

This RMSE was again used to fit a normal distribution which was used in the Monte Carlo simulation.

A total of 19 data points on the cost of hammer mills for feedstock particle size reduction was obtained from five sources **[311, 313, 319–321]**. This data set was used to model the CER between mill capacity and CAPEX. The following steps were taken to normalise all data and ready it for modelling:

1. All data was converted to the same currency by using the average exchange rate of the year the data is from.

2. CEPCI indices were used to convert all data to the same year.

3. Outliers were removed.

The cleaned and normalised data set, alongside a linear fit, are shown in Figure 5.15. To keep the CER consistent with the rest of the model, the hammermill capacity was converted to tonnes/hour. The quality of the fit, given by the following performance metrics: $R^2 = 0.042$, RMSE = 22,871, and MAPE = 72%, is relatively poor. This largely results from the large spread in the data, in particular, the data points with a CAPEX of around £60,000 to £90,000. The straight-line fit may be described by $y = 1,229x + 30,889$.

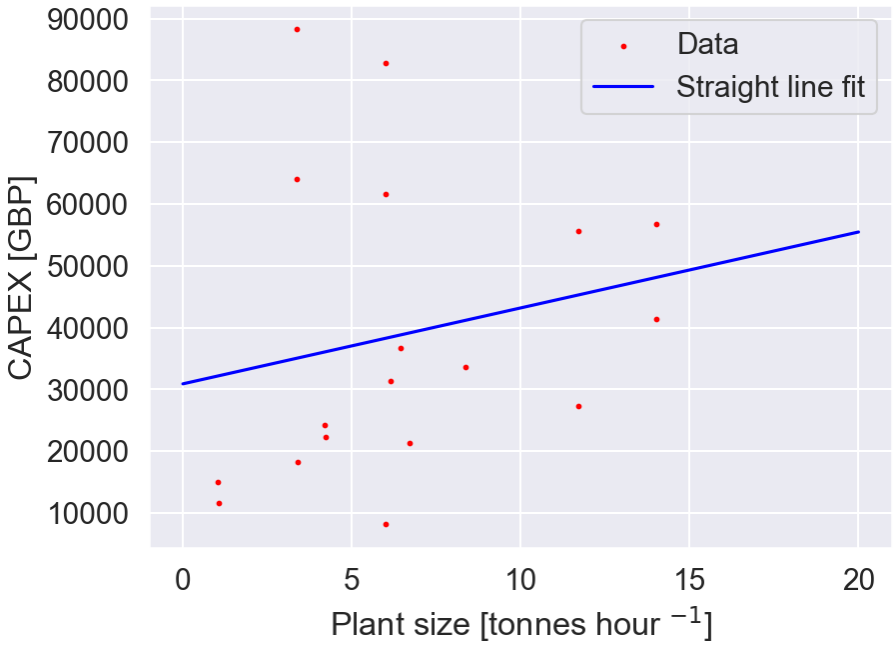


Figure 5.15: Scatter plot and regression fit for the normalised CAPEX of hammer mills.

To account for the large uncertainty in the model's predictions, triangular distributions were used to draw samples from during the Monte Carlo simulation. Like previously described CAPEX models, the CER model's prediction for a given plant size was used as

the triangular distribution's mode, whereas the lower and upper bound were defined by the predicted value ± the predicted value multiplied by the model's MAPE.

5.4.3.4 Pelleting

A common pretreatment step, especially when a feedstock is to be transported, is densification, which refers to the compression of a loose feedstock into a high-density form **[318]**. Typically, a feedstock is turned into pellets, but other shapes are possible. Some of the advantages of densification are (i) reliable and uniform particle size of the feedstock, (ii) improvements to handling and conveyance throughout the supply chain and operation at the plant, (iii) improved feedstock composition due to fractionated structural components, and (iv) conformance to existing standards in the supply chain and gasification plant **[318]**.

Pellets produced by a typical pellet mill are cylindrical with a typical length of 13 to 19 mm and width of 6.3 to 6.4 mm **[318]**. In this work, a width of 6.35 mm post pelleting was assumed based on a study by Wilson *et al.* (2014) from which a large amount of pelleting data was taken **[309]**. Pellet length was taken as 18 mm.

Wilson *et al.* (2014) compared the energy demands of pelleting four types of cellulosic biomass (sorghum stalks, corn stover, wheat straw, and big bluestem), after being processed by a hammer mill with a 3.2 or 6.5 mm screen **[309]**. Finer particle sizes resulting from a smaller mill screen, generally required less energy than larger particle sizes. On average, energy requirements for pelleting increased by 28% from the 3.2 to 6.5 mm screen. However, this increase was somewhat counterbalanced by smaller screen sizes requiring more energy. Energy requirements for pelleting ranged from 92.80 to 122.86 $kWh/tonne$ and 99.42 to 172.87 $kWh/tonne$ for feedstock previously milled using a 3.2 and 6.5 mm screen, respectively. In comparison, Yancey *et al.* (2013) found that pelleting generally required less energy, ranging from 45.20 to 124.13 $kWh/tonne$ **[318]**. Much smaller particle sizes ($< 1\ mm$) before pelleting are a likely reason for the lower energy demands.

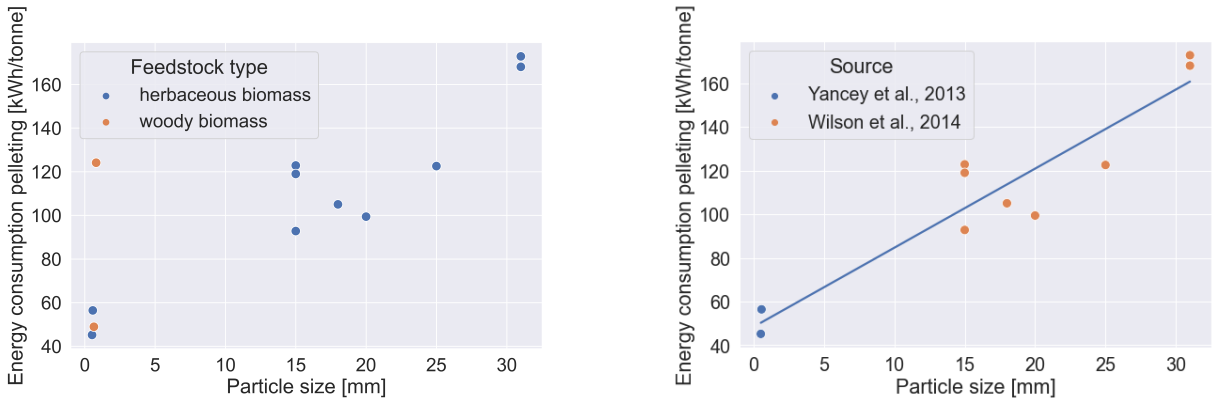
In a review by Tumuluru *et al.* (2011) specific energy consumptions for the pelleting of feedstocks ranging from woody/herbaceous biomass to MSW were compared **[322]**. It was found that pelleting required 16.4 to 74.5 kWh/tonne. These values are significantly lower than the ones found in Wilson *et al.* (2014) **[309]** and Yancey *et al.* (2013) **[318]**. However, the studies considered in the review are mostly from the 1970s and 1980s, which may explain their difference to the values quoted in more recent studies.

The energy requirements for pelleting were modelled based on the studies discussed above. From Figure 5.16, which shows the compiled data, it becomes clear that an increase in particle size is associated with greater energy requirements for pelleting. Too few samples on woody biomass were available to extract any trends from the data. Hence the developed model did not differentiate between woody and herbaceous biomass. As shown

in Figure 5.16b, a linear regression model described by

$$E_p = -5.63 s_p + 48.40 \tag{5.14}$$

was fitted to describe the relationship between the feedstock's particle size (s_p) and energy consumption (E_p) for pelleting. By comparing the linear model to the original data, an RMSE of 13.54 $kWh/tonne$ was calculated. In the greater modelling framework, the linear regression model is used to make a prediction based on the initial particle size. This prediction and the previously calculated RMSE are then used as the mean and standard deviation of a normal distribution. During Monte Carlo simulation it is this normal distribution that samples are drawn from.



(a) (b)

Figure 5.16: Energy consumption for pelleting in relation to feedstock's particle size. Shown for (a) all feedstock types and (b) herbaceous biomass only, where the linear regression fit is shown by the blue line.

The CAPEX of feedstock pelleting was modelled as two distinct processes: the pellet mill itself and a pellet cooler. Most of the data points used for modelling were collected by Sultana *et al.* (2010) **[311]** with a few data points coming from other sources **[313, 320, 321]**. In total, 19 and 18 data points were collected for pellet mills and pellet coolers, respectively. Before establishing a CER, the following steps were taken, for both the pellet mill and cooler data, to normalise all data and ready it for modelling:

1. All data was converted to the same currency by using the average exchange rate of the year the data is from.
2. CEPCI indices were used to convert all data to the same year.
3. Outliers were removed.

The cleaned and normalised data sets for the pellet mill and cooler are shown in Figure 5.17. Note that the plant capacities were converted to tonnes/hour to align with the rest

of the model. Additionally, the figure shows the resulting fits to the data. A linear fit was selected for the pellet mill data (R^2: 0.617; RMSE: 59,981; MAPE: 36%) and a power fit (R^2: 0.518; RMSE: 8,883; MAPE: 31%) was selected for the pellet cooler data. The straight-line fit to the pellet mill data is given by $y = 20,510x + 37,322$ and the power fit to the pellet cooler data is given by $y = 5,243x^{0.749} + 5,690$. In general, the fitted curves predict the pellet mill's and cooler's CAPEX to a reasonable accuracy.

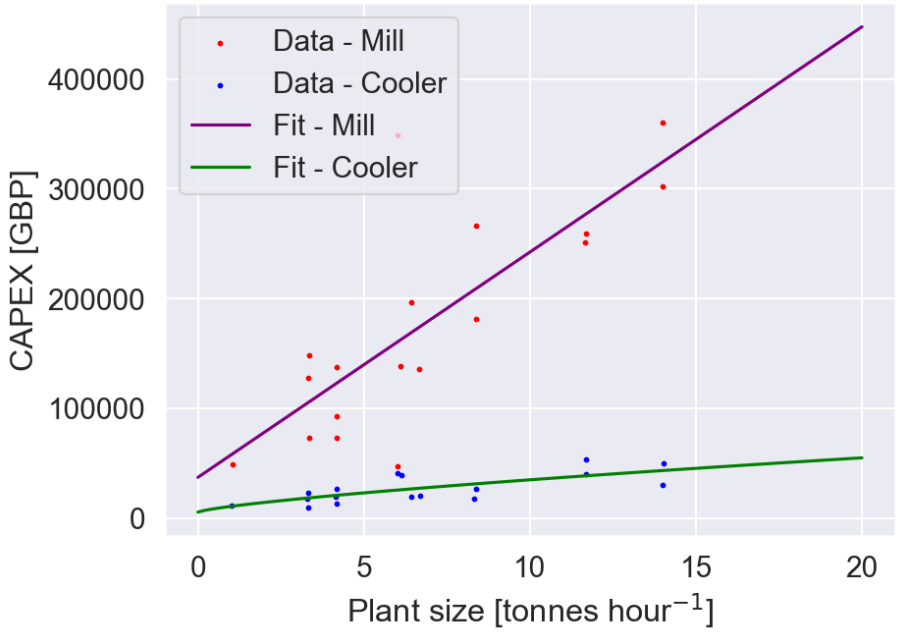


Figure 5.17: Scatter plot and regression fits to the normalised CAPEX of pellet mills and coolers.

To account for the uncertainty in the models' predictions, triangular distributions were used to draw samples from during the Monte Carlo simulation. Like previously described CAPEX models, the CER model's predictions for a given plant size were used as the triangular distributions' modes, whereas the lower and upper bounds were defined by the predicted values ± the predicted values multiplied by the model's MAPE.

5.4.4 Biochar Use

The application of biochar to aid with soil fertility has been used by indigenous populations of the Americas for thousands of years **[323]**. Biochar yield and its properties are predominantly affected by two factors – the employed thermochemical method and its operating conditions, and the selected feedstock. When it comes to thermochemical methods, pyrolysis is typically associated with the production of biochar, with slow pyrolysis having the highest biochar yield **[56]**. However, gasification has more recently also been considered as a viable pathway to co-produce biochar **[27]**. Producing biochar through

gasification has the benefit of producing large portions of a high-value product, in the form of syngas, which improves the system's economics.

In 2009, Brewer *et al.* highlighted that the exact mechanics of how the properties of gasification and fast pyrolysis chars affect their efficacy in soil applications was not yet well understood **[56]**. Whilst the research on gasification biochar remains somewhat limited, significant research on biochar from fast pyrolysis has been conducted since then **[324–326]**. The main potential benefits of applying biochar to soil are the following: (i) soil carbon sequestration due to the stable fraction of carbon in the biochar; (ii) reduction of GHG emissions from soil in the form of N_2O and CH_4; (iii) higher fertiliser efficiency and reduced fertiliser requirements; (iv) improved water retention leading to a reduction in the need for irrigation **[27]**.

This work only considers the effects of soil carbon sequestration and the reduction of N_2O emissions from the soil. Soil carbon sequestration is generally by far the largest factor contributing to the environmental analysis of biochar for soil applications. The reduction in CH_4 emissions is subject to large uncertainty, hence a similar approach to Marzeddu *et al.* (2021) was followed and only the effects due to reduced N_2O emissions were modelled **[27]**. Similarly, improvements to the fertiliser efficiency and the reduction in fertiliser requirements are highly uncertain and subject to a range of factors such as the fertiliser manufacturing method and the soil and crop type the biochar is applied to. Hence these factors were not considered in this work. Effects due to water retention are again highly dependent on the soil type and hence not considered.

One factor thoroughly researched is the effect of different feedstocks on biochar properties. A wide range of feedstocks have successfully been used for the production of biochar - biomass, animal manure and waste plastics to name a few **[326]**. Some feedstocks are better suited to produce biochar, due to their relatively low amount of volatiles and high amount of fixed carbon. Care has to be taken when using some other feedstocks, such as landfill waste, sewage sludge, or poultry litter, as they may contain elevated levels of heavy metals and other contaminants, which will result in biochar not suited for soil application **[326, 327]**.

For the economic analysis, biochar is included up to the point of sale. Biochar prices are highly variable and depend on a multitude of factors such as the quality and properties of the biochar and local market conditions. For this reason, the biochar sales price was chosen on a case-by-case basis and was considered a necessary user input in the greater modelling framework.

5.4.4.1 Soil Carbon Sequestration as a Result of Biochar Application to Soil

Two key factors required to model the soil carbon sequestration resulting from applying biochar to soil are its carbon content and its fraction of recalcitrant carbon. Recalcitrant

carbon is resistant to decomposition and leads to long-term carbon sequestration in soil, whereas labile carbon is released to the atmosphere in the form of CO_2 **[57]**. In a 2016, meta-analysis Wang *et al.* (2016) used 128 observations from 24 studies to estimate the labile and recalcitrant carbon fraction of biochar and their respective mean residence times **[328]**. It was found that 97% of carbon was recalcitrant with a mean residence time of 556 years, whereas 3% of carbon was labile with a residence time of only 108 days. In comparison, Roberts *et al.* (2010) considered 80% of carbon to be recalcitrant a conservative estimate **[84]**. Typically, higher process temperatures, as they are the case for gasification, are associated with higher fused aromatic ring clusters, which lead to a more stable carbon **[56]**. Similarly, the IPCC estimated that 89% of high-temperature pyrolysis and gasification ($>$ 600°C) biochar carbon remained in the soil after 100 years **[329]**. In comparison, medium-temperature pyrolysis (450 to 600°C) and low-temperature pyrolysis (350 to 450°C) resulted in significantly less stable chars with a stability of 80% and 65%, respectively. The IPCC report argued that it was preferable to group carbon stability by categories, instead of using a regression model since the relationship between process temperature and carbon persistence have been shown to be somewhat opaque.

Too few data on the effects of feedstock properties on the recalcitrant carbon fraction of biochar content exist to perfectly model these factors. Hence a general model was implemented which assumes, based on the previously discussed literature, that the recalcitrant carbon fraction of biochar from gasification (on a 100-year basis) can be defined by a triangular distribution with a lower limit, mode, and upper limit of 80%, 90%, and 97%, respectively. CO_2 resulting from labile carbon was treated as if it was emitted upon application, whereas recalcitrant carbon remained in the soil over the studied time horizon.

The biochar's carbon content was primarily estimated from data collected by the IPCC **[329]**. In their report, biochar data from pyrolysis and gasification was presented for the following six feedstock categories: (i) animal manure, (ii) wood, (iii) herbaceous biomass, (iv) rice husk and rice straw, (v) nut shells, pits, and stones, (vi) biosolids (paper sludge and sewage sludge). All data was presented in the form of the mean value $\pm$ the 95% confidence intervals. The IPCC data was validated with data collected from 13 studies by You *et al.* (2017) **[25]**. In their paper, the properties of biochar produced by the gasification of various feedstocks were summarised. In general biochar carbon contents tended to be higher in the data collected by You *et al.*, however too little data concerning certain feedstock types was available to see any clear trends.

In the end, the IPCC data was used to calculate the mean and standard deviation of each feedstock type from their respective confidence intervals to fit normal distributions representing the biochar's carbon content. If the selected feedstock did not fall into any of the categories given by the IPCC a random value between the lowest and highest carbon

content of all collected data was drawn by the model to represent the large uncertainty in the biochar's carbon content.

5.4.4.2 Reductions in Soil N_2O Emissions as a Result of Biochar Application to Soil

N_2O mainly forms in the soil through nitrification and denitrification processes. Marzeddu *et al.* (2021) reviewed numerous scientific articles and concluded that the application of biochar to soil typically reduces N_2O emissions by about 30% when biochar is applied at a rate of 25 $t\ ha^{-1}\ yr^{-1}$ **[27]**. Baseline N_2O emissions were taken from Lugato *et al.* (2017) whose biogeochemistry model covers more than 11,000 sampling points across Europe **[330]**. The authors found that Mediterranean regions typically displayed lower emission rates than central Europe, the UK, and Ireland. The total data set was defined by a mean, 1st quartile, and 3rd quartile of 2.27, 1.18, and 2.63 $kg\ N\ ha^{-1}\ yr^{-1}$.

Ultimately, the GWP resulting from the reduction in N_2O emission is given by

$$GWP_{N_2O-Biochar} = 273\left(\frac{3.286 F_N R_N y_B}{r_B}\right) \tag{5.15}$$

where F_N is a random sample drawn from a distribution describing the baseline soil N_2O flux (given in $kg\ N\ ha^{-1}\ yr^{-1}$) which was estimated based on the data by Lugato *et al.* (2017), R_N which is the above-mentioned reduction factor (30%), y_B which is the biochar yield per FU, and r_B which is the biochar application rate (25 $t\ ha^{-1}\ yr^{-1}$). The factor of 273 converts N_2O emissions to their CO_2 equivalent and the factor of 3.286 is used to scale the flux given in N to N_2O. As emission rates vary significantly even in the same country, spatial factors were not accounted for, and it is assumed that the European data also describes the rest of the world sufficiently.

5.4.5 Carbon Capture and Storage

Bioenergy with carbon capture and storage (BECCS) as a method to produce carbon-negative electricity, heat, or biofuels has gained significant attention after the IPCC included BECCS in 85% of scenarios aimed at limiting global warming to 2°C **[48]**. To achieve this goal, the British government is committed to establishing four CO_2 transport and storage clusters by 2030 and putting a carbon market in place to encourage the development of effective carbon-negative technologies **[243]**.

One major challenge for CCS is the requirement for significant infrastructural development to transport the captured carbon to its final storage location (or utilisation location in the case of CCUS). As the development of their own CO_2 transport network is generally not feasible for small to medium-scale plants, this study follows the approach taken in Oreggioni *et al.* (2017), where the developed plant would join an existing network **[331]**.

Fugitive emissions and global warming impacts associated with the infrastructure of the CO_2 network have been shown to be negligible for large-scale plants **[332, 333]**. Hence, these have been deemed acceptable to ignore in this work.

A range of different carbon capture methods have been developed. One key distinction is between pre-combustion capture, post-combustion capture, and oxy-fuel combustion. During post-combustion capture, the CO_2 is removed from the flue gas which typically leads to high capture rates. However, the flue gas is typically at a low pressure which means that a large volume of gas needs to be processed and additional compression is required before transportation. A key advantage of post-combustion technology is that existing plants can be retrofitted with it.

Pre-combustion capture typically requires the addition of a water-gas shift reactor to increase the CO_2 concentration of the syngas before combustion, by shifting the CO in the syngas to CO_2 and H_2. This means the CO_2 is more concentrated and at higher pressure, making it easier to capture which can lead to a decrease in auxiliary energy demands and carbon capture plant size.

Amine-based carbon capture technologies are the most deployed and arguably most developed technology **[334]**. Amines, such as monoethanolamine are well suited for carbon capture as they can be effectively separated after capture via a reversible reaction. Pressure swing adsorption or vacuum pressure swing adsorption relies on the cyclic adsorption of CO_2 before purging the CO_2 for capture and is typically well suited for gas streams with relatively low CO_2 concentrations **[335]**. Finally, during oxy-fuel combustion a fuel is burned using pure oxygen and the reaction temperature is controlled by recycling some of the flue gases **[336]**. This method can be extended to gasification by using oxygen as the gasifying agent and for syngas combustion **[337]**.

As pre-combustion methods, significantly alter the overall system by requiring a water-gas shift reactor and thus altering the composition of the syngas, only post-combustion methods have been implemented. Firstly, an amine-based carbon capture process using aqueous monoethanolamine as a solvent has been implemented based on data from Ahn *et al.* (2013) and Oreggioni *et al.* (2015) **[338, 339]**. The process has been shown to have a CO_2 capture rate of 90% on a flue gas basis. Electricity and heat requirements have been shown to vary, as shown in Figure 5.18, with various process configurations influencing these. After discarding outliers, the fractions of electricity consumption over total energy consumption have been calculated, resulting in a mean and standard deviation of 0.12 and 0.037, respectively. Similarly, the mean and standard deviation of the total energy consumption were calculated to be 3.36 and 0.32 $MJ/kg\ CO_2$ removed, respectively. During Monte Carlo simulation, a total energy consumption is drawn from a normal distribution defined by the mean and standard deviation given above. Then a similarly defined normal distribution for the electricity fraction is used to obtain thermal energy

and electricity consumption for each Monte Carlo iteration.

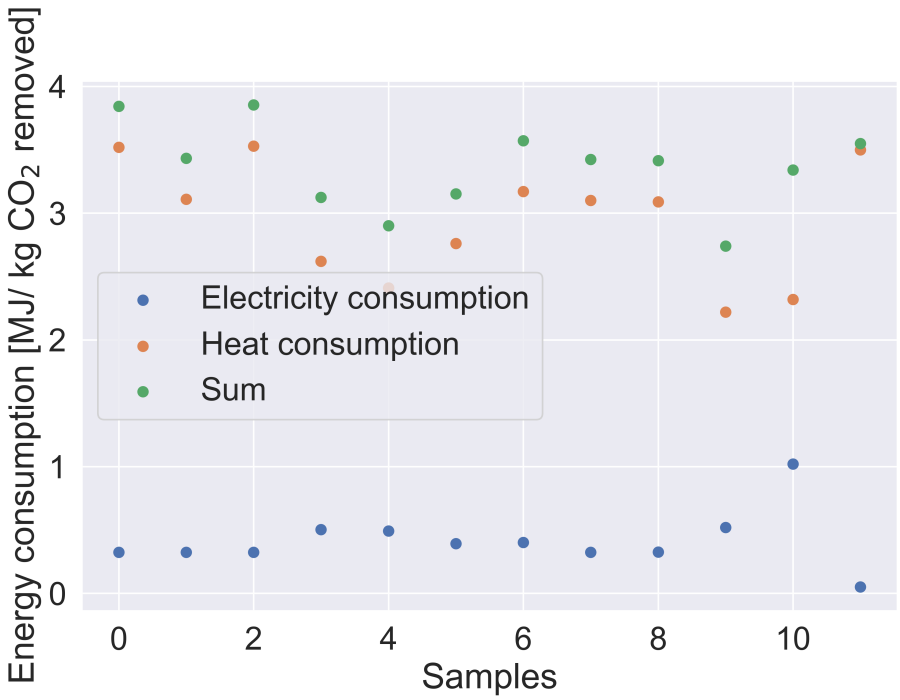
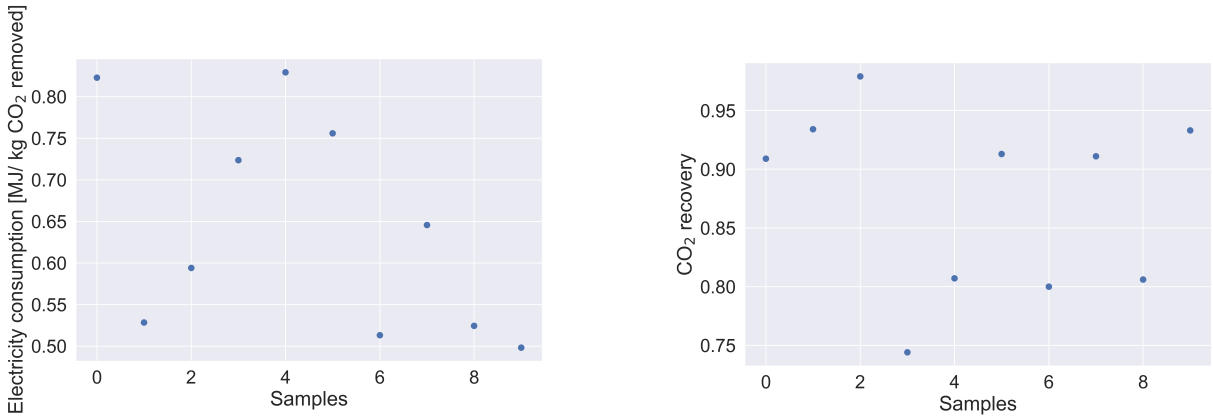


Figure 5.18: Energy consumption for amine-based carbon capture process **[338, 339]**.

Secondly, a rapid vacuum pressure swing adsorption (RVPSA) capture process was implemented based on data from Luberti *et al.* (2017) **[340]**. The variations in electricity consumption and carbon recovery rate are shown in Figure 5.19. The mean and standard deviation of the electricity consumption were calculated to be 0.64 and 0.12 $MJ/kg\ CO_2$ removed, respectively. A mean and standard deviation of 0.87 and 0.073 were calculated for the recovery rate. An increase in energy consumption was not shown to improve the CO_2 recovery rate. Fitting a linear regression model yielded a coefficient of determination of 0.027. Hence no correlation between energy consumption and carbon recovery was assumed in the developed model. Ultimately, normal distributions were defined based on the calculated means and standard deviations, which were then used to draw samples during the Monte Carlo simulation.

As CCS is a relatively recent technology, accurate cost estimates can be difficult to obtain. In particular, data for small-scale systems is scarce **[268]**. Whilst data for larger-scale systems can serve as a reasonable indicator for the approximate costs of small-scale systems, factors such as economies of scale are not always easy to account for. In addition, CCS costs vary greatly depending on a range of other factors such as the choice of technology and design used, requirements for CO_2 transport and storage, and site or location-specific factors **[341]**. In fact, most available CAPEX data in the literature is for large-scale coal power plants **[267, 287, 306, 342]**. For these reasons, a different approach to modelling carbon capture CAPEX was taken, as compared to the approach



(a) (b)

Figure 5.19: (a) Energy consumption and (b) carbon recovery rate for RVPSA carbon capture process **[340]**.

used for the gasifier and CHP CAPEX model.

Some studies quoted carbon capture cost in terms of a flat cost per unit of CO_2 captured. Hence this type of data was used to estimate the CAPEX in terms of an annuity related to the annual amount of CO_2 available for capture. For instance, a discussion paper titled "Realistic Costs of Carbon Capture" estimated representative cost ranges for first-of-a-kind and nth-of-a-kind carbon capture plants **[341]**. Whilst first-of-a-kind costs were estimated to be high at 100 to 150 \$US/tonne CO_2, the authors estimated that abatement costs would see an approximate reduction of 65% by 2030 to 30 to 50 \$US/tonne CO_2 for nth-of-a-kind plants. This agrees well with a study by Jiang *et al.* (2021) which estimated a capture cost of 47.1 \$US/tonne CO_2 for an amine-based absorption process with 90% efficiency for a 550MW coal power plant **[343]**. For these reasons, a uniform distribution with a lower and upper limit of 30 and 50 \$US/tonne CO_2 was selected to estimate the costs for carbon capture.

Due to the lack of available carbon capture cost data, the current simplified model does not differentiate between different carbon capture technologies.

5.4.5.1 CO_2 Transport and Storage Costs

Transport and storage costs are subject to several factors. For one, the utilised reservoir type influences both storage and transport costs. The two main, reservoir types are old oil and gas fields and saline aquifers. In addition, reservoirs can be both onshore and offshore, with the latter typically being the significantly more expensive option **[344]**. However, geological offshore storage options are much more abundant than onshore storage options and some regions may only have access to one of the options **[345]**. Furthermore, for small-scale systems in particular, the construction of dedicated pipelines may not be economically feasible. For this reason, utilising existing CO_2 networks or trucks to

transport CO_2 may be a more realistic option for small-scale systems **[346]**. From this, it becomes clear that a large range of local factors may influence CO_2 transport and storage costs. For this reason, whilst the model incorporates default cost estimates for CO_2 transport and storage, these should be updated with more case-specific data if available.

CO_2 transport and storage costs were estimated based on several literature sources. Amongst other factors, Rubin *et al.* (2015) reviewed CO_2 transport costs across a range of different geographical locations, storage capacities, and for offshore and onshore options **[344]**. They found transport costs as low as 1 \$US/tonne CO_2, in particular for large-scale onshore facilities. The highest estimates were around 20 €/tonne CO_2. The mode of all data points was around 5 \$US/tonne CO_2. When it comes to storage costs, they reviewed four large studies and put them on a common basis (2013 \$US). The lower bound of storage costs across the four studies was between 1 and 7 \$US/tonne CO_2. The upper bound was 12 or 13 \$US/tonne CO_2 for 3 of the studies and 18 \$US/tonne CO_2 for the other reviewed study.

James *et al.* (2019) estimated costs for CO_2 storage in deep saline formations across several different states in the United States **[267]**. Costs varied from 8.32 to 19.84 \$US/tonne CO_2. The estimated transport costs were relatively low at 2.07 \$US/tonne CO_2. Other literature sources combined transport and storage costs for their cost estimates and estimated costs of 19 and 22 £/tonne CO_2 and €/tonne CO_2 respectively **[287, 342]**. The latter was composed of a transport and storage cost of 13 and 9 €/tonne CO_2 respectively.

Based on the reviewed literature, triangular distributions were defined to estimate CO_2 transport and storage costs. Transport costs were defined by a lower limit, mode, and upper limit value of 2, 5, and 15 \$US/tonne CO_2 respectively. Storage costs were defined by a lower limit, mode, and upper limit value of 2, 10, and 20 \$US/tonne CO_2 respectively. For case studies conducted in a different currency, these values were converted to the corresponding currency.

5.4.5.2 Costs due to Monoethanolamine (MEA) Degradation for Amine-based Carbon Capture

Monoethanolamine (MEA) degrades during amine-based carbon capture. For this reason, the running of an amine-based carbon capture plant requires the constant resupply of MEA. The costs incurred due to this are defined by two factors: (i) the specific MEA consumption typically given per tonne of CO_2 captured and (ii) the MEA price.

The literature reviewed by Goff and Rochelle (2004), showed degradation rates ranging from 0.54 to 3.6 kg of MEA per tonne of CO_2 captured **[347]**. However, the data is quite dated and the figure of 3.6 includes other degradation mechanisms, which means modern plants are likely to have a much lower MEA degradation. Other early literature

quotes degradation rates of 1.6 and 2.2 kg of MEA per tonne of CO_2 captured **[348, 349]**. However, more recently, degradation rates as low as 0.3 to 0.7 kg of MEA per tonne of CO_2 captured have been estimated **[350]**. Based on the reviewed literature, it is assumed that the MEA degradation rate can be defined by a triangular distribution defined by a lower bound, mode, and upper bound value of 0.3, 0.7, and 2.2 \$US/tonne CO_2 respectively.

Most MEA price figures tend to be dated, and some more recent studies used figures more than ten years old. For instance, Dinca *et al.* (2018) **[350]** used a price of 2.1 €/kg which has originally been estimated by Kim and Svendsen (2007) **[351]**. Similarly, Zhao *et al.* (2014) **[352]** used an MEA price of 0.97 \$US/kg in their analysis, which dates back to a 2003 study by Singh *et al.* **[348]**. To avoid the reliance on dated literature data, market data of the past 5 years (2018 to 2023) was used to estimate the MEA price **[353]**. Whilst some regional trends exist, they do not paint a clear picture when considering the data of the past five years. Thus, regional factors were not accounted for in the current model. Ultimately, a triangular distribution, given by a lower limit, mode, and upper limit of 0.90, 1.75, and 2.50 \$US/kg MEA respectively, were defined. Again, values were converted to the currency selected in the model framework whenever necessary.

5.5 Complete Gasification System Model

As previously mentioned, the complete gasification system model is made up of the processes described in Section 5.4. The interactions between these processes are illustrated by the LCA system boundary shown in Figure 5.5.

Since gasification schemes and local conditions are extremely varied, a range of user inputs are required to run the model. Many parameters may have default parameters, but providing case-specific data whenever possible improves the model's accuracy for the specific case at hand. To account for uncertainties in the model's inputs, any of the distributions introduced in Section 5.2.2.1 may be used to represent numerical parameters.

Initially, the user may select general simulation parameters such as the number of Monte Carlo iterations and the country the simulation is run for. Next, the feedstock and its ultimate and proximate composition are supplied, as well as the gasifier conditions (such as gasification temperature, reactor type, and ER) and system size. The user then selects which optional processes (shown in yellow in Figure 5.5) are to be included in the scheme and selects a range of economic parameters such as the discount/interest rate i which is used to discount cash flows.

In particular, economic parameters can be fast-evolving and highly country-specific, making it difficult to incorporate default values with a high degree of confidence. For this reason, these types of parameters are recommended to be imputed by the user on a case-by-case basis. These parameters and their most important considerations are listed

below:

- **Electricity wholesale price:** The electricity price is pivotal in determining the economic performance of a gasification scheme as its main product is electricity. The electricity price was considered a user input as country-specific and current prices are typically easy to find. It is important to explore a range of possible parameter values, due to the parameter's vital role in the analysis. This is of particular importance when considering the variability in wholesale electricity prices. Factors such as supply and demand, government policies and subsidies, and global and geopolitical factors can all have drastic effects on a country's electricity price or even cross-national electricity prices. The UK exemplifies this variability in electricity prices well, with a minimum and maximum electricity wholesale price of 24.01 and 362.71 $£/MWh$ between January 2020 and December 2024 **[354]**.

- **Heat/thermal energy wholesale price:** The sale of thermal energy is another important product of a gasification scheme. The fluctuations in prices for thermal energy or natural gas, which remains a primary source of thermal energy in numerous countries, are influenced by many of the same factors that impact electricity prices.

- **Biochar price:** Whilst the sale of biochar is typically not the primary revenue stream for a gasification scheme, it can substantially improve its economics. The cost of biochar exhibits considerable variation across countries and is influenced by a myriad of factors, including the choice of original feedstock and the conditions employed during gasification **[355, 356]**.

- **Gate fee or feedstock price:** The costs for a given feedstock depend on a plethora of factors such as the type of biomass, the geographical location, and market dynamics. When considering European conditions, the import of wood pellets, which are typically from North America, is often costly. Similarly, whilst dedicated energy crops are an effective way to generate large amounts of biomass for bioenergy, their high cost is a major drawback. Alternatively, the waste products, by-products, and residues of industrial and agricultural processes are significantly cheaper or even free when disposal would otherwise be required. Additionally, materials such as wood waste from sawmills and the pulp and paper industry may have other benefits like a low moisture content due to the primary manufacturing process they have already been through **[274]**. Finally, some feedstocks, such as MSW, may even generate revenues for a gasification scheme in the form of a gate fee.

- **Optional carbon tax or carbon cost in an emissions trading scheme:** The effects of a potential carbon tax or other type of carbon trading scheme can be explored in the model. Although carbon markets, such as the EU Emissions Trading

System, have experienced varying degrees of success, numerous nations view them as essential instruments for reducing their GHG emissions. Given the volatility of carbon prices, it is crucial to investigate a range of parameter values to evaluate a spectrum of possible future scenarios.

5.5.1 System Optimisation

The system's environmental and economic performance, given by the GWP and BCR respectively, were taken as the two objectives to be optimised. In cases where the objectives are non-conflicting a single optimal solution can be found. Here this means the dominant solution would minimise the GWP and maximise the BCR. Alternatively, in cases where the objectives are conflicting a trade-off situation occurs where a single optimal solution cannot be found. Here, improving one objective would harm the other and a Pareto front of non-dominant solutions can be witnessed.

A simple optimisation algorithm called grid search was implemented to find the optimal parameter configurations for a given feedstock. In addition, the algorithm can be used to perform a sensitivity analysis and study the effects of certain parameters on the system's overall environmental and economic results. Grid search optimisation is a type of exhaustive search algorithm which explores all model parameters within a predefined search space. This means that during an exhaustive search, every possible combination of values within a specified search space is tried. In the case of the grid search algorithm, the search space is discretized into a grid of parameter values. Each point in the grid represents a combination of parameters that is evaluated. Whilst grid search is easy to implement and guarantees that the best parameter combination within the search space is found, it is also computationally expensive and may not be suitable for a very large search space.

The implemented grid search algorithm allows for the optimisation of the following parameters describing the gasification process and related system components: (i) gasification temperature, (ii) ER, (iii) gasifying agent, (iv) system scale, (v) reactor type, (vi) inclusion of carbon capture. Additionally, a range of parameters with an effect on the economic analysis were implemented: (i) interest/discount rate (i), (ii) system life span (N), (iii) electricity price (iv) thermal energy price, (v) optional carbon tax. Based on initial testing, these parameters were established to most likely change and to have the largest effect on the system's environmental and economic performance. If other parameters were to be included in the future, the grid search algorithm would allow for simple extension.

When the optimisation routine is executed, the simulation first determines the number of unique parameter combinations making up the optimisation grid based on the variables the user selected for optimisation and their possible values. Next, the simulation loads the parameters given by the first point in the grid (i.e. the first unique combination of

model parameters) and uses them to execute all environmental and economic submodels. The simulation then iterates through all grid points, running the model for all unique parameter combinations. Once all grid points have been evaluated, the results for each grid point are saved. Results can then be further analysed or plotted as a scatter plot to show the Pareto front in the case of two conflicting objectives.

In practice, the user may select any number of parameter values for continuous parameters like the gasification temperature. Categorical or boolean parameters, such as the reactor type and the inclusion of carbon capture, are limited to their respective parameter options (i.e. fluidised bed, fixed bed, or other in the case of the "reactor type" parameter or true or false for the "inclusion of carbon capture" parameter. While it is possible to explore any number of parameter options, in reality, one must limit the search space to avoid excessive computational demands.

5.5.2 Decision Support Tool with a Graphical User Interface (GUI) Based on the Developed Framework

As a final step, the developed modelling framework was turned into a decision support tool with an easy-to-use GUI. A GUI was considered important, as it allows for the use of the model without any programming expertise. Figure 5.20 illustrates the basic design of the developed decision support tool. The tool is locally hosted through Streamlit (an open-source Python framework used to build interactive data apps), with the option of hosting it online, and can be accessed through any web browser. The main page of the tool is the *simulation inputs* page, where the user effectively defines the parameters which are to be used in the simulation. User inputs are stored as *.toml* files, a human-readable configuration file format. The file format's simplicity and clarity make it easy for the user to manually edit user inputs if desired. Finally, the simulation is executed for a given user inputs *.toml* file and the simulation results are displayed.

Figure 5.21 shows the page layout of the tool. When the tool is opened the user is greeted by a home page which introduces the purpose of the tool and explains its working. It also contains a video walk-through, which explains how the tool may be used. The next page is the previously mentioned *simulation inputs* page. Here the user can select simulation parameters across the following six broad categories:

- **General simulation parameters:** These are any parameters which affect the overall running of the simulation, such as the number of Monte Carlo iterations, the system's life span, and the country for which the analysis is conducted.
- **Feedstock information:** Here the user can define their feedstock and supply related data such as the feedstock's ultimate and proximate composition, and its energy content.

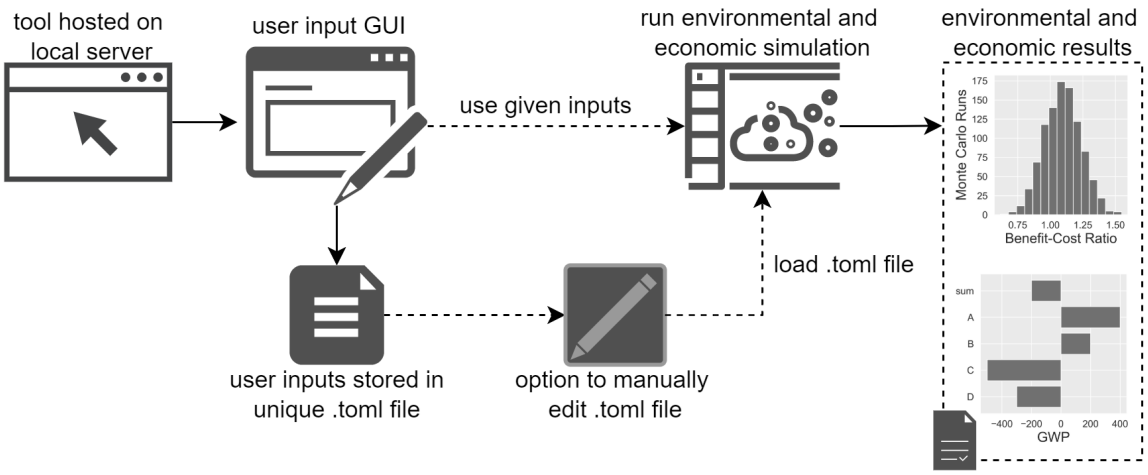


Figure 5.20: Illustration of the design of the developed decision support tool with a GUI. Dashed arrows indicate optional steps.

- **Reactor information and processing conditions:** This category mostly contains information on the type of gasifier and gasifying agent used, and at which gasification conditions the feedstock is converted (e.g. temperature and ER).
- **System size information:** The system's size in terms of at least one of three metrics is defined here. The system size can be defined in terms of feedstock mass (given in tonnes feedstock/hour), feedstock energy supply (given in MW feedstock LHV or MWh feedstock LHV/hour), or system net electric power (given in MW_{el}).
- **Techno-economic analysis choices:** Various parameters important for the TEA stage are given here. General parameters (e.g. year of study or rate of return) are initially defined, before defining prices for energy, biochar, and the feedstock (unless a gate fee is considered when a waste treatment perspective is taken). Finally, optional values for a carbon tax and carbon capture transport and storage costs can be entered if these are considered in the analysis.
- **Process selector:** Here the user can define which subprocesses (e.g. CCS and pretreatment options) are considered in the simulation and supply additional information for some of them

The final page of tool is the the *run simulation* page which is also shown in Figure 5.21. After selecting a *.toml* file of simulation parameters the tool will execute the simulation in the background and display the environmental, economic, and energy generation results. The results figures can then be downloaded if desired. Screenshots detailing all of the tool's pages can be found in Appendix A.

