

DISS. ETH NO. 29259

**COMPUTER-AIDED DESIGN OF
MOLECULES, EXPERIMENTS, AND PROCESSES
FOR A SUSTAINABLE CHEMICAL INDUSTRY**

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

LORENZ HEINRICH JOHANNES FLEITMANN
M. Sc. RWTH, RWTH Aachen University

born on 11.10.1991

accepted on the recommendation of

Prof. Dr.-Ing. André Bardow, examiner
Prof. Dr. rer. nat. Kai Leonhard, co-examiner
Prof. Dr. Gonzalo Guillén Gosálbez, co-examiner

2023

Lorenz Fleitmann: *Computer-aided design of molecules, experiments, and processes for a sustainable chemical industry*, © 2023

DOI: 10.3929/ethz-b-000615039

„Ignoranti quem portum petat nullus suus ventus est.“

*Wer den Hafen nicht kennt, in den er segeln will,
für den ist kein Wind der richtige.*

Seneca, 62 n.Chr.
Epistulae morales ad Lucilium VIII, LXXI

Danksagung

Die vorliegende Arbeit entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter, zunächst am Lehrstuhl für Technische Thermodynamik der RWTH Aachen und anschließend am Lehrstuhl für Energie- und Prozesssystemtechnik der ETH Zürich. Mit der Fertigstellung der Arbeit möchte ich mich bei denjenigen bedanken, die mich in den prägenden und herausfordernden Jahren meiner Promotion unterstützt haben und so zu meinem Erfolg beigetragen haben.

An erster Stelle möchte ich mich bei meinem Doktorvater Prof. Dr.-Ing. André Bardow bedanken, der meine Forschung geduldig gelenkt und durch Anregungen und konstruktive Kritik verbessert hat. Weiterhin danke ich Herrn Prof. Dr. rer. nat. Kai Leonhard für viele wertvolle Einblicke in die molekulare Thermodynamik und die explorativen Projektideen zum Design von Molekülen. Herrn Prof. Dr. Gonzalo Guillén Gosálbez danke ich für Übernahme des Korreferats im Rahmen meiner Doktoratsprüfung sowie Frau Prof. Dr.-Ing. Laura De Lorenzis für die Übernahme des Prüfungsvorsitzes.

Ein großer Dank gilt meinen aktiven und ehemaligen Kollegen am LTT in Aachen und am EPSE in Zürich für das ausgezeichnete Arbeitsklima. Als Teil der Prozessdesigngruppe danke ich besonders meinen Kollegen und Freunden Christoph Gertig, Dr.-Ing. Jan Scheffeyzk, Dr.-Ing. Johannes Schilling und Dominik Tillmanns für die vielen fachlichen Diskussionen, gemeinsame Arbeiten und Unterstützung bei Veröffentlichungen und Vorträgen. Meinen Kollegen aus der SORP und der MBFD Mirko Engelpracht, Dr.-Ing. Carsten Flake, Dr.-Ing. Lukas Krep, Dr.-Ing. Leif Kröger und Marten Lache danke ich für die Aufheiterungen des Promotionsalltags als gute Freunde.

Dem Fuel Science Center in Aachen bin ich dankbar für die Finanzierung meiner Promotionsstelle und die vielen Mitarbeiter des Clusters, mit denen ich gern interdisziplinär geforscht und auf der FSC-Konferenz gefeiert habe. Meinen Studis danke ich für hervorragende Abschlussarbeiten und Leistungen als studentische Hilfskräfte, die einige Ergebnisse dieser Promotion erst ermöglicht haben.

Nicht zuletzt möchte ich mich bei meiner Familie und meinen Freunden außerhalb des Lehrstuhls bedanken. Ihr habt mich immer unterstützt und mich mit eurem Zuspruch motiviert. Ein ganz besonderer Dank gilt dir, Johanna, für deine Geduld, die Unterstützung und den roten Faden auf den letzten Metern!

Zürich, 17.05.2023

Lorenz Fleitmann

Abstract

A key to a sustainable chemical industry is the design of the industry's products and processes. Therefore, chemical products and processes need to be developed with sustainability metrics as objectives. However, integrating sustainability metrics into the design methods is not trivial, as the environmental impacts of chemical products and processes are affected over multiple orders of magnitude: The influences range from the system level, where environmental impacts become apparent, via the technical applications and physical properties, down to the molecule structure at the molecular level. As a result, the design of molecules, experiments, and processes for sustainability is a challenge at multiple levels. To overcome this challenge, targeted solution methods need to be developed that combine environmental assessment, application modelling, and property prediction. In this thesis, the computer-aided design methods for molecules, experiments, and processes are therefore advanced beyond pure technical feasibility by extending the modelling at the system, application, properties, and molecular level.

At the system level, computer-aided molecular and process design is integrated with predictive life cycle assessment to allow minimising environmental impacts as optimisation objective. At the application level, detailed modelling of entire process flowsheets is enabled by including the majority of unit operations and heat integration in the design framework. The integrated framework is demonstrated to minimise the life cycle environmental impacts of solvents in chemical processes. At the property level, the design of experiments is accomplished that maximises the accuracy of predictions on process performance and environmental impacts. Physical property measurements for parametrisation and validation are tailored to their application in molecular and process design using c-optimal experimental design. Finally, at the molecular level, the design scope is extended from processing chemicals towards chemical products since the chemical products represent a major degree of freedom in the design of a sustainable industry. The methods for the integrated design of processing chemicals and processes are extended towards the targeted model-based design of chemical products, forming a starting point for the integrated life cycle design of products and processes in the chemical industry.

Kurzfassung

Ein Schlüssel für eine nachhaltige Chemieindustrie liegt in der Entwicklung nachhaltiger Produkte und Prozesse. Daher müssen insbesondere chemische Produkte und Prozesse unter Berücksichtigung von Nachhaltigkeitsmetriken entwickelt werden. Die Integration von Nachhaltigkeitsmetriken in die Entwicklungsmethoden ist jedoch schwierig, da die Umweltauswirkungen chemischer Produkte und Prozesse über viele Größenordnungen beeinflusst werden: Die Einflüsse reichen von der Systemebene, auf der die Umweltauswirkungen sichtbar werden, über die technischen Prozesse und physikalischen Eigenschaften bis hinunter zur Molekülstruktur auf der Molekularebene. Der Entwurf von Molekülen, Experimenten und Prozessen unter Berücksichtigung der Nachhaltigkeit ist daher eine Herausforderung auf all diesen Ebenen. Um diese Herausforderung zu bewältigen, müssen zielgerichtete Lösungsmethoden entwickelt werden, die Umweltbewertung, Anwendungsmodellierung und Eigenschaftsvorhersage integrieren. In dieser Arbeit werden deshalb computergestützte Methoden für die Entwicklung von Molekülen, Experimenten und Prozessen vorgestellt, die über die rein technische Machbarkeit hinausgehen. Hierzu wird die Modellierung auf der System-, Anwendungs-, Eigenschafts- und Molekülebene erweitert.

Auf der Systemebene wird das computergestützte Molekül- und Prozessdesign mit einer Methode zur prädiktiven Ökobilanz kombiniert, um als Optimierungsziel die Minimierung von Umweltwirkungen zu ermöglichen. Auf der Anwendungsebene wird die detaillierte Modellierung ganzer Prozessfließbilder erreicht, indem die gebräuchlichsten verfahrenstechnischen Grundoperationen einschließlich der Wärmeintegration eingebunden werden. Die integrierte Methode wird für die Minimierung der Umweltauswirkungen von Lösungsmitteln in chemischen Prozessen demonstriert. Auf der Eigenschaftsebene wird die Planung von Versuchen zur Vorhersage der Prozessleistung und Umweltauswirkungen präsentiert. Messungen physikalischer Eigenschaften für Parametrisierung und Validierung werden auf ihre Anwendung im Molekül- und Prozessdesign mit Hilfe der c-optimalen Versuchsplanung zugeschnitten. Auf der molekularen Ebene wird schließlich der Gestaltungsraum von Prozesschemikalien auf chemische Produkte erweitert, da die chemischen Produkte selber einen großen Freiheitsgrad bei der Entwicklung einer nachhaltigen Industrie darstellen. Die Methoden für das integrierte Design von Prozesschemikalien und Prozessen werden auf die gezielte modellbasierte Entwicklung von chemischen Produkten ausgeweitet und bilden so einen Ausgangspunkt für das integrierte Lebenszyklusdesign von Produkten und Prozessen in der chemischen Industrie.