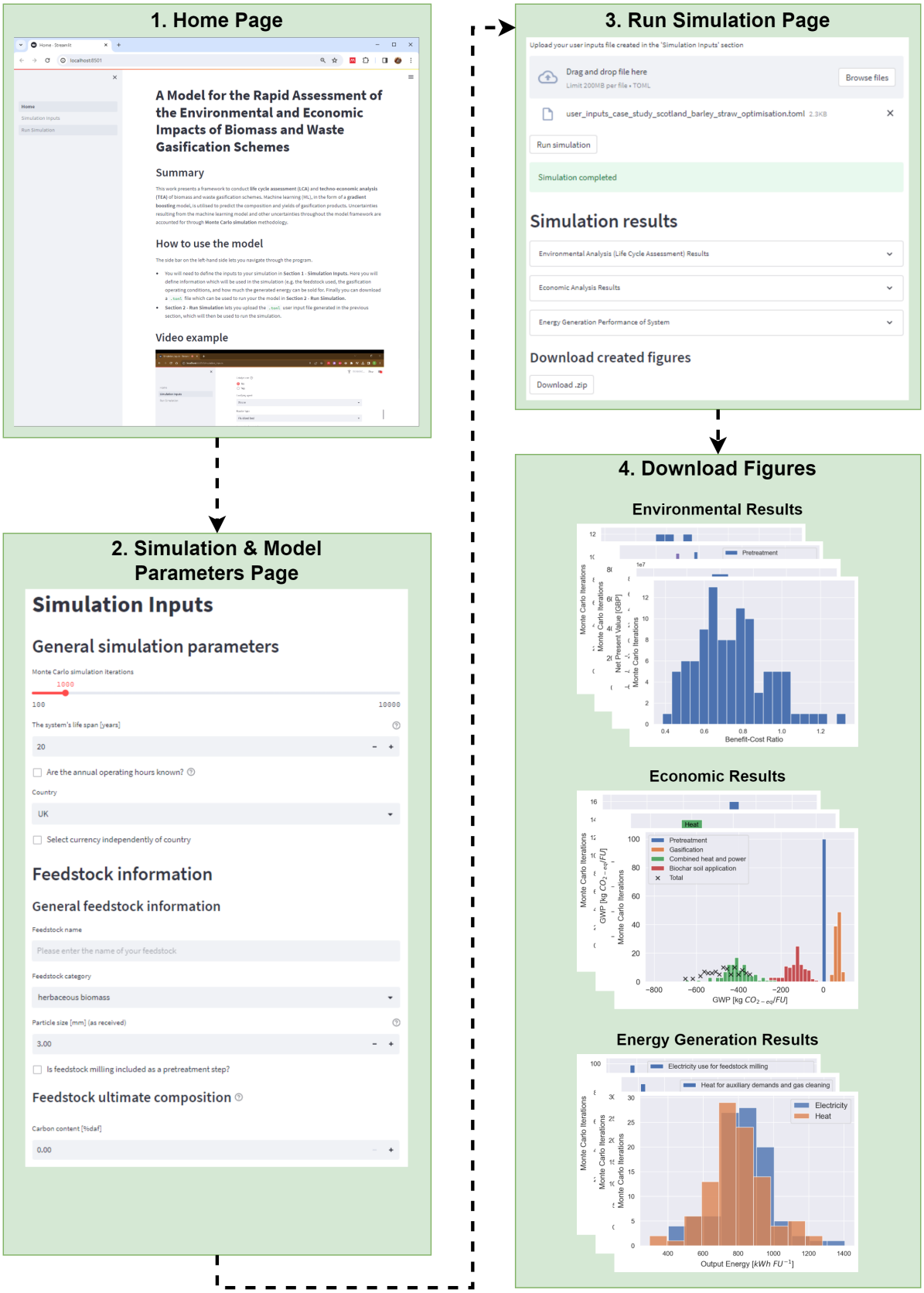


Figure 5.21: Page layout of the developed decision support tool. Full-sized screenshots of the tool can be found in A.

5.6 Chapter Summary and Conclusions

This chapter details methods used to establish a comprehensive biomass and waste gasification modelling framework, based on the ML model introduced in Chapters 3 and 4. The chapter discusses the programming paradigm the modelling framework is founded in and details how uncertainty is accounted for throughout different parts of the framework. Detailed environmental and economic models for the various subprocesses making up the overall gasification scheme were developed in Section 5.4. Finally, the chapter explains how the developed modelling framework was turned into an easy-to-use decision support tool with a GUI. The usefulness of the developed modelling framework in the form of the GUI-based decision support tool is illustrated in Chapter6 in the form of a case study.

Chapter 6

A Case Study Using the Developed Modelling Framework: Gasification of Barley Straw, Forestry Residues, and Distillers' Malt Draff in Scotland

6.1 Chapter Overview

This chapter illustrates the usefulness of the biomass and waste gasification modelling framework developed in Chapter 5 by conducting a case study. The case study examines three promising feedstocks local to Scotland: forestry residues, barley straw, and distillers' malt draff (referred to as "draff" from this point on). Ultimately, the feedstock forestry residues was found to outperform the other feedstocks both environmentally and economically. For this feedstock, the implemented optimisation routine highlighted a maximum carbon avoidance of -2,604 $kg\ CO_2$-eq per tonne of feedstock at a benefit-cost ratio (BCR) of 1.149 when carbon capture and storage (CCS) was considered. When maximising the scheme's economic performance, CCS was not included and a BCR of 1.840 was found for a global warming potential (GWP) of -877 $kg\ CO_2$-eq per tonne of feedstock.

The study's background and reasoning for the feedstock choices are given in Section 6.2. Section 6.3 describes all preliminary analyses and assumptions made to conduct the case study through the developed decision support tool. Thus it illustrates the process a researcher or decision-maker would go through to conduct a comparable case study. Section 6.4 then goes on to show the case study's results for baseline and optimised scenarios and discusses them. Updated results educated by the results of the optimisation and sensitivity analysis are highlighted. Finally, Section 6.5 summarises the findings and gives some concluding remarks.

6.2 Introduction and Background

Scotland's location in the northern part of Great Britain makes it well-situated for both wind and tidal power. Over the last decade, the country has more than doubled its renewable electricity generation, reaching a capacity of 14 GW in 2022. This represents 26% of all renewable energy generation in the UK, despite only making up less than 10% of the UK's population. By 2030, Scotland aims to increase its onshore and offshore wind capacity to 20 GW and 11 GW, respectively. This would represent 10% of the EU's target for 111 GW of offshore renewables by 2030 **[357]**.

As countries cannot overly rely on singular renewable energy solutions, Scotland is looking at expanding its bioenergy potential, as outlined in the 2023 *Biomass Action Plan* **[358]**. In total, Scotland's electricity generation capacity from biomass is estimated at 450 MW. To achieve this capacity, the Scottish government highlighted the importance of making use of various types of biomass. Particularly the importance of considering not only forestry-derived products but also agricultural and other waste products was stressed. Large potential is also seen in the renewable heating sector (a third of primary energy consumption is for heat), where Scotland is currently lagging behind the rest of Europe in terms of renewable heat coming from biomass **[358]**.

Scotland has a long history of deforestation, with deforestation starting around 6,000 years ago with farmers burning forests for livestock grazing. Deforestation continued throughout the medieval period, industrialisation, and Victorian era, resulting in less than 1% of Scotland's native pinewoods remaining **[359]**. In the last 100 years, Scotland's forest cover has increased from 5% to 18.5%, which puts it above the rest of the UK, but well below the EU's average of 38% **[360]**. Today, the sustainable management of forests is an important part of the Scottish economy and forestry contributed around £ 1 billion gross value added and employed over 25,000 people full-time **[360]**. Due to the high priority of reforestation and the sustainable management of forests, forestry residues and by-products represent a future-proof and sustainable feedstock for bioenergy developments.

When it comes to agricultural production, barley is by far the most common cereal crop in Scotland, with nearly 300,000 hectares of barley being grown every year. In fact, 28% of the UK's barley is grown in Scotland and the grown barley is mostly used for animal feed (55%) and malting (35%) **[361]**. This makes barley straw another interesting feedstock to consider for bioenergy production.

Malted barley is primarily used for whisky production, another major industry in Scotland. For example, in 2021 "food, beverages, and tobacco" was the largest industry sector for international exports, accounting for around 20% of Scotland's international export value in 2021 **[362]**. The whisky industry produces a range of by-products and their use for bioenergy has recently overtaken their use for animal feed **[363]**. The main product of malt distilling in terms of dry matter is draff. Draff is produced in the first

stage of the production step when malted barley is steeped in hot water. The spent grain is called draff and whilst it typically has a high moisture content it has successfully been used for bioenergy.

On the whole, waste and byproducts from forestry and agriculture represent an untapped resource for bioenergy developments. In the EU it has been estimated that an improved collection and utilisation strategy of agricultural and forest residues could create up to 83,000 and 50,000 permanent jobs, respectively, with 162,000 additional temporary jobs being created as a result of biorefinery construction **[364]**. Similarly, the better utilisation of brewery and distillery byproducts has recently received increasing research interest **[365]**. This makes forestry residues, barley straw, and draff three promising types of biomass for bioenergy production through gasification in Scotland and they have hence been selected as the three feedstock analysed in this case study. Their environmental and economic feasibility is explored in the rest of this chapter.

6.3 Methodology

6.3.1 General Process Data

In general, UK reference data (e.g. UK electricity grid mix carbon intensity) was used throughout the simulation. The biomass gasification plant's lifespan is estimated to be 20 years and the baseline simulation was repeated for 10,000 Monte Carlo iterations. To estimate the plant's operating hours, literature data on a range of different gasification systems was analysed. Based on this data (shown in Table 6.1), the plant's operating hours were estimated to follow a normal distribution defined by a mean and standard deviation of 7,954 and 443 hours/year, respectively. For the life cycle assessment (LCA), the utilisation of 1 tonne of feedstock is used as the functional unit (FU).

Table 6.1: Annual operating hours of gasification plants.

Operating hours [hours/year]	Country/Region	Year	Reference
8410	USA	2007	**[282]**
7896	UK	2009	**[286]**
7680	Italy	2012	**[290]**
6000	Poland	2014	**[292]**
8000	Europe	2014	**[287]**
7200	Italy	2015	**[291]**
7500	Spain	2018	**[285]**
7500	Europe	2018	**[350]**
8000	Mexico	2019	**[289]**
8766	USA	2020	**[288]**
8000	UK	2022	**[180]**

The baseline scenarios for the three different feedstocks were made up of all the non-optional and some of the optional processes previously shown in Figure 5.5. Particle size reduction in the form of milling was considered necessary for barley straw and forestry residues. The particle size of draff was considered suitable for direct application without further particle size reduction. Drying was included as a pretreatment step for all feedstocks, whereas initial bale shredding and pelleting, steps necessary for some feedstocks, were not deemed necessary for any of the feedstocks considered here. CCS was not included in the baseline scenarios, but the effects of considering CCS were examined during the optimisation routine in Section 6.3.5.

6.3.2 Feedstock Availability and Composition

In 2022 around 10 million green tonnes of round wood was removed from UK forests, with most of it being softwood (9.2 million green tonnes). Of this, Scotland is by far the biggest source, with more than 70% of all privately harvested softwood and more than 50% of all publicly harvested softwood coming from Scottish forests **[366]**. Information on the availability of forestry residues is somewhat unclear, but based on the above figures nearly 6 million green tonnes of roundwood are produced in Scotland every year.

In 2023 Scotland produced 1.6 million tonnes of spring barley and 0.348 million tonnes of winter barley. This makes barley the most produced cereal in Scotland, making up 63% of all cereal production. Wheat and oats mainly make up the remainder of cereal production with 32% and 5% of total production, respectively **[367]**. Typically around half a tonne of barley is produced for every tonne of barley harvested and 57% of all straw is sold off farm. Most of the straw is currently sold for livestock feed or bedding, with only 7% of the total straw production going towards bioenergy **[368]**. Considering these numbers, it would indicate that a minimum of approximately 70,000 tonnes of barley straw would be available for bioenergy annually.

In 2018, on a dry matter basis, 142,000 tonnes of all distillery by-products were used for animal feed, whereas 232,847 and 46,125 tonnes dry matter were used for bioenergy through anaerobic digestion (AD) and combined heat and power (CHP), respectively. Of these by-products draff, with a dry matter content of approximately 22%, makes up approximately 42% **[363]**. Based on these figures, approximately 444,500 and 88,000 tonnes wet basis (wb) of draff are currently used each year for bioenergy through AD and CHP, respectively.

Based on the availability of the three different feedstocks and an estimate in annual plant operating hours of around 8,000 hours/year (as described in Section 6.3.1), a plant size of 10 tonnes/hour is selected as the baseline size for all feedstock types. All plants were chosen to be of the same size to avoid factors such as economies of scale warping the results. However, as some feedstocks may allow for bigger plants, the effect of this is

studied in the form of a sensitivity analysis as described in Section 6.3.5.1.

The ultimate and proximate feedstock composition of the three feedstock types is summarised in Tables 6.2 and 6.3, respectively. Most feedstock data was obtained from the Phyllis2 database, a database containing information on the composition and other data of a wide range of biomass and waste feedstocks **[369]**. Values are quoted on a number of different basis, depending on their definition in the underlying ML prediction model. The following three definitions were used: (i) wet basis (wb) – the material's weight percentage in its original form (including ash and moisture). This is typically the same as "as received" (ar); (ii) dry basis (db) – the materials weight percentage in its original form minus its moisture content; (iii) dry ash-free basis (daf) – the materials weight percentage in dry and ash free form.

Table 6.2: Ultimate composition of forestry residues, barley straw, and draff.

	Forestry residues	Barley Straw	Draff
C [% daf]	51.99	49.09	49.12
H [% daf]	6.18	6.06	6.81
N [% daf]	0.41	0.64	2.27
S [% daf]	0.02	0.08	0.20
O [% daf]	41.4	44.13	41.60
daf - dry ash free basis			

Table 6.3: Proximate composition and lower heating value (LHV) of forestry residues, barley straw, and draff.

	Forestry residues	Barley Straw	Draff
Moisture (as received) [% wb]	20.00	11.53	78
Moisture (post drying) [% wb]	12.00	11.53	15
Ash [% db]	1.33	5.88	3.10
LHV [MJ/kg wb (post drying)]	17.97	15.14	9.95
Mean Particle size ar [mm]	25	10	5
ar - as received; db - dry basis; LHV - lower heating value; wb - wet basis			

6.3.3 Economic Factors

6.3.3.1 Interest/Discount Rate

The interest or discount rate used in the techno-economic analysis (TEA) is an important factor to consider as it is used to discount future cash flows to their present value equivalent. In 2023 the Bank of England's monetary policy committee has set the UK's official interest rate at 5.25% **[370, 371]**. Previously the interest rate has been low at a value of less than

1% for nearly a decade **[370, 371]**. For the baseline scenario, the interest rate is taken as a fixed value at its current value of 5.25%. However, to explore the effect of this variable it is varied to 1% and 6% during the optimisation stage of this work as described in Section 6.3.5.

6.3.3.2 Energy prices

Electricity wholesale prices in the UK and Europe have fluctuated significantly in recent years. As these changes can be caused by external factors, such as Russia's full-scale invasion of Ukraine in February 2022, it is essential to make robust assumptions about electricity prices. In 2022 alone, electricity prices varied from 148.48 to 511.20 £/MWh in the UK **[354]**. In contrast, electricity prices in 2023 started just under 200 £/MWh and decreased to a price of around 80 £/MWh by the end of the year **[354]**. When considering the period before 2022 and looking at long-term predictions, electricity prices are expected to settle around a price of 50 £/MWh **[372]**. Based on these assumptions a conservative approach is taken and a triangular distribution with a lower bound, mode, and upper bound of 40, 70, and 200 £/MWh, respectively, is assumed to represent the electricity price well. However, due to the significant uncertainty in the electricity price, its effect on the model's results is further explored through a sensitivity analysis as described in Section 6.3.5.1.

The price for heat or thermal energy was estimated based on natural gas prices. In general, natural gas prices have seen a similar trend to electricity prices in the recent past. A maximum natural gas price of around 200 £/MWh was reached in 2022 **[354]**. In 2023 prices varied between approximately 68 and 29 £/MWh, whilst prices were around 17 £/MWh in early 2021 **[354]**. Based on this range of values a thermal energy price defined by a lower bound, mode, and upper bound of 17, 29, and 85, respectively, was used for the analysis.

6.3.3.3 Feedstock Prices

Accurate literature data for the price of lower-grade woody biomass, such as forestry residues, in the UK was challenging to obtain. However, the UK-based research organisation *Forest Research* has a good record of prices for both softwood sawlogs and small roundwood **[366]**. Assuming an average softwood density of 500 kg/m^3, the average prices for sawlogs and small roundwood in 2023 were 122 £/tonne and 186 £/tonne, respectively. When considering the past decade, prices were as low as 90 £/tonne and as high as 190 £/tonne for sawlogs and they ranged from 150 to 220 £/tonne for small roundwood **[366]**. However, these woody materials are of much higher value than forestry residues which include branches, stumps, tree tops, bark, etc.

Dulys-Nusbaum *et al.* (2019) studied the willingness of North American landowners to permit the removal of forestry residues at the time of harvest **[373]**. They found that most commercial forest managers were willing to permit residue removal for 4 to 11 $US/tonne. However, these figures do not include to cost of the actual removal. Considering forestry residue prices from across Europe, prices were as low as 25 €/tonne for low labour-cost countries, and higher at 62.5 €/tonne for higher labour-cost countries like Finland **[364]**. In addition, transport costs were estimated to typically add 8 to 12 €/tonne for a transport distance of approximately 100 *km* **[364]**. These values are similar to current market prices for bulk wood chips and bark prices in the UK, with values as low as 53 and 31 £/tonne, respectively **[374]**. Based on the available data the delivered price for forestry residues was estimated to follow a triangular distribution with a lower bound, mode, and upper bound of 40, 65, and 100 £/tonne, respectively. This is in good agreement with Fernanda Rojas Michaga *et al.* (2022) who estimated UK nominal forestry residue prices of 58.53 £/tonne with a chance to decrease or increase by 50% **[180]**.

Straw prices vary significantly across the UK with transport costs often being a significant cost factor for remote regions. However, accurate historical data is available from the *British Hay and Straw Merchants Association* which compiles and supplies trade information to the UK's *Department for Environment Food and Rural Affairs* (DEFRA). Based on their data, barley straw prices varied between 50 to 105 £/tonne (ex-farm) between 2019 and 2024 in southeast Scotland **[375]**. In 2024, straw costs were around 80 £/tonne **[375]**. Additionally, transport costs are typically in the range of 25 £/tonne **[376]**. Based on these historic values, the delivered barley straw price, used in the model, is given by a triangular distribution with a lower bound, mode, and upper bound of 75, 105, and 130 £/tonne, respectively.

The most recently available price data on draff prices is from the 2012 to 2018 period. The *Scottish Rural College* studied the fluctuation in draff prices for this period and found that draff prices ranged from 10.50 to 49.00 £/tonne. The mean and median draff price were both 27.00 £/tonne. Whilst there was a draff oversupply in the 2014 to 2016 period, leading to below-average prices for that period, prices increased towards the end of the 2016 to 2018 period as a greater demand for draff was seen **[363]**. Based on these findings, a triangular distribution with a lower and upper bound of 10.50 and 49.00 £/tonne, respectively, was used to model the draff price. As an increase in draff demand was seen towards the end of the 2012 to 2018 period, and no more recent data was available, a conservative approach was taken and the mode of the distribution was set at 35 £/tonne.

6.3.3.4 Biochar Price

Biochar prices can vary drastically depending on a range of factors, with the local market conditions and the quality of the biochar being among the most important factors.

Additional confusion is introduced, because the price of biochar can refer to biochar as a product for e.g. agricultural use, but also to biochar as a carbon trading commodity.

When it comes to the biochar for carbon credits market, prices have been very dynamic, with prices as high as 525 $US/tonne CO_2-eq for small volumes of credits **[377]**. However, more typically prices were in the 100 to 200 $US/tonne CO_2-eq region **[377]**. In contrast, the *Biochar Demonstrator* is a UK-based leading interdisciplinary research project assessing all matters related to the deployment of biochar. As part of their research, they estimated the price of biochar for soil application to range from 100 to 400 £/tonne **[378]**. Ultimately, this work considers biochar a product for soil application in an agricultural context and the value of the stored carbon in the biochar is accounted for separately in Section 6.3.4. To take a conservative approach, the price of biochar was estimated to follow a triangular distribution with a lower bound, mode, and upper bound of 100, 200, and 400 £/tonne, respectively.

6.3.4 Carbon Tax

A carbon tax or emissions trading schemes (ETSs) are two popular market-based mechanisms aimed at promoting cleaner technologies and driving the transition to a low-carbon economy. The UK previously participated in the EU-ETS but replaced it with its own ETS in January 2021 after the country's exit from the EU. In recent years, the carbon price in the UK-ETS roughly varied between 50 and 100 £/CO_2-eq **[379]**. The carbon price in the EU-ETS has been in a similar region. In this case study, the default scenarios do not consider a carbon price. However, the effects of carbon prices of 50 and 100 £/CO_2-eq are explored during the system optimisation stage of this work as further described in Section 6.3.5

6.3.5 System Optimisation

As previously discussed in Section 5.5.1, a simple optimisation algorithm was implemented and used for this case study. In addition to showing the best-performing simulation parameters in terms of GWP and BCR, the optimisation also acts as a sensitivity analysis which shows the effects of varying certain input parameters. The parameters explored during optimisation and their respective value are shown in Table 6.4. Due to the exhaustive nature of the optimisation algorithm, the parameter combinations shown in Table 6.4 resulted in a total of 1,296 simulation runs during optimisation. To reduce computational time the number of Monte Carlo iterations for each simulation run was reduced to 1,000 iterations. A more granular sensitivity analysis is discussed in Section 6.3.5.1.

Table 6.4: Optimised parameters and their respective values.

Optimised parameter	Parameter values
Gasification temperature [°C]	650, 750, 850
ER	0.25, 0.30
Gasifying agent	air, steam, oxygen
Reactor type	fluidised bed, fixed bed
Interest Rate [%]	1.00, 5.25, 6.00
Carbon capture	considered, not considered
Carbon tax [£/tonne CO_2-eq]	0, 50, 100

6.3.5.1 Sensitivity Analysis

Whilst the general optimisation methodology acts as a type of sensitivity analysis, it does not provide sufficient granularity for some parameters which during preliminary analysis showed a large effect on the simulation outputs. For this reason, the electricity price and system size were added as two additional parameters to the ones already listed in Table 6.4. Furthermore, a greater range of gasification temperatures was explored during the sensitivity analysis.

The system's sensitivity to changes in the gasification temperature was tested by exploring the maximum range of realistic gasification temperatures. More specifically, the gasification temperature was varied in 100°C steps from 600 to 1,200°C.

As electricity is the system's main output, the electricity price was considered another essential variable to consider during sensitivity analysis. As previously discussed in Section 6.3.3.2, electricity prices have been as high as 511.20 £/MWh and are expected to decline and settle around 50 £/MWh in the future. Based on this, the following electricity prices were explored during the sensitivity analysis: 40, 50, 70, 150, 200, 300, 500 £/MWh. This range of values was chosen to analyse the lower range of values, which are considered to represent the most likely future electricity prices, in greater detail, whilst also considering outliers of higher electricity prices based on the recent past.

The size of the system affects both plant efficiency and economic performance because of factors like economies of scale. For this reason, a range of values was considered to quantify the effects of the system size on the system's overall environmental and economic performance. Ultimately, the following system size values in terms of tonnes/hour were considered for the sensitivity analysis: 30, 10, 5, 1, 0.5.

6.4 Results and Discussion

6.4.1 Baseline Results

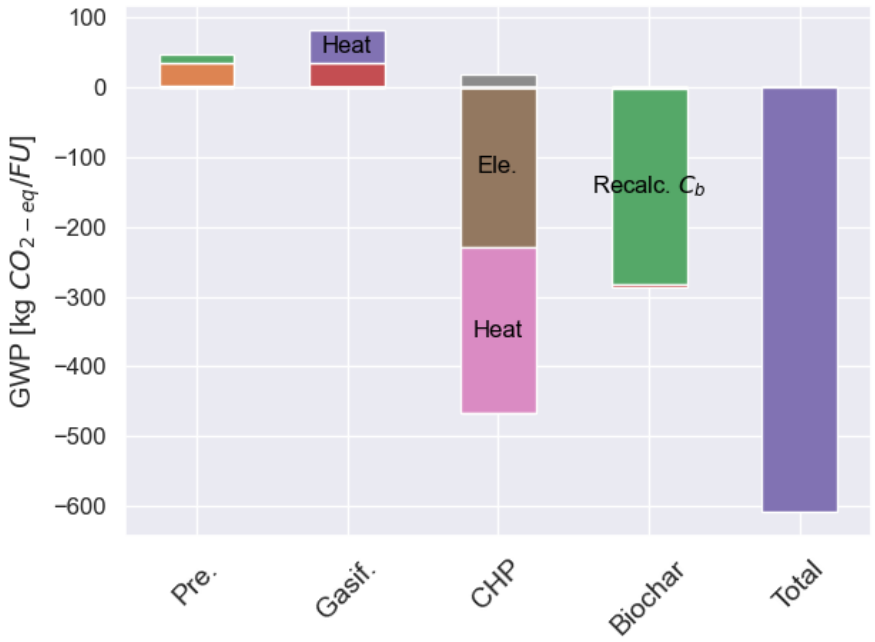
Generally speaking, the system was found to perform best, both in terms of environmental and economic performance, when using forestry residues as the feedstock. Barley Straw performed slightly better than draff in terms of its environmental performance, but slightly worse than draff in terms of its economic performance.

The gasification scheme using forestry residues as a feedstock resulted in an average GWP of -613 $kg\ CO_2$-eq/FU, with the GWP roughly ranging from -850 to -400 $kg\ CO_2$-eq/FU across all Monte Carlo iterations as shown in Figure 6.1. In comparison, using barley straw as a feedstock led to an average GWP of -458 $kg\ CO_2$-eq/FU, with a range of approximately -650 to -300 $kg\ CO_2$-eq/FU. This is illustrated by Figure 6.2. Finally, as shown in Figure 6.3, draff performed worst (in terms of GWP) with a GWP ranging from -500 to -100 $kg\ CO_2$-eq/FU, and a mean GWP of -290 $kg\ CO_2$-eq/FU.

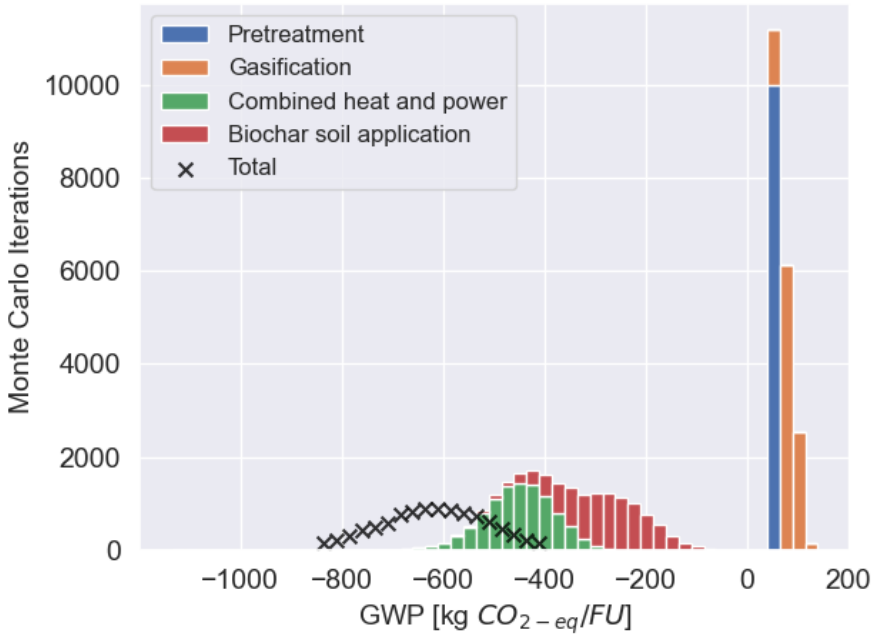
Importantly, all three feedstocks consistently achieved negative emissions across all Monte Carlo simulation runs. The biogenic nature of the feedstocks meant that minimal emissions were associated with the gasification of the feedstock and the subsequent combustion of the resulting syngas. For all three feedstocks, the main drivers for negative emissions were the substitution of grid electricity and conventional heat generation, as well as the long-term storage of biogenic carbon in the form of biochar. The size of the contribution of the three drivers was comparable. In the case of forestry residues, the substitution of electricity and heat both contributed around -220 $kg\ CO_2$-eq/FU, whereas biochar contributed around -280 $kg\ CO_2$-eq/FU. It is important to note that as countries are trying to decarbonise e.g. the electricity grid, the negative emissions associated with displacing the grid mix would reduce due to the lower fraction of fossil fuel-based electricity. However, even when considering a greener grid, the system would still be able to contribute to negative emissions.

Considering Figure 6.3a, it can be seen that thermal energy requirements for the drying of draff, had a substantial environmental impact of around 280 $kg\ CO_2$-eq/FU on average. This came as a result of the draff's high moisture content. In comparison, in the case of forestry residues, feedstock drying had a minimal impact of approximately 35 $kg\ CO_2$-eq/FU on average, as shown in Figure 6.1a. Other pretreatment steps like milling were associated with a minimal GWP of less than 15 $kg\ CO_2$-eq/FU.

When it comes to the scheme's economic performance, both the BCR and net present value (NPV) were used for assessment. As previously mentioned, the feedstock forestry residues was found to have the best economic performance with a mean BCR and NPV of 0.928 and £-17,504,986, respectively. The BCR and NPV distribution for forestry residues is shown in Figure 6.4. It can be seen that the BCR ranged from 0.4 to 1.75 and the NPV

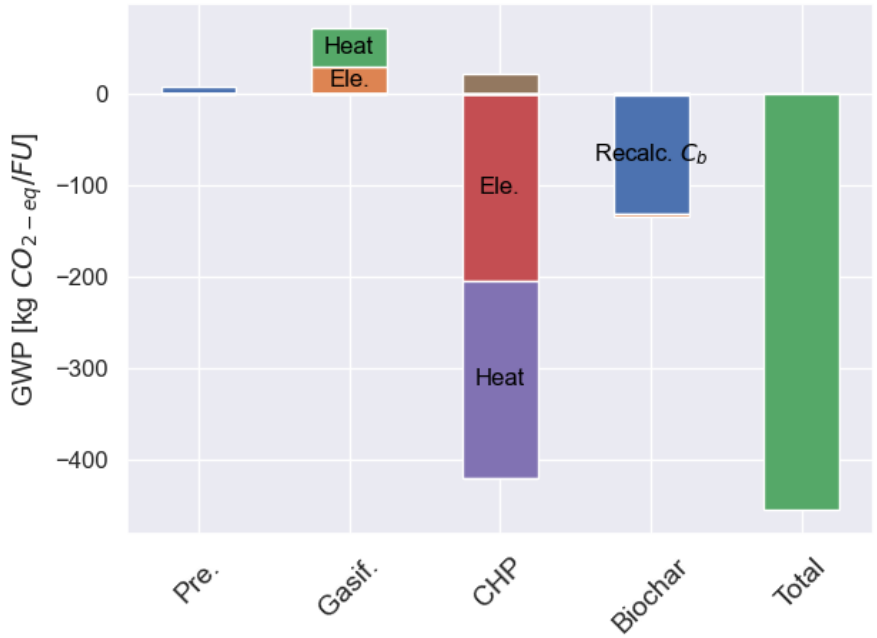


(a)

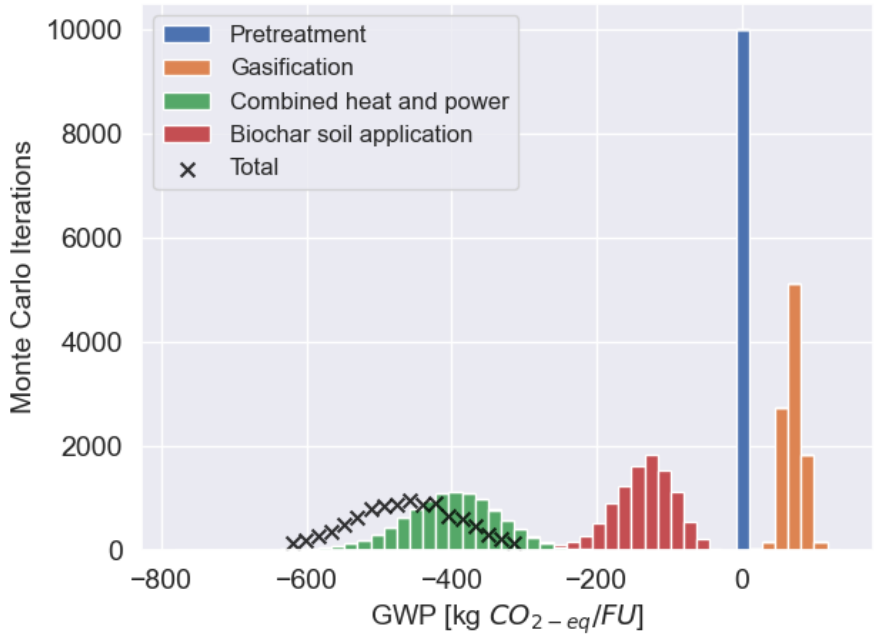


(b)

Figure 6.1: Baseline environmental results for forestry residues as the feedstock. In (a) the results were averaged across all Monte Carlo runs (colours are used to differentiate between subprocesses), whereas (b) shows the results as a histogram based on all Monte Carlo runs. Note: For improved legibility, subprocess labels are only shown for subprocesses contributing more than 3% of the global impact.

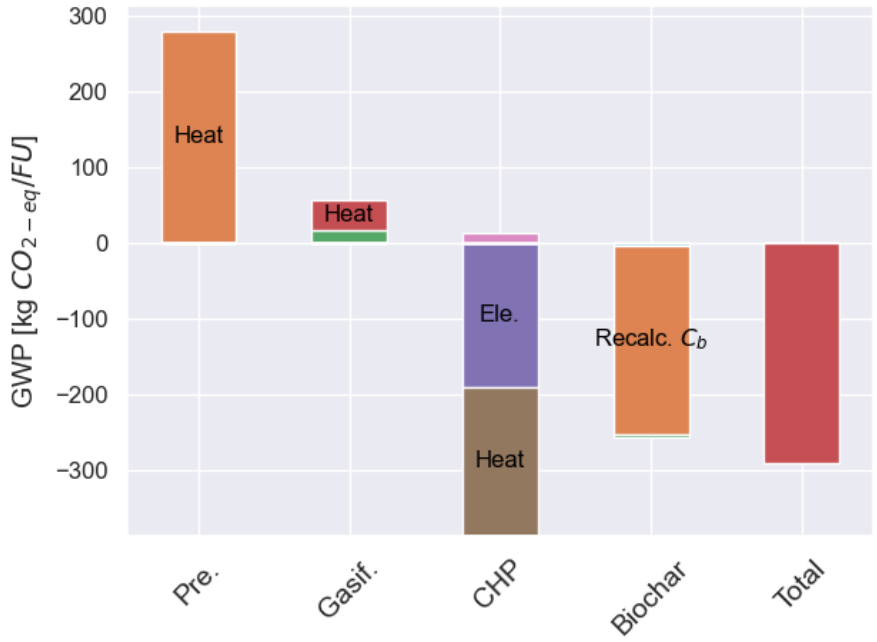


(a)

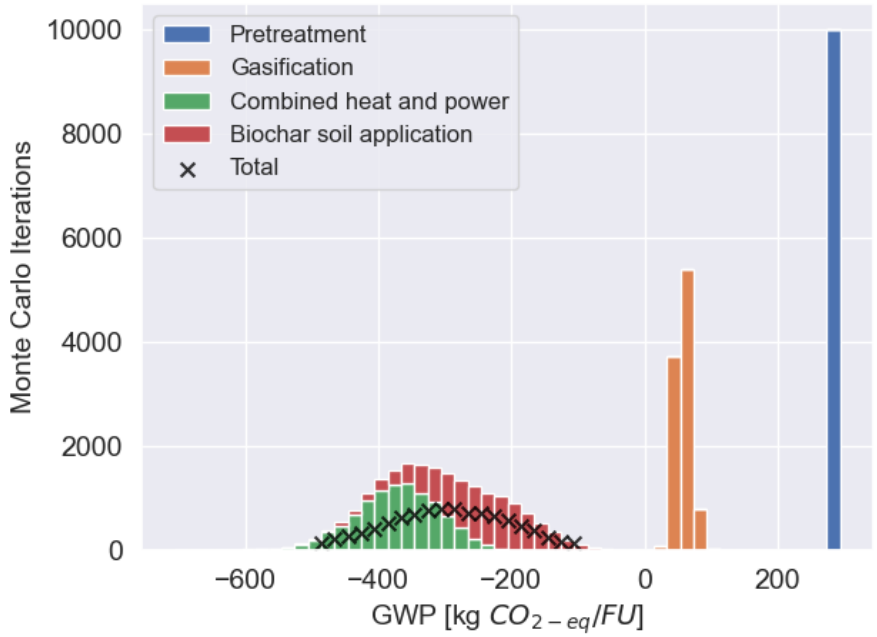


(b)

Figure 6.2: Baseline environmental results for barley straw as the feedstock. In (a) the results were averaged across all Monte Carlo runs (colours are used to differentiate between subprocesses), whereas (b) shows the results as a histogram based on all Monte Carlo runs. Note: For improved legibility, subprocess labels are only shown for subprocesses contributing more than 3% of the global impact.



(a)



(b)

Figure 6.3: Baseline environmental results for draff as the feedstock. In (a) the results were averaged across all Monte Carlo runs (colours are used to differentiate between subprocesses), whereas (b) shows the results as a histogram based on all Monte Carlo runs. Note: For improved legibility, subprocess labels are only shown for subprocesses contributing more than 3% of the global impact.

ranged from around £-150,000,000 to £150,000,000. With an average BCR of slightly below 1 and a negative NPV, the baseline scenario is typically not economically feasible and further government incentives or other changes would be required to ensure such a project's feasibility. However, with 39% of all Monte Carlo runs resulting in a BCR of $>$ 1, small changes could improve the economic feasibility for the project to be feasible more often than not.

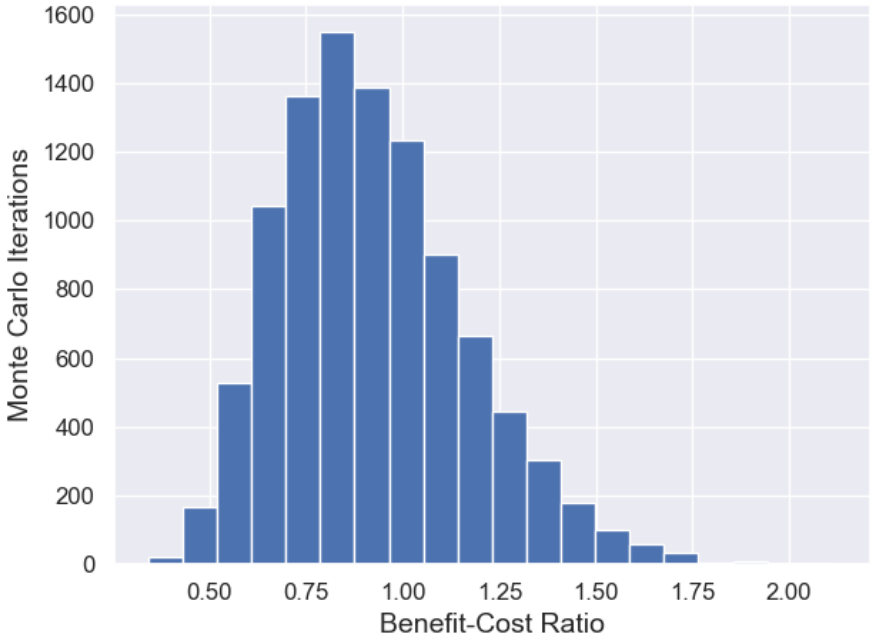
Barley Straw and Draff both performed significantly worse (as shown in Figures 6.5 and 6.6) than forestry residues with an average BCR of 0.770 and 0.779, respectively. Both feedstocks also displayed slightly lower BCR bounds, with their BCR ranging from around 0.3 to 1.6. Similarly, their NPV ranged from approximately £-175,000,000 to £140,000,000 (£125,000,000 in the case of barley straw). Barley straw's mean NPV was £-56,946,732 and draff's NPV was slightly higher at a value of £-52,339,899. Whilst on average barley straw and draff did not exhibit economic feasibility, both feedstocks had the potential to achieve economic feasibility according to the Monte Carlo simulation. In fact, barley straw and draff both achieved a BCR of $>$ 1 in about 15% of all cases as shown in Figures 6.5 and 6.6, respectively.

A more detailed breakdown of the economic performance of the three feedstocks is shown in Figures 6.7, 6.8, and 6.9.

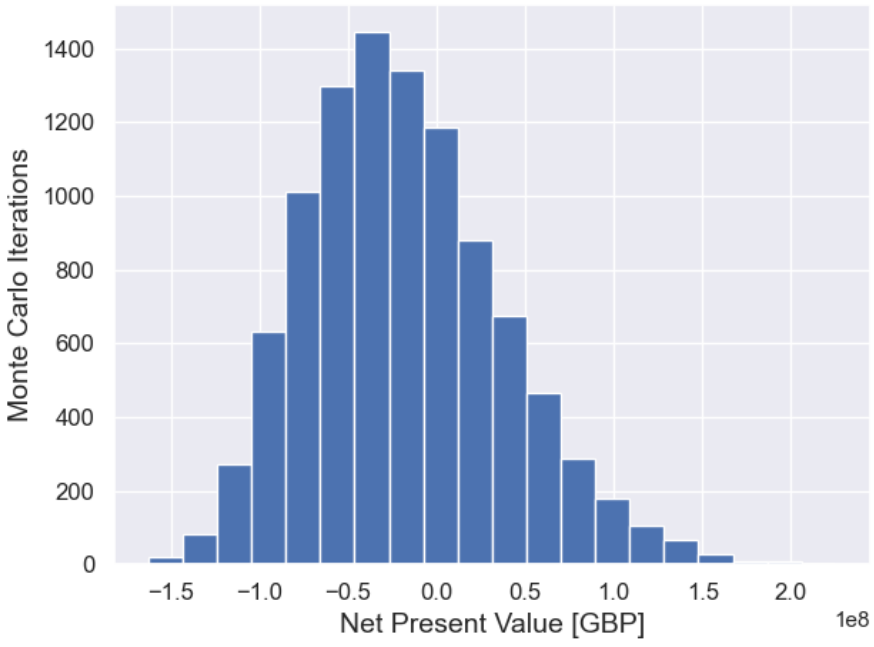
The gasification stage was the biggest cost factor for all feedstocks, with average present values (PVs) in the range of £-103,000,000 to £-112,000,000. The CAPEX and operation and maintenance (O&M) cost for the gasifier were the main cost drivers, with the CAPEX and O&M cost for gas cleaning, and auxiliary energy demands contributing to a lesser amount. When it comes to sources of income, the sale of electricity, heat, and biochar provided positive cash flows. Forestry residues led to the highest average PV resulting from the sale of electricity and heat with values of £140,000,000 and £59,000,000 respectively, as shown in Figure 6.7a. Whilst the positive cash flows due to the sale of electricity and heat were on average only slightly lower for barley straw (£123,000,000 and £53,000,000, respectively), a significant decrease was seen for draff, with PVs of £100,000,000 and £42,000,000, respectively. This reduction in energy sales benefits is strongly linked to the decrease in the respective LHVs of the feedstocks.

The use of forestry residues as a feedstock also led to higher PVs due to the sale of biochar than the other feedstocks, with biochar from forestry residues providing positive cash flows of £38,000,000 on average, across the system's life cycle. Meanwhile, the sale of biochar from barley straw and draff resulted in PVs of around £33,000,000 for both feedstocks.