Contents

List of figures	XIII
List of tables	XIX
Notation	XXI
1 Introduction	1
2 The state of the art in designing molecules, experiments, and processes	5
2.1 Introduction to designing molecules, experiments, and processes . . .	6
2.2 System level: Computer-aided molecular and process design for environmental sustainability	9
2.3 Application level: Process design scope in integrated molecular and process design	11
2.4 Property level: Model-based design of experiments	14
2.5 Molecular level: Molecular design as product design	17
2.6 Contributions of this thesis	18
3 Combining computer-aided molecular and process design with predictive life cycle assessment	23
3.1 A framework for the design of sustainable solvents and processes . . .	24
3.1.1 Implementation of the COSMO-susCAMPD framework	26
3.1.2 Set-up and accuracy of the predictive LCA method	29
3.2 Case study: Benign solvents for the hybrid extraction-distillation of γ -valerolactone	35
3.2.1 Problem specification	35
3.2.2 Results and discussion	37
3.3 Conclusion	43
4 Computer-aided design of solvents and processes using quantum chemistry	45
4.1 COSMO-(sus)CAMPD for heat-integrated processes	46

4.2	Case studies	52
4.2.1	Hybrid extraction-distillation of γ -valerolactone	52
4.2.2	Integrated carbon capture and utilisation for the production of carbon monoxide	56
4.3	Conclusion	64
5	Optimal experimental design for optimal chemical process simulations	67
5.1	Optimal experimental design using the c-optimality criterion	68
5.1.1	Derivation of the c-optimal objective function	68
5.1.2	Solving OED problems	71
5.1.3	Comparison of experimental designs	72
5.2	Application of c-OED for extraction and hybrid extraction-distillation processes	72
5.2.1	OED for estimating isothermal NRTL- τ -parameters for a pinch-based extraction model	73
5.2.2	OED for estimating isothermal NRTL- τ - and diffusion parameters for a countercurrent rate-based extraction column with HTU-NTU sizing	77
5.2.3	OED for estimating temperature-dependent NRTL- τ -parameters for a hybrid extraction-distillation process	80
5.3	Discussion: Uncertainties resulting from the experimental designs . .	83
5.3.1	Accuracy of the extraction process simulation	85
5.3.2	Accuracy of the hybrid extraction-distillation process simulation	86
5.4	Conclusion	89
6	Molecular design of spark-ignition fuels for maximum engine efficiency	93
6.1	Fuel design as a special case of product design	94
6.2	Fuel design method	96
6.2.1	Combustion properties	98
6.2.2	Thermodynamic properties	98
6.2.3	Environment, health, and safety indicators & synthesizability .	99
6.2.4	Engine efficiency merit function as objective function	102
6.2.5	Property constraints	103
6.2.6	Evaluation of mixture properties	104
6.3	Design of pure fuels and fuel blends for spark-ignition engines	106
6.3.1	Pure component fuel design	106
6.3.2	Design of binary blends with ethanol	112
6.4	Conclusion	120

7	Summary, conclusion, and future perspectives	123
7.1	Summary and conclusion	123
7.2	Future perspectives	126
Appendices		131
A Supporting information to Chapter 3		133
A.1	Soft- and hardware	133
A.2	Molecular descriptors selection for the artificial neural network	134
A.3	Accuracy of the artificial neural network predictions for all impact categories	137
A.4	Molecular fragment library for the genetic algorithm LEA3D	146
B Supporting information to Chapter 4		147
B.1	Soft- and hardware	147
B.2	Accuracy of predicted enthalpies of vaporisation and ideal-gas heat capacities	148
B.3	Accuracy of the process model for absorption and the multiphase reactor model	151
B.4	System boundaries and considered impacts for the LCA of the ICCU process	154
B.5	Supplementary COSMO-(sus)CAMPD designs for minimum exergy demand	156
B.6	Specifications of the process models	164
B.7	Molecular fragment libraries for the genetic algorithm LEA3D	166
C Supporting information to Chapter 5		169
C.1	Distribution of experimental effort for the experimental designs	169
C.2	Input parameters for the case studies	171
C.3	Additional results from the Monte Carlo analysis	179
C.4	Additional optimal experimental design for the hybrid extraction-distillation process	187
D Supporting information to Chapter 6		189
D.1	Soft- and hardware	189
D.2	Molecular fragment library for the genetic algorithm LEA3D	190
D.3	Prediction accuracy of environment, health, and safety indicators	191
D.4	Contribution of fuel properties to expected engine efficiency increase	199

D.5 Influence of property constraints on the number of feasible candidate blends	201
D.6 Calculation and property data for the estimation of direct CO ₂ emissions	201
Bibliography	203
Publications	239

List of Figures

1.1	The scales relevant to the design of a sustainable chemical industry and methods to link the individual levels.	2
2.1	Multi-scale design from the molecular to the system level.	6
2.2	Visualisation of the contributions of this thesis	19
3.1	COSMO-susCAMPD framework combining COSMO-CAMPD with predictive LCA to design environmentally beneficial solvents.	25
3.2	Life cycle stages of a solvent and frequently used system boundaries in environmental assessment.	28
3.3	Flow diagram for the automated set-up of the artificial neural network.	30
3.4	Accuracy of the artificial neural network predictions for the LCA impact categories Climate Change and Ozone Depletion.	34
3.5	Flowsheet and contributions to environmental impacts of the hybrid extraction-distillation process for γ -valerolactone purification.	36
3.6	Cradle-to-grave impacts on Climate Change of all solvents designed versus corresponding process energy demand.	38
3.7	(A) Cradle-to-grave impact on Climate Change per kmol γ -valerolactone versus specific impact from solvent production per kilogram solvent. (B) Cradle-to-grave impact on Climate Change versus mole fraction of solvent in the wastewater stream.	40
3.8	Cradle-to-grave impacts on Ozone Depletion of all solvents designed depending on the corresponding process energy demand.	43
4.1	COSMO-(sus)CAMPD method for the design of heat-integrated processes by including thermochemistry from quantum chemistry calculations and process optimisation including heat integration.	51
4.2	Flowsheet and degrees of freedom for the hybrid extraction-distillation of γ -valerolactone with process degrees of freedom highlighted for the process units.	53

4.3	Comparison of the impact on Climate Change of the heat-integrated process with the impact on Climate Change neglecting sensible heat and considering only the reboiler energy demand.	56
4.4	Flowsheet for the integrated carbon capture and utilisation process producing carbon monoxide via the liquid energy carrier dimethylformamide.	58
4.5	Results of the integrated molecular and process design of the heat-integrated integrated carbon capture and utilisation process. . .	61
4.6	Environmental impact on Climate Change depending on process design scope.	63
5.1	Experimental designs of liquid-liquid equilibrium experiments for the extraction process: (A) Location of liquid-liquid equilibrium experiments. (B) Share of experimental effort.	75
5.2	Experimental designs of liquid-liquid equilibrium and diffusion experiments for the extraction model using the HTU-NTU method. .	79
5.3	Experimental designs of liquid-liquid equilibrium experiments for the hybrid extraction-distillation process.	81
5.4	Procedure to determine the uncertainties of the process simulation results for each experimental design.	85
5.5	Uncertainties of the solvent demand in the extraction process for c-optimal, D-optimal, and equidistantly distributed conventional experimental designs.	86
5.6	Uncertainties of the environmental impact on Climate Change in the hybrid extraction-distillation process for c-optimal, D-optimal, and equidistantly distributed conventional experimental designs.	87
5.7	Pareto frontier between c- and D-efficiency for parameter estimation of the hybrid extraction-distillation process.	89
6.1	Fuel design method for maximum engine efficiency considering constraints on environment, health, and safety hazards, as well as synthesizability.	97
6.2	Predicted engine efficiency increase for the pure-component design. .	107
6.3	Influence of property constraints on the number of candidate fuels in the pure-component fuel design.	111
6.4	Predicted engine efficiency increase for the binary blends with ethanol.	113
6.5	Parity plot comparing pure-component and blend merit function value of the candidate fuels.	118

6.6	Direct CO ₂ emissions from fuel combustion compared to predicted engine efficiency increase for each candidate blend.	120
A.1	Parity plot of predicted versus database Agricultural Land Occupation	138
A.2	Parity plot of predicted versus database Climate Change without biogenic carbon	139
A.3	Parity plot of predicted versus database Fossil Depletion	139
A.4	Parity plot of predicted versus database Freshwater Eutrophication .	140
A.5	Parity plot of predicted versus database Freshwater Ecotoxicity . . .	140
A.6	Parity plot of predicted versus database Human Toxicity	141
A.7	Parity plot of predicted versus database Ionising Radiation	141
A.8	Parity plot of predicted versus database Marine Eutrophication . . .	142
A.9	Parity plot of predicted versus database Marine Ecotoxicity	142
A.10	Parity plot of predicted versus database Metal Resource Depletion . .	143
A.11	Parity plot of predicted versus database Particulate Matter Formation	143
A.12	Parity plot of predicted versus database Photochemical Oxidant Formation	144
A.13	Parity plot of predicted versus database Terrestrial Acidification . .	144
A.14	Parity plot of predicted versus database Terrestrial Ecotoxicity	145
A.15	Parity plot of predicted versus database Water Depletion	145
A.16	Molecular fragment library used for the hybrid extraction-distillation of γ -valerolactone.	146
B.1	Comparison of experimental enthalpies of vaporisation with predictions from COSMO-RS on TZVP-MF level.	149
B.2	Comparison of experimental ideal-gas heat capacities with predictions using B3LYP/TZVP with RRHO approximation.	150
B.3	Comparison of minimal solvent demand calculated by the pinch-based process model with rigorous simulations in ASPEN Plus.	151
B.4	Comparison of equilibrium conversion calculated by the vapour-liquid-liquid equilibrium reactor model used in this work and simulations in ASPEN Plus.	152
B.5	Comparison of product mole fraction calculated by the vapour-liquid-liquid equilibrium reactor model used in this work and simulations in ASPEN Plus.	153
B.6	System boundaries and considered impacts for the LCA of the ICCU process and the benchmark CCU process.	154
B.7	Comparison of the total exergy demand of the heat-integrated process and exergy demand from reboiler energy neglecting sensible heat. . .	157