Feedstock costs showed significant variation between the three feedstocks, with draff being by far the cheapest with an average PV of around £-27,000,000 across the system's lifespan, as shown in Figure 6.9a. In contrast, forestry residues and barley straw were on

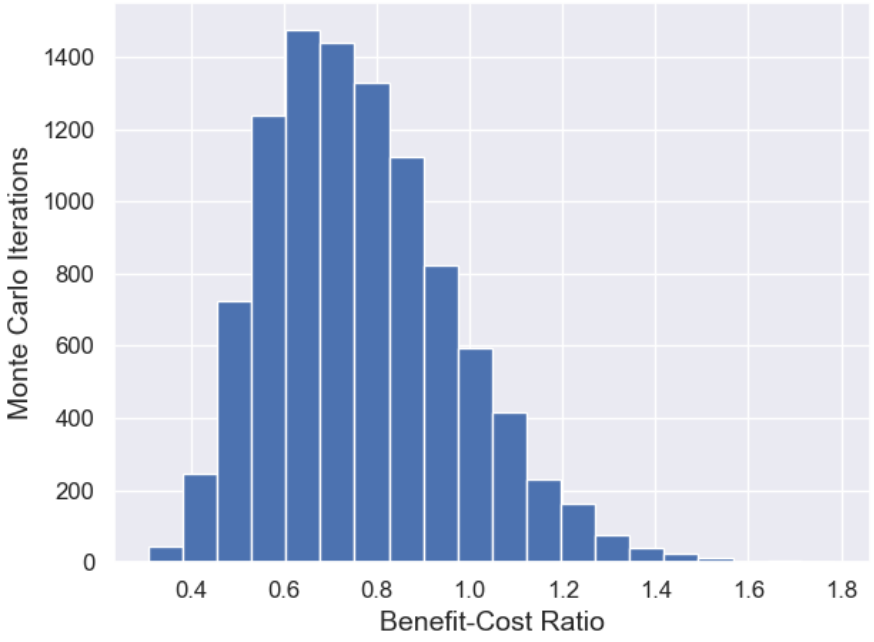


(a)

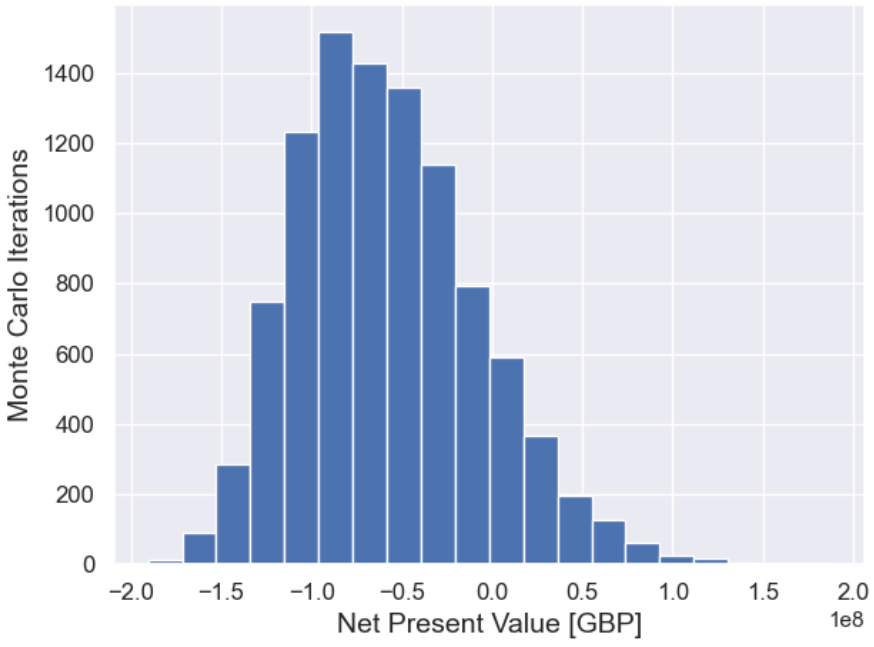


(b)

Figure 6.4: Net economic results, in terms of BCR (shown in (a)) and NPV (shown in (b)), for forestry residues as the feedstock.

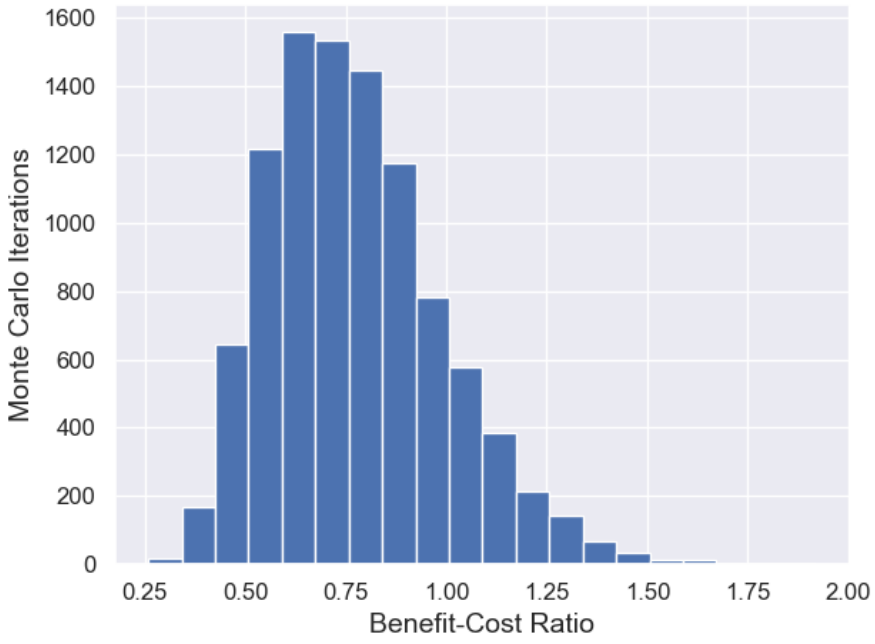


(a)

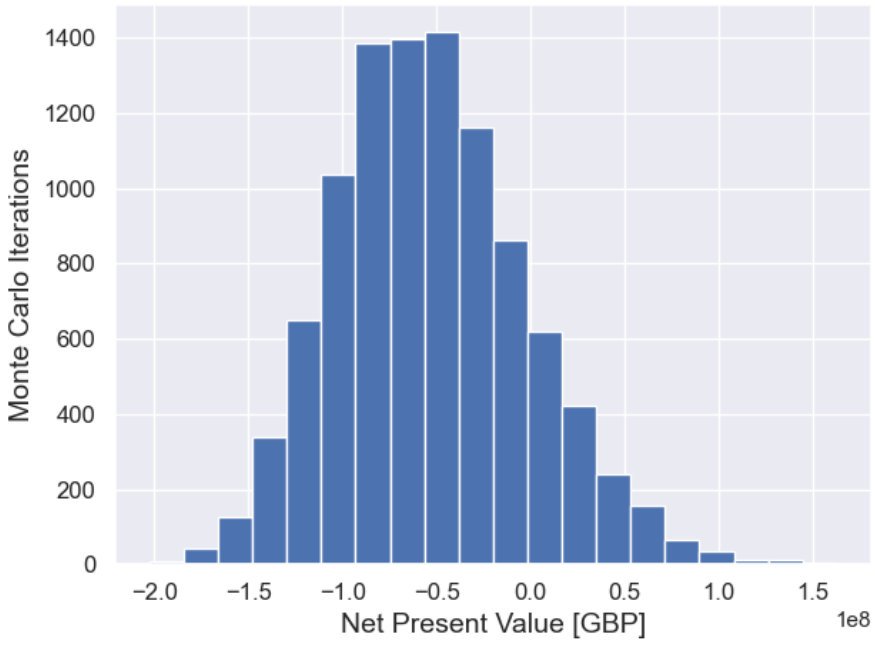


(b)

Figure 6.5: Net economic results, in terms of BCR (shown in (a)) and NPV (shown in (b)), for barley straw as the feedstock.

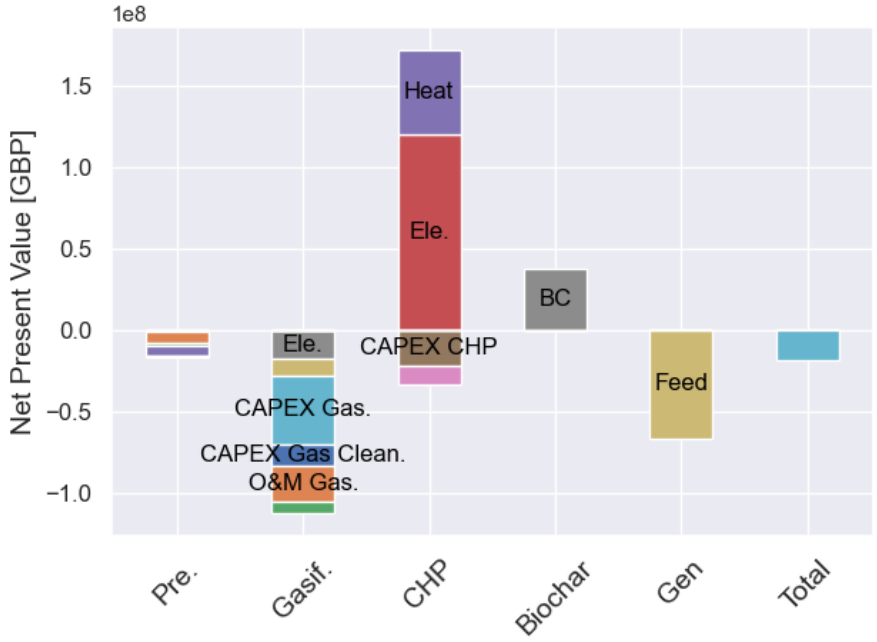


(a)

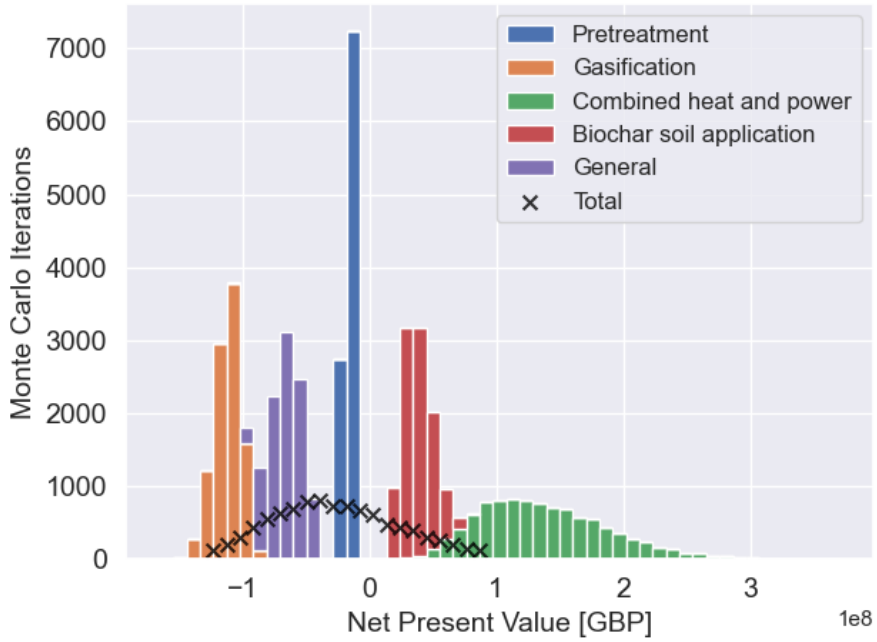


(b)

Figure 6.6: Net economic results, in terms of BCR (shown in (a)) and NPV (shown in (b)), for draff as the feedstock.

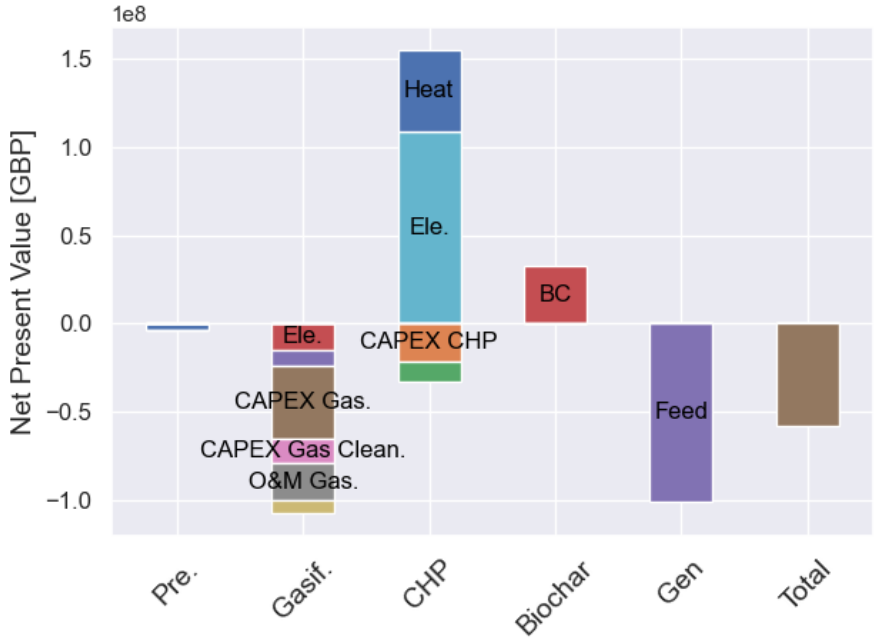


(a)

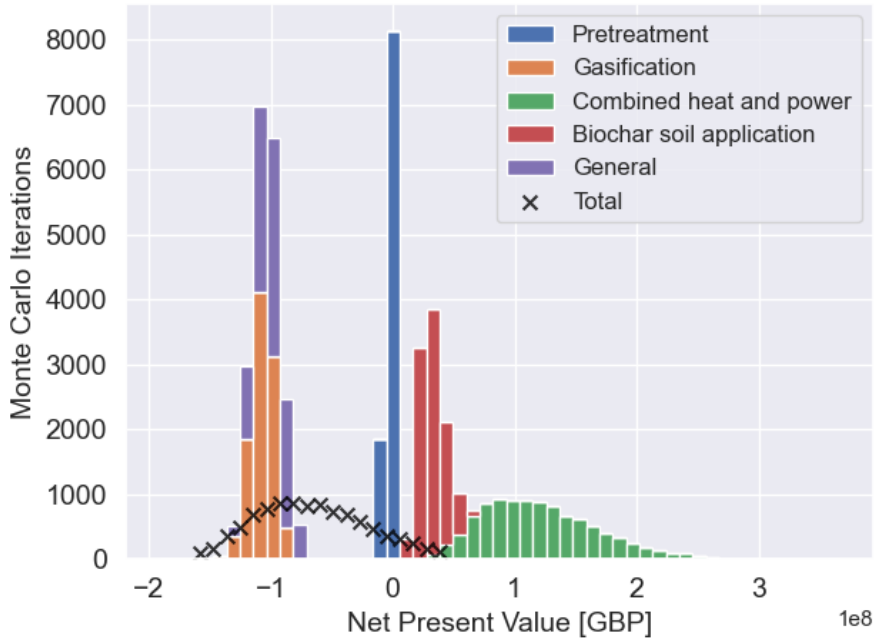


(b)

Figure 6.7: Baseline economic results shown by process for forestry residues as the feedstock. In (a) the results were averaged across all Monte Carlo runs (colours are used to highlight different cost/benefit factors), whereas (b) shows the results as a histogram based on all Monte Carlo runs. Note: For improved legibility, subprocess labels are only shown for subprocesses contributing more than 3% of the total NPV.

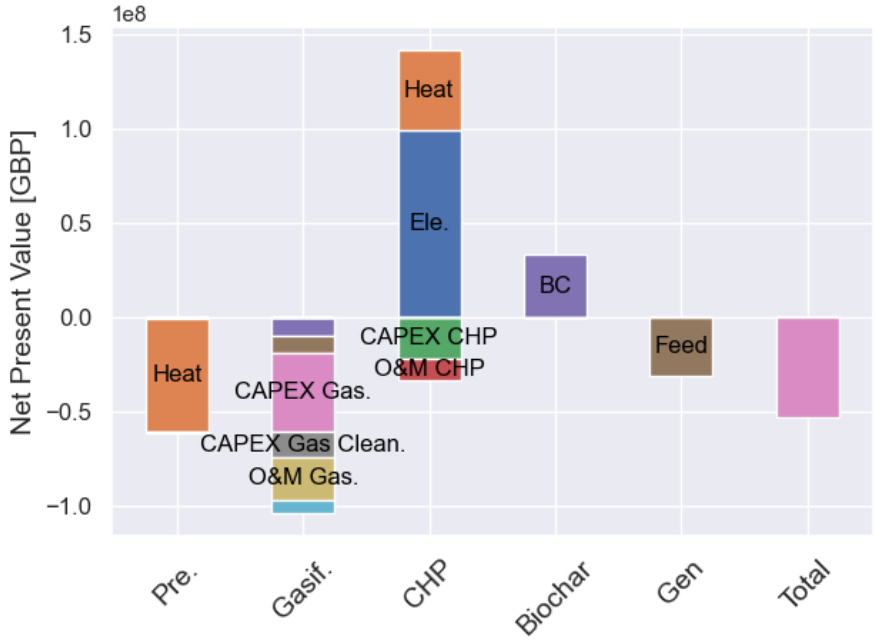


(a)

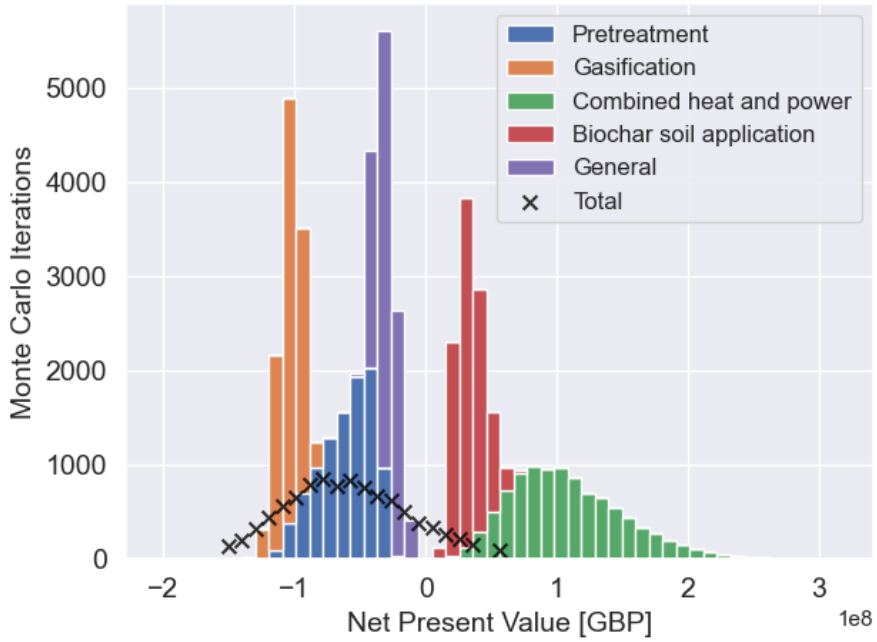


(b)

Figure 6.8: Baseline economic results shown by process for barley straw as the feedstock. In (a) the results were averaged across all Monte Carlo runs (colours are used to highlight different cost/benefit factors), whereas (b) shows the results as a histogram based on all Monte Carlo runs. Note: For improved legibility, subprocess labels are only shown for subprocesses contributing more than 3% of the total NPV.



(a)



(b)

Figure 6.9: Baseline economic results shown by process for draff as the feedstock. In (a) the results were averaged across all Monte Carlo runs (colours are used to highlight different cost/benefit factors), whereas (b) shows the results as a histogram based on all Monte Carlo runs. Note: For improved legibility, subprocess labels are only shown for subprocesses contributing more than 3% of the total NPV.

average more than twice and nearly four times as expensive, respectively. However, these savings in feedstock costs are somewhat offset by higher pretreatment costs due to draff's extensive feedstock drying requirements. The average PV of pretreatment was around £-16,000,000 and £-3,800,000 for forestry residues and barley straw as shown in Figures 6.7a and 6.8a, respectively. In contrast, the average PV of pretreatment, in the form of drying, was approximately £-60,000,000 for draff.

When it comes to the system's net electricity generation potential, forestry residues performed best again. Averaged across all Monte Carlo iterations, 864 kWh/FU of electricity from forestry residues was available for sale after accounting for all auxiliary demands. Barley straw and draff resulted in an electricity generation of 805 and 799 kWh/FU, respectively. Net electricity generation from forestry residues as a feedstock varied between 400 and 1,400 kWh/FU across the 10,000 Monte Carlo iterations. A similar lower bound and slightly lower upper bound of around 1,100 kWh/FU was seen for barley straw and draff.

A different picture was seen for the system's net heat or thermal energy generation. Barley straw with a net heat generation of 777 kWh/FU generated the most heat on average. Forestry residues followed closely behind with an average heat generation of 711 kWh/FU. Barley straw was assumed to be received on a sufficiently dry basis for direct gasification without additional drying, which has been accounted for in the feedstock's price. Utilising forestry residues as a feedstock required a small amount of energy (approximately 150 kWh/FU) for drying, which explains the difference in net heat generation. Draff on the other hand had a large heat demand for drying of approximately 1,250 kWh/FU on average, which was a result of the feedstock's high moisture content. Ultimately this led to a net heat generation of -546 kWh/FU. This means that additional heat would have to be supplied for the operation of the gasification system. Detailed figures showing the net energy generation of the three different feedstocks and their respective system components are shown in Appendix B.

In conclusion, all three feedstocks performed well on an environmental basis. In particular, the feedstock forestry residues showed a strong environmental performance. The high moisture content of draff (78% wb) makes its effective utilisation through thermochemical methods more challenging than other lower moisture content feedstocks. This might indicate that draff is better suited for biochemical conversion (e.g. AD). The high LHV of forestry residues (17.97 MJ/kg wb) and its competitive pricing (triangular distribution mode of 65 £/tonne) also made it the most economical feedstock. However, the economic feasibility of any feedstock option was shown to be uncertain. For this reason, government subsidies or incentives, such as feed-in tariffs or a carbon tax, may be required to ensure the economic feasibility of biomass gasification developments. Additionally, future research and development might bring cost factors such as the CAPEX of various

system components down, allowing the technology to better compete with other energy generation options.

6.4.2 Optimisation Results

A comparison of the optimisation results for the three studied feedstocks is shown in Figure 6.10. The figure shows the results for each feedstock for the 1,296 unique parameter combinations resulting from the parameters shown in Table 6.4. Optimisation results differentiated by the seven parameters varied during the optimisation (as shown in Table 6.4) are shown in Figures 6.11, 6.12, 6.13, 6.14, 6.15, 6.16, and 6.17. Results are only shown for the feedstock forestry residues as it performed the best during the baseline analysis and similar trends were seen for all feedstocks. However, the optimisation results for barley straw and draff can be found in Appendix C.

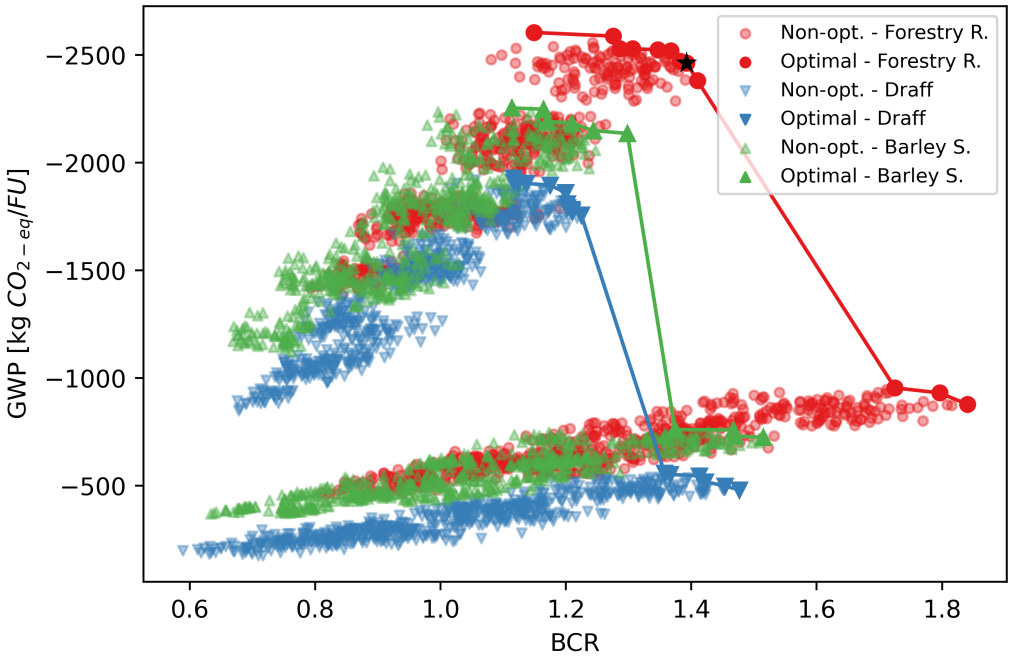


Figure 6.10: Optimisation results differentiated by feedstock. Solid colours indicate results on the Pareto front (shown by a solid line), whereas opaque colours indicate non-Pareto optimal results. The Pareto optimal solution highlighted by a black star is the solution further studied for the best-performing feedstock forestry residues.

From Figure 6.10 it can be seen that the feedstock forestry residues again outperformed both other feedstocks across the board. For forestry residues, the most optimal parameter combination in terms of environmental performance resulted in a mean GWP of -2,604 kg CO_2-eq/FU and a BCR of 1.149. In contrast, the most economically optimal solution was found to have a mean GWP of -877 kg CO_2-eq/FU and BCR of 1.840. Other solutions on the Pareto front show GWP and BCR values between these extreme values. Optimal

values for barley straw ranged between a GWP of -2,254 to -724 $kg\ CO_2$-eq/FU and a BCR of 1.112 to 1.515. Finally, the Pareto optimal solutions for draff showed a GWP from -1,929 to -480 $kg\ CO_2$-eq/FU and a BCR between 1.116 to 1.413.

When studying the effect of the gasification temperature on the system's GWP and BCR, Figure 6.11 makes it obvious that the system is highly sensitive to changes in the gasification temperature and an increase in it led to large improvements in both GWP and BCR. In fact, all Pareto optimal solutions were found for the highest considered temperature of 850°C. For this reason, the effect of a further increase in the gasification temperature was analysed in Section 6.4.3. A similar improvement in environmental and economic performance with an increase in gasification temperature was also seen for the other two feedstocks.

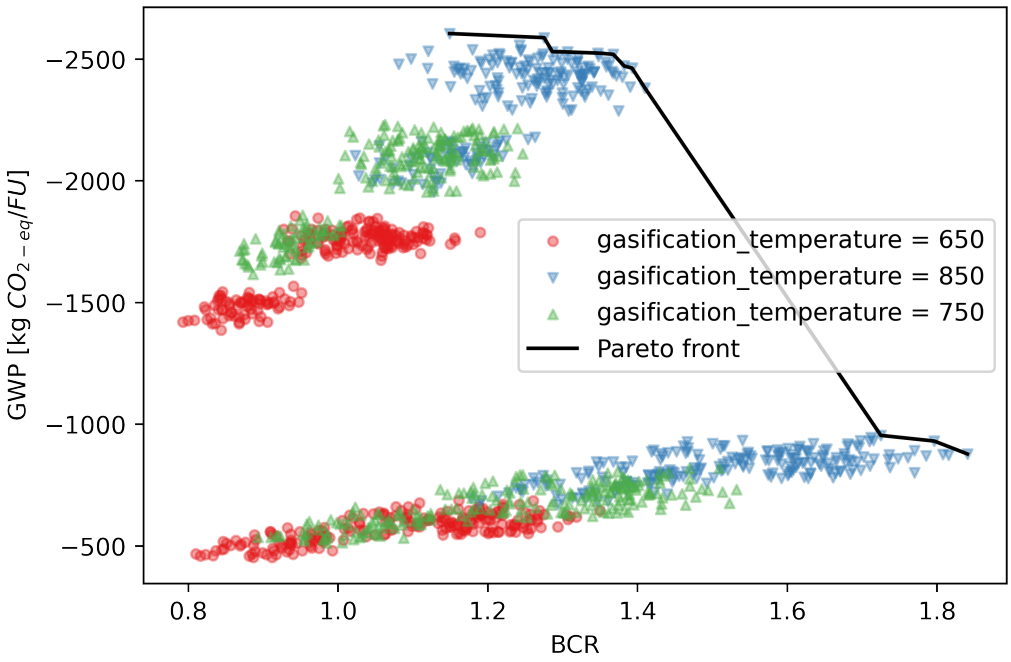


Figure 6.11: Optimisation results for the feedstock forestry residues shown for the model parameter gasification temperature [°C].

Changes in ER were found to have a relatively small effect on the system's performance as shown in Figure 6.12. Small changes were seen between the ER values of 0.25 and 0.30 and preliminary analysis showed a similarly small effect for a larger range of parameter values.

Figure 6.13 shows the effect of the gasifying agent on the system's GWP and BCR. From the figure, it can be seen that oxygen and steam performed better as gasifying agents than air, but overall the effect of the gasifying agent was relatively small and smaller than the effect of for instance the gasification temperature. Oxygen performed slightly better than steam, but both agents performed at a comparable level. When prioritising the GWP, oxygen may be seen as a better agent, whereas steam was preferable when trying

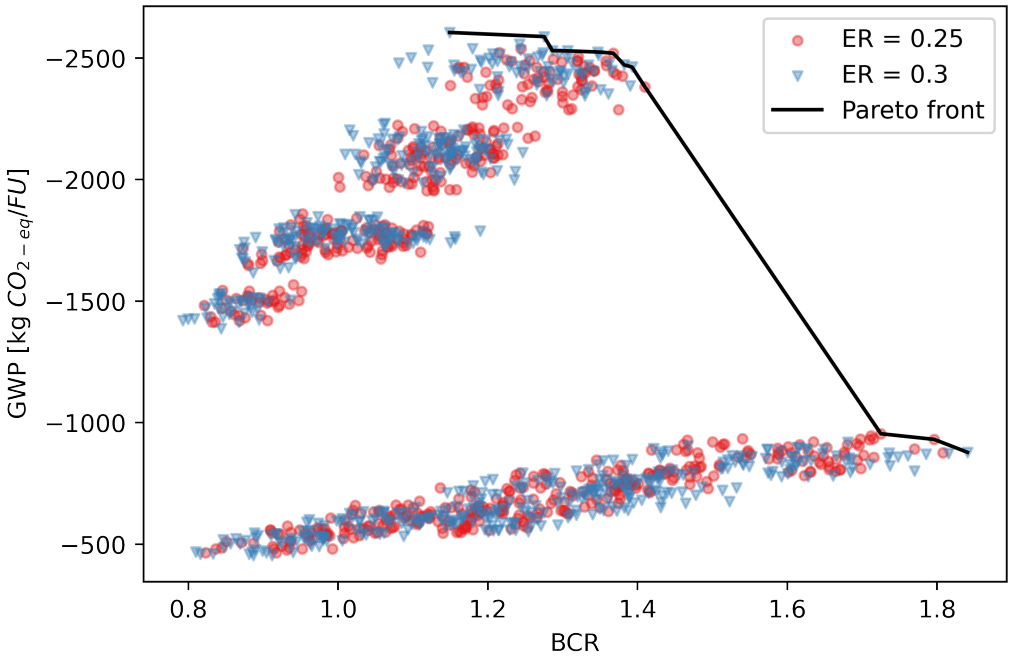


Figure 6.12: Optimisation results for the feedstock forestry residues shown for the model parameter ER.

to maximise the system's economic performance.

The gasification reactor type had a relatively small effect on the optimisation outputs, as can be seen in Figure 6.14. Fluidised bed reactors generally performed better than fixed bed reactors, but some Pareto optimal solutions were also found for fixed bed reactors.

When it comes to the interest rate or rate of return used for the economic analysis, it was found that a higher interest rate improved the economic feasibility of the project. This is shown in Figure 6.15. This can largely be explained by the fact that most future cash flows occurring throughout the system's lifespan are benefits due to the sale of products.

The effect of including carbon capture is shown in Figure 6.16. As shown in the figure, the inclusion (or exclusion) of CCS has the largest impact on the system's overall GWP and BCR. Two distinct groups can be identified on the plot, with solutions including CCS showing significantly greater carbon-saving potentials. However, the maximum BCR was achieved by solutions not considering CCS, which can be explained by the added capital expenditures (CAPEX) and operation and maintenance (O&M) costs due to the CCS system, as well as additional costs for transport and storage. For example, the maximum BCR for a system considering CCS was 1.410, whereas BCRs of up to 1.840 were achieved for systems excluding CCS.

The introduction of a carbon tax was considered a potential option to improve the economic feasibility of a NET. Its effect is shown in Figure 6.17. From the figure, it can be seen that the introduction of a carbon tax had a relatively small effect on the overall feasibility of the system. Whilst, the highest BCR was achieved for the highest carbon

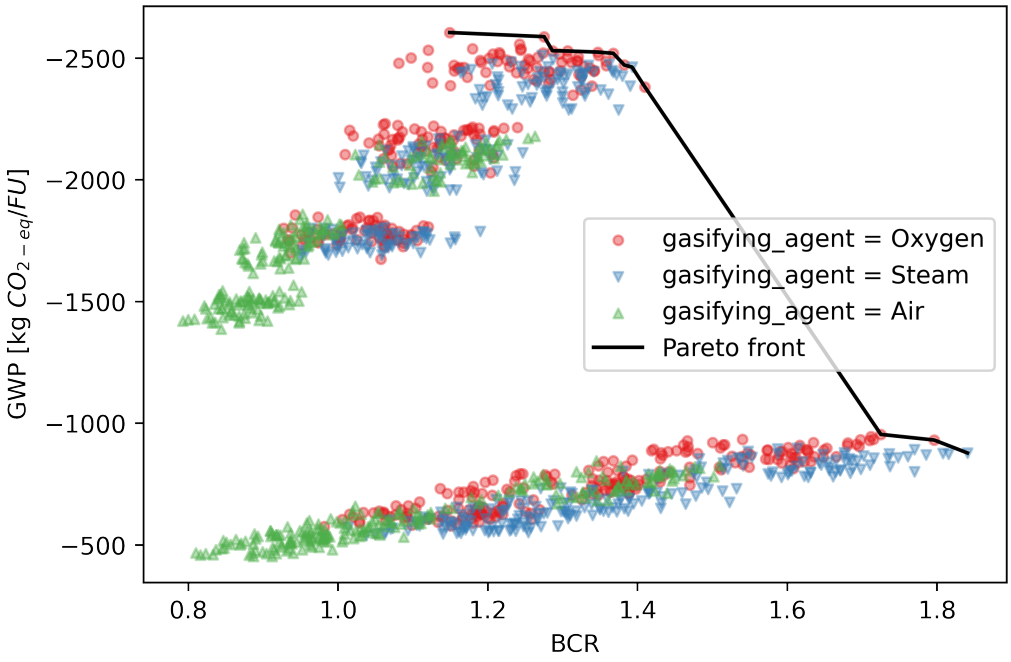


Figure 6.13: Optimisation results for the feedstock forestry residues shown for the model parameter gasifying agent.

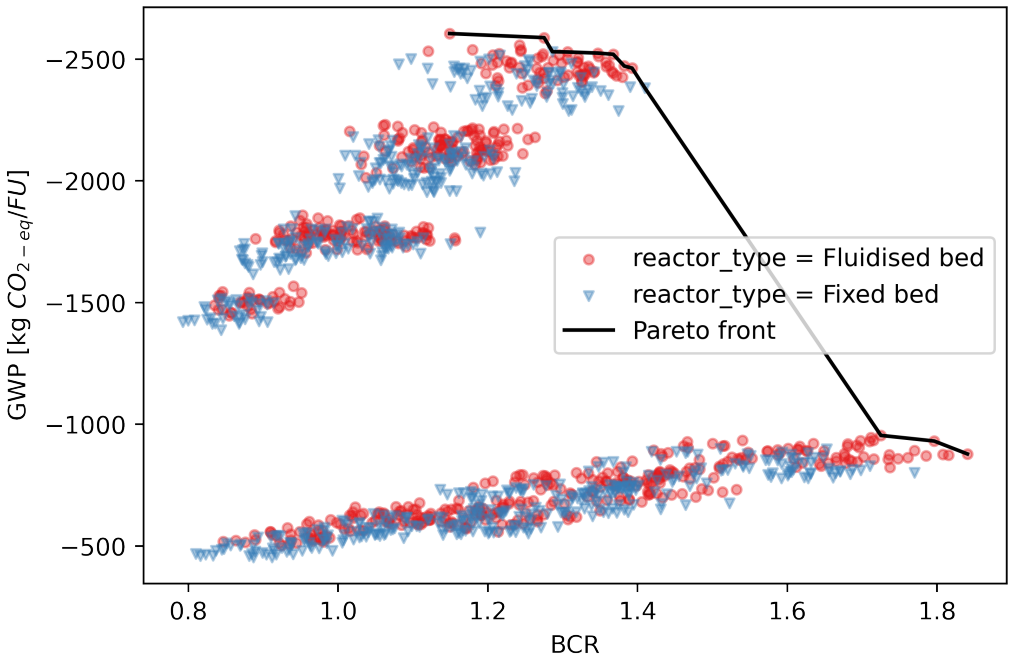


Figure 6.14: Optimisation results for the feedstock forestry residues shown for the model parameter reactor type.

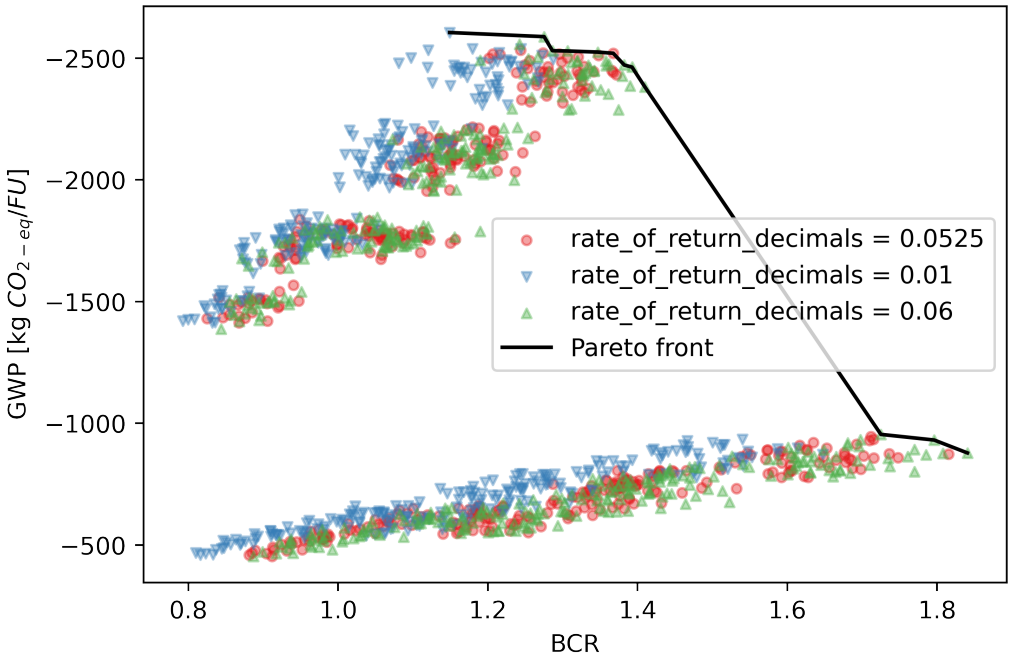


Figure 6.15: Optimisation results for the feedstock forestry residues shown for the model parameter interest rate [%].

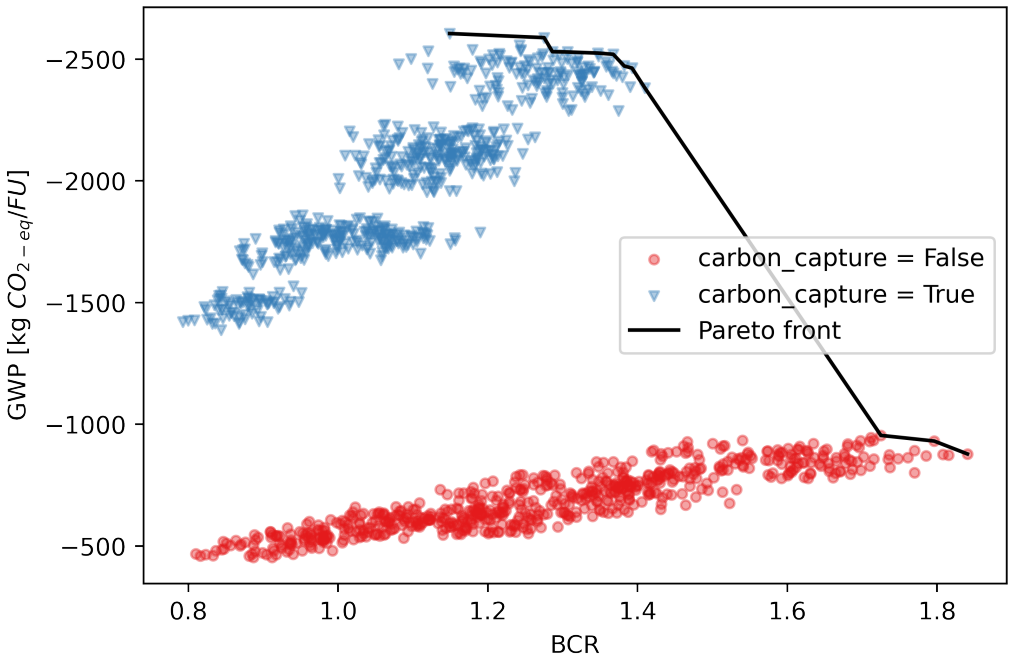


Figure 6.16: Optimisation results for the feedstock forestry residues shown for the model parameter CCS (included or not included).

tax value, most Pareto optimal solutions were found for a carbon tax value of 50 £/tonne CO_2-eq. This indicates that whilst a carbon tax can be an effective means to improve a NET's economic feasibility, in this work other system parameters were found to have a more significant impact.

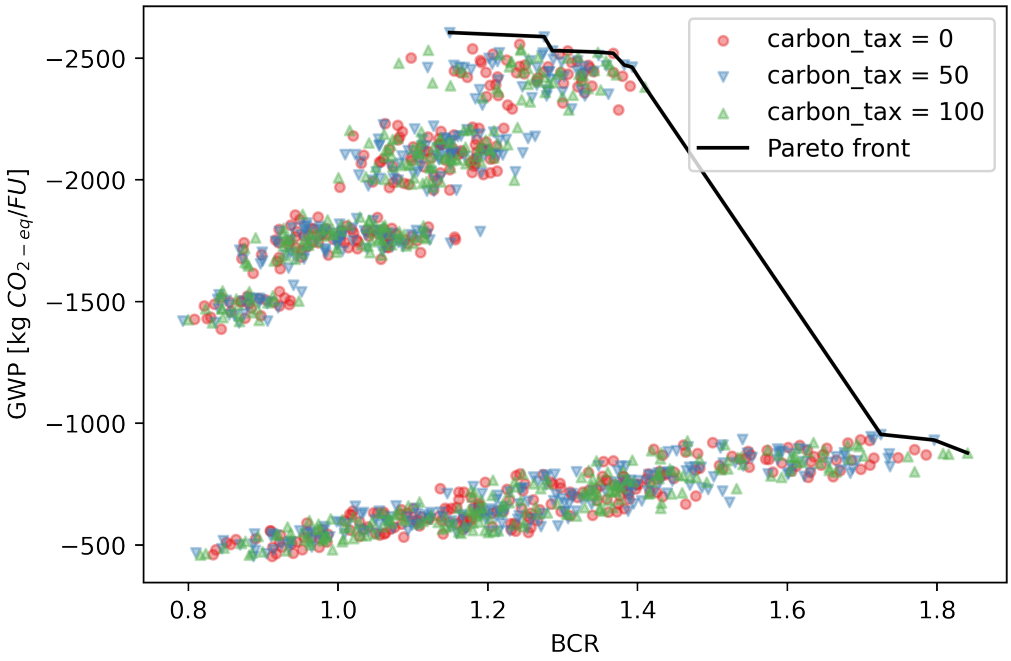


Figure 6.17: Optimisation results for the feedstock forestry residues shown for the model parameter carbon tax [£/tonne CO_2-eq].