B.8	Results of the integrated molecular and process design of the heat-integrated ICCU process.	161
B.9	Comparison of the total process exergy demand considering process optimisation with heat integration compared to a process optimisation without considering heat integration potential and the total reboiler exergy demand only.	163
B.10	Molecular fragment library used for the hybrid extraction-distillation case study.	166
B.11	Molecular fragment library used for the ICCU process case study. . .	167
C.1	Uncertainties of the solvent demand in the extraction process for the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for $\sigma_w = 0.001$	179
C.2	Uncertainties of the solvent demand in the extraction process for the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for $\sigma_w = 0.01$	180
C.3	Parameter accuracy of the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for the extraction process	181
C.4	Uncertainties of the environmental impact on Climate Change in the hybrid extraction-distillation process for the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for $\sigma_w = 0.001$	182
C.5	Uncertainties of the environmental impact on Climate Change in the hybrid extraction-distillation process for the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for $\sigma_w = 0.01$	183
C.6	Effect of number of experiments on the relative RMSE of the experimental designs for the hybrid extraction-distillation process for $\sigma_w = 0.005$	184
C.7	Effect of number of experiments on the relative RMSE of the experimental designs for the hybrid extraction-distillation process for $\sigma_w = 0.01$	185
C.8	Property parameter accuracy of c-optimal, D-optimal, and conventional experimental designs for the hybrid extraction-distillation process. . .	186
C.9	Experimental designs of liquid-liquid equilibrium experiments for the hybrid extraction-distillation process.	187
D.1	Molecular fragment library used for the pure-component and blend design.	190

D.2 Parity plot of predicted versus experimental autoignition temperature of the test set.	192
D.3 Parity plot of predicted versus experimental bioconcentration factor of the test set.	193
D.4 Parity plot of predicted versus experimental aqueous toxicity of the test set.	194
D.5 Parity plot of predicted versus experimental oral toxicity of the test set.	195
D.6 Parity plot of predicted versus experimental permissible exposure limit of the test set.	196
D.7 Parity plot of predicted versus experimental unified yield sooting index of the test set.	197
D.8 Parity plot of predicted versus experimental gas-phase standard enthalpy of formation of the test set.	198
D.9 Contributions of fuel properties to expected engine efficiency increase for the pure-component design	199
D.10 Contributions of fuel properties to expected engine efficiency increase for the blend design and influence of the K-value	200
D.11 Influence of property constraints on the number of feasible candidate blends.	201

List of Tables

3.1	Prediction accuracy of the artificial neural network for the impact categories Climate Change and Ozone Depletion, as well as an average of all 17 regarded impact categories.	33
3.2	Comparison of the candidate solvents with the lowest cradle-to-grave and the lowest cradle-to-gate impact on Climate Change, as well as the solvent with the lowest solvent loss and the benchmark solvent. . . .	42
4.1	Highest ranking solvents for the hybrid extraction-distillation of γ -valerolactone based on the corresponding objective functions. . . .	55
4.2	Ranking of candidate solvents based on the chosen objective function and impact on Climate Change of the corresponding process.	62
5.1	c- and D-efficiencies of the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for the estimation of isothermal NRTL- τ -parameters and use in the pinch-based extraction process model.	77
5.2	c- and D-efficiencies of the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs estimation of isothermal NRTL- τ -and diffusion parameters and use in the countercurrent rate-based extraction process model.	80
5.3	c- and D-efficiencies of the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for the estimation of temperature-dependent NRTL- τ -parameters and use in the pinch-based hybrid extraction-distillation process model.	83
6.1	Data set sizes and prediction accuracies on the test sets of the EHS indicators using group contribution-based GPR models	101
6.2	Reference values for RON95 gasoline used in the engine efficiency merit function.	103
6.3	Property constraints for the fuel design used in this work.	105

6.4	Details on the pure-component fuel design for the 11 candidate fuels predicted to exceed the engine efficiency of RON95 gasoline.	109
6.6	Details on the binary blend design with ethanol for the three highest-ranking candidate blend components, the highest-ranking commercially available blend components, and the optimal pure-component fuel tert-butyl formate.	115
A.1	Molecular descriptors selected for the artificial neural network	135
A.3	Accuracy of the neural network predictions for all impact categories	137
B.1	Highest ranking solvents for the hybrid extraction-distillation of γ -valerolactone based on the corresponding objective functions.	158
B.2	Ranking of candidate solvents based on the chosen objective function and exergy demand of the process.	162
C.1	Distribution of experimental effort for the estimation of isothermal NRTL- τ -parameters for the extraction process	169
C.2	Distribution of experimental effort for the estimation of isothermal NRTL- τ - and diffusion parameters for the extraction process	170
C.3	Distribution of experimental effort for the estimation of temperature-dependent NRTL- τ -parameters for the hybrid extraction-distillation process	170
C.4	c- and D-efficiencies of the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for the estimation of temperature-dependent NRTL- τ -parameters and use in the pinch-based hybrid extraction-distillation process model.	188

Notation

Abbreviations

AiT	Autoignition Temperature
ANN	Artificial Neural Networks
BCF	BioConcentration Factor
CAMD	Computer-Aided Molecular Design
CAMPD	Computer-Aided Molecular and Process Design
CAPD	Computer-Aided Product Design
CC	Climate Change
COSMO-RS	COnductor-like Screening MOdel for Realistic Solvation
DCN	Derived Cetane Number
DFT	Density Functional Theory
EHS	Environmental, Health, and Safety
GA	Genetic Algorithm
GC	Group Contribution
GNN	Graph Neural Network
GPR	Gaussian Process Regression
(I)CCU	(Integrated) Carbon Capture and Utilisation
LBV	Laminar Burning Velocity
LC ₅₀ (FM)	aqueous toxicity as Lethal Concentration for Fathead Minnow fish
LCA	Life Cycle Assessment
LCI	Life Cycle Inventory

LCIA	Life Cycle Impact Assessment
LD ₅₀	oral toxicity as Lethal Dose for rats
LHV	Lower Heating Value
LLE	Liquid-Liquid Equilibrium
MAE	Mean Absolute Error
MON	Motor Octane Number
(n)RMSE	(normalised) Root-Mean-Square Error
NRTL	Non-Random Two Liquid
OD	Ozone Depletion
OED	Optimal Experimental Design
PEL _{OSHA-TWA}	Permissible Exposure Limit using the OSHA time-weighted average
QSAR	Quantitative Structure-Activity Relationship
QSPR	Quantitative Structure-Property Relation
RON	Research Octane Number
SI	Spark Ignition
uYSI	unified Yield Sooting Index
VLE	Vapour-Liquid Equilibrium

Chemicals

CO	carbon monoxide
CO ₂	carbon dioxide
DMA	dimethylamine
DMF	dimethylformamide
GVL	γ -valerolactone

H₂O water

Greek symbols

α scalar measure describing the centre of the tie lines

Δh_{vap} enthalpy of vaporisation

η maximum brake thermal engine efficiency

$\zeta_{\text{c}}, \zeta_{\text{D}}$ c- and D-efficiency, respectively

θ thermodynamic model parameters

ξ design vector describing an experimental design

ρ Pearson correlation coefficient

ρ_{rank} Spearman's rank correlation coefficient

σ standard deviation

Σ_0 parameter variance-covariance matrix from previously performed experiments

χ targets for heat integration

ψ process model results

Ω general variable for properties

ω experimental measurements

Latin symbols

AFR stoichiometric mass air-to-fuel ratio

A matrix containing the sensitivities of the thermodynamic model g with respect to the experimental measurements ω

B matrix containing the sensitivities of the thermodynamic model g with respect to the parameters θ

Notation

c	vector containing the sensitivities of the process model result ψ with respect to the parameters θ
f	objective function
F	Fisher-Information-Matrix
g	thermodynamic model
h	process model
j	heat integration model
K	parameter describing an engine's operating conditions
k	constraints
\dot{m}	mass flow
m	representation of molecules
\dot{n}	molar flow
N	number of distinct experimental settings
N_{exp}	number of experiments
$N_{\text{Train/Test}}$	number of data points in training/test set
p	pressure
P_{engine}	engine power
Q_{reb}	minimum reboiler energy demand calculated by pinch-based distillation model
R^2	coefficient of determination
S_{min}	minimum solvent demand calculated by pinch-based extraction model
T	temperature
v_i	normalised weight of distinct experimental setting i
V_{ω}	variance-covariance matrix of the experimental measurements ω
V_{θ}	variance-covariance matrix of the parameters θ

x	molecular structure
x_i	mole fraction
y	process variables
z	experimental settings

General subscripts and superscripts

\wedge	initial or predefined value
*	optimal
aq	aqueous
boil	boiling point
bubble	bubble point
c	c-optimal
c2grave	cradle-to-grave
con	conventional
D	D-optimal
Dec	decanter
Dist	distillation
exp	experimental
Extr	extraction
FU	functional unit
g2g	gate-to-gate
max	maximum
min	minimum
reb	reboiler

Notation

Rx	reaction
solv	solvent
tot	total
val	validation
woHI	without heat integration

Introduction

The chemical industry is today one of the largest manufacturing industries and is continuously growing. From 2000 to 2017, chemical production almost doubled to 2.3 billion tonnes (United Nations Environment Programme [UNEP], 2019). According to forecasts, sales will further double by 2030 (UNEP, 2019). Today's chemical industry is also a major consumer of fossil fuels and significantly contributes to environmental pollution, climate change, and resource consumption (UNEP, 2019). Consequently, environmental impacts also continue to grow as the production volume of chemicals increases. For example, the chemical industry's oil consumption for energy and raw material supply is projected to account for one-third of the global oil production by 2030 and one-half by 2050 (International Energy Agency [IEA], 2018, 2021). As a result, the direct CO₂ emissions of the chemical industry will increase by 20% until 2030 and by 30% until 2050. Besides CO₂ emissions, the chemical industry will increasingly cause additional burdens for the environment, e.g. by air pollution or water consumption (IEA, 2018).