6.4.3 Sensitivity Analysis Results

The sensitivity analysis results are again only shown for the best-performing feedstock choice of forestry residues, but similar trends were seen for the other studied feedstocks.

In the previous section (Section 6.4.2) it was found that an increase in gasification temperature in the 650 to 850°C range led to an increase in both BCR and GWP. Figure 6.18 further explores the effect of the gasification temperature on the simulation outputs across a wider range of temperatures. At low gasification temperatures of 600°C and 700°C a minimal change on the simulation outputs was seen. In the 700 to 1,000°C range, significant improvements in both BCR and GWP can be identified from the figure. Interestingly, increasing the temperature beyond 1,000°C did not yield further improvements, but rather decreased the system's performance. These results illustrate the system's sensitivity to the gasification temperature and highlight the importance of finding an optimal gasification temperature to maximise its environmental and economic performance.

As previously discussed in Section 6.3.5.1, electricity prices have been prone to large variations in the last couple of years and are subject to significant uncertainties. Their

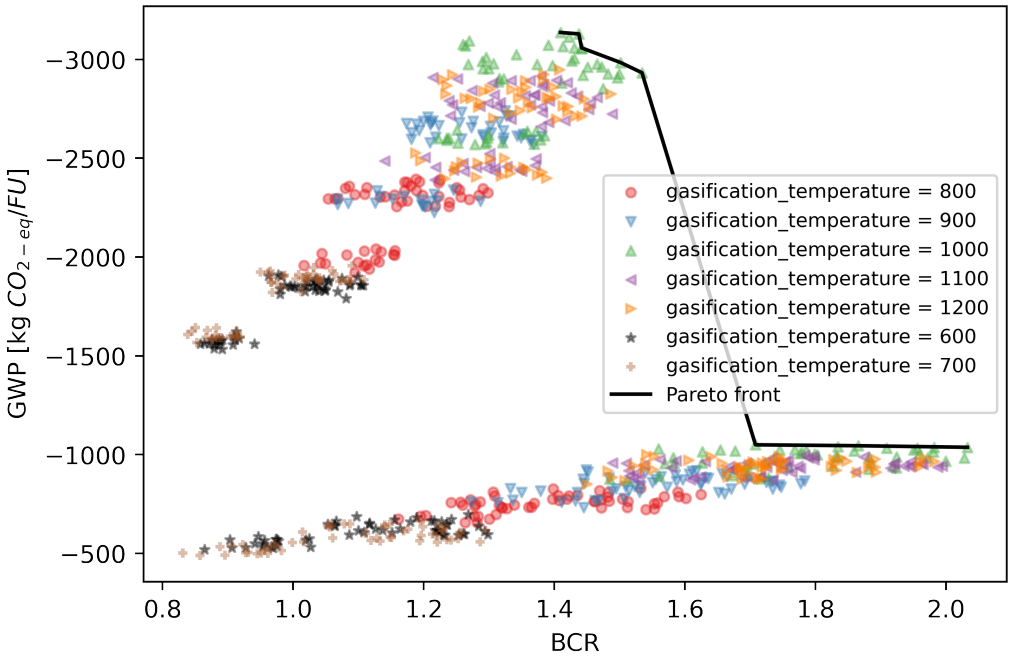


Figure 6.18: Sensitivity analysis results for the feedstock forestry residues shown for the model parameter gasification temperature [°C].

impact on the system's economic performance is shown in Figure 6.19. The figure clearly shows an increase in BCR for higher electricity prices, which is logical as the system's main source of income is the sale of electricity. At a price of 150 £/MWh all solutions not considering CCS were found to achieve a BCR > 1. Whilst electricity prices have reached these prices in the last couple of years, due to factors like political unrest, it is unlikely that they will settle at this value. However, it was also shown that the system can achieve economic feasibility (i.e. BCR > 1) for lower electricity prices. Ultimately, this indicates that by wisely selecting operating parameters the simulated gasification development was able to operate economically, with further government involvement through e.g. feed-in tariffs having the power to further improve its economics.

The final parameter considered as part of the sensitivity analysis was the system's size in terms of hourly feedstock supply. Figure 6.20 indicates that system size had a small effect on the system's environmental performance, but a large effect on its economics. Unexpectedly, the figure does not show a monotonic increase in BCR with an increase in system scale. The BCR improved in the 1 to 30 tonnes/hour region, with a large jump being seen from a system size of 10 tonnes/hour to 30 tonnes/hour. This trend aligns well with what economies of scale would dictate. Importantly, at the baseline system size of 10 tonnes/hour more than 80% of solutions were found to have a mean BCR > 1. Solutions for the smallest considered system size of 0.5 tonnes/hour fell between solutions for the system sizes of 10 and 30 tonnes/hour. This indicates that whilst economies of scale generally hold true, at vastly different system sizes other factors such as the choice

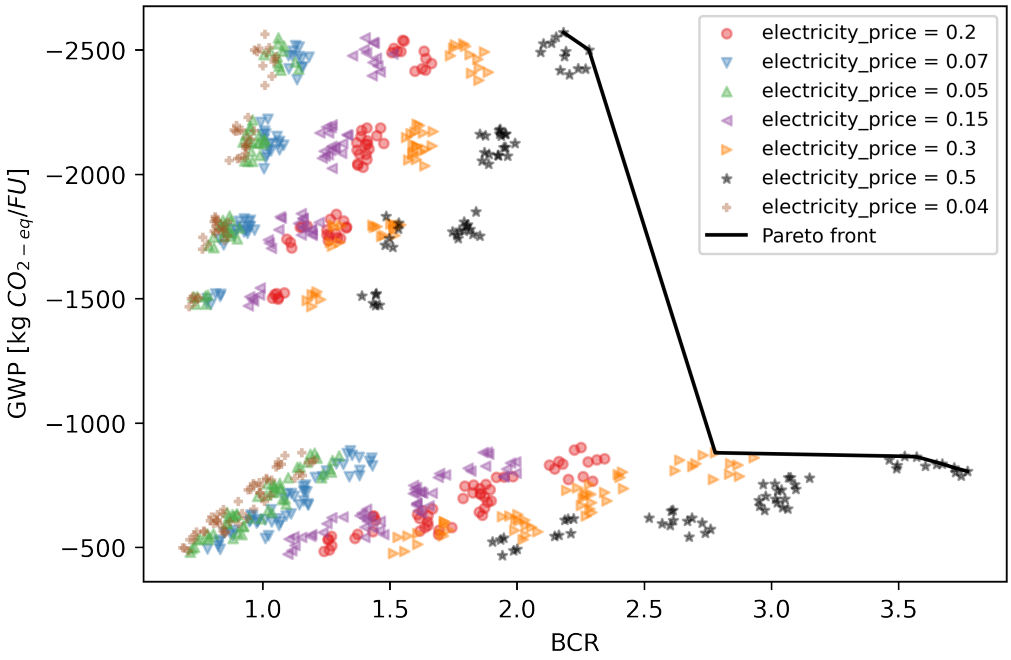


Figure 6.19: Sensitivity analysis results for the feedstock forestry residues shown for the model parameter electricity price [£/kWh].

of different system components may mean that smaller systems can be economically more feasible than larger systems.

6.4.4 Updated Baseline Results for the Feedstock Forestry Residues

Whilst all Pareto solutions are considered optimal, a decision maker has to typically select one solution based on the present trade-offs. In this instance, a compromise between environmental and economic factors must be found. This process can be somewhat subjective, as different decision-makers may prioritise different objectives. In this instance, the Pareto optimal solution indicated by the black star in Figure 6.10 was considered the most suitable solution for the following reasons: (i) The solution resulted in a mean BCR of > 1, indicating its economic feasibility; (ii) The inclusion of CCS resulted in a significantly improved environmental performance, so it was prioritised over more economical solutions with worse environmental performance; (iii) The solution shows a small decrease in environmental performance compared to the environmentally best-performing solution for a significant increase in BCR. Thus, this section shows the simulation outputs for the feedstock forestry residues after modifying parameters based on the findings in Sections 6.4.2 and 6.4.3.

To ensure a fair comparison to the baseline results, the system size remained unchanged at 10 tonnes/hour. However, the following parameters were changed in comparison to the baseline results:

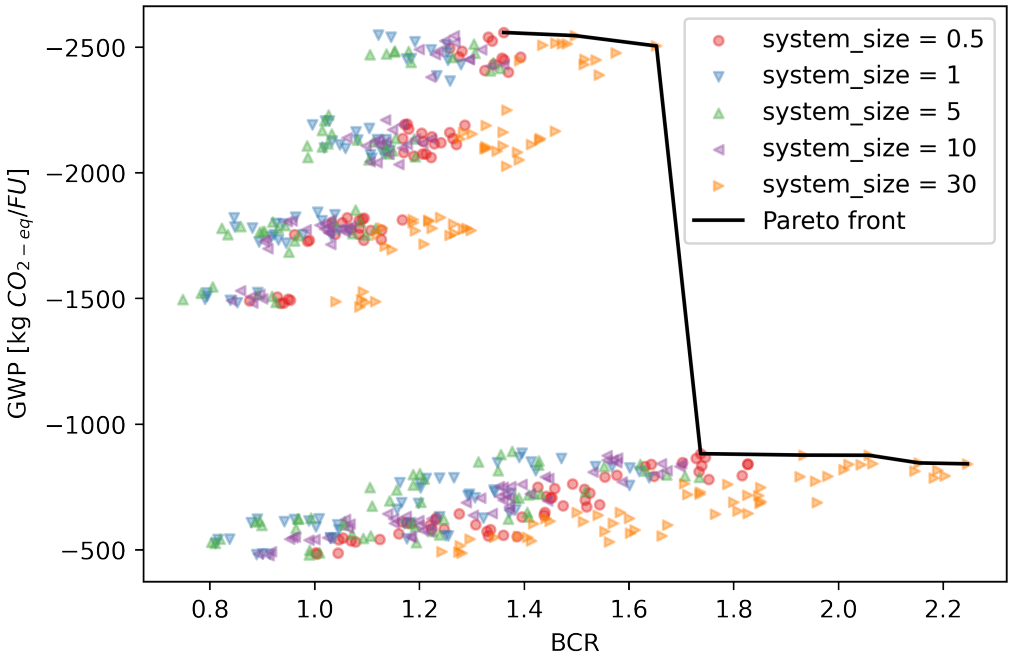
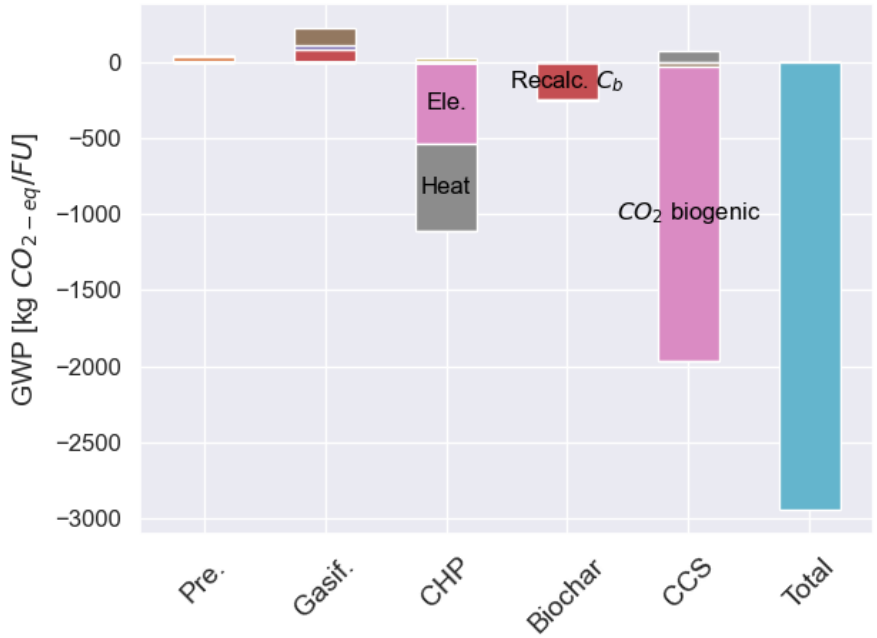


Figure 6.20: Sensitivity analysis results for the feedstock forestry residues shown for the model parameter system size [tonnes/hour].

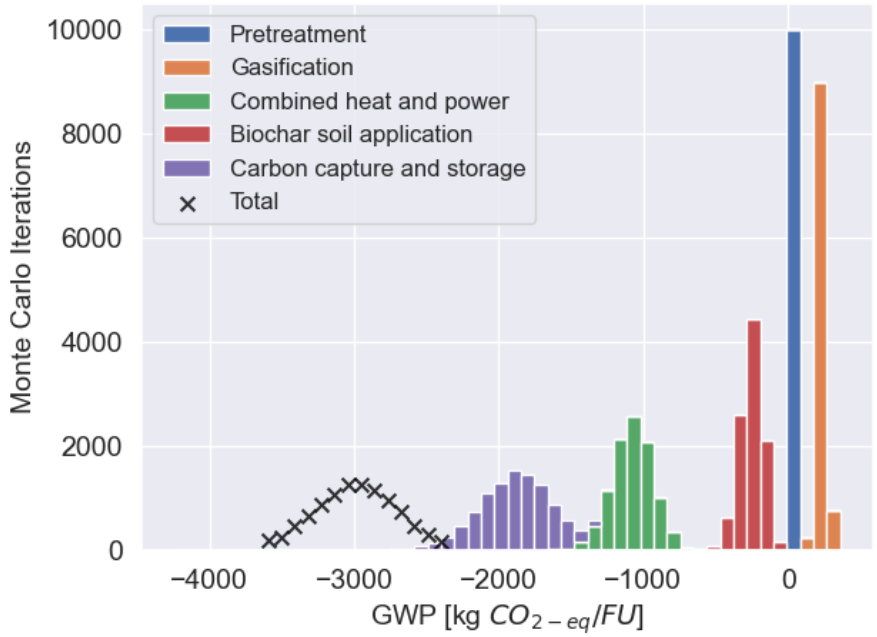
- The gasification temperature was increased from 750°C to 1,000°C, based on the findings in Section 6.4.3. It should be highlighted that in Figure 6.10, the solution marked by the black star utilised a temperature of 850°C. As a result, the GWP and BCR were further enhanced in comparison to the values displayed in the figure, due to the use of the best temperature value found during the sensitivity analysis.
- The gasifying agent was changed from air to steam based on the optimisation results.
- CCS was included based on the optimisation results.
- The carbon tax was changed from 100 £/tonne CO_2-eq to its default value of 0 £/tonne CO_2-eq.
- The interest rate was changed from 6% to its default value of 5.25%.

The simulation showed that the model post-optimisation performed significantly better than the baseline model. The mean GWP changed from -613 kg CO_2-eq/FU to -2,948 kg CO_2-eq/FU, with CCS being the main factor leading to this improvement with a GWP of nearly - 2,000 kg CO_2-eq/FU as shown in Figure 6.21a. Figure 6.21b shows that the net GWP ranged from approximately -3,500 to -2,400 kg CO_2-eq/FU across the 10,000 Monte Carlo iterations.

A lot of improvement was also seen in the economic performance of the model. Previously, a mean BCR of < 1 and a negative NPV showed that on average the system was not



(a)



(b)

Figure 6.21: Environmental results for the baseline scenario, updated according to the optimisation and sensitivity analysis findings for the feedstock forestry residues. In (a) the results were averaged across all Monte Carlo runs, whereas (b) shows the results as a histogram based on all Monte Carlo runs. Note: For improved legibility, subprocess labels are only shown for subprocesses contributing more than 3% of the global impact.

economically feasible. After optimisation, both metrics indicated economic feasibility with a mean BCR of 1.101 and NPV of £37,074,711. Now more than 55% of all Monte Carlo iterations resulted in a BCR of > 1 and the inclusion of e.g. a carbon tax would further improve this margin, in particular due to the large amount of negative emissions resulting from the CCS system. When considering the contribution of different subprocesses to the updated model's NPV, as shown in Figure 6.22, it becomes clear that CCS is expensive, resulting in a greater cost than any other subprocess. However, improved efficiency, resulting from better choices in factors such as the gasifying agent and gasification temperature, resulted in increased positive cash flows from the sale of electricity and heat.

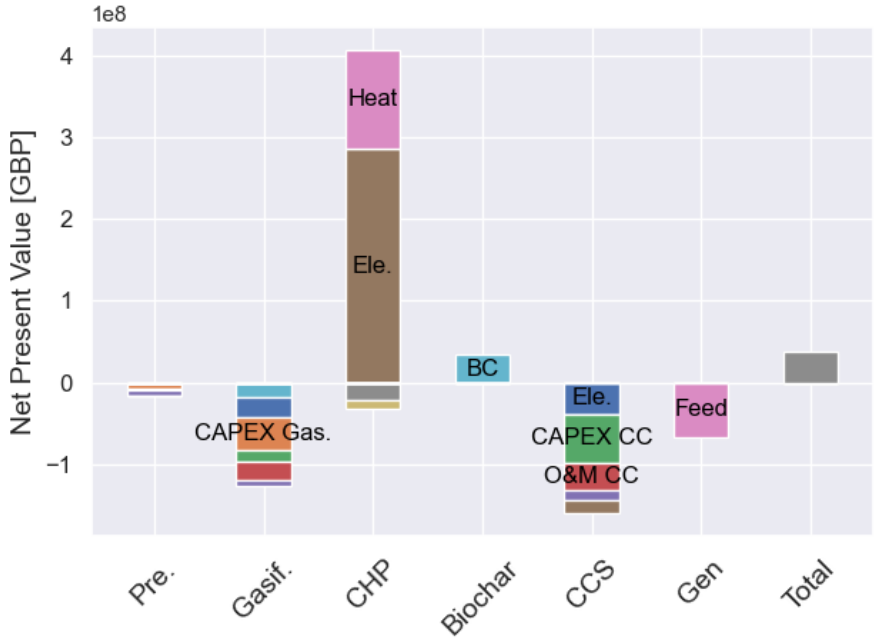
6.5 Chapter Summary and Conclusions

In this chapter, the comprehensive biomass and waste gasification modelling framework developed in Chapter 5 was utilised to conduct a case study of three promising feedstocks (forestry residues, barley straw, and draff) for a gasification development in Scotland. Section 6.3 outlines how various model parameters were selected and illustrates how a researcher or decision-maker could conduct their own case study. It also explains how the system was optimised and the sensitivity of important parameters was studied.

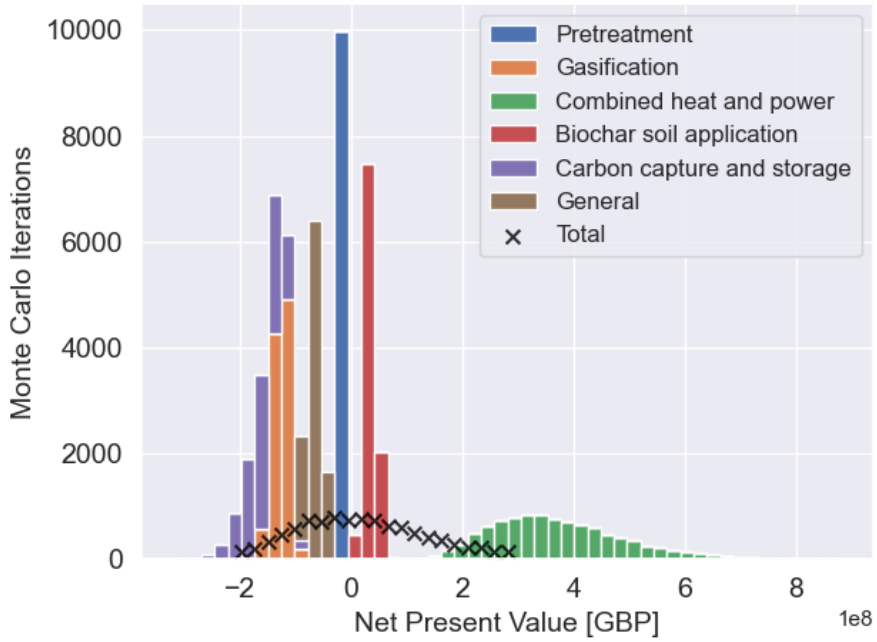
Section 6.4 shows the case study's results and discusses them. In general, the feedstock forestry residues was shown to outperform barley straw and draff. However, it was found that all feedstocks could effectively act as NETs through two main mechanisms, namely the displacement of conventional electricity and heat generation and the production of biochar. When considering the economic feasibility, it was found that whilst the baseline scenarios could operate economically, their BCR was < 1 for a large fraction of Monte Carlo iterations.

During system optimisation great improvements in both the environmental and economic performance of the system were shown. For the feedstock forestry residues around 80% of all solutions were found to achieve a mean BCR of > 1. Considering Pareto optimal solutions for forestry residues, a maximum carbon avoidance of -2,604 $kg\ CO_2$-eq/FU at a BCR of 1.149 was found for solutions considering CCS. The most economically optimal solution did not consider CCS and showed a mean GWP of -877 $kg\ CO_2$-eq/FU and a BCR of 1.840. After rerunning the model with the best parameters found during both the optimisation and sensitivity analysis stage, even further improvements were made, resulting in a solution boasting a mean GWP of -2,948 $kg\ CO_2$-eq/FU and a mean BCR of 1.101.

The case study effectively illustrates how the developed modelling framework can be used to rapidly and simply compare a range of feedstocks and gasification system choices. In the past, this process was labour- and time-intensive, making it inaccessible to many



(a)



(b)

Figure 6.22: Economic results for the baseline scenario, updated according to the optimisation and sensitivity analysis findings for the feedstock forestry residues. In (a) the results were averaged across all Monte Carlo runs, whereas (b) shows the results as a histogram based on all Monte Carlo runs. Note: For improved legibility, subprocess labels are only shown for subprocesses contributing more than 3% of the total NPV.

policy- and decision-makers. Through the use of the developed tool, the entire process has been streamlined and a much larger number of technological alternatives can be compared. The implemented optimisation routine enables users to simultaneously maximise both carbon savings and economic performance. In the future, holistic models larger than the one that has been developed here could consider further bioenergy options to ensure our biomass resources are utilised by the most suitable bioenergy choice.

Chapter 7

Conclusions, Limitations, and Future Work

7.1 Summary and Conclusions

This work set out to establish a comprehensive biomass and waste gasification modelling framework to holistically assess the environmental and economic performance of a gasification scheme. As a first step, machine learning (ML) models to predict gasification product yields and properties were developed. For this, a varied data set of 312 data points containing data concerning a range of different gasifier types, gasification conditions, and feedstocks was collected. Initially, an artificial neural network (ANN), which novelly utilised categorical predictor parameters (e.g. operation mode (batch/continuous), gasifying agent, and catalyst use) to include information omitted by previous models, was developed. The ANN model showed excellent prediction accuracy, with a coefficient of determination (R^2) of 0.86 when averaged across all model outputs.

To improve stakeholders' trust in the ML models underpinning the overarching modelling framework, the initial ANN's lack of interpretability was addressed by exploring a range of other ML algorithms and model interpretability methods. After comparing and optimising seven ML models, gradient boosting regression (GBR) was found to outperform other model types, and the previously developed ANN model, with a 5-fold cross-validated R^2 of 0.90. By using the Gini, permutation, and SHapley Additive exPlanations (SHAP) interpretability methods, additional information was extracted from the developed models. For instance, the feedstock's particle size and the gasifying agent were found to be among the top predictor variables influencing the models, highlighting their importance on the yield and properties of gasification products.

The developed ML models formed the foundation for further system analysis in the form of life cycle assessment (LCA) and techno-economic analysis (TEA). Environmental and economic models were created for all subprocesses making up the modelled gasifi-

cation scheme. By incorporating ML predictions and literature data, a robust Monte Carlo approach to account for all uncertainty within the modelling framework was developed. Accurate and transparent accounting of uncertainties within the framework was deemed essential to further improve stakeholders' trust. Finally, an exhaustive grid-search algorithm was implemented to optimise the joined objectives of optimal economic and environmental performance in terms of global warming potential (GWP) and benefit-cost ratio (BCR).

The usefulness of the developed modelling framework and decision support tool was demonstrated in the form of a case study considering three potential feedstocks for Scottish conditions, namely barley straw, forestry residues, and distillers' malt draff. In the past, comparing various feedstocks and gasification conditions was time-consuming and labour-intensive. With the modelling framework developed as part of this study, case studies can be implemented quickly by non-experts, thanks to a graphical user interface (GUI). Optimisation results are shown as a set of non-dominant solutions forming a Pareto front, clearly highlighting the potential trade-offs between objectives.

Whilst the case study found forestry residues to be the preferred feedstock, all feedstocks could aid in limiting global warming. This was done through two main mechanisms, (i) the displacement of conventional electricity and heat generation and through (ii) the application of biochar to soil. After optimisation 80% of all Monte Carlo scenarios were found to be economical with a BCR > 1 for the feedstock forestry residues. Thus, the case study effectively illustrates how the developed modelling framework could make such analysis more accessible to policy- and decision-makers.

In conclusion, this study successfully established a new high-quality data set on biomass and waste gasification product yields and properties. Utilising the data set, ML models which were more comprehensive and interpretable than previous models were developed. By combining the developed ML models and LCA and TEA, a comprehensive biomass and waste gasification modelling framework to assess and optimise the environmental and economic performance of a gasification scheme was thus developed. The developed framework can be used to further our knowledge of gasification for bioenergy and encourage its adoption at commercial and industrial scales. Furthermore, the implementation of an easy-to-use GUI can enable stakeholders to rapidly assess the performance of a range of different gasification conditions and feedstock combinations.

7.2 Limitations and Future Work

Data availability is a major limitation for any ML development. In this work, particularly industrial scale gasifier data was found to be rare, with the data set collected to train ML models in Chapters 3 and 4 being limited to lab and pilot scale data. Whilst the

interpretability analysis conducted in Chapter 4 only found minor differences between predictions for lab and pilot scale data, certain factors could differ significantly for larger systems (e.g. economics will change due to factors like economies of scale). Thus, it is important to incorporate industrial-scale data into the modelling framework as it becomes available to ensure the model's accuracy for industrial-scale gasifiers. Similarly, since most gasification studies focused on the production of syngas rather than biochar, very limited data is available on the compositions of biochar produced from gasification. As more data becomes available the compositions of the produced biochar can be predicted to greater accuracy, leading to more information being available regarding the best application of the generated biochar. This could open up new potential markets for biochar, potentially leading to increased revenues from gasification developments.

The lack of available training data also leads to limitations in which types of ML models can be implemented. For instance, deep learning techniques have shown excellent prediction accuracies in other fields, but require far larger amounts of data than available for model training here.

Another limitation of the implemented ML methods is the difficulty of accounting for the uncertainty in particular predictions made by the models. The *local error method* approach introduced in Section 5.2.2.2 represents a step towards explaining the uncertainty of a certain group of predictions, but is unable to quantify the uncertainty in individual predictions. Other model types, such as Gaussian mixture models may be capable of assessing the model's uncertainty on a prediction-by-prediction basis. However, other model types may require large amounts of data, making them less suited, in terms of prediction accuracy, than the models implemented in Chapters 3 and 4. As larger amounts of data become available the performance of other ML models may be explored.

Like many previous studies, the developed modelling framework omits social considerations. Social factors are important for engineering projects because they influence stakeholder engagement, community acceptance, and the overall success and sustainability of the project. As noted in Section 2.4, life cycle sustainability assessment (LCSA) is a suitable method to effectively integrate economic and social aspects into a system's or product's LCA. However, social considerations are highly case-specific, making it challenging to incorporate this type of analysis in a general tool like the one developed here. Although social considerations were not incorporated into the modelling framework at this stage, the environmental and techno-economic results provided by the framework can still provide useful information, forming the baseline of social analysis.

Another limitation of the developed modelling framework is that it only considers the use of syngas for energy generation via a combined heat and power (CHP) plant and the use of biochar for soil application. Whilst the use of syngas for CHP is most common, it can also be upgraded to a range of other products such as biohydrogen or sustainable

aviation fuel **[66, 380]**. Similarly, biochar is most commonly applied to soil, but it also has other potential applications. As previously mentioned, too little data on the detailed properties of biochar from gasification was available to train an ML model, however, as more data becomes available alternative biochar uses could be incorporated into the model. The modular nature of the modelling framework, founded in object-oriented programming (Section 5.2.1), makes it straightforward to extend the framework with new processes.

At the current model complexity, the implemented algorithm to optimise the modelling framework's environmental and economic performance was deemed satisfactory. However, the exhaustive nature of the optimisation algorithm limits its usefulness if the modelling framework were to be extended and its computational demand were to increase significantly as a result. In such a case, other non-exhaustive optimisation algorithms would need to be explored to ensure that the computational requirements for model optimisation would stay within satisfactory limits.

For the environmental analysis, only the impact category GWP was considered at this stage. Whilst global warming is one of the most imminent problems society has ever seen, and many existing studies also focused on this impact category only **[66, 163, 165, 166]**, more information on impact categories related to the ecosystem or human health could expand this work's scope. The utilised life cycle assessment (LCA) coding framework was specifically developed for this work, allowing for easy connection between LCA, the developed ML models, and TEA, as well as subsequent system optimisation. Commercial LCA software like *GaBi*, *SimaPro*, or *openLCA* make it more challenging to integrate various models but typically support a range of impact categories by default. As the gasification process matures and process data becomes more widely available, its subprocesses may be standardly implemented in commercial LCA software, enabling the easier implementation of additional impact categories.

This work follows the European Commission's and the International Organization for Standardization's (ISO's) recommendation for carbon neutrality of biogenic carbon **[261, 381, 382]**. This means that the GWP of biogenic carbon was considered to be zero. Even though this is the approach typically used and recommended in the LCA community, it does not always capture the short/medium-term dynamics of the biogenic carbon cycle. When considering a 100-year time horizon, the impact of the biogenic carbon cycle is irrelevant, particularly when compared to the release of fossil carbon. However, global warming is an issue that needs to be addressed in a much shorter time window than the next 100 years. Thus the carbon neutrality assumption of biogenic carbon may lead to misleading results in the short term, particularly when considering the use of slow-growing feedstocks such as virgin wood. Future models could incorporate the time-dependent nature of the biogenic carbon cycle depending on the feedstock and regrowing regime.

Data availability was also a limitation for the TEA. An attempt was made to use

the most recent cost data where possible, but sometimes only somewhat dated data was available. Methods to update older data and the cost estimation relationship (CER) method introduced in Section 5.3.2.3 were used whenever possible to not overly rely on individual data points. Again, as more up-to-date data becomes available, this data could be added to the TEA, to ensure factors such as process costs remain relevant. Similarly, the technology readiness levels (TLRs) of various subprocesses are bound to change, typically reducing their costs. Finally, as the technology matures, more detailed data on the operation and maintenance (O&M) cost will become available, allowing for the implementation of more detailed cost models. The modular nature of this work would allow for the replacement of the current O&M cost models by improved models considering factors such as country-specific wages.

References

[1] Simon Ascher, William Sloan, Ian Watson, and Siming You. "A comprehensive artificial neural network model for gasification process prediction". In: *Appl. Energy* 320 (2022), p. 119289. DOI: 10.1016/j.apenergy.2022.119289.

[2] Simon Ascher, Xiaonan Wang, Ian Watson, William Sloan, and Siming You. "Interpretable machine learning to model biomass and waste gasification". In: *Bioresour. Technol.* 364 (2022), p. 128062. DOI: 10.1016/j.biortech.2022.128062.

[3] Simon Ascher, Ian Watson, and Siming You. "Machine learning methods for modelling the gasification and pyrolysis of biomass and waste". In: *Renew. Sustain. Energy Rev.* 155 (2022), p. 111902. DOI: 10.1016/J.RSER.2021.111902.

[4] Matthew W. Jones, Glen P. Peters, Thomas Gasser, Robbie M. Andrew, Clemens Schwingshackl, Johannes Gütschow, Richard A. Houghton, Pierre Friedlingstein, Julia Pongratz, and Corinne Le Quéré. "National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850". In: *Sci. Data* 10.1 (2023), pp. 1–23. DOI: 10.1038/s41597-023-02041-1.

[5] Intergovernmental Panel on Climate Change (IPCC). "The Earth's Energy Budget, Climate Feedbacks and Climate Sensitivity". In: *Clim. Chang. 2021 – Phys. Sci. Basis Work. Gr. I Contrib. to Sixth Assess. Rep. Intergov. Panel Clim. Chang.* Ed. by Intergovernmental Panel on Climate Change (IPCC). Cambridge: Cambridge University Press, 2023, pp. 923–1054. DOI: DOI:10.1017/9781009157896.009. https://www.cambridge.org/core/product/AE57C97E588FF3060C7C7E47DD4F3C6E.

[6] IPCC. *Sixth Assessment Report (AR6)*. Tech. rep. IPCC, 2023.

[7] Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press, 2023. DOI: DOI:10.1017/9781009157896. https://www.cambridge.org/core/product/415F29233B8BD19FB55F65E3DC67272B.

[8] IPCC. *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Tech. rep. Geneva, Switzerland: Intergovernmental Panel on Climate Change, 2023. DOI: 10.59327/IPCC/AR6-9789291691647.

[9] UNFCCC. *Paris Climate Change Conference - COP 21*. 2015.

[10] UNFCCC. *Katowice Climate Package - COP 24*. 2018.

[11] Mark Poynting. *World's first year-long breach of key 1.5C warming limit*. 2024. https://www.bbc.co.uk/news/science-environment-68110310.

[12] Hannah Ritchie and Max Roser. *CO2 emissions*. 2024. https://ourworldindata.org/co2-emissions (visited on 02/08/2024).

[13] IPCC. *IPCC First Assessment Report Overview and Policymaker Summaries and 1992 IPCC Supplement*. Tech. rep. Intergovernmental Panel on Climate Change (IPCC), 1992.

[14] Felix Creutzig, Karl Heinz Erb, Helmut Haberl, Christian Hof, Carol Hunsberger, and Stephanie Roe. "Considering sustainability thresholds for BECCS in IPCC and biodiversity assessments". In: *GCB Bioenergy* 13.4 (2021), pp. 510–515. DOI: 10.1111/GCBB.12798.

[15] S. J. Mandley, V. Daioglou, H. M. Junginger, D. P. van Vuuren, and B. Wicke. "EU bioenergy development to 2050". In: *Renew. Sustain. Energy Rev.* 127.July 2019 (2020). DOI: 10.1016/j.rser.2020.109858.

[16] IEA Bioenergy. *Bioenergy*. 2023. https://www.iea.org/energy-system/renewables/bioenergy (visited on 02/08/2024).

[17] Alison Mohr and Sujatha Raman. "Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels". In: *Energy Policy* 63 (2013), pp. 114–122. DOI: 10.1016/j.enpol.2013.08.033.

[18] Sze Ying Lee, Revathy Sankaran, Kit Wayne Chew, Chung Hong Tan, Rambabu Krishnamoorthy, Dinh-Toi Chu, and Pau-Loke Show. "Waste to bioenergy: a review on the recent conversion technologies". In: *BMC Energy* 1.1 (2019), pp. 1–22. DOI: 10.1186/s42500-019-0004-7.

[19] Hossain M. Zabed, Suely Akter, Junhua Yun, Guoyan Zhang, Faisal N. Awad, Xianghui Qi, and J. N. Sahu. "Recent advances in biological pretreatment of microalgae and lignocellulosic biomass for biofuel production". In: *Renew. Sustain. Energy Rev.* 105.June 2018 (2019), pp. 105–128. DOI: 10.1016/j.rser.2019.01.048.

[20] Philipp Cavelius, Selina Engelhart-Straub, Norbert Mehlmer, Johannes Lercher, Dania Awad, and Thomas Brück. "The potential of biofuels from first to fourth generation". In: *PLOS Biol.* 21.3 (2023). DOI: 10.1371/JOURNAL.PBIO.3002063.

[21] Pritam Bardhan, Anuron Deka, Satya S. Bhattacharya, Manabendra Mandal, and Rupam Kataki. "Economical aspect in biomass to biofuel production". In: *Value-Chain Biofuels Fundam. Technol. Stand.* INC, 2021. Chap. 18, pp. 395–427. DOI: 10.1016/B978-0-12-824388-6.00003-8. http://dx.doi.org/10.1016/B978-0-12-824388-6.00003-8.

[22] Ottmar Edenhofer, Ramón Pichs Madruga, Youba Sokona, Kristin Seyboth, Patrick Matschoss, Susanne Kadner, Timm Zwickel, Patrick Eickemeier, Gerrit Hansen, Steffen Schlömer, and Christoph von Stechow. *Renewable energy sources and climate change mitigation: Special report of the intergovernmental panel on climate change*. Tech. rep. IPCC, 2011, pp. 1–1075. DOI: 10.1017/CBO9781139151153.

[23] Sonil Nanda, Javeed Mohammad, Sivamohan N. Reddy, Janusz A. Kozinski, and Ajay K. Dalai. "Pathways of lignocellulosic biomass conversion to renewable fuels". In: *Biomass Convers. Biorefinery* 4.2 (2014), pp. 157–191. DOI: 10.1007/s13399-013-0097-z.

[24] Prabir Basu. *Biomass gasification, pyrolysis and torrefaction: Practical design and theory*. 2nd. Academic Press, 2018, pp. 1–564. DOI: 10.1016/C2016-0-04056-1.

[25] Siming You, Huanhuan Tong, Joel Armin-Hoiland, Yen Wah Tong, and Chi Hwa Wang. "Techno-economic and greenhouse gas savings assessment of decentralized biomass gasification for electrifying the rural areas of Indonesia". In: *Appl. Energy* 208.June (2017), pp. 495–510. DOI: 10.1016/j.apenergy.2017.10.001.

[26] Luca Campion, Madina Bekchanova, Robert Malina, and Tom Kuppens. "The costs and benefits of biochar production and use: A systematic review". In: *J. Clean. Prod.* 408.March (2023), p. 137138. DOI: 10.1016/j.jclepro.2023.137138.

[27] Simone Marzeddu, Andrea Cappelli, Andrea Ambrosio, María Alejandra Décima, Paolo Viotti, and Maria Rosaria Boni. "A life cycle assessment of an energy-biochar chain involving a gasification plant in Italy". In: *Land* 10 (2021). DOI: 10.3390/land10111256.

[28] Sabine Fuss, William F. Lamb, Max W. Callaghan, Jérôme Hilaire, Felix Creutzig, Thorben Amann, Tim Beringer, Wagner De Oliveira Garcia, Jens Hartmann, Tarun Khanna, Gunnar Luderer, Gregory F. Nemet, Joeri Rogelj, Pete Smith, Joséluis Vicente Vicente, Jennifer Wilcox, Maria Del Mar Zamora Dominguez, and Jan C. Minx. "Negative emissions - Part 2: Costs, potentials and side effects". In: *Environ. Res. Lett.* 13.6 (2018), pp. 2016–2050. DOI: 10.1088/1748-9326/aabf9f.

[29] Felix Creutzig, N. H. Ravindranath, Göran Berndes, Simon Bolwig, Ryan Bright, Francesco Cherubini, Helena Chum, Esteve Corbera, Mark Delucchi, Andre Faaij, Joseph Fargione, Helmut Haberl, Garvin Heath, Oswaldo Lucon, Richard Plevin,

Alexander Popp, Carmenza Robledo-Abad, Steven Rose, Pete Smith, Anders Stromman, Sangwon Suh, and Omar Masera. "Bioenergy and climate change mitigation: An assessment". In: *GCB Bioenergy* 7.5 (2015), pp. 916–944. DOI: 10.1111/GCBB.12205.

[30] Carmenza Robledo-Abad, Hans Jörg Althaus, Göran Berndes, Simon Bolwig, Esteve Corbera, Felix Creutzig, John Garcia-Ulloa, Anna Geddes, Jay S. Gregg, Helmut Haberl, Susanne Hanger, Richard J. Harper, Carol Hunsberger, Rasmus K. Larsen, Christian Lauk, Stefan Leitner, Johan Lilliestam, Hermann Lotze-Campen, Bart Muys, Maria Nordborg, Maria Ölund, Boris Orlowsky, Alexander Popp, Joana Portugal-Pereira, Jürgen Reinhard, Lena Scheiffle, and Pete Smith. "Bioenergy production and sustainable development: science base for policymaking remains limited". In: *GCB Bioenergy* 9.3 (2017), pp. 541–556. DOI: 10.1111/GCBB.12338.

[31] Claudia Kammann, Jim Ippolito, Nikolas Hagemann, Nils Borchard, Maria Luz Cayuela, José M. Estavillo, Teresa Fuertes-Mendizabal, Simon Jeffery, Jürgen Kern, Jeff Novak, Daniel Rasse, Sanna Saarnio, Hans Peter Schmidt, Kurt Spokas, and Nicole Wrage-Mönnig. "Biochar as a tool to reduce the agricultural greenhouse-gas burden – knowns, unknowns and future research needs". In: *J. Environ. Eng. Landsc. Manag.* 25.2 (2017), pp. 114–139. DOI: 10.3846/16486897.2017.1319375.

[32] Chris Bamminger, Christian Poll, Christina Sixt, Petra Högy, Dominik Wüst, Ellen Kandeler, and Sven Marhan. "Short-term response of soil microorganisms to biochar addition in a temperate agroecosystem under soil warming". In: *Agric. Ecosyst. Environ.* 233 (2016), pp. 308–317. DOI: 10.1016/J.AGEE.2016.09.016.

[33] Arvind Singh Bisht and N. S. Thakur. "Small scale biomass gasification plants for electricity generation in India: Resources, installation, technical aspects, sustainability criteria & policy". In: *Renew. Energy Focus* 28.March (2019), pp. 112–126. DOI: 10.1016/j.ref.2018.12.004.

[34] IEA Bioenergy. *India Energy Outlook 2021*. 2021. https://www.iea.org/reports/india-energy-outlook-2021 (visited on 02/09/2024).

[35] Sunil L Narnaware and N L Panwar. "Biomass gasification for climate change mitigation and policy framework in India: A review". In: *Bioresour. Technol. Reports* 17 (2022), p. 100892. DOI: 10.1016/j.biteb.2021.100892.

[36] M I Jordan and T M Mitchell. "Machine learning: Trends, perspectives, and prospects". In: *Science (80-.).* 349.6245 (2015), pp. 255–260.

[37] Christoph Molnar. *Interpretable Machine Learning - A Guide for Making Black Box Models Explainable*. 2019. https://christophm.github.io/interpretable-ml-book/.

[38] Finale Doshi-Velez and Been Kim. *Towards A Rigorous Science of Interpretable Machine Learning*. 2017. arXiv: 1702.08608. http://arxiv.org/abs/1702.08608.

[39] Somayeh Farzad, Mohsen Ali Mandegari, and Johann F. Görgens. "A critical review on biomass gasification, co-gasification, and their environmental assessments". In: *Biofuel Res. J.* 3.4 (2016), pp. 483–495. DOI: 10.18331/BRJ2016.3.4.3.

[40] Özgün Tezer, Nazlıcan Karabağ, Atakan Öngen, Can Özgür Çolpan, and Azize Ayol. "Biomass gasification for sustainable energy production: A review". In: *Int. J. Hydrogen Energy* 47.34 (2022), pp. 15419–15433. DOI: 10.1016/j.ijhydene.2022.02.158.

[41] Zahir Barahmand and Marianne S. Eikeland. "A Scoping Review on Environmental, Economic, and Social Impacts of the Gasification Processes". In: *Environments* 9.7 (2022). DOI: 10.3390/environments9070092.

[42] Walter V. Reid, Mariam K. Ali, and Christopher B. Field. "The future of bioenergy". In: *Glob. Chang. Biol.* 26.1 (2020), pp. 274–286. DOI: 10.1111/gcb.14883.

[43] Arjay A. Arpia, Wei Hsin Chen, Su Shiung Lam, Patrick Rousset, and Mark Daniel G. de Luna. "Sustainable biofuel and bioenergy production from biomass waste residues using microwave-assisted heating: A comprehensive review". In: *Chem. Eng. J.* 403.April 2020 (2021), p. 126233. DOI: 10.1016/j.cej.2020.126233.