To reduce its environmental impact, the chemical industry needs to move towards sustainable operations. A key to sustainability is the underlying structure of the chemical industry, i.e. which products are manufactured with which processes (Bakshi, 2019; Grossmann and Harjunkoski, 2019; Martín and Adams II, 2019; Pistikopoulos et al., 2021). Therefore, the methods for product and process design play a particularly critical role for sustainable chemicals and processes. However, present design methods frequently approximate performance and environmental impacts by characteristic properties instead of performing integrated assessments (Gertig et al., 2020b; Adjiman et al., 2021). As a result, the simplified representation of design goals as objectives leads to suboptimal designs. Without precise design targets, the computer-aided design of products and processes is inefficient, leading to unnecessary resource consumption and waste, higher hazards and environmental impacts, as well as avoidable costs (Bakshi, 2019).

Setting tailored metrics as design objectives is a challenging task as the design of products and processes spans multiple orders of magnitude, from the molecules via the

physical properties and applications to the assessment of economic and environmental impacts (cf. Fig 1.1; Zhou et al., 2019a; Adjiman et al., 2021). Consequently, advancing the design methodology needs to address the requirements at the levels of molecules, properties, applications, and system, as well as the links between them. These links are formed by property prediction and measurement, energy and mass balances, and assessment models. The design of sustainable products and processes thus also includes the prediction and experimental measurement of physico-chemical properties, application modelling, and environmental assessment.

To aid the transformation of the chemical industry, the established design methods need to be advanced to enable targeted designs beyond purely technical function (Zimmerman et al., 2020). For a chemical industry with sustainable products and processes, the design objectives need to be formulated in terms of sustainability metrics, e.g. environmental impacts determined through life cycle assessment (ISO 14040, 2006). In this thesis, the computer-aided design of molecules, experiments, and processes is advanced to design sustainable chemical products and processes tailored to the corresponding applications. For this purpose, the design of molecules, experiments, and processes with respect to environmental assessment is enabled at each level, and the methods are refined towards tailor-made design.

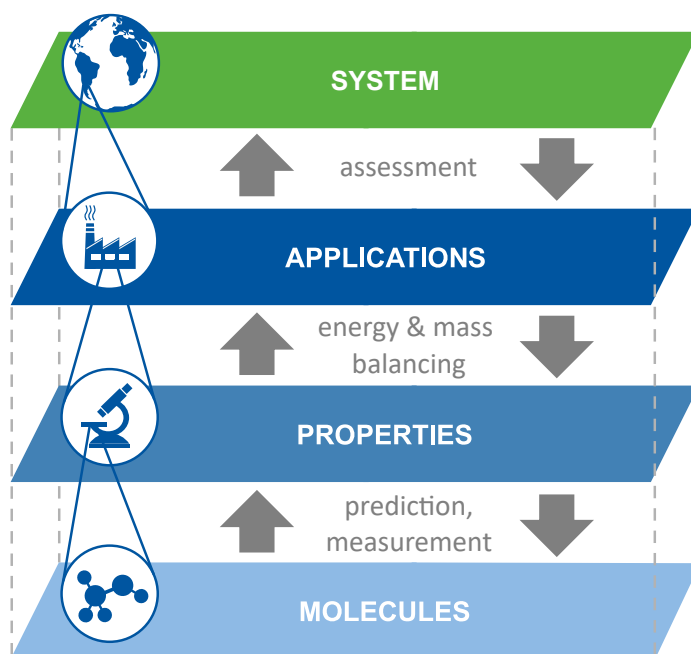


Figure 1.1: The scales relevant to the design of a sustainable chemical industry and methods to link the individual levels: The design space spans from molecules at the smallest scale to environmental impacts at the largest scale.

Structure of this thesis

This thesis is structured in seven chapters: **Chapter 2** provides an overview of the state of the art in the computer-aided design of molecules, experiments, and processes. Building on the basic relationships between molecules, properties, and applications, design opportunities are highlighted, and current design methods are described. Based on the literature review, current challenges and limitations of the state-of-the-art methods are identified, and the scientific contribution of this work is outlined.

In **Chapter 3**, a Computer-Aided Molecular and Process Design (CAMPD) framework is proposed that allows for minimising of environmental impacts. The CAMPD framework integrates Life Cycle Assessment (LCA) of solvents from cradle to grave as a holistic environmental design objective.

Chapter 4 expands the process design scope in CAMPD. The CAMPD framework is extended to include models for the most common unit operations in chemical engineering and process optimisation for minimum utility demand. As a result, the method yields optimal combinations of solvents and process settings considering heat integration.

In **Chapter 5**, tailored design of experiments is investigated that considers the application of property parameter estimation: c-optimal experimental design (c-OED). c-OED designs experiments that minimise the uncertainties of a process model leading to the most accurate process simulations.

In **Chapter 6**, the focus in the design of molecules is shifted from processing chemicals, e.g. solvents, to chemical products, e.g. fuels for internal combustion engines. An optimisation-based method is developed to design fuels for spark-ignition engines that includes a model predicting engine efficiency as objective function.

Finally, **Chapter 7** summarises this thesis and draws conclusions on the results of the individual chapters. Based on the knowledge gained in this thesis, future research perspectives are outlined.

The state of the art in designing molecules, experiments, and processes

In this chapter, the state of the art in computer-aided design of molecules, experiments, and processes is reviewed, and current challenges and limitations are highlighted. The chapter starts with an introduction to designing molecules, experiments, and processes (Section 2.1). Section 2.2 reviews current environmental objectives and constraints in CAMPD. In Section 2.3, the current process design scope and the modelling resolution of chemical processes in molecular design are discussed. Section 2.4 describes the state of the art in designing experiments for chemical engineering problems. Section 2.5 reviews the application of CAMD to product design problems, in which the focus is shifted from processing chemicals to chemical end products. Finally, Section 2.6 summarises the literature review and outlines the contributions of this thesis.

Major parts of this chapter are reproduced by permission of Elsevier, John Wiley & Sons, Inc., and the American Chemical Society from:

Fleitmann, L.; Kleinekorte, J.; Leonhard, K. and Bardow, A. (2021). COSMO-susCAMPD: Sustainable Solvents from Combining Computer-Aided Molecular and Process Design with Predictive Life Cycle Assessment. *Chemical Engineering Science*, 245, 116836.

Fleitmann, L.; Gertig, C.; Scheffczyk, J.; Schilling, J.; Leonhard, K. and Bardow, A. (2023). From molecules to heat-integrated processes: Computer-aided design of solvents and processes using quantum chemistry. *Chemie Ingenieur Technik*, 95(3), 368–380.

Fleitmann, L.; Pyschik, J.; Wolff, L.; Schilling, J. and Bardow, A. (2022). Optimal experimental design of physical property measurements for optimal chemical process simulations. *Fluid Phase Equilibria*, 557, 113420.

Fleitmann, L.; Ackermann, P.; Schilling, J.; Kleinekorte, J.; Rittig, J.G.; vom Lehn, F.; Schweidtmann, A.M.; Pitsch, H.; Leonhard, K.; Mitsos, A.; Bardow, A. and Dahmen, M. (2023). Molecular design of spark-ignition fuels for maximum engine efficiency by combining predictive thermodynamics and machine learning. *Energy & Fuels*, 37(3), 2213–2229.

The author of this thesis contributed to the literature research and wrote the first draft of the papers as the principal author. The conceptualisation of the methods was jointly developed in discussion with the co-authors. More details on the methods are provided in later chapters of this thesis.

2.1 Introduction to designing molecules, experiments, and processes

The design of molecules, experiments, and processes for a sustainable chemical industry is a challenging task since the design problem spans multiple orders of magnitude (Zhou et al., 2019a; Adjiman et al., 2021): From the system level, where environmental impacts become apparent, via the application and property level down to the molecular level (Figure 2.1). The multi-scale approach is required as the micro scale strongly influences the macro scale: The molecules employed in a technical application, e.g. a chemical process, determine the physico-chemical and thermodynamic properties. Following the physico-chemical and thermodynamic laws, these properties describe the performance of the technical application. The macro-scale effects of the application finally determine the environmental impacts.

The description of macro-scale effects caused by micro-scale decisions is key to the multi-scale design and corresponds to the so-called **analysis** or **direct problem** (Figure 2.1). Solutions to analysis or direct problems are usually predictive models connecting adjacent levels. In the design of molecules, experiments, and processes, these connections are established by predictive property, application, and assessment models: To link the molecular and the property level, physico-chemical and thermodynamic properties are estimated given a molecular structure. For this purpose, various property prediction methods have been developed for pure components and mixtures, e.g. using group contribution (GC; Gani, 2019), quantum chemistry (Gertig et al., 2020b), or machine learning-based methods (Zhou et al., 2019a, 2021). Given property data,

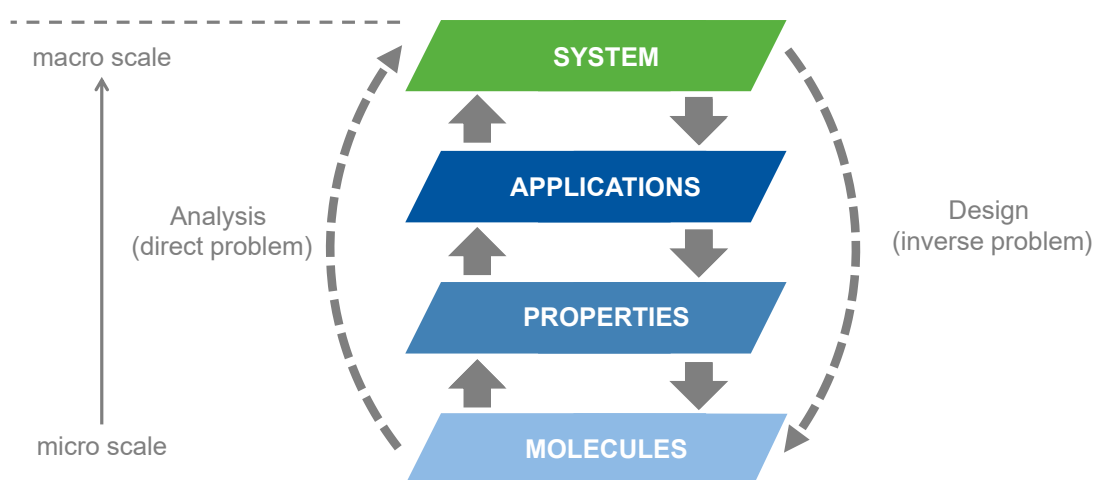


Figure 2.1: Multi-scale design from the molecular to the system level.

property and application level are linked by application models, e.g. a chemical process model such as a flowsheet model. Finally, the system level can be reached by an assessment model of economics (e.g. net present value), resources (e.g. exergy analysis), or environmental impacts (e.g. LCA).

Inverting the analysis leads to the so-called **inverse** or **design problem**: Given a macro-scale objective, the inverse problem is to determine the micro-scale decisions that lead to the desired macro-scale effects by formulating an optimisation problem. To formulate and solve the design problem, the fundamental cause-effect relationships between the objective and the degrees of freedom need to be known as inputs, i.e. the direct problems connecting adjacent levels have been solved before and resulted in predictive models describing the effect of the micro-scale decisions on the macro-scale objective. Using the predictive models, the design problem is formulated as an optimisation problem and solved with an appropriate solution strategy, e.g. based on generate-and-test, metaheuristic or deterministic optimisation algorithms (Sun et al., 2020; Pistikopoulos et al., 2021).

Depending on the problem scope, the design problem formulation spans one or more levels and includes one or more predictive models. In the literature, various design problems have been studied: For example, molecular design problems cover the molecular and properties level with design objective reaching the applications level (Papadopoulos et al., 2018); experimental design mainly focuses on the properties level (Franceschini and Macchietto, 2008); and process synthesis and design frequently address application and system level (Chen and Grossmann, 2017). In this thesis, the design of molecules, experiments, and processes is investigated by methods of computer-aided molecular and process design, and experimental design.

Methods for **Computer-Aided Molecular Design (CAMD)** optimise the molecular structure of candidate molecules, frequently by focussing on finding molecules with characteristic thermodynamic properties that are heuristically related to superior performance in an application (Gertig et al., 2020b; Adjiman et al., 2021). In these methods, the design is carried out at the property level. More advanced CAMD methods assess the molecules at the application level by integrating a process model to define a proper objective function and constraints relevant to the application. If the design of molecules is integrated with process design at the application level, integrated **Computer-Aided Molecular and Process Design (CAMPD)** problems are formed, i.e. process settings are simultaneously optimised with the molecular structure. CAMPD allows for the targeted exploration of vast molecular design spaces and corresponding optimal processes without experimentation and is thus increasingly investigated in literature (Papadopoulos et al., 2018; Chai et al., 2022).

The most commonly designed molecules in CAMPD are processing chemicals such as solvents and working fluids for application in energy and chemical conversion processes. Examples for CAMPD of processing chemicals are the design of working fluids for Organic Rankine Cycles (Schilling et al., 2017; White et al., 2018) or solvents for separation problems, e.g. liquid-liquid extraction or CO₂ absorption, including the design of non-classical components such as ionic liquids (Zhou et al., 2020; Chai et al., 2022). CAMPD has also been successfully applied to optimise reactions through the design of reaction solvents (Liu et al., 2019a; Gertig et al., 2020a) and even catalysts (Gertig et al., 2021).

Besides processing chemicals, the design of chemical products is recognized as a key part of chemical engineering (Adjiman et al., 2021), since the molecular structure of a product determines its properties and functionalities (Gani, 2004). **Computer-aided product design (CAPD)** is a collective term for designing components for various applications and ranges from small molecules over solids and nanoparticles to pharmaceuticals and formulated products (Uhlemann et al., 2019). The design of some chemical end products, e.g. fuels, is closely related to CAMD for processing chemicals (Gani and Zhang, 2020; Zhang et al., 2020a). However, in contrast to CAMD for processing chemicals, a challenge in CAPD is the prediction of properties describing the application, frequently going far beyond thermodynamic properties (Ng et al., 2015b; Zhang et al., 2016). Moreover, application models in CAPD that solve the direct problem between properties and application are frequently rare, while the effects of processing chemicals in process models can usually be well described using existing chemical engineering knowledge.

Experimental design complements CAMPD/CAPD and process and product development since an experimental design does not optimise the chemical process or product but rather its description. For the design of chemical products and processes, physical properties are crucial for model parametrisation and validation (Mitsos et al., 2018). In particular, thermodynamic properties, e.g. describing phase behaviour, largely influence predicted performance in application (Mathias, 2016). Therefore, high-quality property data are required for accurate results. The need for high-quality property data has led to the development of model-based **Optimal Experimental Design (OED)** (Atkinson et al., 2006; Franceschini and Macchietto, 2008). State-of-the-art OED identifies optimal experimental settings with respect to parameter accuracy by analysing the uncertainty propagation from experimental measures to estimated property parameters through a predefined model of the experiment. OED has already been applied for the estimation of important thermodynamic properties in the chemical industry, such as reaction kinetics (Forte et al., 2017; Walz et al., 2018), phase equilibria

(Dechambre et al., 2014b; Duarte et al., 2021), diffusion coefficients (Wolff et al., 2016), or adsorption isotherms (Walz et al., 2018).

Based on these general concepts for the design of molecules, experiments, and processes, the individual design methods are detailed in the following Sections 2.2-2.5, and the state of the art in modelling and design at each level is investigated.

2.2 System level: Computer-aided molecular and process design for environmental sustainability

To date, CAMPD methods have mainly focused on economics and technical process performance for assessing candidate molecules and processes at the system level (Zhou et al., 2020; Gertig et al., 2020a). However, the design of sustainable chemical processes needs a more holistic objective with an environmental dimension (Brown et al., 1987). Both economics and environmental impacts of many chemical processes depend strongly on the employed molecules (Clarke et al., 2018; Jimenez-Gonzalez, 2019; Zhou et al., 2020). Therefore, not only process performance and economics but also environmental impacts need to be optimised (Zimmerman et al., 2020). To capture environmental impacts at the system level, CAMPD needs to integrate environmental assessment (Zhou et al., 2020; Gertig et al., 2020a; Adjiman et al., 2021).

A few CAMPD methods already integrated environmental assessment, i.e. the assessment of environmental impact potentials or hazards. Many of these approaches are based on metrics and guidelines for green molecules (Soh and Eckelman, 2016). In particular, indicators for Environmental, Health, and Safety hazards (EHS; Adu et al., 2008) have been successfully integrated in CAMPD problems. For example, systematic screening approaches evaluate candidates based on environmental databases and Quantitative Structure-Activity Relationship (QSAR) toolboxes (McBride et al., 2018; Linke et al., 2020; Song et al., 2020). If CAMPD problems are formulated and solved as a mathematical optimisation problem, solution algorithms require an automated, integrated evaluation of EHS criteria. For this purpose, predictive models are frequently employed, e.g. GC models fitted to experimental data (Papadopoulos et al., 2010; Schilling et al., 2017; Ten et al., 2017; Ooi et al., 2018; Jonuzaj et al., 2019; Ten et al., 2020, 2021). These approaches have in common that they evaluate environmental impact potentials from the molecular properties of the candidate molecules.

However, environmental assessment has to go beyond the environmental impact potential of the molecules, which is a molecular property, such as the global warming potential (Hellweg et al., 2004). For a holistic assessment, CAMPD needs to consider

the environmental impacts of the full life cycle of a molecule, including emissions caused during production, use and disposal (Jimenez-Gonzalez, 2019; Chemmangattualappil, 2020). A broadly accepted method for the holistic environmental assessment is Life Cycle Assessment (LCA). LCA is an ISO-normed method (ISO 14040, 2006) considering emissions of all life cycle stages from cradle to grave of a substance. As a consequence of the holistic analysis, LCA helps to avoid problem shifting between life cycle stages or environmental impacts. However, a cradle-to-grave LCA generally requires much information on a substance, i.e. detailed mass and energy balances of all flows from and to the environment during production, use and disposal (Hellweg and Milà i Canals, 2014).

In CAMPD, available data on candidate molecules is usually minimal, in particular on *in silico* designed molecules. For economic objectives, CAMPD methods have already been equipped with predictive tools to close data gaps: Predictive thermodynamic models estimate thermodynamic properties so that process simulation can be performed for economic assessment. Likewise, CAMPD needs to integrate predictive LCA approaches for environmental assessment. Similarly to the prediction of thermodynamic properties from thermodynamic models, the environmental impacts of candidate molecules need to be predicted given their molecular structure (Kleinekorte et al., 2020).