[44] Prangan Duarah, Dibyajyoti Haldar, Anil Kumar Patel, Cheng Di Dong, Reeta Rani Singhania, and Mihir Kumar Purkait. "A review on global perspectives of sustainable development in bioenergy generation". In: *Bioresour. Technol.* 348.January (2022). DOI: 10.1016/j.biortech.2022.126791.

[45] Mathias Fridahl and Mariliis Lehtveer. "Bioenergy with carbon capture and storage (BECCS): Global potential, investment preferences, and deployment barriers". In: *Energy Res. Soc. Sci.* 42.October 2017 (2018), pp. 155–165. DOI: 10.1016/j.erss.2018.03.019.

[46] S. V. Hanssen, V. Daioglou, Z. J.N. Steinmann, J. C. Doelman, D. P. Van Vuuren, and M. A.J. Huijbregts. "The climate change mitigation potential of bioenergy with carbon capture and storage". In: *Nat. Clim. Chang.* 10.11 (2020), pp. 1023–1029. DOI: 10.1038/s41558-020-0885-y.

[47] Tabbi Wilberforce, A. G. Olabi, Enas Taha Sayed, Khaled Elsaid, and Mohammad Ali Abdelkareem. "Progress in carbon capture technologies". In: *Sci. Total Environ.* 761 (2021), p. 143203. DOI: 10.1016/j.scitotenv.2020.143203.

[48] Paul Adams, Tony Bridgwater, Amanda Lea-Langton, Andrew Ross, and Ian Watson. "Biomass Conversion Technologies". In: *Greenh. Gas Balanc. Bioenergy Syst.* Elsevier Inc., 2017. Chap. 8, pp. 107–139. DOI: 10.1016/B978-0-08-101036-5.00008-2.

[49] Raffaella Petruccelli and Sara Di Lonardo. "Role of biochars in soil fertility management of fruit crops". In: *Fruit Crop. Diagnosis Manag. Nutr. Constraints*. 2019. Chap. 31, pp. 431–444. DOI: 10.1016/B978-0-12-818732-6.00031-9.

[50] "Biochar in agriculture for achieving sustainable development goals". In: ed. by Daniel CW Tsang and Yong Sik Ok. Academic Press, 2022.

[51] C. E. Brewer and R. C. Brown. "Biochar". In: *Compr. Renew. Energy*. Vol. 5. 2012. Chap. 5.18, pp. 357–384. DOI: 10.1016/B978-0-08-087872-0.00524-2.

[52] Kezhen Qian, Ajay Kumar, Hailin Zhang, Danielle Bellmer, and Raymond Huhnke. "Recent advances in utilization of biochar". In: *Renew. Sustain. Energy Rev.* 42 (2015), pp. 1055–1064. DOI: 10.1016/j.rser.2014.10.074.

[53] Andrew Cross and Saran P. Sohi. "A method for screening the relative long-term stability of biochar". In: *GCB Bioenergy* 5.2 (2013), pp. 215–220. DOI: 10.1111/gcbb.12035.

[54] Jan Matuštík, Michael Pohořelý, and Vladimír Kočí. "Is application of biochar to soil really carbon negative? The effect of methodological decisions in Life Cycle Assessment". In: *Sci. Total Environ.* 807 (2022). DOI: 10.1016/j.scitotenv.2021.151058.

[55] Siming You, Yong Sik Ok, Season S. Chen, Daniel C.W. Tsang, Eilhann E. Kwon, Jechan Lee, and Chi Hwa Wang. "A critical review on sustainable biochar system through gasification: Energy and environmental applications". In: *Bioresour. Technol.* 246 (2017), pp. 242–253. DOI: 10.1016/j.biortech.2017.06.177.

[56] Catherine E. Brewer, Klaus Schmidt-Rohr, Justinus A. Satrio, and Robert C. Brown. "Characterization of Biochar from Fast Pyrolysis and Gasification Systems". In: *Environ. Prog. Sustain. Energy* 28.3 (2009), pp. 386–396. DOI: 10.1002/ep.

[57] Siming You and Xiaonan Wang. "On the carbon abatement potential and economic viability of biochar production systems: Cost-benefit and life cycle assessment". In: *Biochar from Biomass Waste Fundam. Appl.* Elsevier Inc., 2018. Chap. 20, pp. 385–408. DOI: 10.1016/B978-0-12-811729-3.00020-0. http://dx.doi.org/10.1016/B978-0-12-811729-3.00020-0.

[58] Veronika Hansen, Dorette Müller-Stöver, Jesper Ahrenfeldt, Jens Kai Holm, Ulrik Birk Henriksen, and Henrik Hauggaard-Nielsen. "Gasification biochar as a valuable by-product for carbon sequestration and soil amendment". In: *Biomass and Bioenergy* 72.1 (2015), pp. 300–308. DOI: 10.1016/j.biombioe.2014.10.013.

[59] Jianfen Li, Rong Yan, Bo Xiao, Xiaoling Wang, and Haiping Yang. "Influence of temperature on the formation of oil from pyrolyzing palm oil wastes in a fixed bed reactor". In: *Energy and Fuels* 21 (2007), pp. 2398–2407. DOI: 10.1021/ef060548c.

[60] Tao Kan, Vladimir Strezov, and Tim J. Evans. "Lignocellulosic biomass pyrolysis: A review of product properties and effects of pyrolysis parameters". In: *Renew. Sustain. Energy Rev.* 57 (2016), pp. 1126–1140. DOI: 10.1016/j.rser.2015.12.185.

[61] A. V. Bridgwater. "Review of fast pyrolysis of biomass and product upgrading". In: *Biomass and Bioenergy* 38 (2012), pp. 68–94. DOI: 10.1016/j.biombioe.2011.01.048.

[62] M. N. Uddin, Kuaanan Techato, Juntakan Taweekun, Md Mofijur Rahman, M. G. Rasul, T. M.I. Mahlia, and S. M. Ashrafur. "An overview of recent developments in biomass pyrolysis technologies". In: *Energies* 11 (2018), p. 3115. DOI: 10.3390/en11113115.

[63] Shin Ying Foong, Rock Keey Liew, Yafeng Yang, Yoke Wang Cheng, Peter Nai Yuh Yek, Wan Adibah Wan Mahari, Xie Yi Lee, Chai Sean Han, Dai Viet N. Vo, Quyet Van Le, Mortaza Aghbashlo, Meisam Tabatabaei, Christian Sonne, Wanxi Peng, and Su Shiung Lam. "Valorization of biomass waste to engineered activated biochar by microwave pyrolysis: Progress, challenges, and future directions". In: *Chem. Eng. J.* 389 (2020), p. 124401. DOI: 10.1016/j.cej.2020.124401.

[64] Szymon Sobek and Sebastian Werle. "Solar pyrolysis of waste biomass: Part 1 reactor design". In: *Renew. Energy* 143 (2019), pp. 1939–1948. DOI: 10.1016/j.renene.2019.06.011.

[65] Wan Adibah Wan Mahari, Elfina Azwar, Shin Ying Foong, Ashfaq Ahmed, Wanxi Peng, Meisam Tabatabaei, Mortaza Aghbashlo, Young Kwon Park, Christian Sonne, and Su Shiung Lam. "Valorization of municipal wastes using co-pyrolysis for green energy production, energy security, and environmental sustainability: A review". In: *Chem. Eng. J.* 421 (2021), p. 129749. DOI: 10.1016/J.CEJ.2021.129749.

[66] J. M. Bermudez and B. Fidalgo. "Production of bio-syngas and bio-hydrogen via gasification". In: *Handb. Biofuels Prod. Process. Technol. Second Ed.* 2016. Chap. 15, pp. 431–494. DOI: 10.1016/B978-0-08-100455-5.00015-1.

[67] Richard Thomson, Philip Kwong, Ejaz Ahmad, and K. D.P. Nigam. "Clean syngas from small commercial biomass gasifiers; a review of gasifier development, recent advances and performance evaluation". In: *Int. J. Hydrogen Energy* 45.41 (2020), pp. 21087–21111. DOI: 10.1016/j.ijhydene.2020.05.160.

[68] Antonio Molino, Vincenzo Larocca, Simeone Chianese, and Dino Musmarra. "Biofuels Production by Biomass Gasification: A Review". In: *Energies* 11.4 (2018), p. 811. DOI: 10.3390/en11040811.

[69] Leichang Cao, Iris K.M. Yu, Xinni Xiong, Daniel C.W. Tsang, Shicheng Zhang, James H. Clark, Changwei Hu, Yun Hau Ng, Jin Shang, and Yong Sik Ok. "Biorenewable hydrogen production through biomass gasification: A review and future prospects". In: *Environ. Res.* 186 (2020), p. 109547. DOI: 10.1016/j.envres.2020.109547.

[70] E. Shayan, V. Zare, and I. Mirzaee. "Hydrogen production from biomass gasification; a theoretical comparison of using different gasification agents". In: *Energy Convers. Manag.* 159.December 2017 (2018), pp. 30–41. DOI: 10.1016/j.enconman.2017.12.096.

[71] Jie Ren, Jing Pei Cao, Xiao Yan Zhao, Fei Long Yang, and Xian Yong Wei. "Recent advances in syngas production from biomass catalytic gasification: A critical review on reactors, catalysts, catalytic mechanisms and mathematical models". In: *Renew. Sustain. Energy Rev.* 116 (2019), p. 109426. DOI: 10.1016/j.rser.2019.109426.

[72] Martha Lucia Valderrama Rios, Aldemar Martínez González, Electo Eduardo Silva Lora, and Oscar Agustin Almazán del Olmo. "Reduction of tar generated during biomass gasification: A review". In: *Biomass and Bioenergy* 108 (2018), pp. 345–370. DOI: 10.1016/j.biombioe.2017.12.002.

[73] Md Mosleh Uddin and Mark Mba Wright. "Anaerobic digestion fundamentals, challenges, and technological advances". In: *Phys. Sci. Rev.* 8.9 (2023), pp. 2819–2837. DOI: 10.1515/psr-2021-0068.

[74] Roger C. Prince and Haroon S. Kheshgi. "The photobiological production of hydrogen: Potential efficiency and effectiveness as a renewable fuel". In: *Crit. Rev. Microbiol.* 31.1 (2005), pp. 19–31. DOI: 10.1080/10408410590912961.

[75] Alastair J. Ward, Phil J. Hobbs, Peter J. Holliman, and David L. Jones. "Optimisation of the anaerobic digestion of agricultural resources". In: *Bioresour. Technol.* 99.17 (2008), pp. 7928–7940. DOI: 10.1016/j.biortech.2008.02.044.

[76] M. Rasapoor, B. Young, R. Brar, A. Sarmah, W. Q. Zhuang, and S. Baroutian. "Recognizing the challenges of anaerobic digestion: Critical steps toward improving biogas generation". In: *Fuel* 261.July 2019 (2020), p. 116497. DOI: 10.1016/j.fuel.2019.116497.

[77] Ianny Andrade Cruz, Wachiranon Chuenchart, Fei Long, K. C. Surendra, Larissa Renata Santos Andrade, Muhammad Bilal, Hong Liu, Renan Tavares Figueiredo, Samir Kumar Khanal, and Luiz Fernando Romanholo Ferreira. "Application of machine learning in anaerobic digestion: Perspectives and challenges". In: *Bioresour. Technol.* 345.September 2021 (2022), p. 126433. DOI: 10.1016/j.biortech.2021.126433.

[78] Jade Lui, Wei Hsin Chen, Daniel C.W. Tsang, and Siming You. "A critical review on the principles, applications, and challenges of waste-to-hydrogen technologies". In: *Renew. Sustain. Energy Rev.* 134 (2020). DOI: 10.1016/j.rser.2020.110365.

[79] A V Bridgwater, A J Toft, and J G Brammer. "A techno-economic comparison of power production by biomass fast pyrolysis with gasification and combustion". In: *Renew. Sustain. Energy Rev.* 6 (2002), pp. 181–248.

[80] Timothy D. Searchinger. "Biofuels and the need for additional carbon". In: *Environ. Res. Lett.* 5 (2010), p. 24007. DOI: 10.1088/1748-9326/5/2/024007.

[81] John A. Mathews. "Carbon-negative biofuels". In: *Energy Policy* 36.3 (2008), pp. 940–945. DOI: 10.1016/j.enpol.2007.11.029.

[82] Joseph Fargione, Jason Hill, David Tilman, Stephen Polasky, and Peter Hawthorne. "Land clearing and the biofuel carbon debt". In: *Science (80-.).* 319 (2008), pp. 1235–1238. DOI: 10.1126/science.1152747.

[83] Johannes Lehmann. "A handful of carbon". In: *Nature* 447.7141 (2007), pp. 143–144. DOI: 10.1038/447143a.

[84] Kelli G. Roberts, Brent A. Gloy, Stephen Joseph, Norman R. Scott, and Johannes Lehmann. "Life cycle assessment of biochar systems: Estimating the energetic, economic, and climate change potential". In: *Environ. Sci. Technol.* 44 (2010), pp. 827–833. DOI: 10.1021/es902266r.

[85] Darko Matovic. "Biochar as a viable carbon sequestration option: Global and Canadian perspective". In: *Energy* 36.4 (2011), pp. 2011–2016. DOI: 10.1016/j.energy.2010.09.031.

[86] Gary C. Young. "Introduction to Gasification/Pyrolysis and Combustion Technology(s)". In: *Munic. Solid Waste to Energy Convers. Process. Econ. Tech. Renew. Comp.* Wiley, 2010. Chap. 1.

[87] L. Fiori and L. Florio. "Gasification and combustion of grape marc: Comparison among different scenarios". In: *Waste and Biomass Valorization* 1.2 (2010), pp. 191–200. DOI: 10.1007/s12649-010-9025-7.

[88] Jun Dong, Yuanjun Tang, Ange Nzihou, Yong Chi, Elsa Weiss-Hortala, and Mingjiang Ni. "Life cycle assessment of pyrolysis, gasification and incineration waste-to-energy technologies: Theoretical analysis and case study of commercial plants". In: *Sci. Total Environ.* 626 (2018), pp. 744–753. DOI: 10.1016/j.scitotenv.2018.01.151.

[89] Umberto Arena, Filomena Ardolino, and Fabrizio Di Gregorio. "A life cycle assessment of environmental performances of two combustion- and gasification-based waste-to-energy technologies". In: *Waste Manag.* 41 (2015), pp. 60–74. DOI: 10.1016/j.wasman.2015.03.041.

[90] Clarivate Analytics. *Web of Science*. 2021. https://apps.webofknowledge.com/UA%7B%5C_%7DGeneralSearch%7B%5C_%7Dinput.do?product=UA%7B%5C&%7Dsearch%7B%5C_%7Dmode=GeneralSearch%7B%5C&%7DSID=F5PfPgYdxUvvmDBoMwA%7B%5C&%7DpreferencesSaved= (visited on 03/17/2021).

[91] Tapas Kumar Patra and Pratik N. Sheth. *Biomass gasification models for downdraft gasifier: A state-of-the-art review*. 2015. DOI: 10.1016/j.rser.2015.05.012.

[92] Sahar Safarian, Rúnar Unnþórsson, and Christiaan Richter. "A review of biomass gasification modelling". In: *Renew. Sustain. Energy Rev.* 110 (2019), pp. 378–391. DOI: 10.1016/j.rser.2019.05.003.

[93] Maria Puig-Arnavat and Joan Carles Bruno. "Artificial Neural Networks for Thermochemical Conversion of Biomass". In: *Recent Adv. Thermochem. Convers. Biomass*. 2015. Chap. 5, pp. 133–156. DOI: 10.1016/B978-0-444-63289-0.00005-3.

[94] Furkan Elmaz, Özgün Yücel, and Ali Yener Mutlu. "Predictive modeling of biomass gasification with machine learning-based regression methods". In: *Energy* 191 (2020), p. 116541. DOI: 10.1016/j.energy.2019.116541.

[95] Maria Puig-Arnavat, Joan Carles Bruno, and Alberto Coronas. "Review and analysis of biomass gasification models". In: *Renew. Sustain. Energy Rev.* 14.9 (2010), pp. 2841–2851. DOI: 10.1016/j.rser.2010.07.030.

[96] Dipal Baruah and D. C. Baruah. "Modeling of biomass gasification: A review". In: *Renew. Sustain. Energy Rev.* 39 (2014), pp. 806–815. DOI: 10.1016/j.rser.2014.07.129.

[97] Ana Ramos, Eliseu Monteiro, and Abel Rouboa. "Numerical approaches and comprehensive models for gasification process: A review". In: *Renew. Sustain. Energy Rev.* 110 (2019), pp. 188–206. DOI: 10.1016/j.rser.2019.04.048.

[98] C. C. Sreejith, C. Muraleedharan, and P. Arun. "Performance prediction of fluidised bed gasification of biomass using experimental data-based simulation models". In: *Biomass Convers. Biorefinery* 3 (2013), pp. 283–304. DOI: 10.1007/s13399-013-0083-5.

[99] Pavlo Kostetskyy and Linda J Broadbelt. "Progress in Modeling of Biomass Fast Pyrolysis: A Review". In: *Energy & Fuels* 34 (2020), pp. 15195–15216. DOI: 10.1021/acs.energyfuels.0c02295.

[100] Samreen Hameed, Abhishek Sharma, Vishnu Pareek, Hongwei Wu, and Yun Yu. "A review on biomass pyrolysis models: Kinetic, network and mechanistic models". In: *Biomass and Bioenergy* 123 (2019), pp. 104–122. DOI: 10.1016/j.biombioe.2019.02.008.

[101] Zuzanna Kaczor, Zbigniew Buliński, and Sebastian Werle. "Modelling approaches to waste biomass pyrolysis: a review". In: *Renew. Energy* 159 (2020), pp. 427–443. DOI: 10.1016/j.renene.2020.05.110.

[102] Abhishek Sharma, Vishnu Pareek, and Dongke Zhang. "Biomass pyrolysis - A review of modelling, process parameters and catalytic studies". In: *Renew. Sustain. Energy Rev.* 50 (2015), pp. 1081–1096. DOI: 10.1016/j.rser.2015.04.193.

[103] B. V. Babu. *Biomass pyrolysis: A state-of-the-art review*. 2008. DOI: 10.1002/bbb.92.

[104] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Second. Springer Series in Statistics. New York, NY: Springer New York, 2009. DOI: 10.1007/978-0-387-84858-7. http://link.springer.com/10.1007/978-0-387-84858-7.

[105] Jürgen Schmidhuber. "Deep Learning in neural networks: An overview". In: *Neural Networks* 61 (2015), pp. 85–117. DOI: 10.1016/j.neunet.2014.09.003. arXiv: 1404.7828.

[106] Michael A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015. http://neuralnetworksanddeeplearning.com/index.html.

[107] Daniel Serrano and David Castelló. "Tar prediction in bubbling fluidized bed gasification through artificial neural networks". In: *Chem. Eng. J.* 402 (2020), p. 126229. DOI: 10.1016/j.cej.2020.126229.

[108] Daya Shankar Pandey, Saptarshi Das, Indranil Pan, James J. Leahy, and Witold Kwapinski. "Artificial neural network based modelling approach for municipal solid waste gasification in a fluidized bed reactor". In: *Waste Manag.* 58 (2016), pp. 202–213. DOI: 10.1016/j.wasman.2016.08.023.

[109] Jie Li, Lanjia Pan, Manu Suvarna, Yen Wah Tong, and Xiaonan Wang. "Fuel properties of hydrochar and pyrochar: Prediction and exploration with machine learning". In: *Appl. Energy* 269 (2020), p. 115116. DOI: 10.1016/j.apenergy.2020.115166.

[110] Julian D. Olden, Michael K. Joy, and Russell G. Death. "An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data". In: *Ecol. Modell.* 178.3-4 (2004), pp. 389–397. DOI: 10.1016/j.ecolmodel.2004.03.013.

[111] Leo Breiman. "Random Forests". In: *Mach. Learn.* 45 (2001), pp. 5–32.

[112] Harris Drucker, Chris J C Burges, Linda Kaufman, Alex Smola, and Vladimir Vapoik. "Support Vector Regression Machines". In: *Adv. Neural Inf. Process. Syst.* 9 (1996), pp. 155–161.

[113] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer New York, 2000. DOI: 10.1007/978-1-4757-3264-1.

[114] R. Noori, M. A. Abdoli, A. Ameri Ghasrodashti, and M. Jalili Ghazizade. "Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: A case study of mashhad". In: *Environ. Prog. Sustain. Energy* 28.2 (2009), pp. 249–258. DOI: 10.1002/ep.10317.

[115] Dipal Baruah, D. C. Baruah, and M. K. Hazarika. "Artificial neural network based modeling of biomass gasification in fixed bed downdraft gasifiers". In: *Biomass and Bioenergy* 98 (2017), pp. 264–271. DOI: 10.1016/j.biombioe.2017.01.029.

[116] Joel George, P. Arun, and C. Muraleedharan. "Assessment of producer gas composition in air gasification of biomass using artificial neural network model". In: *Int. J. Hydrogen Energy* 43 (2018), pp. 9558–9568. DOI: 10.1016/j.ijhydene.2018.04.007.

[117] Ali Yener Mutlu and Ozgun Yucel. "An artificial intelligence based approach to predicting syngas composition for downdraft biomass gasification". In: *Energy* 165 (2018), pp. 895–901. DOI: 10.1016/j.energy.2018.09.131.

[118] Ozgun Yucel, Ebubekir Siddik Aydin, and Hasan Sadikoglu. "Comparison of the different artificial neural networks in prediction of biomass gasification products". In: *Int. J. Energy Res.* 43.11 (2019), pp. 5992–6003. DOI: 10.1002/er.4682.

[119] Furkan Elmaz and Özgün Yücel. "Data-driven identification and model predictive control of biomass gasification process for maximum energy production". In: *Energy* 195 (2020), p. 117037. DOI: 10.1016/j.energy.2020.117037.

[120] Daniel Serrano, Iman Golpour, and Sergio Sánchez-Delgado. "Predicting the effect of bed materials in bubbling fluidized bed gasification using artificial neural networks (ANNs) modeling approach". In: *Fuel* 266 (2020), p. 117021. DOI: 10.1016/j.fuel.2020.117021.

[121] Sahar Safarian, Seyed Mohammad, Ebrahimi Saryazdi, Runar Unnthorsson, and Christiaan Richter. "Artificial neural network integrated with thermodynamic equilibrium modeling of downdraft biomass gasification-power production plant". In: *Energy* 213 (2020), p. 118800. DOI: 10.1016/j.energy.2020.118800.

[122] Bamidele Victor Ayodele, Siti Indati Mustapa, Ramesh Kanthasamy, Mohammed Zwawi, and Chin Kui Cheng. "Modeling the prediction of hydrogen production by co-gasification of plastic and rubber wastes using machine learning algorithms". In: *Int. J. Energy Res.* 45 (2021), pp. 9580–9594. DOI: 10.1002/er.6483.

[123] Shribalaji Shenbagaraj, Pankaj Kumar Sharma, Amit Kumar Sharma, Geetanjali Raghav, Karthikeya Bharadwaj Kota, and Ashokumar Veeramuthu. "Gasification of food waste in supercritical water: An innovative synthesis gas composition prediction model based on Artificial Neural Networks". In: *Int. J. Hydrogen Energy* 46.24 (2021), pp. 12739–12757. DOI: 10.1016/j.ijhydene.2021.01.122.

[124] Navid Kardani, Annan Zhou, Majidreza Nazem, and Xiaoshan Lin. "Modelling of municipal solid waste gasification using an optimised ensemble soft computing model". In: *Fuel* 289 (2021), p. 119903. DOI: 10.1016/j.fuel.2020.119903.

[125] Hongliang Cao, Ya Xin, and Qiaoxia Yuan. "Prediction of biochar yield from cattle manure pyrolysis via least squares support vector machine intelligent approach". In: *Bioresour. Technol.* 202 (2016), pp. 158–164. DOI: 10.1016/j.biortech.2015.12.024.

[126] Sasithorn Sunphorka, Benjapon Chalermsinsuwan, and Pornpote Piumsomboon. "Artificial neural network model for the prediction of kinetic parameters of biomass pyrolysis from its constituents". In: *Fuel* 193 (2017), pp. 142–158. DOI: 10.1016/j.fuel.2016.12.046.

[127] Xing Chen, Huiyan Zhang, Yao Song, and Rui Xiao. "Prediction of product distribution and bio-oil heating value of biomass fast pyrolysis". In: *Chem. Eng. Process. - Process Intensif.* 130 (2018), pp. 36–42. DOI: 10.1016/j.cep.2018.05.018.

[128] Özge Çepelioğullar, İlhan Mutlu, Serdar Yaman, and Hanzade Haykiri-Acma. "Activation energy prediction of biomass wastes based on different neural network topologies". In: *Fuel* 220 (May 2018), pp. 535–545. DOI: 10.1016/j.fuel.2018.02.045.

[129] Mortaza Aghbashlo, Meisam Tabatabaei, Mohammad Hossein Nadian, Vandad Davoodnia, and Salman Soltanian. "Prognostication of lignocellulosic biomass pyrolysis behavior using ANFIS model tuned by PSO algorithm". In: *Fuel* 253 (Oct. 2019), pp. 189–198. DOI: 10.1016/j.fuel.2019.04.169.

[130] Xinzhe Zhu, Yinan Li, and Xiaonan Wang. "Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions". In: *Bioresour. Technol.* 288 (Sept. 2019), p. 121527. DOI: 10.1016/j.biortech.2019.121527.

[131] Qinghui Tang, Yingquan Chen, Haiping Yang, Ming Liu, Haoyu Xiao, Ziyue Wu, Hanping Chen, and Salman Raza Naqvi. "Prediction of Bio-oil Yield and Hydrogen Contents Based on Machine Learning Method: Effect of Biomass Compositions and Pyrolysis Conditions". In: *Energy & Fuels* 34.9 (Sept. 2020), pp. 11050–11060. DOI: 10.1021/acs.energyfuels.0c01893.

[132] Fangwei Cheng, Hongxi Luo, and Lisa M. Colosi. "Slow pyrolysis as a platform for negative emissions technology: An integration of machine learning models, life cycle assessment, and economic analysis". In: *Energy Convers. Manag.* 223 (Nov. 2020), p. 113258. DOI: 10.1016/j.enconman.2020.113258.

[133] Chunhan Gu, Xiaohan Wang, Qianshi Song, Haowen Li, and Yu Qiao. "Prediction of gas-liquid-solid product distribution after solid waste pyrolysis process based on artificial neural network model". In: *Int. J. Energy Res.* (2021), pp. 1–15. DOI: 10.1002/er.6707.

[134] Maria Puig-Arnavat, J. Alfredo Hernández, Joan Carles Bruno, and Alberto Coronas. "Artificial neural network models for biomass gasification in fluidized bed gasifiers". In: *Biomass and Bioenergy* 49 (2013), pp. 279–289. DOI: 10.1016/j.biombioe.2012.12.012.

[135] Robert Mikulandrić, Dražen Lončar, Dorith Böhning, Rene Böhme, and Michael Beckmann. "Artificial neural network modelling approach for a biomass gasification process in fixed bed gasifiers". In: *Energy Convers. Manag.* 87 (2014), pp. 1210–1223. DOI: 10.1016/j.enconman.2014.03.036.

[136] Kedar Potdar, Taher S. Pardawala, and Chinmay Pai. "A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers". In: *Int. J. Comput. Appl.* 175.4 (2017), pp. 975–8887. DOI: 10.5120/ijca2017915495.

[137] Musa Buyukada. "Co-combustion of peanut hull and coal blends: Artificial neural networks modeling, particle swarm optimization and Monte Carlo simulation". In: *Bioresour. Technol.* 216 (Sept. 2016), pp. 280–286. DOI: 10.1016/j.biortech.2016.05.091.

[138] Olatunji Obafemi, Akinlabi Stephen, Oluseyi Ajayi, and Madushele Nkosinathi. "A survey of artificial neural network-based prediction models for thermal properties of biomass". In: *Procedia Manuf.* Vol. 33. Elsevier B.V., 2019, pp. 184–191. DOI: 10.1016/j.promfg.2019.04.103.

[139] G David Garson. "A Comparison of Neural Network and Expert Systems Algorithms with Common Multivariate Procedures for Analysis of Social Science Data". In: *Soc. Sci. Comput. Rev.* 9.3 (1991), pp. 399–434.

[140] Muriel Gevrey, Ioannis Dimopoulos, and Sovan Lek. "Review and comparison of methods to study the contribution of variables in artificial neural network models". In: *Ecol. Modell.* 160 (2003), pp. 249–264. DOI: 10.1016/S0304-3800(02)00257-0.

[141] Julian D Olden and Donald A Jackson. "Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks". In: *Ecol. Modell.* 154 (2002), pp. 135–150.

[142] Bing Guo, Dingkai Li, Congming Cheng, Zi An Lü, and Youting Shen. "Simulation of biomass gasification with a hybrid neural network model". In: *Bioresour. Technol.* 76.2 (Jan. 2001), pp. 77–83. DOI: 10.1016/S0960-8524(00)00106-1.

[143] Frederik vom Scheidt, Hana Medinová, Nicole Ludwig, Bent Richter, Philipp Staudt, and Christof Weinhardt. "Data analytics in the electricity sector – A quantitative and qualitative literature review". In: *Energy AI* 1 (Aug. 2020), p. 100009. DOI: 10.1016/j.egyai.2020.100009.

[144] Cristian Chiñas-Palacios, Carlos Vargas-Salgado, Jesus Aguila-Leon, and Elias Hurtado-Pérez. "A cascade hybrid PSO feed-forward neural network model of a biomass gasification plant for covering the energy demand in an AC microgrid". In: *Energy Convers. Manag.* 232 (Mar. 2021), p. 113896. DOI: 10.1016/j.enconman.2021.113896.

[145] Hossein Shahbeig and Mohsen Nosrati. "Pyrolysis of biological wastes for bioenergy production: Thermo-kinetic studies with machine-learning method and Py-GC/MS analysis". In: *Fuel* 269 (June 2020), p. 117238. DOI: 10.1016/j.fuel.2020.117238.

[146] Jerome H. Friedman. "Greedy function approximation: A gradient boosting machine". In: *Ann. Stat.* 29.5 (2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451.

[147] E. Alfassio Grimaldi, F. Grimaccia, M. Mussetta, and R. E. Zich. "PSO as an effective learning algorithm for neural network applications". In: *ICCEA 2004 - 2004 3rd Int. Conf. Comput. Electromagn. its Appl. Proc.* 2004, pp. 557–560. DOI: 10.1109/iccea.2004.1459416.

[148] Maurice Clerc. *Particle swarm optimization*. Wiley & Sons, 2010.

[149] Ugur Ozveren. "An artificial intelligence approach to predict gross heating value of lignocellulosic fuels". In: *J. Energy Inst.* 90.3 (June 2017), pp. 397–407. DOI: 10.1016/j.joei.2016.04.003.

[150] Shu Gong, Jafar Sasanipour, Mohammad Reza Shayesteh, Mohammad Eslami, and Alireza Baghban. "Radial basis function artificial neural network model to estimate higher heating value of solid wastes". In: *Energy Sources, Part A Recover. Util. Environ. Eff.* 39.16 (Aug. 2017), pp. 1778–1784. DOI: 10.1080/15567036.2017.1370513.

[151] Beatriz A. Garro and Roberto A. Vázquez. "Designing Artificial Neural Networks Using Particle Swarm Optimization Algorithms". In: *Comput. Intell. Neurosci.* 2015 (2015). DOI: 10.1155/2015/369298.

[152] Laurens Van Der Maaten, Eric Postma, and Jaap Van Den Herik. "Dimensionality Reduction: A Comparative Review". In: *J Mach Learn Res* 10.13 (2009), pp. 66–71.

[153] Soleiman Hosseinpour, Mortaza Aghbashlo, Meisam Tabatabaei, and Mehdi Mehrpooya. "Estimation of biomass higher heating value (HHV) based on the proximate analysis by using iterative neural network-adapted partial least squares (INNPLS)". In: *Energy* 138 (Nov. 2017), pp. 473–479. DOI: 10.1016/j.energy.2017.07.075.

[154] J. A.K. Suykens and J. Vandewalle. "Least squares support vector machine classifiers". In: *Neural Process. Lett.* 9.3 (1999), pp. 293–300. DOI: 10.1023/A:1018628609742.

[155] Hao-nan Guo, Shu-biao Wu, Ying-jie Tian, Jun Zhang, and Hong-tao Liu. "Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review". In: *Bioresour. Technol.* 319 (Jan. 2021), p. 124114. DOI: 10.1016/j.biortech.2020.124114.

[156] Yeshona Sewsynker-Sukai, Funmilayo Faloye, and Evariste Bosco Gueguim Kana. "Artificial neural networks: an efficient tool for modelling and optimization of biofuel production (a mini review)". In: *Biotechnol. Biotechnol. Equip.* 31.2 (Mar. 2017), pp. 221–235. DOI: 10.1080/13102818.2016.1269616.

[157] Amir Mosavi, Mohsen Salimi, Sina Faizollahzadeh Ardabili, Timon Rabczuk, Shahaboddin Shamshirband, and Annamaria R. Varkonyi-Koczy. "State of the art of machine learning models in energy systems, a systematic review". In: *Energies* 12 (2019), p. 1301. DOI: 10.3390/en12071301.

[158] Changqing Dong, Baosheng Jin, and Daji Li. "Predicting the heating value of MSW with a feed forward neural network". In: *Waste Manag.* 23.2 (2003), pp. 103–106. DOI: 10.1016/S0956-053X(02)00162-9.

[159] Shagufta U. Patel, B. Jeevan Kumar, Yogesh P. Badhe, B. K. Sharma, Sujan Saha, Subhasish Biswas, Asim Chaudhury, Sanjeev S. Tambe, and Bhaskar D. Kulkarni. "Estimation of gross calorific value of coals using artificial neural networks". In: *Fuel* 86.3 (2007), pp. 334–344. DOI: 10.1016/j.fuel.2006.07.036.

[160] G. J. Postma, P. W.T. Krooshof, and L. M.C. Buydens. "Opening the kernel of kernel partial least squares and support vector machines". In: *Anal. Chim. Acta* 705.1-2 (Oct. 2011), pp. 123–134. DOI: 10.1016/j.aca.2011.04.025.

[161] International Organization for Standardization (ISO). *ISO 14044*. 2006.

[162] International Organization for Standardization (ISO). *ISO 14040:2006*. 2006.

[163] Madhumita Patel, Xiaolei Zhang, and Amit Kumar. "Techno-economic and life cycle assessment on lignocellulosic biomass thermochemical conversion technologies: A review". In: *Renew. Sustain. Energy Rev.* 53 (Jan. 2016), pp. 1486–1499. DOI: 10.1016/j.rser.2015.09.070.

[164] Simon Ascher, Ian Watson, Xiaonan Wang, and Siming You. "Township-based bioenergy systems for distributed energy supply and efficient household waste re-utilisation: Techno-economic and environmental feasibility". In: *Energy* 181 (Aug. 2019), pp. 455–467. DOI: 10.1016/j.energy.2019.05.191.

[165] Hatem Alhazmi and Adrian Chun Minh Loy. "A review on environmental assessment of conversion of agriculture waste to bio-energy via different thermochemical routes: Current and future trends". In: *Bioresour. Technol. Reports* 14.March (2021). DOI: 10.1016/j.biteb.2021.100682.

[166] Alejandro Padilla-Rivera, María Guadalupe Paredes, and Leonor Patricia Güereca. "A systematic review of the sustainability assessment of bioenergy: The case of gaseous biofuels". In: *Biomass and Bioenergy* 125 (June 2019), pp. 79–94. DOI: 10.1016/j.biombioe.2019.03.014.

[167] Luis López-Bellido, Jacques Wery, and Rafael J. López-Bellido. "Energy crops: Prospects in the context of sustainable agriculture". In: *Eur. J. Agron.* 60 (2014), pp. 1–12. DOI: 10.1016/j.eja.2014.07.001.

[168] F. Sebastián, J. Royo, and M. Gómez. "Cofiring versus biomass-fired power plants: GHG (Greenhouse Gases) emissions savings comparison by means of LCA (Life Cycle Assessment) methodology". In: *Energy* 36.4 (2011), pp. 2029–2037. DOI: 10.1016/j.energy.2010.06.003.

[169] Sjouke Beemsterboer, Henrikke Baumann, and Holger Wallbaum. "Ways to get work done: a review and systematisation of simplification practices in the LCA literature". In: *Int. J. Life Cycle Assess.* 25.11 (2020), pp. 2154–2168. DOI: 10.1007/s11367-020-01821-w.

[170] A. Shah, N. R. Baral, and A. Manandhar. "Technoeconomic Analysis and Life Cycle Assessment of Bioenergy Systems". In: *Adv. Bioenergy*. Vol. 1. Elsevier, 2016. Chap. 4, pp. 189–247. DOI: 10.1016/bs.aibe.2016.09.004. http://dx.doi.org/10.1016/bs.aibe.2016.09.004.

[171] Fangwei Cheng, Arthur A Small, and Lisa M Colosi. "The levelized cost of negative CO 2 emissions from thermochemical conversion of biomass coupled with carbon capture and storage". In: *Energy Convers. Manag.* 237.March (2021), p. 114115. DOI: 10.1016/j.enconman.2021.114115.

[172] M. Shahabuddin, Bhavya B. Krishna, Thallada Bhaskar, and Greg Perkins. "Advances in the thermo-chemical production of hydrogen from biomass and residual wastes: Summary of recent techno-economic analyses". In: *Bioresour. Technol.* 299.September 2019 (2020), p. 122557. DOI: 10.1016/j.biortech.2019.122557.

[173] Ahmed M. Elgarahy, Ahmed Hammad, Dina M. El-Sherif, Mohamed Abouzid, Mohamed S. Gaballah, and Khalid Z. Elwakeel. "Thermochemical conversion strategies of biomass to biofuels, techno-economic and bibliometric analysis: A conceptual review". In: *J. Environ. Chem. Eng.* 9.6 (2021), p. 106503. DOI: 10.1016/j.jece.2021.106503.

[174] Yetunde Sorunmu, Pieter Billen, and Sabrina Spatari. "A review of thermochemical upgrading of pyrolysis bio-oil: Techno-economic analysis, life cycle assessment, and technology readiness". In: *GCB Bioenergy* 12.1 (2020), pp. 4–18. DOI: 10.1111/gcbb.12658.

[175] Walter Klöpffer and Andreas Ciroth. "Is LCC relevant in a sustainability assessment?" In: *Int. J. Life Cycle Assess.* 16 (2011), pp. 99–101. DOI: 10.1007/s11367-011-0249-y.

[176] Simon Shackley, Sarah Carter, Tony Knowles, Erik Middelink, Stephan Haefele, and Stuart Haszeldine. "Sustainable gasification-biochar systems? A case-study of rice-husk gasification in Cambodia, Part II: Field trial results, carbon abatement, economic assessment and conclusions". In: *Energy Policy* 41 (2012), pp. 618–623. DOI: 10.1016/j.enpol.2011.11.023.

[177] Simon Shackley, Sarah Carter, Tony Knowles, Erik Middelink, Stephan Haefele, Saran Sohi, Andrew Cross, and Stuart Haszeldine. "Sustainable gasification-biochar systems? A case-study of rice-husk gasification in Cambodia, Part I: Context, chemical properties, environmental and health and safety issues". In: *Energy Policy* 42 (2012), pp. 49–58. DOI: 10.1016/j.enpol.2011.11.026.

[178] Enze Jin and John W. Sutherland. "An integrated sustainability model for a bioenergy system: Forest residues for electricity generation". In: *Biomass and Bioenergy* 119 (Dec. 2018), pp. 10–21. DOI: 10.1016/j.biombioe.2018.09.005.

[179] Dipal Baruah and D. C. Baruah. "Decision support system based planning of biomass gasification system for decentralised energy generation". In: *Renew. Energy Focus* 38.September (2021), pp. 22–35. DOI: 10.1016/j.ref.2021.05.002.

[180] Maria Fernanda Rojas Michaga, Stavros Michailos, Muhammad Akram, Evelyn Cardozo, Kevin J. Hughes, Derek Ingham, and Mohamed Pourkashanian. "Bioenergy with carbon capture and storage (BECCS) potential in jet fuel production from forestry residues: A combined Techno-Economic and Life Cycle Assessment approach". In: *Energy Convers. Manag.* 255.November 2021 (2022), p. 115346. DOI: 10.1016/j.enconman.2022.115346.

[181] Morena Bruno, Michela Marchi, Nicolò Ermini, Valentina Niccolucci, and Federico Maria Pulselli. "Life Cycle Assessment and Cost – Benefit Analysis as Combined Economic – Environmental Assessment Tools : Application to an Anaerobic Digestion Plant". In: *Energies* 16.3686 (2023). DOI: https://doi.org/10.3390/en16093686.

[182] Valentina Bisinella, Knut Conradsen, Thomas Højlund Christensen, and Thomas Fruergaard Astrup. "A global approach for sparse representation of uncertainty in Life Cycle Assessments of waste management systems". In: *Int. J. Life Cycle Assess.* 21.3 (2016), pp. 378–394. DOI: 10.1007/s11367-015-1014-4.

[183] Robert M. Campbell, Nathaniel M. Anderson, Daren E. Daugaard, and Helen T. Naughton. "Financial viability of biofuel and biochar production from forest biomass in the face of market price volatility and uncertainty". In: *Appl. Energy* 230.August 2018 (2018), pp. 330–343. DOI: 10.1016/j.apenergy.2018.08.085.

[184] Yi Fang, Li Ma, Zhiyi Yao, Wangliang Li, and Siming You. "Process optimization of biomass gasification with a Monte Carlo approach and random forest algorithm". In: *Energy Convers. Manag.* 264 (2022), p. 115734. DOI: 10.1016/j.enconman.2022.115734.

[185] Guiyan Zang, Junxi Jia, Sharma Tejasvi, Albert Ratner, and Electo Silva Lora. "Techno-economic comparative analysis of Biomass Integrated Gasification Combined Cycles with and without CO2 capture". In: *Int. J. Greenh. Gas Control* 78 (2018), pp. 73–84. DOI: 10.1016/j.ijggc.2018.07.023.

[186] Shouheng Sun and Myriam Ertz. "Life cycle assessment and Monte Carlo simulation to evaluate the environmental impact of promoting LNG vehicles". In: *MethodsX* 7.August (2020), p. 101046. DOI: 10.1016/j.mex.2020.101046.

[187] Laetitia Uwineza, Hyun Goo Kim, and Chang Ki Kim. "Feasibilty study of integrating the renewable energy system in Popova Island using the Monte Carlo model and HOMER". In: *Energy Strateg. Rev.* 33 (2021), p. 100607. DOI: 10.1016/j.esr.2020.100607.

[188] World Bank Group. *What a Waste 2.0 - A global snapshot of solid waste management to 2050.*

[189] IEA Bioenergy. *The Availability of Biomass Resources for Energy: Summary and Conclusions from the IEA Bioenergy ExCo58 Workshop*. Tech. rep. IEA Bioenergy, 2008, p. 12. http://www.ieabioenergy.com/wp-content/uploads/2013/10/The-Availability-of-Biomass-Resources-for-Energy-Sumamry-and-Conclusions1.pdf.

[190] Kangcheng Wang, Jie Zhang, Chao Shang, and Dexian Huang. "Operation optimization of Shell coal gasification process based on convolutional neural network models". In: *Appl. Energy* 292.January (2021), p. 116847. DOI: 10.1016/j.apenergy.2021.116847.

[191] Qing Yang, Hewen Zhou, Pietro Bartocci, Francesco Fantozzi, Ondřej Mašek, Foster A Agblevor, Zhiyu Wei, Haiping Yang, Hanping Chen, Xi Lu, Guoqian Chen, Chuguang Zheng, Chris P Nielsen, and Michael B McElroy. "Prospective contributions of biomass pyrolysis to China's 2050 carbon reduction and renewable energy goals". In: *Nat. Commun.* 21.1698 (2021). DOI: 10.1038/s41467-021-21868-z.

[192] Qinghui Tang, Yingquan Chen, Haiping Yang, Ming Liu, and Haoyu Xiao. "Machine learning prediction of pyrolytic gas yield and compositions with feature reduction methods : Effects of pyrolysis conditions and biomass characteristics". In: *Bioresour. Technol.* 339 (2021), p. 125581. DOI: 10.1016/j.biortech.2021.125581.

[193] Yngve Ögren, Pál Tóth, Attila Garami, Alexey Sepman, and Henrik Wiinikka. "Development of a vision-based soft sensor for estimating equivalence ratio and major species concentration in entrained flow biomass gasification reactors". In: *Appl. Energy* 226.April (2018), pp. 450–460. DOI: 10.1016/j.apenergy.2018.06.007.

[194] Chao Ning and Fengqi You. "Data-driven Wasserstein distributionally robust optimization for biomass with agricultural waste-to-energy network design under uncertainty". In: *Appl. Energy* 255.September (2019), p. 113857. DOI: 10.1016/j.apenergy.2019.113857.

[195] Jie Li, Manu Suvarna, Lanjia Pan, Yingru Zhao, and Xiaonan Wang. "A hybrid data-driven and mechanistic modelling approach for hydrothermal gasification". In: *Appl. Energy* 304.September (2021), p. 117674. DOI: 10.1016/j.apenergy.2021.117674.

[196] Hannah O. Kargbo, Jie Zhang, and Anh N. Phan. "Optimisation of two-stage biomass gasification for hydrogen production via artificial neural network". In: *Appl. Energy* 302.July (2021), p. 117567. DOI: 10.1016/j.apenergy.2021.117567.

[197] Umberto Arena and Fabrizio Di Gregorio. "Gasification of a solid recovered fuel in a pilot scale fluidized bed reactor". In: *Fuel* 117 (2014), pp. 528–536. DOI: 10.1016/j.fuel.2013.09.044.

[198] Umberto Arena, Lucio Zaccariello, and Maria Laura Mastellone. "Fluidized bed gasification of waste-derived fuels". In: *Waste Manag.* 30 (July 2010), pp. 1212–1219. DOI: 10.1016/j.wasman.2010.01.038.

[199] Marco Baratieri, Elisa Pieratti, Thomas Nordgreen, and Maurizio Grigiante. "Biomass gasification with dolomite as catalyst in a small fluidized bed experimental and modelling analysis". In: *Waste and Biomass Valorization* 1 (2010), pp. 283–291. DOI: 10.1007/s12649-010-9034-6.

[200] Siyi Luo, Yangmin Zhou, and Chuijie Yi. "Syngas production by catalytic steam gasification of municipal solid waste in fixed-bed reactor". In: *Energy* 44.1 (2012), pp. 391–395. DOI: 10.1016/j.energy.2012.06.016.

[201] Zakir Khan, Suzana Yusup, Murni Melati Ahmad, and Bridgid Lai Fui Chin. "Hydrogen production from palm kernel shell via integrated catalytic adsorption (ICA) steam gasification". In: *Energy Convers. Manag.* 87 (2014), pp. 1224–1230. DOI: 10.1016/j.enconman.2014.03.024.

[202] Jingbo Wang, Gong Cheng, Yanli You, Bo Xiao, Shiming Liu, Piwen He, Dabin Guo, Xianjun Guo, and Guijuan Zhang. "Hydrogen-rich gas production by steam gasification of municipal solid waste (MSW) using NiO supported on modified dolomite". In: *Int. J. Hydrogen Energy* 37.8 (Apr. 2012), pp. 6503–6510. DOI: 10.1016/j.ijhydene.2012.01.070.

[203] Hakan Karatas and Fehmi Akgun. "Experimental results of gasification of walnut shell and pistachio shell in a bubbling fluidized bed gasifier under air and steam atmospheres". In: *Fuel* 214 (Feb. 2018), pp. 285–292. DOI: 10.1016/j.fuel.2017.10.061.

[204] Cuong Van Huynh and Song Charng Kong. "Performance characteristics of a pilot-scale biomass gasifier using oxygen-enriched air and steam". In: *Fuel* 103 (2013), pp. 987–996. DOI: 10.1016/j.fuel.2012.09.033.

[205] Aitziber Erkiaga, Gartzen Lopez, Maider Amutio, Javier Bilbao, and Martin Olazar. "Influence of operating conditions on the steam gasification of biomass in a conical spouted bed reactor". In: *Chem. Eng. J.* 237 (Feb. 2014), pp. 259–267. DOI: 10.1016/j.cej.2013.10.018.

[206] G. Ruoppolo, P. Ammendola, R. Chirone, and F. Miccio. "H2-rich syngas production by fluidized bed gasification of biomass and plastic fuel". In: *Waste Manag.* 32.4 (Apr. 2012), pp. 724–732. DOI: 10.1016/j.wasman.2011.12.004.

[207] Huili Liu, Jianhang Hu, Hua Wang, Chao Wang, and Juanqin Li. "Experimental studies of biomass gasification with air". In: *J. Nat. Gas Chem.* 21 (2012), pp. 374–380. DOI: 10.1016/S1003-9953(11)60379-4.

[208] Chao Gai and Yuping Dong. "Experimental study on non-woody biomass gasification in a downdraft gasifier". In: *Int. J. Hydrogen Energy* 37.6 (Mar. 2012), pp. 4935–4944. DOI: 10.1016/j.ijhydene.2011.12.031.

[209] Pooya Lahijani and Zainal Alimuddin Zainal. "Gasification of palm empty fruit bunch in a bubbling fluidized bed: A performance and agglomeration study". In: *Bioresour. Technol.* 102 (2011), pp. 2068–2076. DOI: 10.1016/j.biortech.2010.09.101.

[210] Ian Narváez, Alberto Orío, Maria P Aznar, and José Corella. "Biomass Gasification with Air in an Atmospheric Bubbling Fluidized Bed. Effect of Six Operational Variables on the Quality of the Produced Raw Gas". In: *Ind. Eng. Chem. Res.* 35 (1996), pp. 2110–2120.

[211] Renzhan Yin, Ronghou Liu, Jinkai Wu, Xiaowu Wu, Chen Sun, and Ceng Wu. "Influence of particle size on performance of a pilot-scale fixed-bed gasification system". In: *Bioresour. Technol.* 119 (Sept. 2012), pp. 15–21. DOI: 10.1016/j.biortech.2012.05.085.

[212] Tao Song, Jiahua Wu, Laihong Shen, and Jun Xiao. "Experimental investigation on hydrogen production from biomass gasification in interconnected fluidized beds". In: *Biomass and Bioenergy* 36 (Jan. 2012), pp. 258–267. DOI: 10.1016/j.biombioe.2011.10.021.

[213] Enrico Biagini, Federica Barontini, and Leonardo Tognotti. "Development of a bi-equilibrium model for biomass gasification in a downdraft bed reactor". In: *Bioresour. Technol.* 201 (2016), pp. 156–165. DOI: 10.1016/j.biortech.2015.11.057.

[214] P. P. Dutta, V. Pandey, A. R. Das, S. Sen, and D. C. Baruah. "Down draft gasification modelling and experimentation of some indigenous biomass for thermal applications". In: *Energy Procedia*. Vol. 54. Elsevier Ltd, Jan. 2014, pp. 21–34. DOI: 10.1016/j.egypro.2014.07.246.

[215] Javier Herguido, Jose Corella, and Jose Gonzalez-Saiz. "Steam Gasification of Lignocellulosic Residues in a Fluidized Bed at a Small Pilot Scale. Effect of the Type of Feedstock". In: *Ind. Eng. Chem. Res* 31 (1992), pp. 2295–2310.

[216] Sylvain Fremaux, Sayyed Mohsen Beheshti, Hojat Ghassemi, and Rasoul Shahsavan-Markadeh. "An experimental study on hydrogen-rich gas production via steam gasification of biomass in a research-scale fluidized bed". In: *Energy Convers. Manag.* 91 (2015), pp. 427–432. DOI: 10.1016/j.enconman.2014.12.048.

[217] Ziyin Zhang and Shusheng Pang. "Experimental investigation of biomass devolatilization in steam gasification in a dual fluidised bed gasifier". In: *Fuel* 188 (2017), pp. 628–635. DOI: 10.1016/j.fuel.2016.10.074.

[218] P. Iovane, A. Donatelli, and A. Molino. "Influence of feeding ratio on steam gasification of palm shells in a rotary kiln pilot plant. Experimental and numerical investigations". In: *Biomass and Bioenergy* 56 (2013), pp. 423–431. DOI: 10.1016/j.biombioe.2013.05.025.

[219] P. M. Lv, Z. H. Xiong, J. Chang, C. Z. Wu, Y. Chen, and J. X. Zhu. "An experimental study on biomass air-steam gasification in a fluidized bed". In: *Bioresour. Technol.* 95 (Oct. 2004), pp. 95–101. DOI: 10.1016/j.biortech.2004.02.003.

[220] Juan Manuel De Andrés, Adolfo Narros, and María Encarnación Rodríguez. "Air-steam gasification of sewage sludge in a bubbling bed reactor: Effect of alumina as a primary catalyst". In: *Fuel Process. Technol.* 92.3 (Mar. 2011), pp. 433–440. DOI: 10.1016/j.fuproc.2010.10.006.

[221] Sommas Kaewluan and Suneerat Pipatmanomai. "Gasification of high moisture rubber woodchip with rubber waste in a bubbling fluidized bed". In: *Fuel Process. Technol.* 92 (2011), pp. 671–677. DOI: 10.1016/j.fuproc.2010.11.026.

[222] Shiplu Sarker, Fernando Bimbela, José Luis Sánchez, and Henrik Kofoed Nielsen. "Characterization and pilot scale fluidized bed gasification of herbaceous biomass: A case study on alfalfa pellets". In: *Energy Convers. Manag.* 91 (2015), pp. 451–458. DOI: 10.1016/j.enconman.2014.12.034.

[223] K G Mansaray, A E Ghaly, A M Al-Taweel, F Hamdullahpur, and V I Ugursal. "Air gasification of rice husk in a dual distributor type fluidized bed gasifier". In: *Biomass and Bioenergy* 17 (1999), pp. 315–332.

[224] Chanchal Loha, Himadri Chattopadhyay, and Pradip K. Chatterjee. "Energy generation from fluidized bed gasification of rice husk". In: *J. Renew. Sustain. Energy* 5 (2013), p. 043111. DOI: 10.1063/1.4816496.

[225] Lee Walker, Wipa Charles, and Ralf Cord-Ruwisch. "Comparison of static, in-vessel composting of MSW with thermophilic anaerobic digestion and combinations of the two processes". In: *Bioresour. Technol.* 100.16 (2009), pp. 3799–3807. DOI: 10.1016/j.biortech.2009.02.015.

[226] I. S. Antonopoulos, A. Karagiannidis, A. Gkouletsos, and G. Perkoulidis. "Modelling of a downdraft gasifier fed by agricultural residues". In: *Waste Manag.* 32.4 (2012), pp. 710–718. DOI: 10.1016/j.wasman.2011.12.015.

[227] C. Franco, F. Pinto, I. Gulyurtlu, and I. Cabrita. "The study of reactions influencing the biomass steam gasification process". In: *Fuel* 82.7 (2003), pp. 835–842. DOI: 10.1016/S0016-2361(02)00313-7.

[228] Juan Manuel De Andrés, Elena Roche, Adolfo Narros, and María Encarnación Rodríguez. "Characterisation of tar from sewage sludge gasification. Influence of gasifying conditions: Temperature, throughput, steam and use of primary catalysts". In: *Fuel* 180 (2016), pp. 116–126. DOI: 10.1016/j.fuel.2016.04.012.

[229] Keiichi Tomishige, Mohammad Asadullah, and Kimio Kunimori. "Syngas production by biomass gasification using Rh/CeO2/SiO 2 catalysts and fluidized bed reactor". In: *Catal. Today* 89 (2004), pp. 389–403. DOI: 10.1016/j.cattod.2004.01.002.

[230] Stef van Buuren. *Flexible Imputation of Missing Data*. 2nd ed. CRC Press, 2018.

[231] Jie Li, Lanjia Pan, Manu Suvarna, and Xiaonan Wang. "Machine learning aided supercritical water gasification for H2-rich syngas production with process optimization and catalyst screening". In: *Chem. Eng. J.* 426.July (2021), p. 131285. DOI: 10.1016/j.cej.2021.131285.

[232] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5 (1989), pp. 359–366. DOI: 10.1016/0893-6080(89)90020-8.

[233] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323.6088 (1986), pp. 533–536. DOI: 10.1038/323533a0.

[234] Donald W Marquardt. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters". In: *J. Soc. Ind. Appl. Math.* 11 (1963), pp. 431–441.

[235] M. T. Hagan and M.B. Menhaj. "Training feedforward networks with the Marquardt algorithm". In: *IEEE Trans. neural networks* 5 (1994), pp. 989–993. DOI: 10.1109/72.329697.

[236] F Dan Foresee and Martin T Hagan. "Gauss-Newton approximation to bayesian learning". In: *Proc. Int. Jt. Conf. Neural Networks*. 1997.

[237] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc.198 Madison Ave. New York, NY, United States, 1995, p. 482.

[238] Marcus W. Beck. "NeuralNetTools: Visualization and analysis tools for neural networks". In: *J. Stat. Softw.* 85.11 (2018), pp. 1–20. DOI: 10.18637/jss.v085.i11.

[239] MathWorks. *MathWorks - MATLAB*. 2022. https://uk.mathworks.com/products/matlab.html (visited on 04/18/2022).

[240] Nor Anisa Nasrudin, Jegalakshimi Jewaratnam, Md Arafat Hossain, and Poo Balan Ganeson. "Performance comparison of feedforward neural network training algorithms in modelling microwave pyrolysis of oil palm fibre for hydrogen and biochar production". In: *Asia-Pacific J. Chem. Eng.* 15.1 (2020), pp. 1–16. DOI: 10.1002/apj.2388.

[241] Uğur Özveren, Furkan Kartal, Senem Sezer, and Z. Sibel Özdoğan. "Investigation of steam gasification in thermogravimetric analysis by means of evolved gas analysis and machine learning". In: *Energy* 239 (Jan. 2022), p. 122232. DOI: 10.1016/J.ENERGY.2021.122232.

[242] Ludovic Lacrosse, Martin Englisch, Katharina Danner, and Harry Stokes. *The role of bioenergy in the clean energy transition and sustainable development – lessons from developing countries*. Tech. rep. United Nations Industrial Development Organization, 2021.

[243] BEIS. *Biomass Policy Statement*. Tech. rep. Department for Business Energy & Industrial Strategy, 2021.

[244] Pengfei Wei, Zhenzhou Lu, and Jingwen Song. "Variable importance analysis: A comprehensive review". In: *Reliab. Eng. Syst. Saf.* 142 (2015), pp. 399–432. DOI: 10.1016/j.ress.2015.05.018.

[245] Tonghuan Zhang, Danyang Cao, Xin Feng, Jiahua Zhu, Xiaohua Lu, Liwen Mu, and Hongliang Qian. "Machine learning prediction of bio-oil characteristics quantitatively relating to biomass compositions and pyrolysis conditions". In: *Fuel* 312 (2022), p. 122812. DOI: 10.1016/j.fuel.2021.122812.

[246] Scott M. Lundberg and Su In Lee. "A unified approach to interpreting model predictions". In: *Adv. Neural Inf. Process. Syst. (NIPS 2017)*. 2017. arXiv: 1705.07874.

[247] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Model-Agnostic Interpretability of Machine Learning". In: *ICML Work. Hum. Interpret. Mach. Learn.* (2016). DOI: 10.48550/arxiv.1606.05386. arXiv: 1606.05386.

[248] Balázs Kégl. "The return of ADABOOST.MH: multi-class Hamming trees". In: *arXiv* (2013). DOI: 10.48550/ARXIV.1312.6086. arXiv: 1312.6086v1.

[249] Steven L. Brunton and J. Nathan Kutz. *Data-driven science and engineering: machine learning, dynamical systems, and control*. 2017. DOI: 10.1109/mcs.2021.3076544.

[250] Mark J. Van Der Laan, Eric C. Polley, and Alan E. Hubbard. "Super learner". In: *Stat. Appl. Genet. Mol. Biol.* 6.1 (2007). DOI: 10.2202/1544-6115.1309.

[251] Sheng Zhao, Jian Li, Chao Chen, Beibei Yan, Junyu Tao, and Guanyi Chen. "Interpretable machine learning for predicting and evaluating hydrogen production via supercritical water gasification of biomass". In: *J. Clean. Prod.* 316.July (2021), p. 128244. DOI: 10.1016/j.jclepro.2021.128244.

[252] Chuanjia Sun, Lishen Ai, and Ting Liu. "The PSO-ANN modeling study of highly valuable material and energy production by gasification of solid waste: an artificial intelligence algorithm approach". In: *Biomass Convers. Biorefinery* (2022). DOI: 10.1007/s13399-022-02342-2.

[253] Vineet Singh Sikarwar, Ming Zhao, Paul S. Fennell, Nilay Shah, and Edward J. Anthony. "Progress in biofuel production from gasification". In: *Prog. Energy Combust. Sci.* 61 (July 2017), pp. 189–248.

[254] Charu C. Aggarwal. *Neural Networks and Deep Learning - A Textbook*. Springer, 2018. DOI: 10.1107/978-3-319-94463-0.

[255] Xiangzhou Yuan, Manu Suvarna, Sean Low, Pavani Dulanja Dissanayake, Ki Bong Lee, Jie Li, Xiaonan Wang, and Yong Sik Ok. "Applied Machine Learning for Prediction of CO2 Adsorption on Biomass Waste-Derived Porous Carbons". In: *Environ. Sci. Technol.* 55.17 (Sept. 2021), pp. 11925–11936. DOI: 10.1021/ACS.EST.1C01849.

[256] Jie Li, Xinzhe Zhu, Yinan Li, Yen Wah Tong, Yong Sik Ok, and Xiaonan Wang. "Multi-task prediction and optimization of hydrochar properties from high-moisture municipal solid waste: Application of machine learning on waste-to-resource". In: *J. Clean. Prod.* 278 (Jan. 2021), p. 123928. DOI: 10.1016/J.JCLEPRO.2020.123928.

[257] Dylan Slack, Sophie Hilgard, Emily Jia, Sameer Singh, and Himabindu Lakkaraju. "Fooling LIME and SHAP". In: *Proc. AAAI/ACM Conf. AI, Ethics, Soc.* 2020, pp. 180–186. DOI: 10.1145/3375627.3375830.

[258] Bertrand Meyer. *Object oriented software construction*. Vol. 2. Englewood Cliffs: Prentice Hall, 1997. DOI: 10.1016/0167-6423(89)90034-8.

[259] Patricia Thornley. "Scope of System for Analysis". In: *Greenh. Gas Balanc. Bioenergy Syst.* Elsevier Inc., 2017. Chap. 4, pp. 43–53. DOI: 10.1016/B978-0-08-101036-5.00004-5. https://doi.org/10.1016/B978-0-08-101036-5.00004-5.

[260] Anna Skorek-Osikowska, Janusz Kotowicz, and Wojciech Uchman. "Thermodynamic assessment of the operation of a self-sufficient, biomass based district heating system integrated with a Stirling engine and biomass gasification". In: *Energy* 141 (2017), pp. 1764–1778. DOI: 10.1016/j.energy.2017.11.106.

[261] Davide Tonini, Dieuwertje Schrijvers, Simone Nessi, Pelayo Garcia-Gutierrez, and Jacopo Giuntoli. "Carbon footprint of plastic from biomass and recycled feedstock: methodological insights". In: *Int. J. Life Cycle Assess.* 26.2 (2021), pp. 221–237. DOI: 10.1007/s11367-020-01853-2.

[262] Alessandro Agostini, Jacopo Giuntoli, Luisa Marelli, and Stefano Amaducci. "Flaws in the interpretation phase of bioenergy LCA fuel the debate and mislead policy-makers". In: *Int. J. Life Cycle Assess.* 25.1 (2020), pp. 17–35. DOI: 10.1007/s11367-019-01654-2.

[263] Shveta Soam, Pal Borjesson, Pankaj K. Sharma, Ravi P. Gupta, Deepak K. Tuli, and Ravindra Kumar. "Life cycle assessment of rice straw utilization practices in India". In: *Bioresour. Technol.* 228 (2017), pp. 89–98. DOI: 10.1016/j.biortech.2016.12.082.

[264] Stefan Muench and Edeltraud Guenther. "A systematic review of bioenergy life cycle assessments". In: *Appl. Energy* 112 (2013), pp. 257–273. DOI: 10.1016/j.apenergy.2013.06.001.

[265] Reinout Heijungs, Karen Allacker, Enrico Benetto, Miguel Brandão, Jeroen Guinée, Simon Schaubroeck, Thomas Schaubroeck, and Alessandra Zamagni. "System Expansion and Substitution in LCA: A Lost Opportunity of ISO 14044 Amendment 2". In: *Front. Sustain.* 2.June (2021), pp. 1–3. DOI: 10.3389/frsus.2021.692055.

[266] William G. Sullivan, Elin M. Wicks, and C. Patrick Koelling. *Engineering Economy*. 17th. Pearson Education, Limited, 2019.

[267] Robert James, Alexander Zoelle, Dale Keairns, Marc Turner, Mark Woods, and Norma Kuehn. *Cost and Performance Baseline for Fossil Energy Plants Volume 1 : Bituminous Coal and Natural Gas to Electricity*. Tech. rep. 2019.

[268] Wahiba Yaïci, Evgueniy Entchev, and Michela Longo. "Recent Advances in Small-Scale Carbon Capture Systems for Micro-Combined Heat and Power Applications". In: *Energies* 15.8 (2022). DOI: 10.3390/en15082938.

[269] Eric D. Larson, Haiming Jin, and Fuat E. Celik. "Large-scale gasification-based co-production of fuels and electricity from switchgrass". In: *Biofuels, Biorprod. Bioref.* 3.2 (2009), pp. 174–194. DOI: 10.1002/bbb.137.

[270] Robin Smith. *Chemical Process Design and Integration*. 2005, p. 712.

[271] Max S. Peters and Klaus D. Timmerhaus. *Plant Design and Economics for Chemical Engineers*. Ed. by McGraw-Hill Inc. 5th. New York, NY, 2004.

[272] Costas Christou, Ioannis Hadjipaschalis, and Andreas Poullikkas. "Assessment of integrated gasification combined cycle technology competitiveness". In: *Renew. Sustain. Energy Rev.* 12.9 (2008), pp. 2459–2471. DOI: 10.1016/j.rser.2007.06.010.

[273] Amit Kumar, Jay B. Cameron, and Peter C. Flynn. "Biomass power cost and optimum plant size in western Canada". In: *Biomass and Bioenergy* 24.6 (June 2003), pp. 445–464. DOI: 10.1016/S0961-9534(02)00149-6.

[274] IRENA. *Biomass for Power Generation*. Tech. rep. 1/5. IRENA, 2012.

[275] Friedemann G. Albrecht, Daniel H. König, Nadine Baucks, and Ralph Uwe Dietrich. "A standardized methodology for the techno-economic evaluation of alternative fuels – A case study". In: *Fuel* 194 (2017), pp. 511–526. DOI: 10.1016/j.fuel.2016.12.003.

[276] ChemEngOnline. *Chemical Engineering Plant Cost Index*. 2024. https://www.chemengonline.com/ (visited on 08/01/2019).

[277] Siming You, Wei Wang, Yanjun Dai, Yen Wah Tong, and Chi Hwa Wang. "Comparison of the co-gasification of sewage sludge and food wastes and cost-benefit analysis of gasification- and incineration-based waste treatment schemes". In: *Bioresour. Technol.* 218 (2016), pp. 595–605. DOI: 10.1016/j.biortech.2016.07.017.

[278] Harsh Pershad, Eleanor Standen, Emrah Durusut, and Shane Slater. *The costs of Carbon Capture and Storage (CCS) for UK Industry - A high level review*. Tech. rep. Element Energy Limited, 2013. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment%7B%5C_%7Ddata/file/181161/bis-13-745-the-costs-of-carbon-capture-and-storage-for-uk-industry-a-high-level-review.pdf.

[279] ESPE. *Co-generator and biomass*. 2024. https://www.espegroup.com/en/biomass/co-generator-and-biomass/%7B%5C#%7D1614073795532-4ecca7b6-c1aa (visited on 02/08/2024).

[280] Richard Toonssen, Nico Woudstra, and Adrian H.M. Verkooijen. "Exergy analysis of hydrogen production plants based on biomass gasification". In: *Int. J. Hydrogen Energy* 33.15 (2008), pp. 4074–4082. DOI: 10.1016/j.ijhydene.2008.05.059.

[281] Alexander Tremel, Matthias Gaderer, and Hartmut Spliethoff. "Small-scale production of synthetic natural gas by allothermal biomass gasificatio". In: *Int. J. Energy Res.* 37.11 (2013), pp. 1318–1330. DOI: 10.1002/er.2933.

[282] A Dutta, M Talmadge, J Hensley, M Worley, and D Dudgeon. *Process design and economics for conversion of lignocellulosic biomass to ethanol - Thermochemical Pathway by Indirect Gasification and Mixed Alcohol Synthesis*. Tech. rep. National Renewable Energy Laboratory (NREL), 2011. http://www.nrel.gov/docs/fy11osti/51400.pdf%7B%5C%%7D5Cnpapers2://publication/uuid/49A5007E-9A58-4E2B-AB4E-4A4428F6EA66.

[283] T. G. Kreutz, E. D. Larson, G. Liu, and R. H. Williams. *Fischer-tropsch fuels from coal and biomass*. Tech. rep. August. 2008.

[284] Lasse R. Clausen, Brian Elmegaard, and Niels Houbak. "Technoeconomic analysis of a low CO2 emission dimethyl ether (DME) plant based on gasification of torrefied biomass". In: *Energy* 35.12 (2010), pp. 4831–4842. DOI: 10.1016/j.energy.2010.09.004.

[285] Cristina Aracil, Pedro Haro, Diego Fuentes-Cano, and Alberto Gómez-Barea. "Implementation of waste-to-energy options in landfill-dominated countries: Economic evaluation and GHG impact". In: *Waste Manag.* 76 (2018), pp. 443–456. DOI: 10.1016/j.wasman.2018.03.039.

[286] Liban Yassin, Paola Lettieri, Stefaan J.R. Simons, and Antonino Germanà. "Techno-economic performance of energy-from-waste fluidized bed combustion and gasification processes in the UK context". In: *Chem. Eng. J.* 146.3 (2009), pp. 315–327. DOI: 10.1016/j.cej.2008.06.014.

[287] Stefan Heyne and Simon Harvey. "Impact of choice of CO2 separation technology on thermo-economic performance of Bio-SNG production processes". In: *Int. J. Energy Res.* 38.3 (2014), pp. 299–318. DOI: 10.1002/er.3038.

[288] Ikhlas Ghiat, Ahmed AlNouss, Gordon McKay, and Tareq Al-Ansari. "Biomass-based integrated gasification combined cycle with post-combustion CO2 recovery by potassium carbonate: Techno-economic and environmental analysis". In: *Comput. Chem. Eng.* 135 (2020), p. 106758. DOI: 10.1016/j.compchemeng.2020.106758.

[289] M. M. Parascanu, M. Puig-Gamero, G. Soreanu, J. L. Valverde, and L. Sanchez-Silva. "Comparison of three Mexican biomasses valorization through combustion and gasification: Environmental and economic analysis". In: *Energy* 189 (Dec. 2019), p. 116095. DOI: 10.1016/j.energy.2019.116095.

[290] F. Di Gregorio and Lucio Zaccariello. "Fluidized bed gasification of a packaging derived fuel: Energetic, environmental and economic performances comparison for waste-to-energy plants". In: *Energy* 42.1 (2012), pp. 331–341. DOI: 10.1016/j.energy.2012.03.048.

[291] Umberto Arena, Fabrizio Di Gregorio, Gianfranco De Troia, and Alessandro Saponaro. "A techno-economic evaluation of a small-scale fluidized bed gasifier for solid recovered fuel". In: *Fuel Process. Technol.* 131 (2015), pp. 69–77. DOI: 10.1016/j.fuproc.2014.11.003.

[292] Anna Skorek-Osikowska, Łukasz Bartela, Janusz Kotowicz, Aleksander Sobolewski, Tomasz Iluk, and Leszek Remiorz. "The influence of the size of the CHP (combined heat and power) system integrated with a biomass fueled gas generator and piston engine on the thermodynamic and economic effectiveness of electricity and heat generation". In: *Energy* 67 (2014), pp. 328–340. DOI: 10.1016/j.energy.2014.01.015.

[293] A. R. Smith and J. Klosek. "A review of air separation technologies and their integration with energy conversion processes". In: *Fuel Process. Technol.* 70.2 (May 2001), pp. 115–134. DOI: 10.1016/S0378-3820(01)00131-X.

[294] Masaaki Taniguchi, Hitoshi Asaoka, and Toshiyuki Ayuhara. "Energy saving air-separation plant based on exergy analysis". In: *KOBELCO Technol. Rev.* (2015).

[295] Calin-Cristian Cormos, Ana-Maria Cormos, and Serban Agachi. "Power generation from coal and biomass based on integrated gasification combined cycle concept with pre- and post-combustion carbon capture methods". In: *Asia-Pacific J. Chem. Eng.* 4 (2009), pp. 870–877. DOI: 10.1002/apj.

[296] Yas A. Alsultanny and Nayef N. Al-Shammari. "Oxygen specific power consumption comparison for air separation units". In: *Eng. J.* 18.2 (2014), pp. 67–80. DOI: 10.4186/ej.2014.18.2.67.

[297] Dursun Can Ozcan, Hyungwoong Ahn, and Stefano Brandani. "Process integration of a Ca-looping carbon capture process in a cement plant". In: *Int. J. Greenh. Gas Control* 19 (2013), pp. 530–540. DOI: 10.1016/j.ijggc.2013.10.009.

[298] A Wienese. "Boilers, Boiler Fuel and Boiler Efficiency". In: *Proc S Afr Sug Technol Ass* 75 (2001), pp. 275–281.

[299] Sunit Shah and D.M. Adhyaru. "Boiler Efficiency Analysis Using Direct and Indirect Method". In: *Int. Conf. Curr. TRENDS Technol. 'NUiCONE – 2011'* 2 (2023), pp. 721–730. DOI: 10.1007/978-981-19-1457-7_55.

[300] Samah M. Said, Asmaa S. Hamouda, Hamada Mahmaoud, and Said Abd-Elwahab. "Computer-based boiler efficiency improvement". In: *Environ. Prog. Sustain. Energy* 38.5 (2019), pp. 1–9. DOI: 10.1002/ep.13161.

[301] Esa Kari Vakkilainen. "Boiler Processes". In: *Steam Gener. from Biomass.* 2017, pp. 57–86. DOI: 10.1016/b978-0-12-804389-9.00003-4.

[302] Paweł Madejski and Piotr Żymełka. "Calculation methods of steam boiler operation factors under varying operating conditions with the use of computational thermodynamic modeling". In: *Energy* 197 (2020). DOI: 10.1016/j.energy.2020.117221.

[303] Martina Pöschl, Shane Ward, and Philip Owende. "Evaluation of energy efficiency of various biogas production and utilization pathways". In: *Appl. Energy* 87 (2010), pp. 3305–3321. DOI: 10.1016/j.apenergy.2010.05.011.

[304] Sara Evangelisti, Roland Clift, Carla Tagliaferri, and Paola Lettieri. "A life cycle assessment of distributed energy production from organic waste: Two case studies in Europe". In: *Waste Manag.* 64 (2017), pp. 371–385. DOI: 10.1016/j.wasman.2017.03.028.

[305] Jenbacher. *Jenbacher Type 6*. 2024. https://www.innio.com/en/jenbacher/gas-engines/type-6 (visited on 02/08/2024).

[306] Wood. *Assessing the Cost Reduction Potential and Competitiveness of Novel (Next Generation) UK Carbon Capture Technology - Benchmarking State-of-the-art and Next Generation Technologies*. Tech. rep. 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment%7B%5C_%7Ddata/file/800680/Literature%7B%5C_%7DReview%7B%5C_%7DReport%7B%5C_%7DRev%7B%5C_%7D2A%7B%5C_%7D%7B%5C_%7D1%7B%5C_%7D.pdf.

[307] EPA. *Emission Factors for Greenhouse Gas Inventories*. 2018. https://www.epa.gov/climateleadership/ghg-emission-factors-hub (visited on 02/08/2024).

[308] Jun Dong, Yuanjun Tang, Ange Nzihou, Yong Chi, Elsa Weiss-Hortala, Mingjiang Ni, and Zhaozhi Zhou. "Comparison of waste-to-energy technologies of gasification and incineration using life cycle assessment: Case studies in Finland, France and China". In: *J. Clean. Prod.* 203 (2018), pp. 287–300. DOI: 10.1016/J.JCLEPRO.2018.08.139.

[309] J. M. Wilson, L. J. McKinney, K. Theerarattananoon, T. C. Ballard, D. Wang, S. A. Staggenborg, and P. V. Vadlani. "Energy and cost for pelleting and transportation of select cellulosic biomass feedstocks for ethanol production". In: *Appl. Eng. Agric.* 30.1 (2014), pp. 77–85. DOI: 10.13031/aea.30.9719.

[310] R Jannasch, Y Quan, and R Samson. *A Process and Energy Analysis of Pelletizing Switchgrass*. Tech. rep. Resource Efficient Agricultural Production (REAP-Canada), 2001, p. 16. https://reap-canada.com/online%7B%5C_%7Dlibrary/feedstock%7B%5C_%7Dbiomass/11%20A%20Process.pdf.

[311] Arifa Sultana, Amit Kumar, and Don Harfield. "Development of agri-pellet production cost and optimum size". In: *Bioresour. Technol.* 101.14 (2010), pp. 5609–5621. DOI: 10.1016/j.biortech.2010.02.011.

[312] Vulcan Drying Systems. *Personal Communication - Vulcan Drying Systems*. 2023.

[313] Shahab Sokhansanj and Erin Webb. "Evaluating industrial drying of cellulosic feedstock for bioenergy: A systems approach". In: *Biofuels, Bioprod. Biorefining* 10.1 (2016), pp. 47–55. DOI: 10.1002/bbb.1619.

[314] Robert S. Cherry, Rick A. Wood, and Tyler L. Westover. "Analysis of the Production Cost for Various Grades of Biomass Thermal Treatment". In: *U.S. Dep. Energy, Idaho Natl. Lab. INL/EXT-13-30348, Idaho Falls, Idaho* December (2013).

[315] Nawshad Haque and Michael Somerville. "Techno-economic and environmental evaluation of biomass dryer". In: *Procedia Eng.* 56 (2013), pp. 650–655. DOI: 10.1016/j.proeng.2013.03.173.

[316] S. Sokhansanj and J. Fenton. "Cost benefit of biomass supply and pre-processing". In: *Res. Integr. Progr.* March (2006).

[317] L. Tabil and S. Sokhansanj. "Process conditions affecting the physical quality of alfalfa pellets". In: *Appl. Eng. Agric.* 12.3 (1996), pp. 345–350. DOI: 10.13031/2013.25658.

[318] Neal A. Yancey, Jaya Shankar Tumuluru, and Christopher T. Wright. "Drying, grinding and pelletization studies on raw and formulated biomass feedstock's for bioenergy applications". In: *J. Biobased Mater. Bioenergy* 7 (2013), pp. 549–558. DOI: 10.1166/jbmb.2013.1390.

[319] Hassan Shahrukh, Adetoyese Olajire Oyedun, Amit Kumar, Bahman Ghiasi, Linoj Kumar, and Shahab Sokhansanj. "Techno-economic assessment of pellets produced from steam pretreated biomass feedstock". In: *Biomass and Bioenergy* 87 (2016), pp. 131–143. DOI: 10.1016/j.biombioe.2016.03.001.

[320] Sudhagar Mani, Shahab Sokhansanj, Xiaotao Bi, and Anthony Turhollow. "Economics of producing fuel pellets from biomass". In: *Appl. Eng. Agric.* 22.3 (2006), pp. 421–426. DOI: 10.13031/2013.20447.

[321] Anthony Nolan, Kevin Mc Donnell, Ger J. Devlin, John P. Carroll, and John Finnan. "Economic Analysis of Manufacturing Costs of Pellet Production in the Republic of Ireland Using Non-Woody Biomass". In: *Open Renew. Energy J.* 3.1 (2010), pp. 1–11. DOI: 10.2174/1876387101003010001.

[322] Jaya Shankar Tumuluru, Christopher T. Wright, J. Richard Hess, and Kevin L. Kenney. "A review of biomass densification systems to develop uniform feedstock commodities for bioenergy application". In: *Biofuels, Bioprod. Biorefining* 5.6 (Nov. 2011), pp. 683–707. DOI: 10.1002/BBB.324.

[323] S. M. Haefele, Y. Konboon, W. Wongboon, S. Amarante, A. A. Maarifat, E. M. Pfeiffer, and C. Knoblauch. "Effects and fate of biochar from rice residues in rice-based systems". In: *F. Crop. Res.* 121.3 (2011), pp. 430–440. DOI: 10.1016/j.fcr.2011.01.014.

[324] K. Vijayaraghavan. "Recent advancements in biochar preparation, feedstocks, modification, characterization and future applications". In: *Environ. Technol. Rev.* 8.1 (2019), pp. 47–64. DOI: 10.1080/21622515.2019.1631393.

[325] Aisha Al-Rumaihi, Muhammad Shahbaz, Gordon Mckay, Hamish Mackey, and Tareq Al-Ansari. "A review of pyrolysis technologies and feedstock: A blending approach for plastic and biomass towards optimum biochar yield". In: *Renew. Sustain. Energy Rev.* 167.May (2022), p. 112715. DOI: 10.1016/j.rser.2022.112715.

[326] Gajanan Sampatrao Ghodake, Surendra Krushna Shinde, Avinash Ashok Kadam, Rijuta Ganesh Saratale, Ganesh Dattatraya Saratale, Manu Kumar, Ramasubba Reddy Palem, Hind A. AL-Shwaiman, Abdallah M. Elgorban, Asad Syed, and Dae Young Kim. "Review on biomass feedstocks, pyrolysis mechanism and physicochemical properties of biochar: State-of-the-art framework to speed up vision of circular bioeconomy". In: *J. Clean. Prod.* 297 (2021), p. 126645. DOI: 10.1016/j.jclepro.2021.126645.

[327] Prakash Parthasarathy, Tareq Al-Ansari, Hamish R. Mackey, K. Sheeba Narayanan, and Gordon McKay. "A review on prominent animal and municipal wastes as potential feedstocks for solar pyrolysis for biochar production". In: *Fuel* 316 (2022), p. 123378. DOI: 10.1016/j.fuel.2022.123378.

[328] Jinyang Wang, Zhengqin Xiong, and Yakov Kuzyakov. "Biochar stability in soil: Meta-analysis of decomposition and priming effects". In: *GCB Bioenergy* 8 (2016), pp. 512–523. DOI: 10.1111/gcbb.12266.

[329] IPCC. *Method for estimating the change in mineral soil organic carbon stocks from biochar amendments: basis for future methodological development*. Tech. rep. 2019. https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4%7B%5C_%7DVolume4/19R%7B%5C_%7DV4%7B%5C_%7DCh02%7B%5C_%7DAp4%7B%5C_%7DBiochar.pdf.

[330] Emanuele Lugato, Lily Paniagua, Arwyn Jones, Wim De Vries, and Adrian Leip. "Complementing the topsoil information of the Land Use/Land Cover Area Frame Survey (LUCAS) with modelled N2O emissions". In: *PLoS One* 12.4 (2017), pp. 1–16. DOI: 10.1371/journal.pone.0176111.

[331] Gabriel D. Oreggioni, Bhawna Singh, Francesco Cherubini, Geoffrey Guest, Carine Lausselet, Mauro Luberti, Hyungwoong Ahn, and Anders H. Strømman. "Environmental assessment of biomass gasification combined heat and power plants with

absorptive and adsorptive carbon capture units in Norway". In: *Int. J. Greenh. Gas Control* 57 (2017), pp. 162–172. DOI: 10.1016/j.ijggc.2016.11.025.

[332] Bhawna Singh, Anders H. Strømman, and Edgar G. Hertwich. "Comparative life cycle environmental assessment of CCS technologies". In: *Int. J. Greenh. Gas Control* 5 (2011), pp. 911–921. DOI: 10.1016/j.ijggc.2011.03.012.

[333] Joris Koornneef, Tim van Keulen, André Faaij, and Wim Turkenburg. "Life cycle assessment of a pulverized coal power plant with post-combustion capture, transport and storage of CO2". In: *Int. J. Greenh. Gas Control* 2 (2008), pp. 448–467. DOI: 10.1016/j.ijggc.2008.06.008.

[334] Hidetaka Yamada. "Amine-based capture of CO2 for utilization and storage". In: *Polym. J.* 53 (2021), pp. 93–102. DOI: 10.1038/s41428-020-00400-y.

[335] Rafael M. Siqueira, Geovane R. Freitas, Hugo R. Peixoto, Jailton F.Do Nascimento, Ana Paula S. Musse, Antonio E.B. Torres, Diana C.S. Azevedo, and Moises Bastos-Neto. "Carbon Dioxide Capture by Pressure Swing Adsorption". In: *Energy Procedia* 114 (2017), pp. 2182–2192. DOI: 10.1016/j.egypro.2017.03.1355.

[336] Rohan Stanger, Terry Wall, Reinhold Spörl, Manoj Paneru, Simon Grathwohl, Max Weidmann, Günter Scheffknecht, Denny McDonald, Kari Myöhänen, Jouni Ritvanen, Sirpa Rahiala, Timo Hyppänen, Jan Mletzko, Alfons Kather, and Stanley Santos. "Oxyfuel combustion for CO2 capture in power plants". In: *Int. J. Greenh. Gas Control* 40 (2015), pp. 55–125. DOI: 10.1016/j.ijggc.2015.06.010.

[337] Michael J Greencorn, S David Jackson, Justin S J Hargreaves, Souvik Datta, and C Paul. "Enhancement of gasification in oxyfuel BECCS cycles employing a direct recycling CO2 utilisation process". In: *Energy Convers. Manag.* 277 (2023), p. 116601. DOI: 10.1016/j.enconman.2022.116601.

[338] Hyungwoong Ahn, Mauro Luberti, Zhengyi Liu, and Stefano Brandani. "Process configuration studies of the amine capture process for coal-fired power plants". In: *Int. J. Greenh. Gas Control* 16 (2013), pp. 29–40. DOI: 10.1016/j.ijggc.2013.03.002.

[339] Gabriel David Oreggioni, Stefano Brandani, Mauro Luberti, Yusuf Baykan, Daniel Friedrich, and Hyungwoong Ahn. "CO2 capture from syngas by an adsorption process at a biomass gasification CHP plant: Its comparison with amine-based CO2 capture". In: *Int. J. Greenh. Gas Control* 35 (2015), pp. 71–81. DOI: 10.1016/j.ijggc.2015.01.008.

[340] Mauro Luberti, Gabriel David Oreggioni, and Hyungwoong Ahn. "Design of a rapid vacuum pressure swing adsorption (RVPSA) process for post-combustion CO2 capture from a biomass-fuelled CHP plant". In: *J. Environ. Chem. Eng.* 5 (2017), pp. 3973–3982. DOI: 10.1016/j.jece.2017.07.029.

[341] Mohammed Al-Juaied and Adam Whitmore. "Realistic Costs of Carbon Capture". In: *Discuss. Pap. 2009-08, Cambridge, Mass. Belfer Cent. Sci. Int. Aff.* 2009. http://belfercenter.ksg.harvard.edu/files/2009%7B%5C_%7DAlJuaied%7B%5C_%7DWhitmore%7B%5C_%7DRealistic%7B%5C_%7DCosts%7B%5C_%7Dof%7B%5C_%7DCarbon%7B%5C_%7DCapture%7B%5C_%7Dweb.pdf.

[342] Ricardo Energy & Environment. *Analysing the potential of bioenergy with carbon capture in the UK to 2050*. Tech. rep. Ricardo Energy & Environment, 2020.