In literature, predictive LCA has been approached by two main routes: (1) the prediction of Life Cycle Inventory (LCI) and (2) the direct prediction of the Life Cycle Impact Assessment (LCIA). The LCI is the basis for life cycle impact assessment and provides the bill of materials of the life cycle. To yield ultimately environmental impacts, the LCI needs to be multiplied by characterisation factors (Hauschild and Huijbregts, 2015). LCI is frequently predicted from estimates for energy and mass flow from generic flowsheets (Righi et al., 2018; Parvatker and Eckelman, 2020). In contrast, the direct prediction of the LCIA has been investigated by multi-linear regression (Calvo-Serrano et al., 2018; Calvo-Serrano and Guillén-Gosálbez, 2018) and artificial neural networks (Wernet et al., 2008, 2009; Song et al., 2017; Kleinekorte et al., 2019; Karka et al., 2022).

Recently, predictive LCA has successfully been combined with molecular design for the first time: Papadopoulos et al. (2020) formulated an integrated CAMD problem including predictive LCA and predictive EHS scores. For the predictive LCA, the authors use the ANN-based FineChem model (Wernet et al., 2009) to estimate the specific impacts of solvent production per kilogram solvent. For the predictive EHS scores, Papadopoulos et al. (2020) employ group contribution and molecular similarity approaches. By combining these prediction approaches into one integrated multi-

2.3 Application level: Process design scope in integrated molecular and process design

objective CAMD problem, desired solvent properties are optimised simultaneously with environmental impact scores, e.g. maximising specific solvent density and minimising specific environmental impact on Climate Change and EHS scores.

A similar approach was proposed by Baxevanidis et al. (2021). The authors developed a group contribution-based LCA model and integrated the GC-LCA model into a CAMD formulation. The CAMD problem was solved considering the cradle-to-gate environmental impacts of the solvents to design solvents for liquid-liquid extraction based on performance indicators.

However, the current approaches limit the LCA scope to a so-called cradle-to-gate system boundary, considering only emissions caused during the solvent production per kilogram solvent. The amount of solvent required by the process cannot be considered in these approaches, although the amount of solvent varies substantially depending on the solvent’s performance in the process. Moreover, the process corresponds to the use phase of the solvent life cycle, and the solvent properties directly impact the process performance and the emissions of the use phase. Finally, the emissions from solvent disposal depend on the solvent loss during the use phase. Thus, a cradle-to-gate assessment does not capture the full environmental impacts of the candidate solvents. To avoid problem shifting between life cycle stages, CAMPD needs to consider all solvent-related emissions within a cradle-to-grave system boundary.

2.3 Application level: Process design scope in integrated molecular and process design

Accurate environmental assessment of each candidate molecule at the system level requires appropriate modelling and design at the application level to provide life cycle inventory. Therefore, systematic process design is required in CAMPD for maximum performance and minimum environmental impact. For systematic process design, various optimisation-based methods have been developed in process systems engineering (Pistikopoulos et al., 2021). A chemical process involves various units and auxiliaries to transform raw materials into products. Today, process design therefore includes the optimisation of process settings and unit operations as well as heat recovery and utility systems (Smith, 2005) and the selection of molecules as auxiliaries such as solvents or working fluids (Gertig et al., 2020b; Zhou et al., 2020; Adjiman et al., 2021; Chai et al., 2022).

Traditionally, process design follows a sequential approach from the reactor to separation and recycle systems and the heat exchanger network (Douglas, 1985).

However, this sequential approach does not account for the considerable interactions between the entire process system and each unit or the heat recovery subsystem. Moreover, the selection of auxiliaries is not considered as an explicit step in the design approach. Since the optimal process system performance cannot always be achieved by separately optimising the process subsystems and auxiliaries, advanced design methods integrate the individual design steps, e.g. by mathematical optimisation (Chen and Grossmann, 2017). However, current methods focus on either (1) energy and mass integration or (2) molecule selection.

1. Energy and mass integration: Several solutions are presented to combine process optimisation with the design of the heat recovery network (Kong et al., 2016; Elsidio et al., 2017, 2019; Dong et al., 2020; Ryu et al., 2020; Kruber et al., 2021). These methods simultaneously design the (reaction-) separation process and heat exchange by solving large superstructure optimisation problems. Superstructure problems usually contain non-convexities and many discrete degrees of freedom to model the process synthesis decisions and are thus challenging to solve (Chen and Grossmann, 2017). Therefore, the solution methods often require tailored solution algorithms for computational efficiency. Recently, Liesche et al. (2019) and Schack et al. (2020) presented the superstructure-based process synthesis method FluxMax that avoids non-linearities in the optimisation problem by discretising the thermodynamic state space before optimisation. Thereby, the non-linear process synthesis problem is reduced to a linear flux optimisation of elementary process functions on the thermodynamic grid. While these methods master the complexity of energy and mass integration, they are not capable of simultaneously optimising the employed molecules as auxiliaries as well. Process design approaches considering heat integration usually assume a fixed selection of molecules or a small preselected set to avoid the problem complexity due to the large molecular design space (Adjiman et al., 2021).

2. Molecule selection: Aside from process settings and heat integration, the performance of chemical and energy conversion processes is also substantially impacted by molecules used as auxiliaries, in particular solvents (Chemmanurathuvalappil, 2020; Zhou et al., 2020). Solvents influence process conditions and optimal settings of unit operations and thus even heat integration and utility consumption. Consequently, process and solvent cannot be optimised independently, but their design needs to be integrated to for successful process design (Gertig et al., 2020b; Adjiman et al., 2021; Chai et al., 2022; cf. Section 2.1). However, similar to integrated energy and mass integration problems, CAMPD problems are highly non-linear and challenging to solve, since CAMPD problems usually consider non-ideal thermodynamics and contain integer decision on the molecular structure (Samudra and Sahinidis, 2013;

Papadopoulos et al., 2018). Therefore, CAMPD methods commonly simplify either the process design scope or the molecular design space accessible by the employed predictive thermodynamic models.

Many CAMPD methods approximate the solvent influence on the process using simplified performance indicators such as partition coefficients or relative volatilities (Austin et al., 2017; Scheffczyk et al., 2017b; Ooi et al., 2019; Zhou et al., 2019b; Papadopoulos et al., 2020). Other CAMPD methods assess only single process units or flowsheet subsystems (Austin et al., 2017; Jonuzaj and Adjiman, 2017; Jonuzaj et al., 2018; Scheffczyk et al., 2018; Gertig et al., 2020a; Zhang et al., 2020a; Fleitmann et al., 2021a; Watson et al., 2021). However, limiting the process modelling to simplified performance indicators, single-unit operations, or small subsystems of the process flowsheet does not capture all flowsheet-inherent trade-offs. Simplifying the process design scope can thus lead to suboptimal solvent selection for the final optimised overall process flowsheet (Gertig et al., 2020b; Chai et al., 2022). Even more, the influence of heat recovery has been neglected in CAMPD (Adjiman et al., 2021).

CAMPD methods modelling the entire process flowsheet typically simplify the molecular design scope by simplifying property prediction (Zhou et al., 2017; Zhang et al., 2021a) or limiting the molecular design space to specific molecular groups (Pereira et al., 2011; Burger et al., 2015; Gopinath et al., 2016; Schilling et al., 2017; White et al., 2017, 2018; van Kleef et al., 2019; Schilling et al., 2020; Keßler et al., 2021). These CAMPD methods rely on one or more GC methods for predicting the thermodynamic properties of candidate molecules. GC methods have been shown to accurately predict various thermodynamic and environmental properties (Marrero and Gani, 2001; Hukkerikar et al., 2012b). However, group parameters are usually parametrised from experimental data and are not available for all kinds of molecules, in particular for higher-order groups (Gani, 2019). Moreover, several GC methods are usually employed to cover all thermodynamic properties required for process design, e.g. for ideal-gas heat capacities, activity coefficients or enthalpies of vaporisation. However, combining several GC methods and parameter sets can lead to inconsistent predictions and contradictions (Gani, 2019). Thus, CAMPD preferentially requires consistent property prediction that does not limit the molecular design space, e.g. based on quantum chemistry (Gertig et al., 2020b).

In conclusion, systematic process design in CAMPD needs to extend beyond simplifying process models and neglecting heat recovery to model the entire process flowsheet with heat integration, while building on reliable thermodynamic data from a large molecular design space. However, because of the CAMPD problem complexity, models

for thermodynamic property prediction and process design need to be carefully selected to balance details, scope, and computational effort.

2.4 Property level: Model-based design of experiments

In product and process design, property data needed for validation and parametrisation still relies mostly on experimentation (Kontogeorgis et al., 2021). However, experiments consume time and large amounts of materials causing high costs and environmental impacts for estimating parameters in property models. Therefore, experimental effort should be minimised by selecting only the experiments that provide the most information and thus lead to the most accurate simulation. These optimal experiments can be designed by Optimal Experimental Design (OED).

Generally, two approaches for OED can be distinguished: (1) statistical OED (Franceschini and Macchietto, 2008) and (2) bounded-error OED (Pronzato and Walter, 1990). Statistical OED minimises the parameter variances considering a statistical error distribution (Walz et al., 2018). In contrast, bounded-error OED minimises the feasible parameter set consistent with the measurement uncertainty given by upper and lower bounds on the errors (Walz et al., 2018). As a result, bounded-error OED requires fewer assumptions on errors than statistical OED but instead needs to solve a challenging bilevel optimisation problem. For many experiments in chemical engineering problems, the measurement uncertainty is known (Dong et al., 2005) and justifies the use of statistical OED. Thus, the focus of this work is on the more popular statistical OED.

In statistical OED, the objective function is usually a scalar measure of the parameter variances representing parameter uncertainty (Franceschini and Macchietto, 2008). Several well-known objective functions have been developed to determine the experimental designs leading to the most accurate parameters (Franceschini and Macchietto, 2008), e.g. minimising the average uncertainty of all parameters (A-optimality); minimising the uncertainty of the most uncertain parameter (E-optimality); or minimising a generalised variance of the parameters (D-optimality).

However, in chemical engineering, the primary purpose of experiments is rarely to gain knowledge of parameters themselves. Instead, chemical engineers seek to gain thermodynamic insights, predict phase behaviours or simulate a process, etc. Thus, the experimental design needs to reflect the model application (Gevers and Ljung, 1986). Recently, OED methods have focused on incorporating the purpose of parameter estimation. Dechambre et al. (2014b) employed G-optimal experimental design that

minimises the expected variance in the model predictions of the experiments instead of uncertainties in property parameters. In particular, Dechambre et al. (2014b) minimised the predicted variance of phase compositions calculated from a liquid-liquid equilibrium model instead of the property parameters used in the activity coefficient model.

Similarly, for process simulations, the impact of the property parameters on the simulation results is usually more important than the uncertainty of the property parameters. If the governing phenomena of the chemical system are known, and a thermodynamic model capable of describing these phenomena is selected, the purpose of experimentation is to increase accuracy of the simulation through (re-)parametrisation. However, an experimental design for the most accurate property parameters does not ensure the lowest uncertainty in process simulation. Thus, the property parameter use in a process model needs to be considered within the optimal experimental design.

For this purpose, Asprión et al. (2019) recently presented OED for experiments in a plant or mini plant using a flowsheet simulator. In their work, the optimal experimental design considers property parameter use by employing the process model already for the parameter estimation. The authors show that their method improves model discrimination and parameter estimation. However, the method requires expensive and time-consuming plant experiments instead of small lab-scale experiments.

For lab-scale experiments and bounded-error OED, Walz and coworkers accounted for property parameter use in process simulation and design (Walz et al., 2018, 2019). The authors successfully show how to reduce experimental effort without changing the reliability of the process model results. However, their method requires solving a challenging bi- or trilevel optimisation problem and is currently limited to small process models.

For statistical OED and lab-scale experiments, a first approach was published by Recker et al. (2013). The authors considered the sensitivities of the process to the property parameters by heuristically scaling the A-optimality criterion and successfully optimised the experimental design to estimate reaction kinetics for a reaction-separation process. A similar approach was proposed by Lucia and Paulen (2014) for robust non-linear model predictive control. Using the sensitivities of the optimal robust economic objective value to parametric uncertainty, the authors scaled a modified E-criterion. Kaiser and Engell (2020) and Kaiser et al. (2021) linked OED for parameter estimation with superstructure optimisation of early process design stages (Kaiser and Engell, 2020; Kaiser et al., 2021). For this purpose, the authors perform global sensitivity analysis of optimisation results towards the uncertain parameters using heuristically

scaled D-optimality (Kaiser and Engell, 2020) and heuristically scaled A-optimality (Kaiser et al., 2021).

However, even though these heuristic approaches provide a breakthrough by combining OED and process simulation, heuristic designs likely differ from optimal designs with full consideration of the process (Fleitmann et al., 2021b). Instead, full consideration of process information requires uncertainties of property parameters to propagate through the process model, and the uncertainties of the process model results should be used as the OED objective.

In pioneering work, the van Impe group integrated experimental design and non-linear model predictive control (Houska et al., 2015; Telen et al., 2016, 2017). The authors mathematically derived an economic process objective function for experimental design by weighted A-optimality. They defined the OED objective as the minimisation of the expected optimality gap of the parametric optimal control problem via second-order derivatives of the Lagrange function (Houska et al., 2015). The approach was demonstrated successfully to tailor experimental designs for estimating reaction rate constants to control problems of bioreactors.

Similarly, for the most accurate chemical process simulations, the OED objective needs to be defined in terms of process uncertainties to capture the property parameter use in the process simulation. The idea of optimising the uncertainty of a simulation output as the objective for OED can be formulated as the so-called c-optimal experimental design (c-OED; Atkinson et al., 2006). In general, c-OED minimises a linear combination of model parameter variances as the optimisation objective (Atkinson et al., 2006). A linear combination of model parameters corresponds to the linear variance propagation of these parameters through a model if the weights of the linear combination are the first-order derivatives of the model with respect to the model parameters. Therefore, c-OED can reflect the property parameter use in a chemical process simulation directly in the objective, e.g. the impact of NRTL-parameters on the total process energy demand.

Interestingly, c-optimality is mathematically a special case of weighted A-optimality (Fedorov and Leonov, 2014). Thus, c-OED is connected to the modified A-optimal criterion from Houska et al. (2015). In contrast to Houska et al. (2015), c-OED weights parameter uncertainties by first-order derivatives instead of scaling the OED problem by second-order derivatives of the Lagrange function of an optimisation problem. Therefore, c-OED is suitable for chemical process simulations, while the method from Houska et al. (2015) is tailored to equation-based optimisation problems and requires the Lagrange function of the optimisation problem.

To date, c-optimality has only been applied for the optimal experimental design of clinical trials for dose-finding in the area of toxicology studying (Holland-Letz, 2017; Holland-Letz et al., 2018; Holland-Letz and Kopp-Schneider, 2018) or the description of viral dynamics and pharmacokinetics (Han and Chaloner, 2003) but not in chemical engineering for process flowsheet simulation. However, in particular for physical properties for process flowsheet simulations, experiments for parameter estimation serve a purpose beyond the pure parameter knowledge, which needs to be reflected by the OED objective. Thus, future OED in chemical engineering needs to consider the accuracy of performance evaluation and environmental assessment at the application and system level as objectives.

2.5 Molecular level: Molecular design as product design

Apart from processes and processing chemicals, chemical end products are also frequently a degree of freedom in product and process development at the molecular level. Similar to the design of processing chemicals, the molecular structure of chemical end products can be tailored *in silico* by CAMD as a special case of Computer-Aided Product Design (CAPD; Gani and Zhang, 2020; Zhang et al., 2020a).

Today, various product design frameworks exist that can identify molecules with desired product properties, e.g. the OptCAMD (Liu et al., 2019b) or ProCAPD frameworks (Kalakul et al., 2018; Chai et al., 2021). In particular, the ProCAPD framework aims to consider not only properties related to technical performance but also related to environmental hazards, quality, or cost via various submodules (Chai et al., 2021). By multi-objective optimisation, environmental hazards can also be considered explicitly besides technical performance (Jonuzaj et al., 2019). For example, Jonuzaj et al. (2019) designed the active ingredient and the solvent for an adhesive product by optimising simultaneously for minimum toxicity and maximum solubility of the active ingredient in a solvent.

However, these CAPD methods mainly rely on assessing the candidates' physico-chemical properties rather than evaluating the product's performance in application using a model of the application (Gani, 2004; Gani and Ng, 2015; Zhang et al., 2016, 2020a). Frequently, the objective function is a weighted sum of target properties. These target properties are determined by experts translating the product needs into physico-chemical properties (Zhang et al., 2020a), or the target properties are assumed from an existing product, e.g. that needs to be replaced (Zhang et al., 2016; Jhamb et al., 2019). Thus, the design procedure usually involves three separated steps: target

properties definition, model-based design, and experimental verification (Kontogeorgis et al., 2019). In this design procedure, the actual product’s performance in application beyond target properties is only evaluated during experimental verification. However, as CAMD for processing chemicals, CAPD ideally requires a model of the application as objective function that accurately reflects the product use to assess a candidate product’s performance.

A challenge for an application-level objective function is the availability of application-specific properties beyond thermodynamics (Ng et al., 2015b; Zhang et al., 2016), which requires tailored property prediction. Tailored property prediction has recently gained momentum through the availability of advanced Machine Learning (ML) models, e.g. deep learning by artificial neural networks or Bayesian regression using Gaussian process regression (Zhou et al., 2019a, 2021). These ML-based models allow for accurate modelling of phenomena previously inaccessible through rigorous modelling (Alshehri et al., 2020). Therefore, products can be optimised *in silico* for applications such as cosmetics or fragrances. Zhang et al. (2018b) and Ooi et al. (2022) demonstrated fragrance design using an ML-based model maximising odour attributes of molecules such as odour character or pleasantness. Similarly, Zhang et al. (2019) and Zhang et al. (2021b) designed food products and cosmetics, respectively, by maximising the sensorial rating of the products predicted from an ML-based model. These approaches highlight the opportunities available with extended, ML-based property prediction and demonstrate an important step towards CAPD optimising an application model rather than molecular properties.

CAMPD has already shown the potential of integrating molecular design with property prediction and mechanistic process application modelling. Similarly, CAPD should now exploit ML-based property prediction to use in mechanistic application models for targeted product design that directly addresses product application.

2.6 Contributions of this thesis

The literature review reveals that the state of the art in designing molecules, experiments, and processes lacks a systematic integration of sustainability assessment and targeting towards applications. To drive the design of molecules, experiments, and processes towards sustainable chemical products and processes, the design methodology requires the following advances on the four levels:

1. *System level:*

Integrating a life cycle environmental objective into CAMPD.

2. *Application level:*

Expanding the process design scope to modelling and optimising of process flowsheets including heat integration in CAMPD.

3. *Property level:*

Extending integrated property prediction in CAMPD and tailoring experiments for process-level and environmental objectives.

4. *Molecular level:*

Designing products maximising an application-level objective function.