[343] Yuan Jiang, Paul M. Mathias, Charles J. Freeman, Joseph A. Swisher, Richard F. Zheng, Greg A. Whyatt, and David J. Heldebrant. "Techno-economic comparison of various process configurations for post-combustion carbon capture using a single-component water-lean solvent". In: *Int. J. Greenh. Gas Control* 106.October 2020 (2021). DOI: 10.1016/j.ijggc.2021.103279.

[344] Edward S. Rubin, John E. Davison, and Howard J. Herzog. "The cost of CO2 capture and storage". In: *Int. J. Greenh. Gas Control* 40 (2015), pp. 378–400. DOI: 10.1016/j.ijggc.2015.05.018.

[345] Jordan Kearns, Gary Teletzke, Jeffrey Palmer, Hans Thomann, Haroon Kheshgi, Yen Heng Henry Chen, Sergey Paltsev, and Howard Herzog. "Developing a Consistent Database for Regional Geologic CO2 Storage Capacity Worldwide". In: *Energy Procedia* 114.617 (2017), pp. 4697–4709. DOI: 10.1016/j.egypro.2017.03.1603.

[346] Tryfonas Pieri, Alexandros Nikitas, Arturo Castillo-Castillo, and Athanasios Angelis-Dimakis. "Holistic assessment of carbon capture and utilization value chains". In: *Environ. - MDPI* 5.10 (2018), pp. 1–17. DOI: 10.3390/environments5100108.

[347] George S. Goff and Gary T. Rochelle. "Monoethanolamine degradation: O2 mass transfer effects under CO2 capture conditions". In: *Ind. Eng. Chem. Res.* 43.20 (2004), pp. 6400–6408. DOI: 10.1021/ie0400245.

[348] D. Singh, E. Croiset, P. L. Douglas, and M. A. Douglas. "Techno-economic study of CO2 capture from an existing coal-fired power plant: MEA scrubbing vs. O2/CO2 recycle combustion". In: *Energy Convers. Manag.* 44.19 (2003), pp. 3073–3091. DOI: 10.1016/S0196-8904(03)00040-2.

[349] Brian R. Strazisar, Richard R. Anderson, and Curt M. White. "Degradation pathways for monoethanolamine in a CO2 capture facility". In: *Energy and Fuels* 17.4 (2003), pp. 1034–1039. DOI: 10.1021/ef020272i.

[350] Cristian Dinca, Nela Slavu, Călin Cristian Cormoş, and Adrian Badea. "CO2 capture from syngas generated by a biomass gasification power plant with chemical absorption process". In: *Energy* 149 (2018), pp. 925–936. DOI: 10.1016/j.energy.2018.02.109.

[351] Inna Kim and Hallvard F. Svendsen. "Heat of absorption of carbon dioxide (CO2) in monoethanolamine (MEA) and 2-(aminoethyl)ethanolamine (AEEA) solutions". In: *Ind. Eng. Chem. Res.* 46.17 (2007), pp. 5803–5809. DOI: 10.1021/ie0616489.

[352] Li Zhao, Ruikai Zhao, Shuai Deng, Yuting Tan, and Yinan Liu. "Integrating solar Organic Rankine Cycle into a coal-fired power plant with amine-based chemical absorption for CO2 capture". In: *Int. J. Greenh. Gas Control* 31 (2014), pp. 77–86. DOI: 10.1016/j.ijggc.2014.09.025.

[353] Businessanalytiq.com. *Monoethanolamine price index*. 2023. https://businessanalytiq.com/procurementanalytics/index/monoethanolamine-price-index/ (visited on 09/13/2023).

[354] Ofgem. *Wholesale market indicators*. 2024. https://www.ofgem.gov.uk/energy-data-and-research/data-portal/wholesale-market-indicators (visited on 02/08/2024).

[355] A. Ahamed, K. Yin, B. J.H. Ng, F. Ren, V. W.C. Chang, and J. Y. Wang. "Life cycle assessment of the present and proposed food waste management technologies from environmental and economic impact perspectives". In: *J. Clean. Prod.* 131 (2016), pp. 607–614. DOI: 10.1016/j.jclepro.2016.04.127.

[356] A. Wrobel-Tobiszewska, M. Boersma, J. Sargison, P. Adams, and S. Jarick. "An economic analysis of biochar production using residues from Eucalypt plantations". In: *Biomass and Bioenergy* 81 (2015), pp. 177–182. DOI: 10.1016/j.biombioe.2015.06.015.

[357] Neil Gray. *Update on Scotland's renewables and wind power potential - Energy Secretary writes to the UK Scottish Affairs Committee*. 2023. https://www.gov.scot/publications/renewables-and-wind-power-update-to-scottish-affairs-committee/%7B%5C#%7Dpage-top.

[358] Forestry Commission Scotland. *Biomass Action Plan for Scotland*. Tech. rep. 2007, pp. 41–51.

[359] T.C. Smout. *Scottish Woodland History*. Edinburgh: Scottish Cultural Press, 1997.

[360] Scottish Government. *Scotland's Forestry Strategy*. Tech. rep. 2019, p. 60. https://www.gov.scot/binaries/content/documents/govscot/publications/publication/2019/02/scotlands-forestry-strategy-20192029/documents/scotlands-forestry-strategy-2019-2029/scotlands-forestry-strategy-2019-2029/govscot%7B%5C%%7D3Adocument.

[361] NFUScotland. *Farming Facts - What We Produce*. 2024. https://www.nfus.org.uk/farming-facts/what-we-produce.aspx (visited on 02/22/2024).

[362] Scottish Government. *Exports statistics Scotland 2021*. 2021. https://www.gov.scot/publications/exports-statistics-scotland-2021/ (visited on 02/22/2024).

[363] Julian Bell, John Farquhar, and Mary Mcdowell. *Distillery by-products, livestock feed and bio-energy use in Scotland*. Tech. rep. 2019. https://www.gov.scot/publications/distillery-products-livestock-feed-bio-energy-use-scotland/pages/12/%7B%5C#%7D:%7B~%7D:text=Anaerobic%20Digestion%20plants%20use%20a,and%20purpose%20grown%20energy%20crops.%7B%5C&%7Dtext=Energy%20can%20be%20in%20the,methane%20(%7B~%7D96%7B%5C%%7D25%20methane)..

[364] Chris Malins, Stephanie Searle, Anil Baral, David Turley, and Lucy Hopwoodi. *Wasted - Europe's untapped resource: An Assessment of Advanced Biofuels from Wastes and Residues*. Tech. rep. International Council on Clean Transportation, 2014, pp. 1–29. https://europeanclimate.org/wp-content/uploads/2014/02/WASTED-final.pdf.

[365] Jane S. White, Kelly L. Stewart, Dawn L. Maskell, Aboubakry Diallo, Julio E. Traub-Modinger, and Nik A. Willoughby. "Characterization of Pot Ale from a Scottish Malt Whisky Distillery and Potential Applications". In: *ACS Omega* 5.12 (2020), pp. 6429–6440. DOI: 10.1021/acsomega.9b04023.

[366] Simon Maxwell. *Forestry Statistics 2023*. Forest Research, 2023. https://www.forestresearch.gov.uk/tools-and-resources/statistics/forestry-statistics/.

[367] Scottish Government. *Cereal and oilseed rape harvest: final estimates - 2023*. 2023. https://www.gov.scot/publications/cereal-and-oilseed-rape-harvest-2023-final-estimates/pages/composition-of-cereal-production-in-2023/ (visited on 02/22/2024).

[368] DEFRA. *Crops Grown For Bioenergy in the UK: 2018*. Tech. rep. 10th December 2020. Department for Environment, Food and Rural Affairs, 2019, p. 36. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment%7B%5C_%7Ddata/file/856695/nonfood-statsnotice2018-08jan2

0.pdf%7B%5C%7D0Ahttps://www.gov.uk/government/statistics/area-of-crops-grown-for-bioenergy-in-england-and-the-uk-2008-2019.

[369] Phyllis. *Phyllis2*. 2024. https://phyllis.nl/Browse/Standard/ECN-Phyllis%7B%5C#%7D1882 (visited on 02/23/2024).

[370] Daniel Harari. *Interest Rates and Monetary Policy: Key Economic Indicators*. 2024. https://commonslibrary.parliament.uk/research-briefings/sn02802/ (visited on 02/23/2023).

[371] Bank of England. *Official Bank Rate History*. 2024. https://www.bankofengland.co.uk/boeapps/database/Bank-Rate.asp (visited on 02/23/2024).

[372] Statista. *Wholesale price projection of electricity in the United Kingdom (UK) from 2022 to 2040*. 2024. https://www.statista.com/statistics/496328/electricity-wholesale-prices-projection-uk/ (visited on 02/23/2024).

[373] Elena Dulys-Nusbaum, Sarah S.H. Klammer, and Scott M. Swinton. "How willing are different types of landowner to supply hardwood timber residues for bioenergy?" In: *Biomass and Bioenergy* 122.January (2019), pp. 45–52. DOI: 10.1016/j.biombioe.2019.01.026.

[374] Fasque Timber Products Ltd. *Bulk products*. 2024. https://fasquetimberproducts.co.uk/collections/bulk-products (visited on 02/24/2024).

[375] AHDB. *Hay and Straw Prices*. 2024. https://ahdb.org.uk/dairy/hay-and-straw-prices (visited on 02/23/2024).

[376] Julian Bell, Gillian Innman, and Mary McDowell. *Straw and Forage Study*. Tech. rep. March. Scottish Government, 2018.

[377] Quantum Commodity Intelligence. *ANALYSIS: Long-term biochar carbon price already below $100/tCO2e, say developersQuantum*. 2023. https://www.qcintel.com/carbon/article/analysis-long-term-biochar-carbon-price-already-below-100-tco2e-say-developers-17670.html (visited on 02/26/2024).

[378] Biochar Demonstrator. *Information for Farmers*. 2024. https://biochardemonstrator.ac.uk/info-for-farmers/%7B%5C#%7D (visited on 02/26/2024).

[379] Department for Energy Security & Net Zero. *UK ETS: Carbon prices for use in civil penalties, 2024*. 2024. https://www.gov.uk/government/publications/determinations-of-the-uk-ets-carbon-price/uk-ets-carbon-prices-for-use-in-civil-penalties-2024 (visited on 02/26/2024).

[380] Chelsea Baldino, Rosalie Berg, Nikita Pavlenko, and Stephanie Searle. *Advanced alternative fuel pathways: Technology overview and status*. Tech. rep. 13. The international council on clean transportation, 2019, pp. 1–31. https://theicct.org/publications/advanced-alternative-fuel-pathways.

[381] European Commission. “Recommendation: 2013/179/EU: on the use of common methods to measure and communicate the life cycle environmental performance of products and organisations”. In: *Off. J. Eur. Union* (2013), p. 216.

[382] International Organization for Standardization (ISO). *Greenhouse gases — Carbon footprint of products — Requirements and guidelines for quantification and communication*. Tech. rep. 2013.

Appendix

A Detailed Screenshots of the Developed Decision Support Tool with a Graphical User Interface (GUI)

This appendix contains screenshots detailing the exact page layout and contents of the developed decision support tool's GUI.

A Model for the Rapid Assessment of the Environmental and Economic Impacts of Biomass and Waste Gasification Schemes

Summary

This work presents a framework to conduct **life cycle assessment** (LCA) and **techno-economic analysis** (TEA) of biomass and waste gasification schemes. Machine learning (ML), in the form of a **gradient boosting** model, is utilised to predict the composition and yields of gasification products. Uncertainties resulting from the machine learning model and other uncertainties throughout the model framework are accounted for through **Monte Carlo simulation** methodology.

How to use the model

The side bar on the left-hand side lets you navigate through the program.

- You will need to define the inputs to your simulation in **Section 1 - Simulation Inputs.** Here you will define information which will be used in the simulation (e.g. the feedstock used, the gasification operating conditions, and how much the generated energy can be sold for. Finally you can download a `.toml` file which can be used to run your the model in **Section 2 - Run Simulation.**
- **Section 2 - Run Simulation** lets you upload the `.toml` user input file generated in the previous section, which will then be used to run the simulation.

Figure A1: Home screen - page 1.

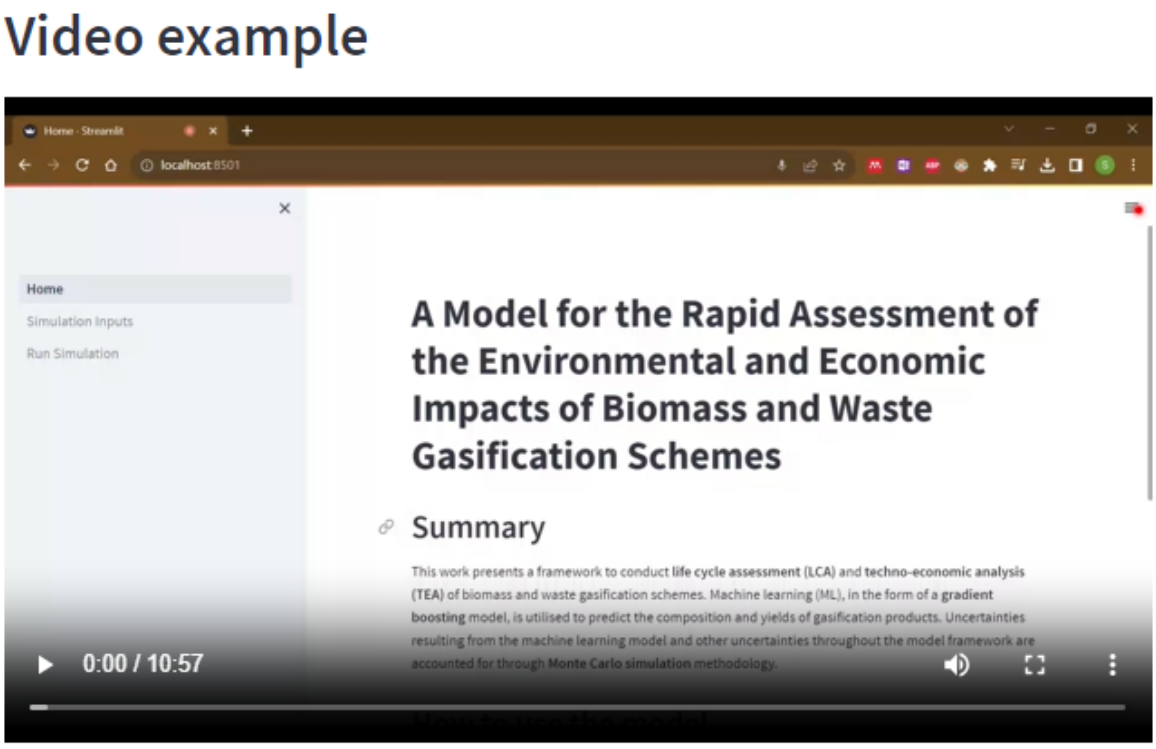


Figure A2: Home screen - page 2.

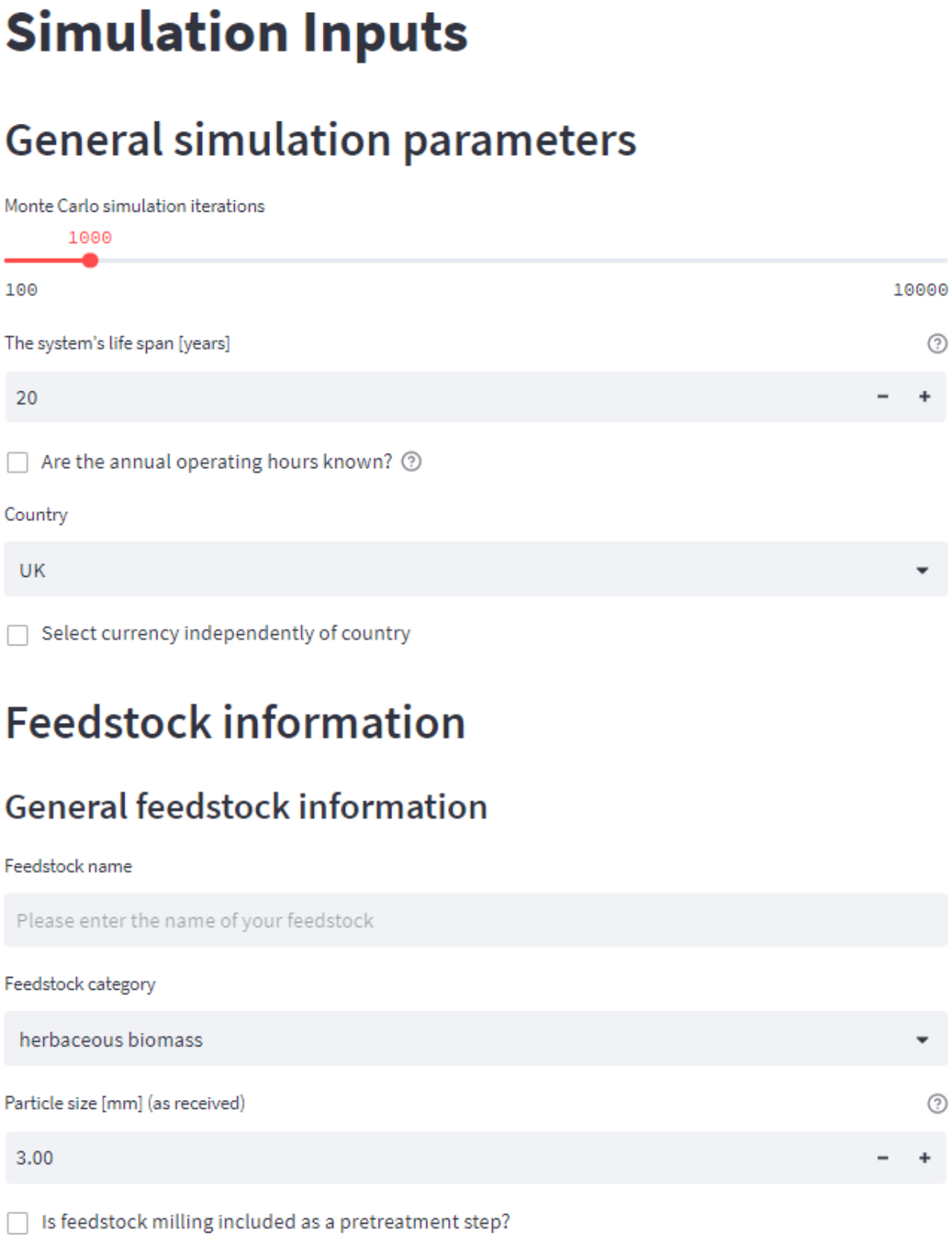


Figure A3: Simulation inputs screen - page 1.

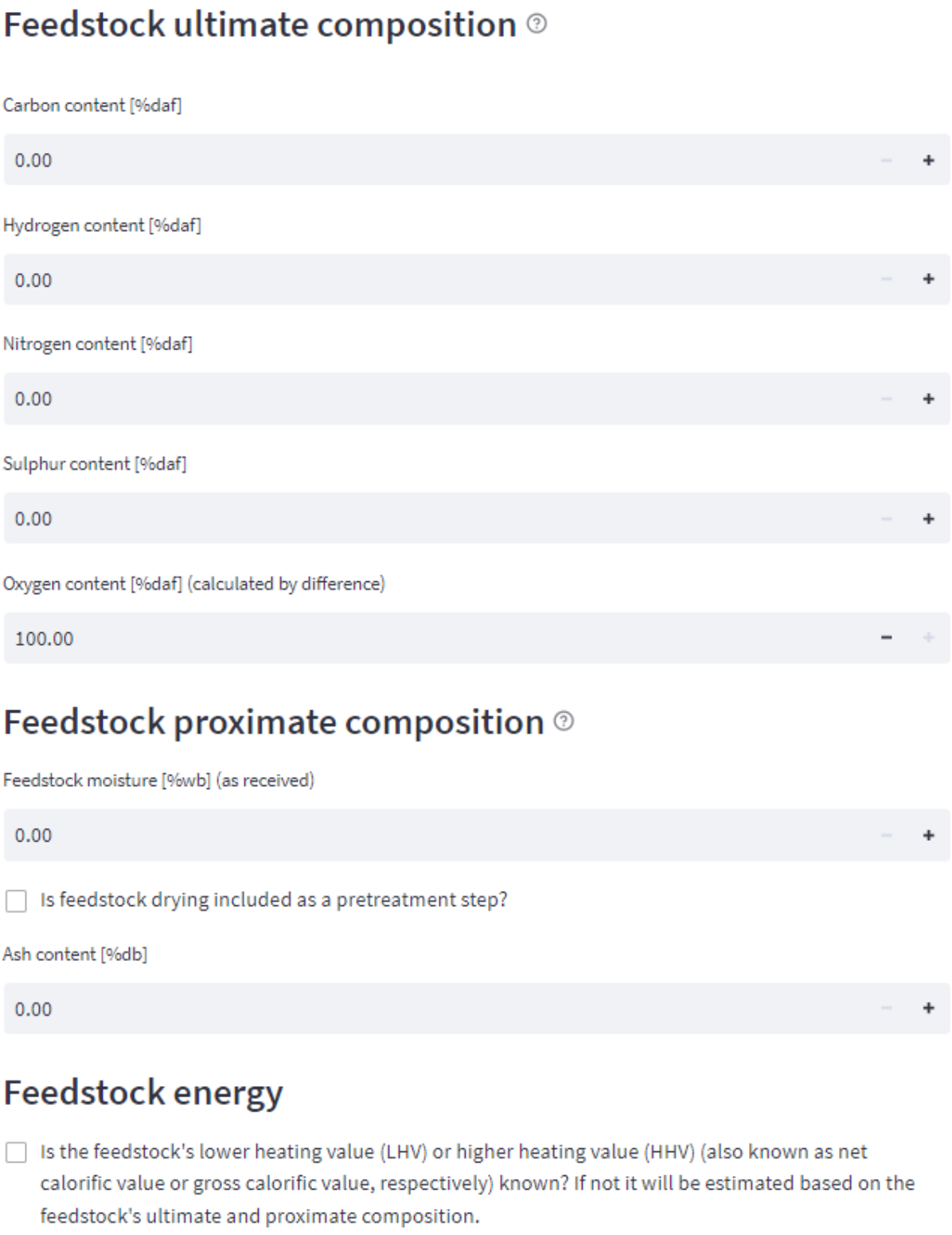


Figure A4: Simulation inputs screen - page 2.

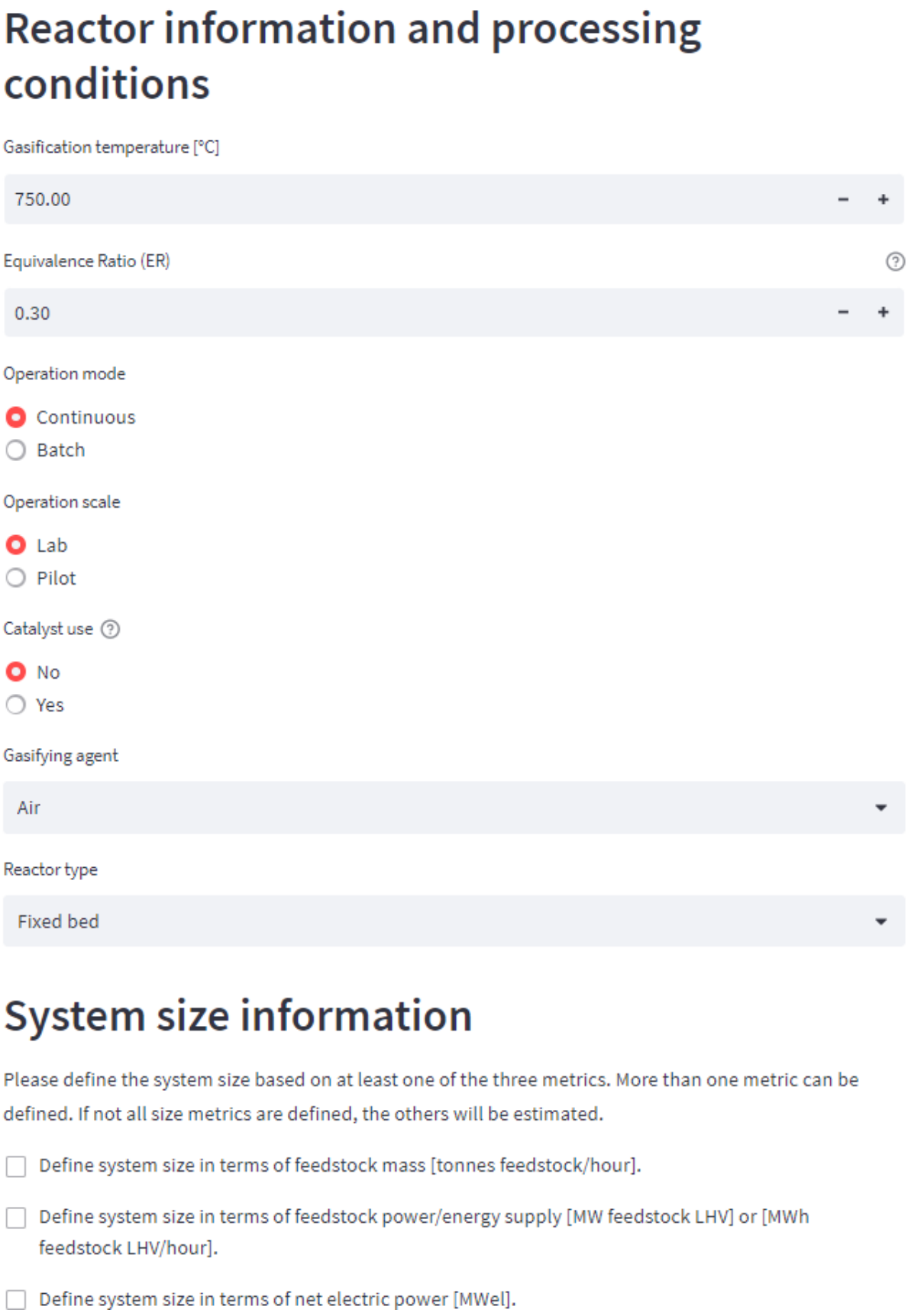


Figure A5: Simulation inputs screen - page 3.

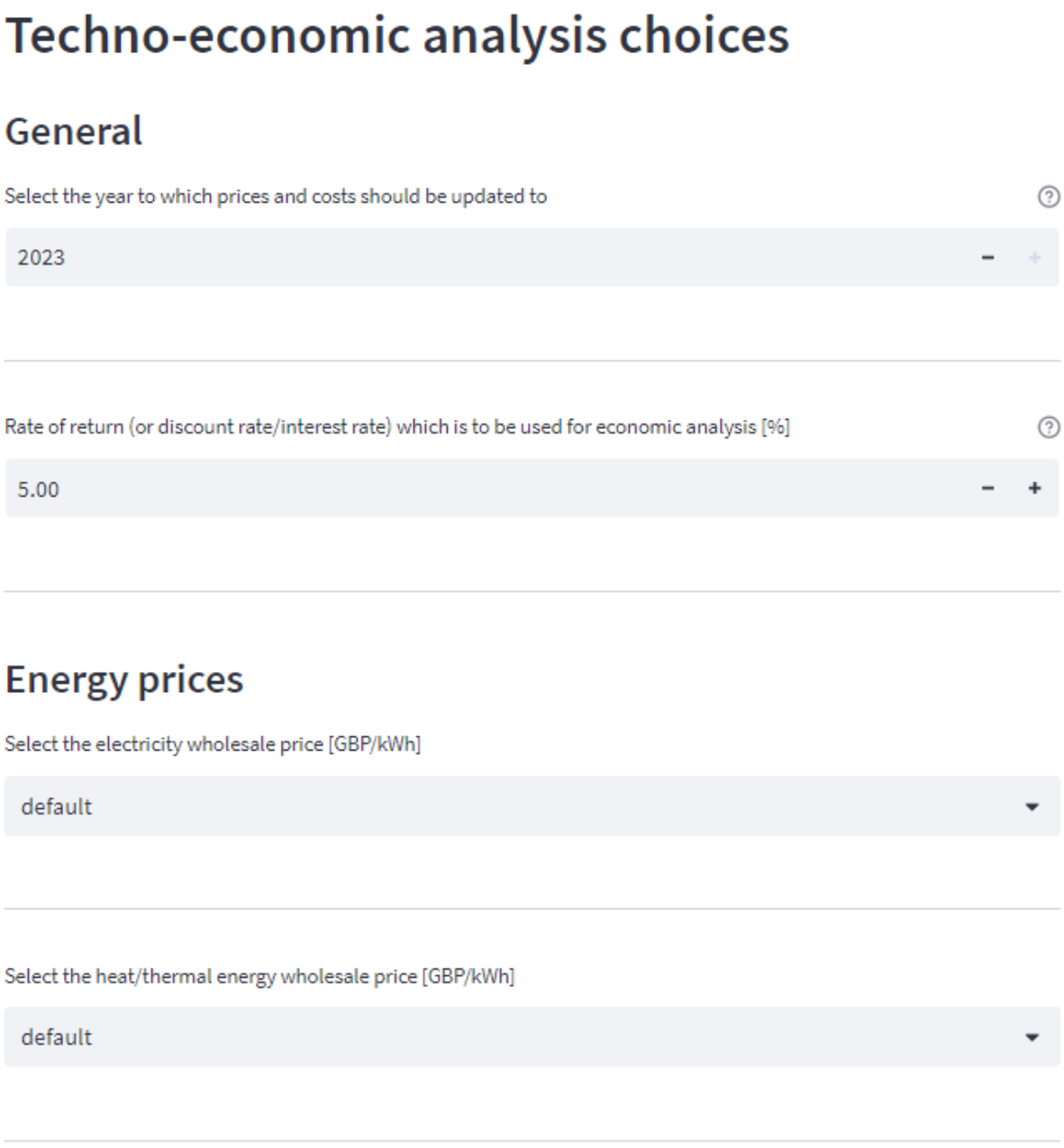


Figure A6: Simulation inputs screen - page 4.

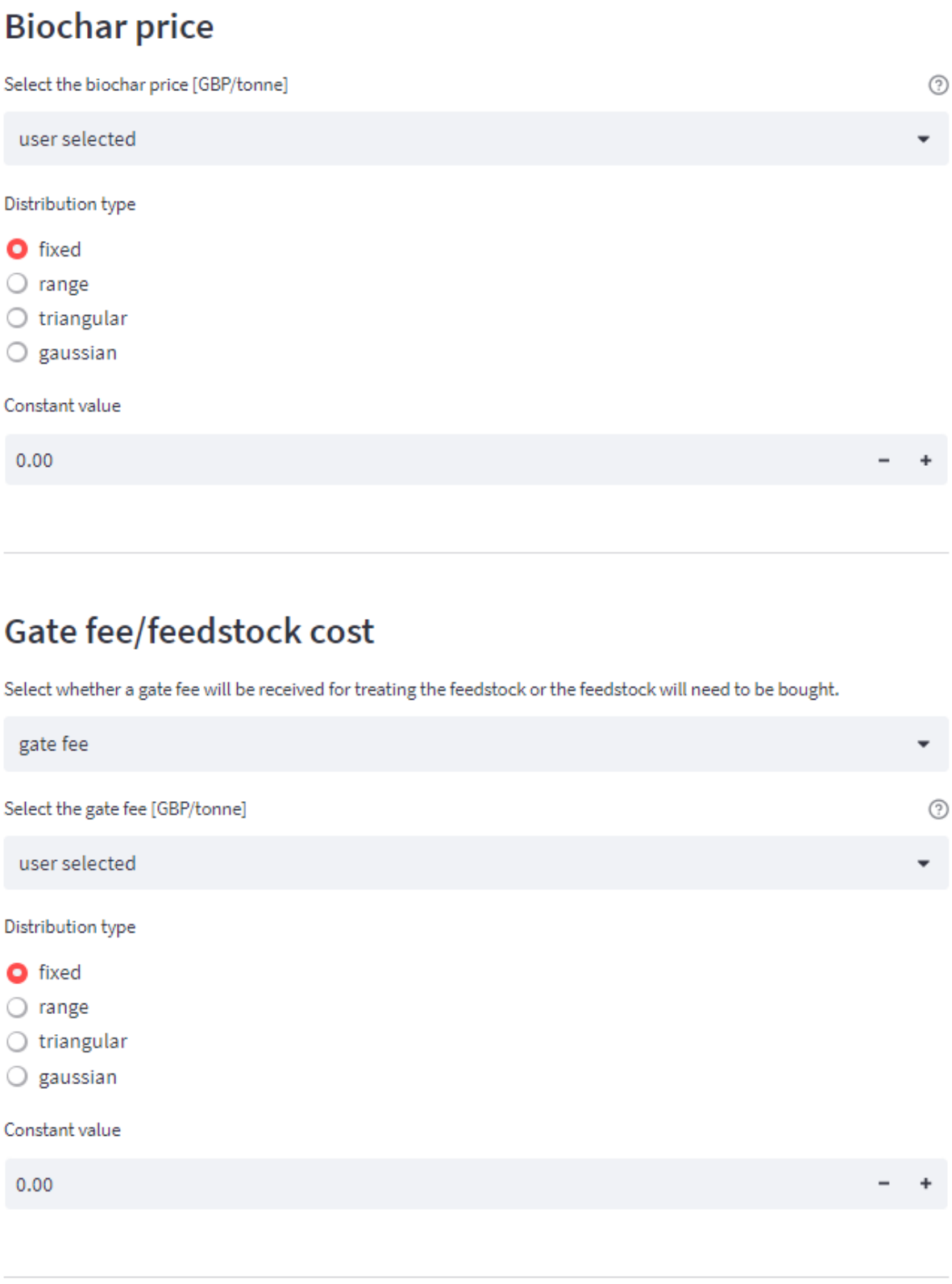


Figure A7: Simulation inputs screen - page 5.

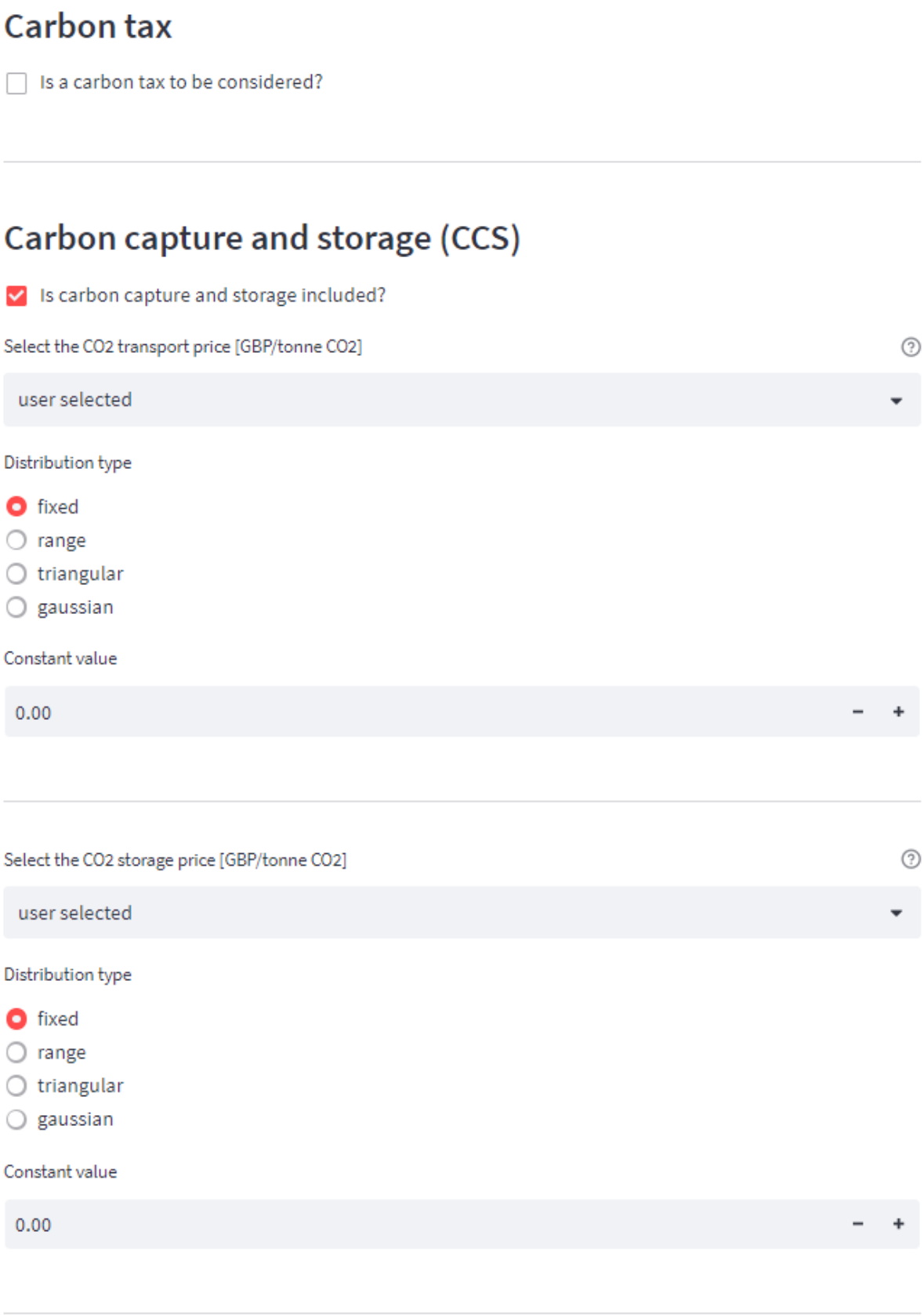


Figure A8: Simulation inputs screen - page 6.

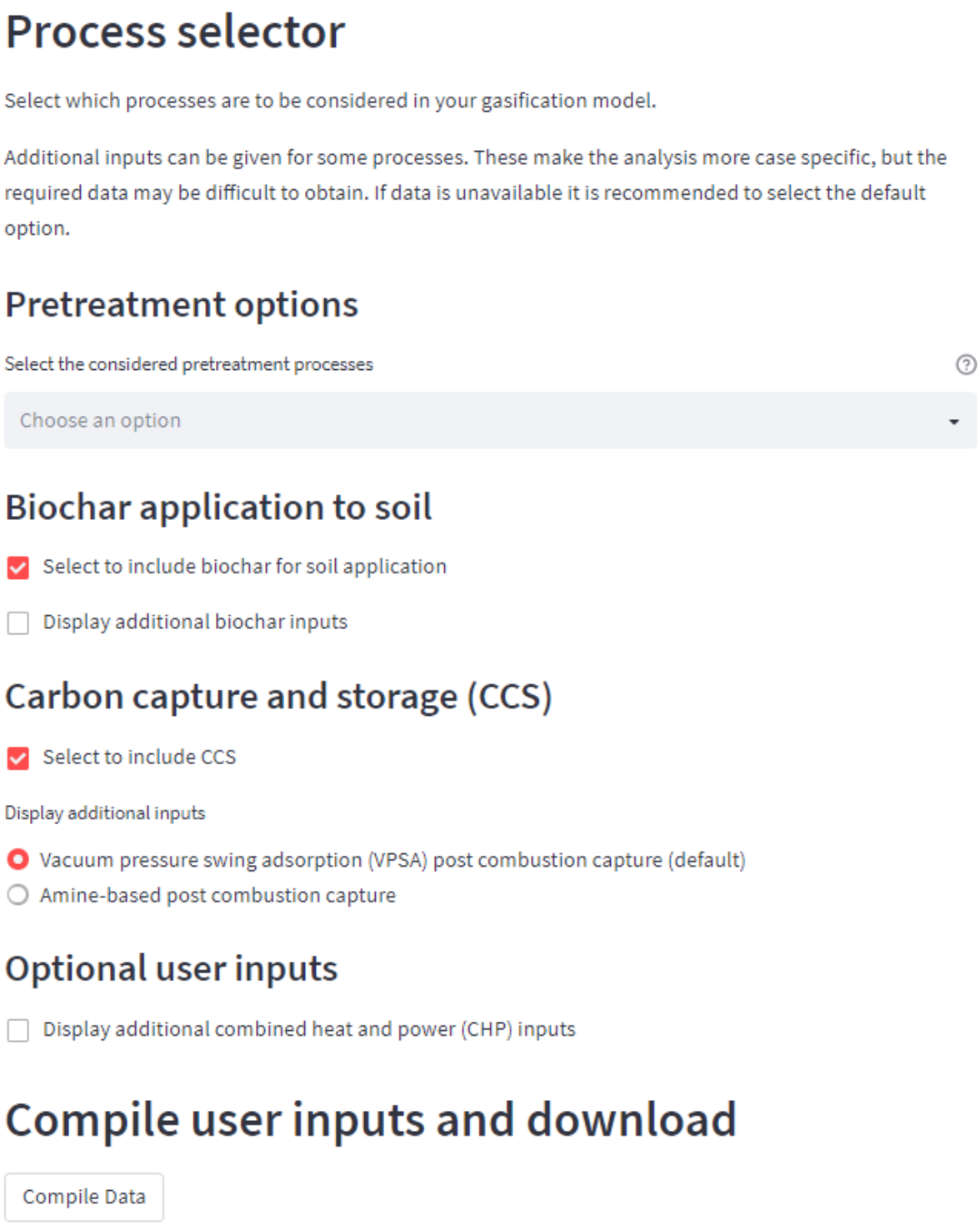


Figure A9: Simulation inputs screen - page 7.

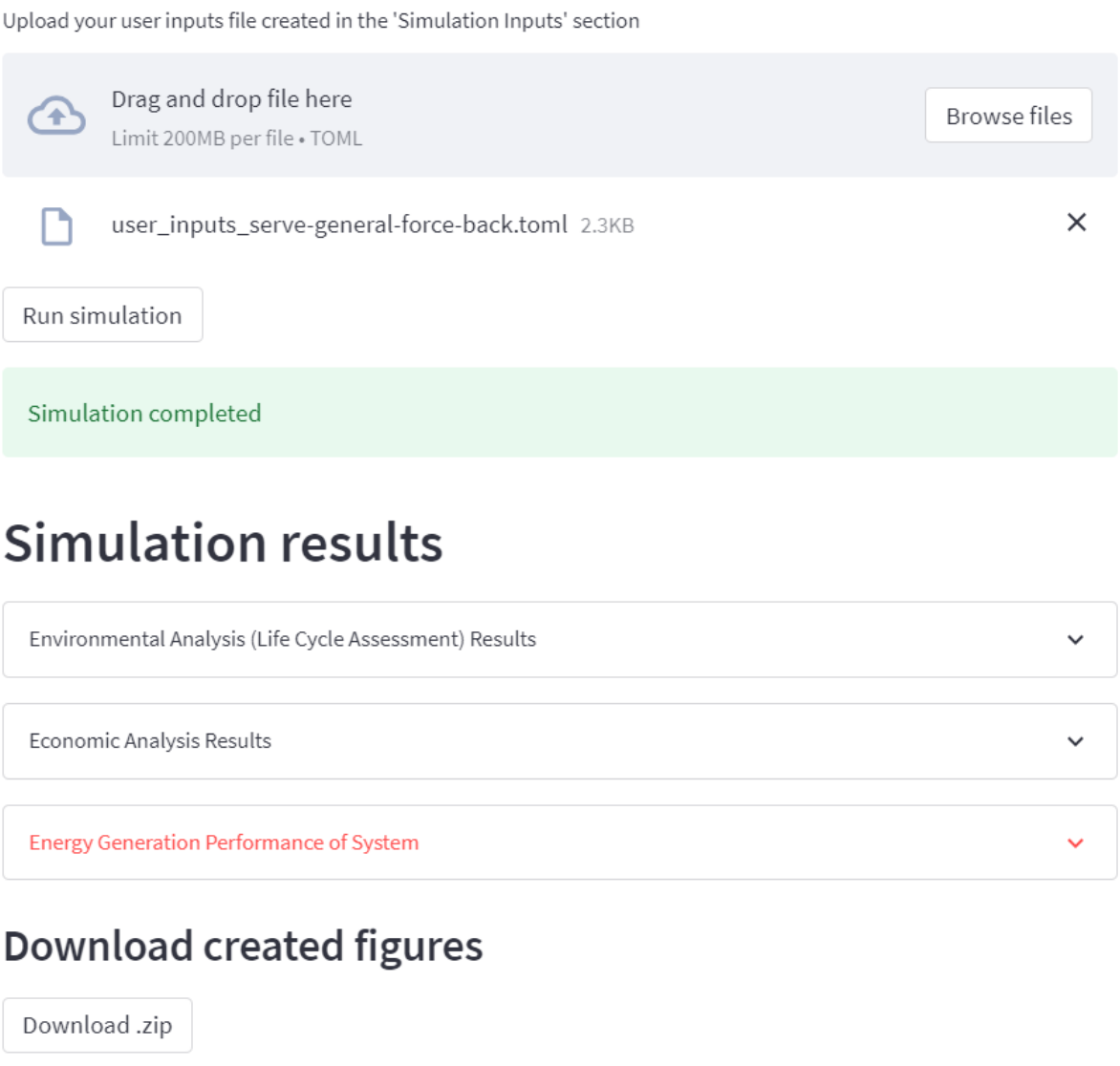
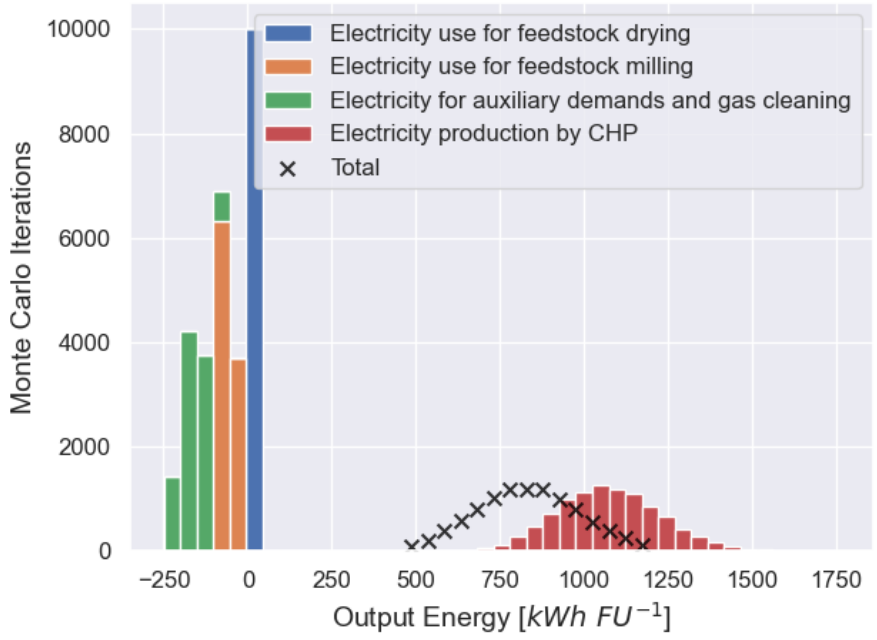


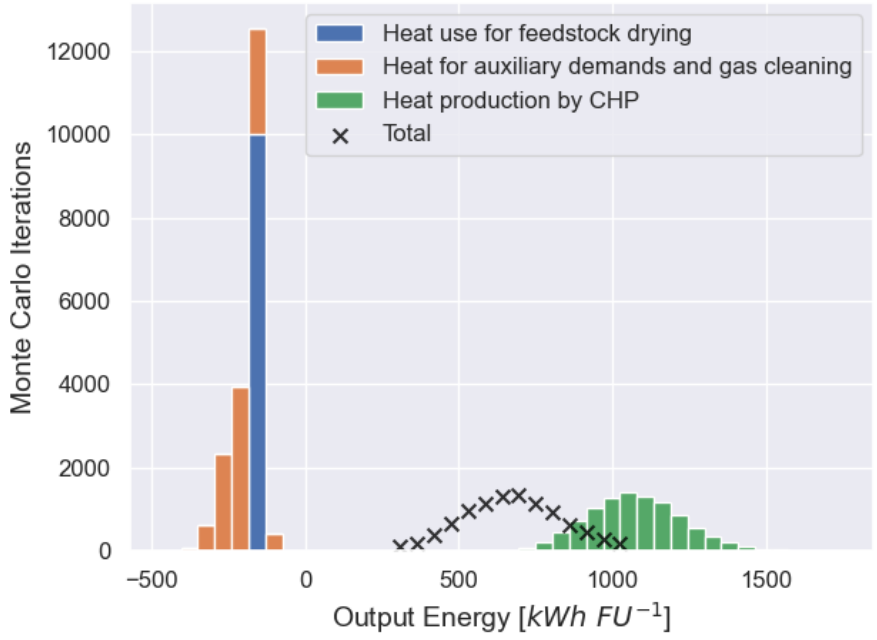
Figure A10: Run simulation screen.

B Baseline Energy Generation Results for Forestry Residues, Barley Straw, and Draff

This appendix contains additional figures showing the energy generation potential of the three feedstocks studied in the case study introduced in Chapter 6. Figures are shown for the baseline scenarios discussed in Section 6.4.1.

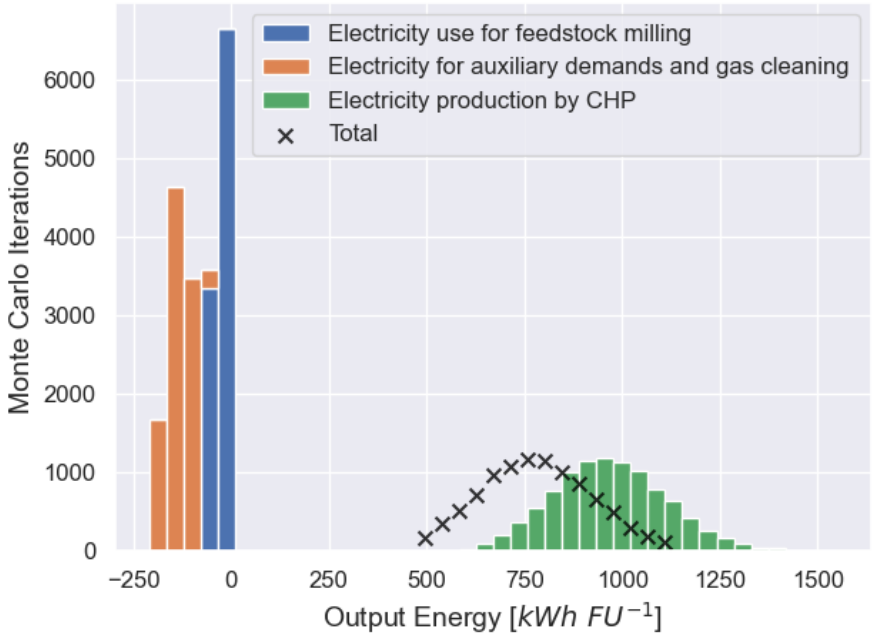


(a)

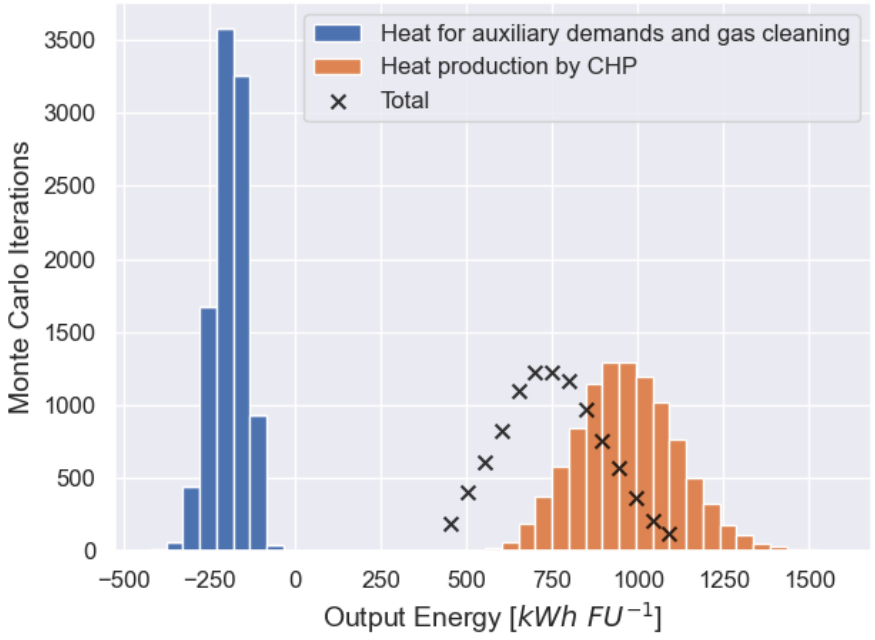


(b)

Figure B1: Baseline energy generation results for forestry residues as the feedstock. Electricity generation is shown by (a) and heat or thermal energy generation by (b)

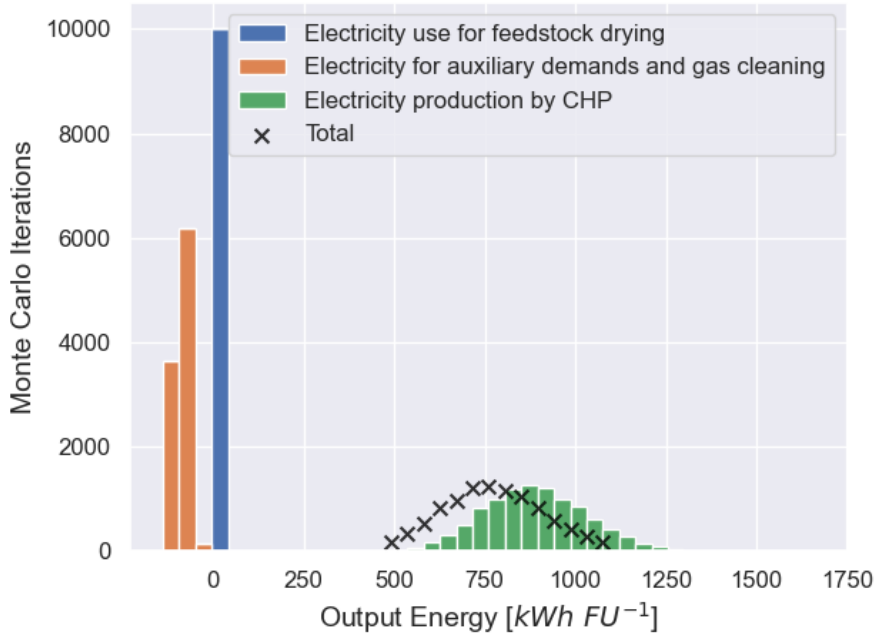


(a)

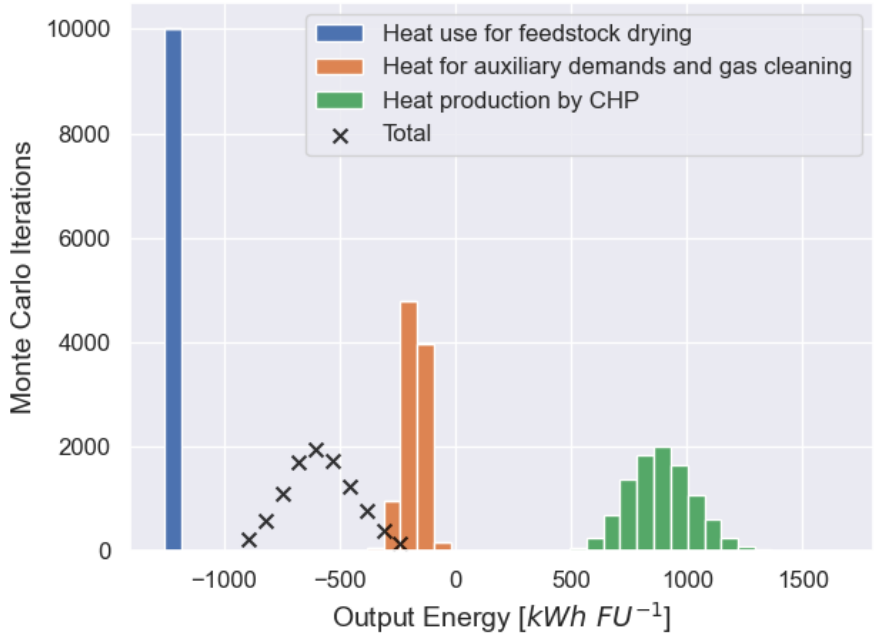


(b)

Figure B2: Baseline energy generation results for barley straw as the feedstock. Electricity generation is shown by (a) and heat or thermal energy generation by (b)



(a)



(b)

Figure B3: Baseline energy generation results for draff as the feedstock. Electricity generation is shown by (a) and heat or thermal energy generation by (b)

C Optimisation Results for Barley Straw and Draff

This appendix contains additional figures illustrating the optimisation results for the case study introduced in Chapter 6. Figures are shown for the two feedstock not discussed in Section 6.4.2 (i.e. barley straw and draff).

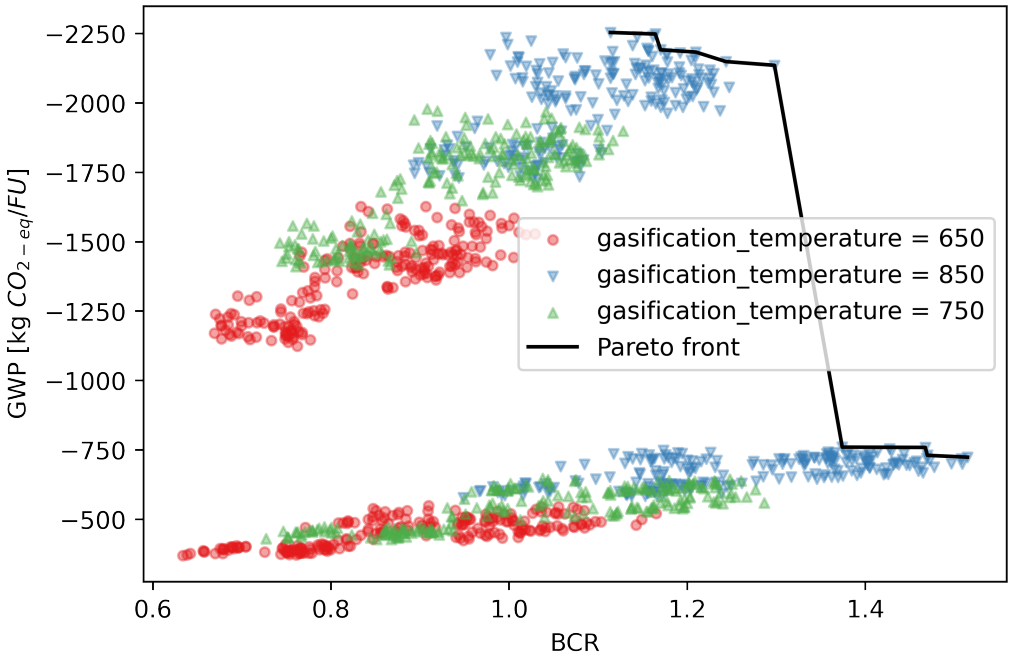


Figure C1: Optimisation results for the feedstock barley straw shown for the model parameter gasification temperature [°C].

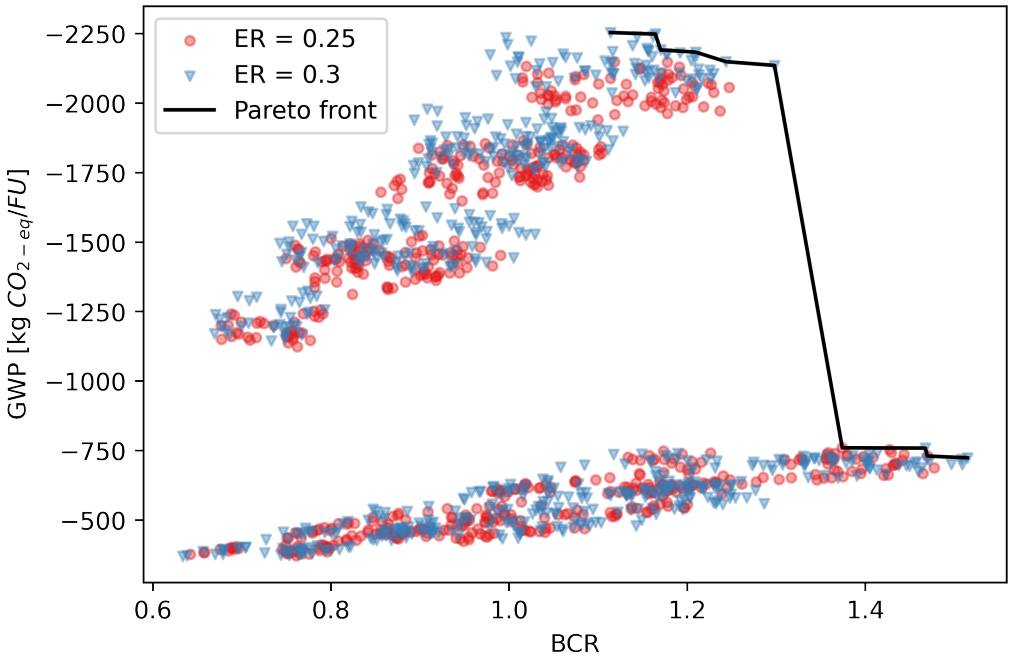


Figure C2: Optimisation results for the feedstock barley straw shown for the model parameter ER.

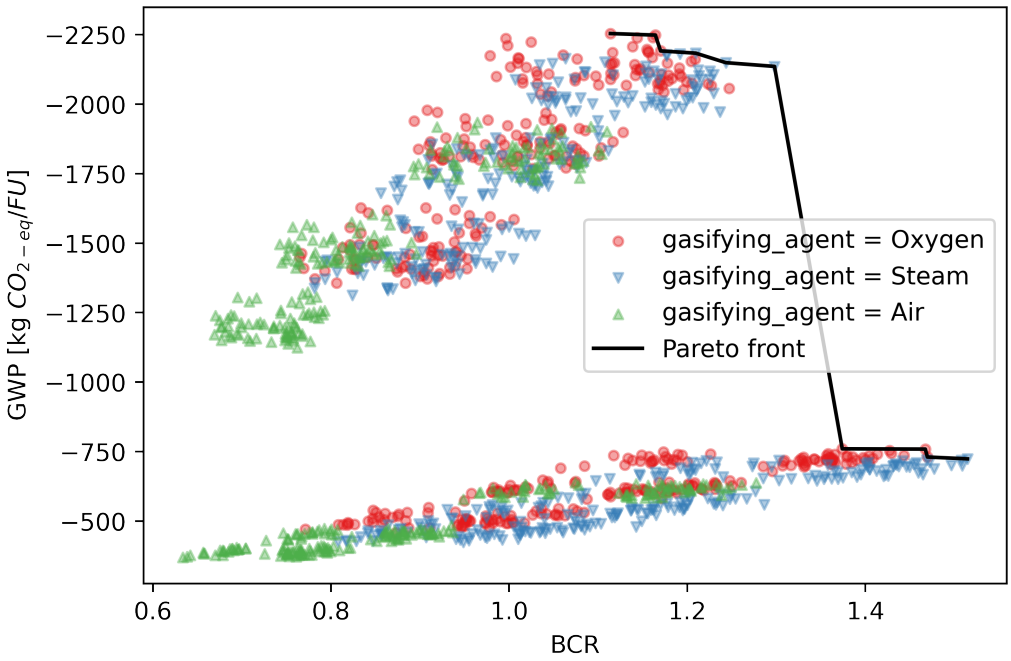


Figure C3: Optimisation results for the feedstock barley straw shown for the model parameter gasifying agent.

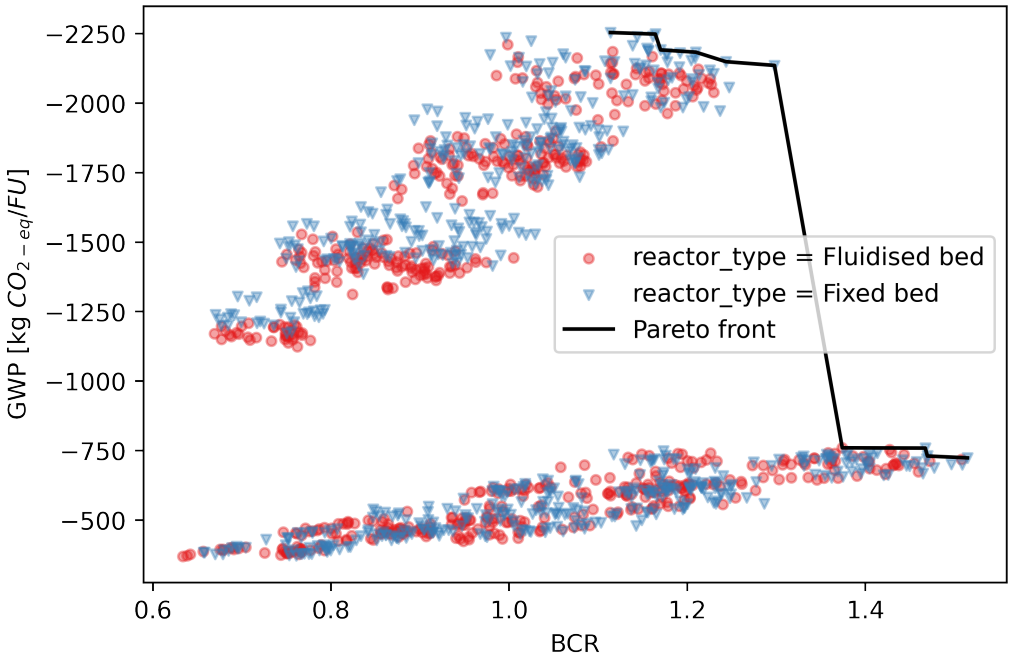


Figure C4: Optimisation results for the feedstock barley straw shown for the model parameter reactor type.

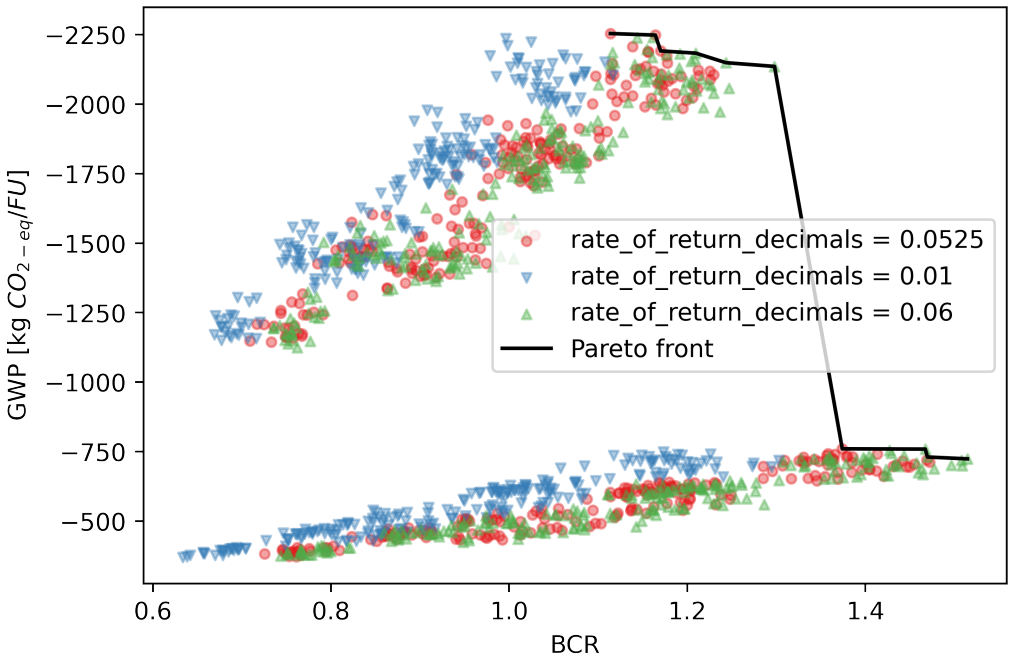


Figure C5: Optimisation results for the feedstock barley straw shown for the model parameter interest rate [%].

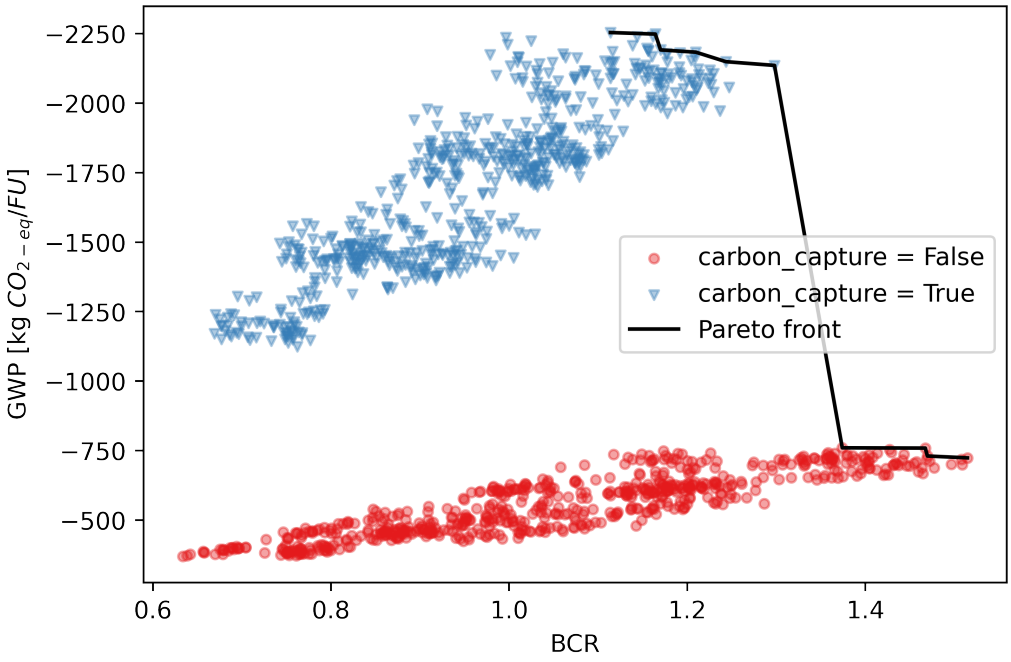


Figure C6: Optimisation results for the feedstock barley straw shown for the model parameter CAPEX (included or not included).

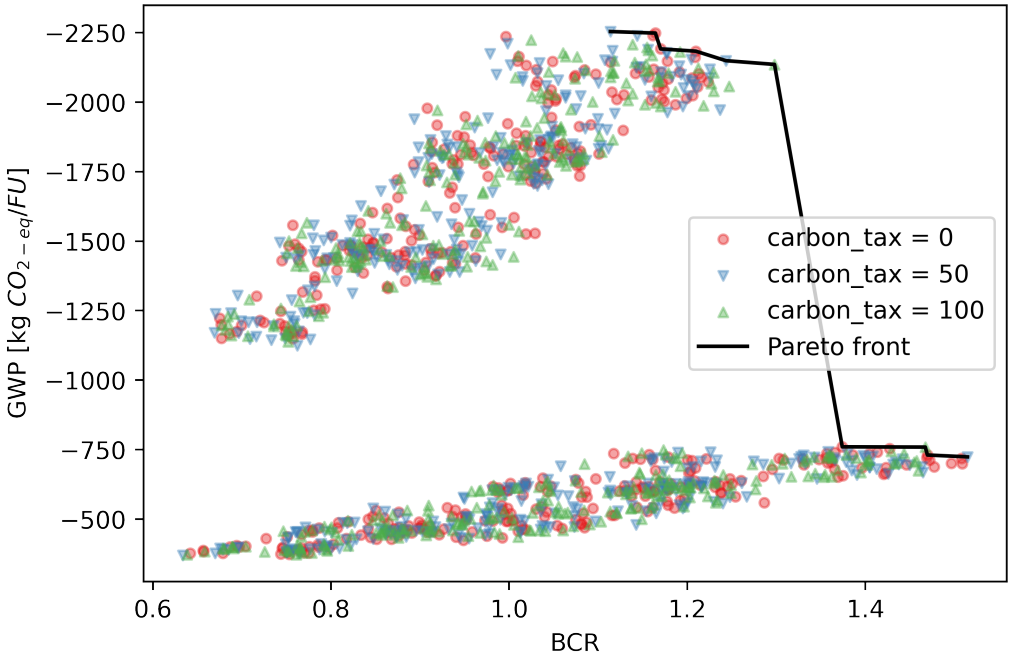


Figure C7: Optimisation results for the feedstock barley straw shown for the model parameter carbon tax [£/tonne CO_2-eq].

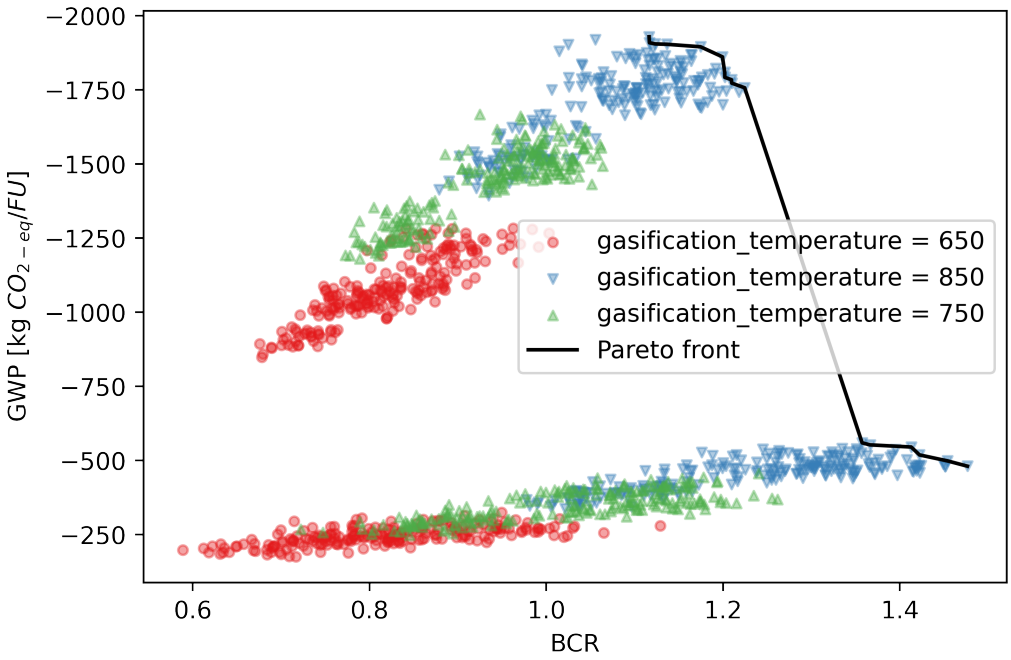


Figure C8: Optimisation results for the feedstock draff shown for the model parameter gasification temperature [°C].

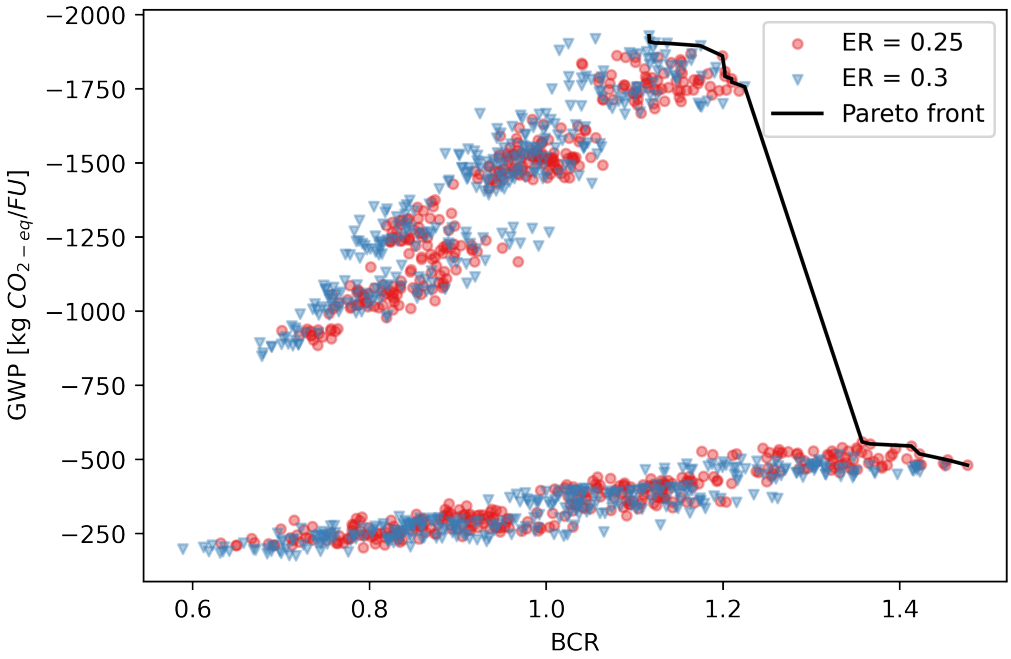


Figure C9: Optimisation results for the feedstock draff shown for the model parameter ER.

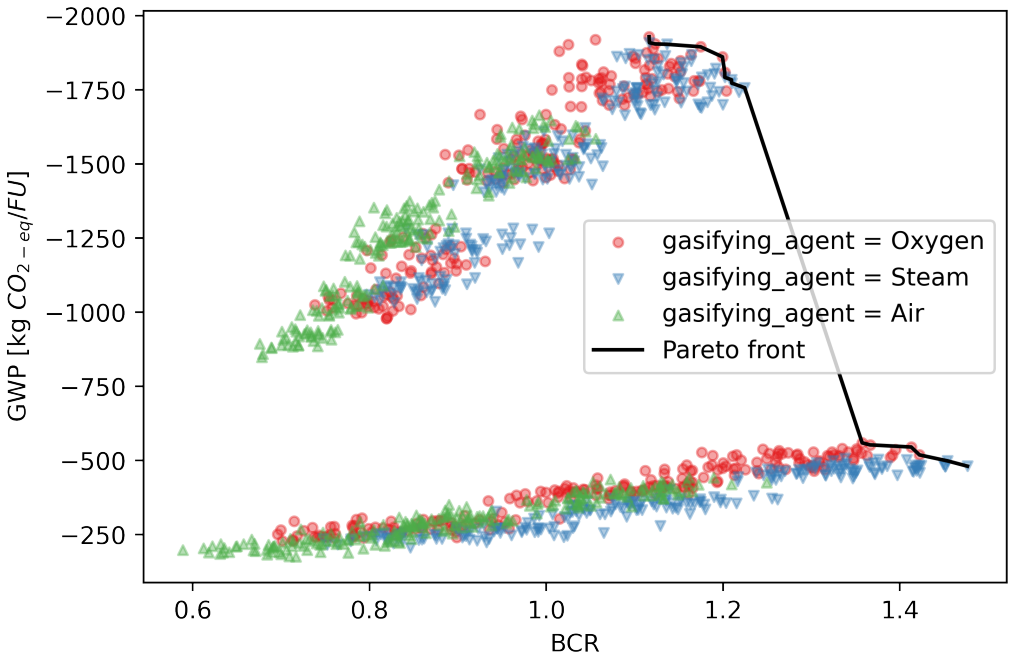


Figure C10: Optimisation results for the feedstock draff shown for the model parameter gasifying agent.

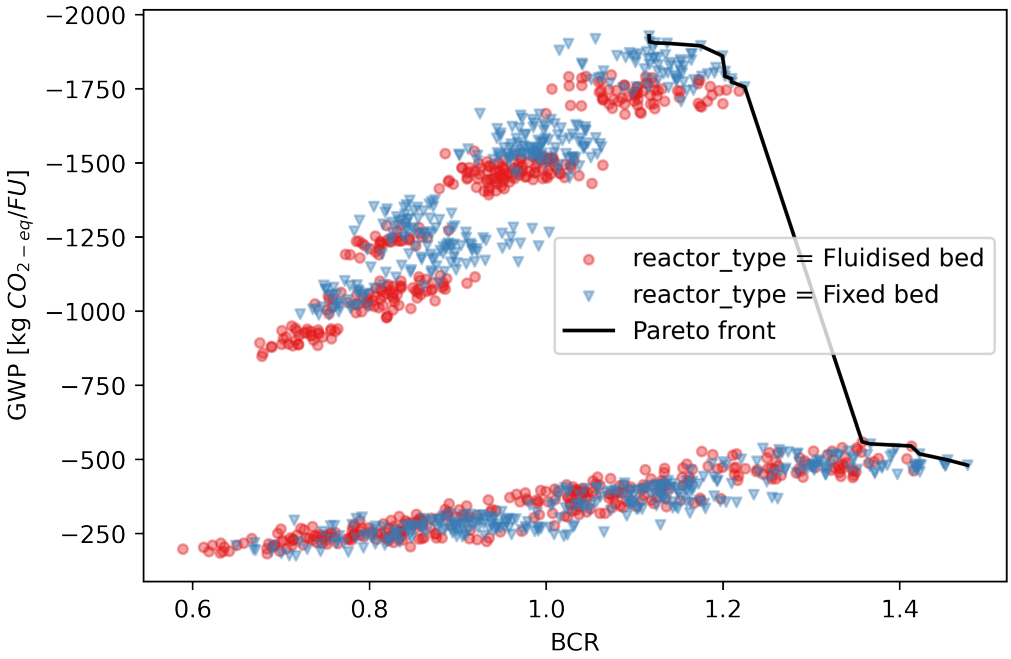


Figure C11: Optimisation results for the feedstock draff shown for the model parameter reactor type.

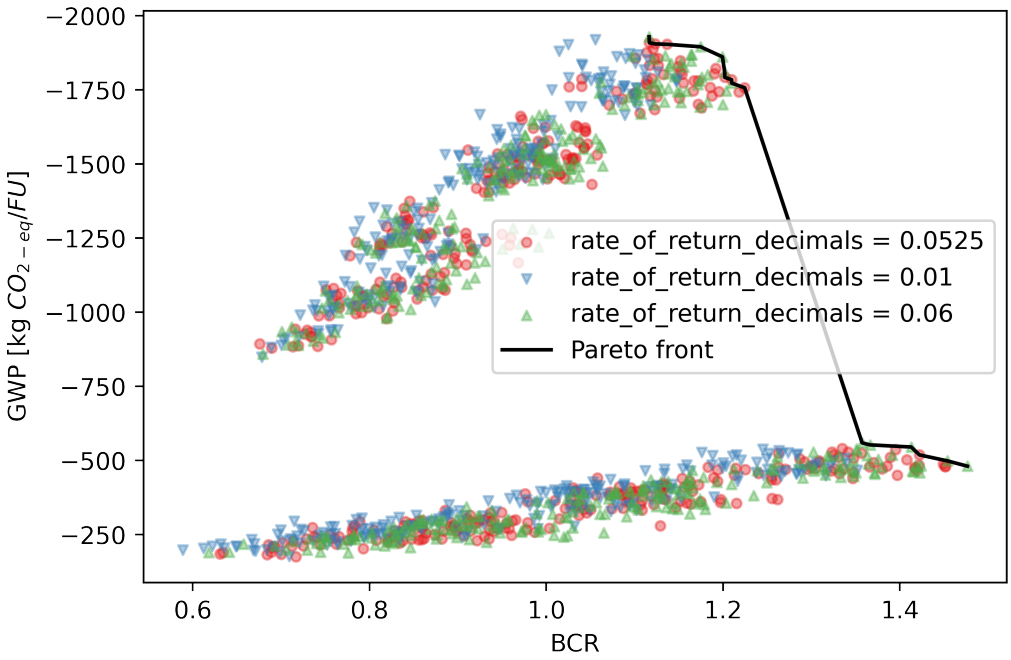


Figure C12: Optimisation results for the feedstock draff shown for the model parameter interest rate [%].

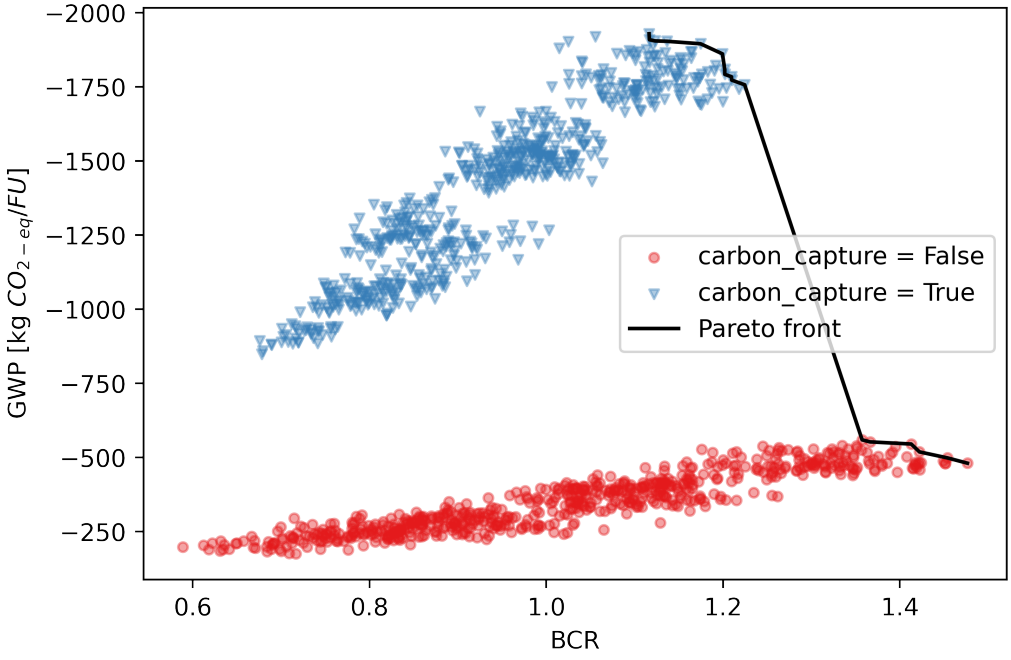


Figure C13: Optimisation results for the feedstock draff shown for the model parameter CAPEX (included or not included).

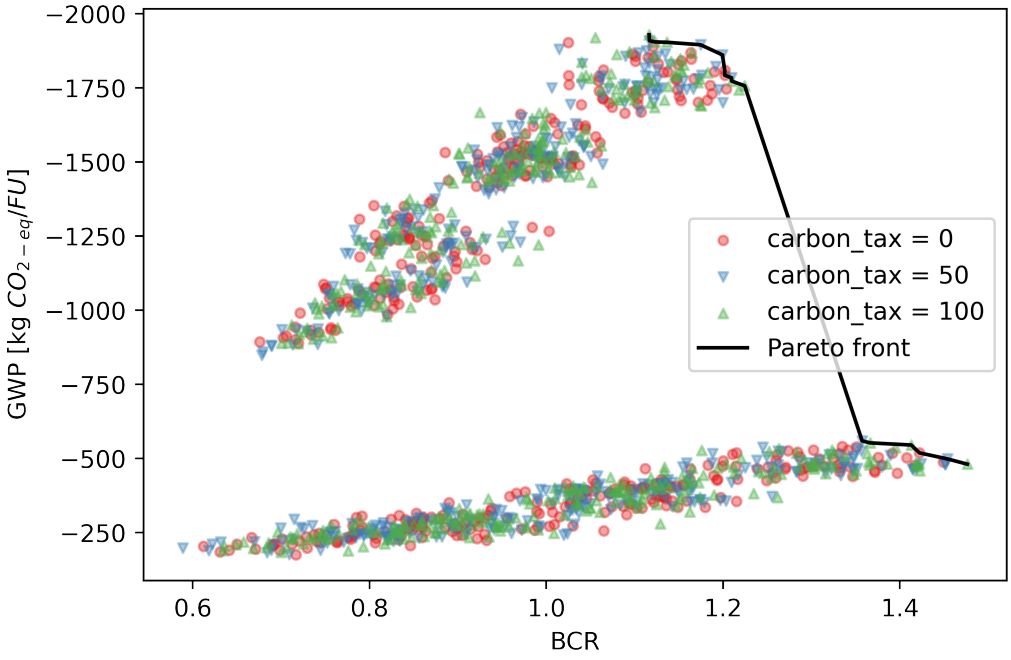


Figure C14: Optimisation results for the feedstock draff shown for the model parameter carbon tax [£/tonne CO_2-eq].