This thesis advances the design of molecules, experiments, and processes by integrating new methods and models, and by linking the levels from the molecular to the system level (Figure 2.2). Therefore, the main contributions of this thesis are the following:

CAMPD with a life cycle environmental objective

At the system level, environmental assessment has to overcome the limitations of merely considering environmental impact potentials and cradle-to-gate system boundaries (cf. Section 2.2). However, current CAMPD methods, e.g. for solvent design, only consider emissions caused during solvent production per kilogram solvent and neglect emissions from the use phase and solvent disposal. Therefore, in Chapter 3,

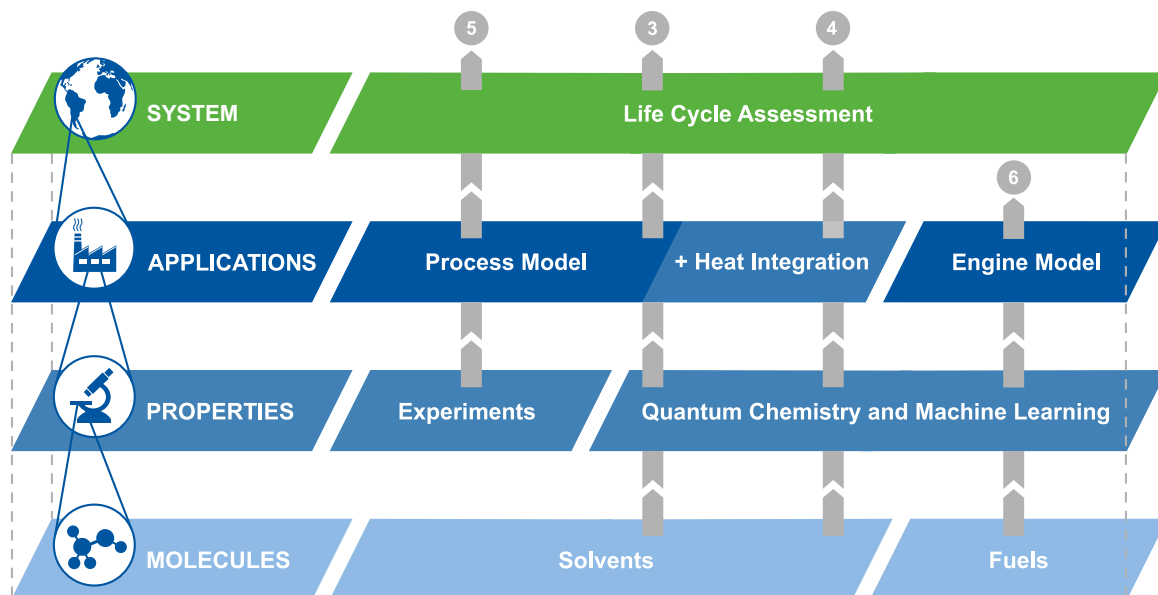


Figure 2.2: Visualisation of the contributions of this thesis. The numbers at the end of each arrow represent the chapters in which the corresponding design is described.

a CAMPD framework for solvent design is presented that considers cradle-to-grave LCA including production, use, and disposal: COSMO-susCAMPD. COSMO-susCAMPD overcomes the limitations of previous CAMPD approaches by predicting the specific cradle-to-gate impacts of solvent production using an artificial neural network and by exploiting process data from the process model as Life Cycle Inventory (LCI) for solvent use and disposal.

CAMPD framework from molecules to heat recovery systems

At the application level, comprehensive modelling and optimisation of the application's degrees of freedom are required within CAMPD (cf. Section 2.3). However, current CAMPD methods either simplify process representation using performance indicators or flowsheet subsystems or simplify the molecular design scope. Therefore, in Chapter 4, an extended COSMO-(sus)CAMPD method is presented that moves beyond simplified flowsheet subsystems while still building on reliable thermodynamic data from a large molecular design space. Process design in CAMPD is extended to incorporate modelling and optimisation of entire process flowsheets, including the most common unit operations as well as the heat recovery and utility system.

Property prediction using quantum chemistry and machine learning

Formulating environmental objectives at the system level and expanding the scope of modelling at the application level requires additional property data of candidate molecules and applications. This property data needs to be provided within the design algorithms by predictive and automated methods using molecular descriptors. However, application-specific property prediction beyond thermodynamics is rarely integrated into CAMPD and CAPD methods. Moreover, currently used methods for thermodynamic properties are usually limited to specific groups of molecules, limiting the molecular design scope. Therefore, in Chapters 3, 4 and 6, various methods for property prediction are integrated into the molecular design algorithms for candidate assessment: automated quantum chemistry and thermochemistry calculations for thermodynamic properties, as well as machine learning-based models for environmental impacts and hazards, combustion properties, and synthesisability prediction.

Experimental design for process-level and environmental objectives

For parameter estimation and validation in process and product design, experimentation focuses today mainly on general parameter accuracy at the property level and not on the use of the parameters (cf. Section 2.4). Current methods for experimental design used in chemical engineering rarely reflect the purpose of experimentation beyond pure parameter knowledge. However, for physical properties for process flowsheet

simulations, the OED objective needs to be defined in terms of process uncertainties to achieve accurate chemical process simulations. Therefore, Chapter 5 demonstrates OED of physical property measurements that considers the subsequent parameter use in simulation and assessment at the application and system level: c-optimal experimental design (c-OED).

Fuel design for maximum engine efficiency

At the molecular level, the progress in CAMPD can be adapted to advance the computer-aided design of products (cf. Section 2.5). In particular, the most recent improvements in CAMPD can be transferred to the tailored design of products: CAPD should use a model of the application as objective function that accurately reflects the product use. However, current methods in CAPD design for surrogate measures of a candidate product's performance such as favourable molecular properties. Therefore, Chapter 6 demonstrates CAPD evaluating the product's performance in application through the design of fuels for maximum engine efficiency. The developed optimisation-based fuel design algorithm employs an empirical model of spark-ignition engine efficiency as objective function to explicitly design for maximum engine efficiency.

In summary, this thesis advances the design of molecules, experiments, and processes towards designing for sustainability and tailoring for applications. The presented methods allow formulating integrated optimisation problems and solving complex multi-scale problems with application- and system-level objectives. Thereby, this thesis contributes to designing a sustainable chemical industry.

Combining computer-aided molecular and process design with predictive life cycle assessment

To minimise environmental impacts, CAMPD needs to include an environmental assessment method that closes the gap between the application and the system level and quantifies the environmental impacts. For this purpose, this chapter proposes a CAMPD framework that integrates LCA of solvents from cradle to grave: COSMO-susCAMPD. The framework builds on the COSMO-CAMPD method for predictive design of solvents using COSMO-RS and pinch-based process models (Scheffczyk et al., 2018). Cradle-to-grave LCA is enabled by combining predictive LCA from cradle-to-gate using an artificial neural network with gate-to-grave Life Cycle Inventory data from the process models.

The methodology of the COSMO-susCAMPD framework is described in Section 3.1. The details of the framework and the set-up of the artificial neural network for predictive LCA are explained, and the accuracy of the predictive LCA is discussed. The framework is applied to design solvents in a hybrid extraction-distillation process in Section 3.2. Results of the optimisation are presented, and the advantages of the integrated design are discussed. Finally, conclusions for the design of sustainable solvents and processes are drawn in Section 3.3.

Major parts of this chapter are reproduced by permission of Elsevier from:

Fleitmann, L.; Kleinekorte, J.; Leonhard, K. and Bardow, A. (2021). COSMO-susCAMPD: Sustainable Solvents from Combining Computer-Aided Molecular and Process Design with Predictive Life Cycle Assessment. *Chemical Engineering Science*, 245, 116836.

The author of this thesis integrated the cradle-to-gate predictive life cycle assessment into the solvent design method and wrote the first draft as the principal author. The author investigated, validated and visualised the results of the molecular design. The conceptualisation and methodology were jointly developed in discussion with the co-authors.

3.1 COSMO-susCAMPD: A framework for the design of sustainable solvents and processes

The design of optimal solvents for maximal process performance and minimal environmental impacts requires the combination of two methods: (1) Integrated molecular and process design (CAMPD) and (2) predictive cradle-to-grave life cycle assessment (LCA).

1. CAMPD methods usually combine three steps (Papadopoulos et al., 2018):
 - a) First, candidate molecules are generated by an algorithm, e.g. by a combination of functional groups or molecular fragments. The algorithm needs to be able to change the molecular structures systematically to explore a given design space, while structural feasibility is ensured for all candidates.
 - b) Secondly, for each candidate molecule, thermodynamic properties are predicted using predictive thermodynamic models. Thermodynamic properties are required to bridge the order of magnitude between the candidate molecules and the process, e.g. by prediction of activity coefficients or vapour pressures.
 - c) Finally, the candidate molecules are evaluated by an objective. The objective function quantifies the fit of the candidate molecules to the process application, e.g. by a particular thermodynamic property, a process variable or an economic metric.
2. A method for predictive LCA of candidate molecules requires the assessment of all stages of a molecule's life cycle: the production, the use phase and the disposal. In literature, predictive methods for particular life cycle stages have been proposed:
 - a) The environmental impacts from the production can be estimated by molecular structure models that use molecular descriptors to predict the cradle-to-gate LCIA, e.g. using multi-linear regression (Calvo-Serrano et al., 2018) or ANN (Song et al., 2017).
 - b) The use phase of a molecule can be modelled by generalised flowsheets estimating, e.g. the gate-to-gate energy demand of processes (Jiménez-González et al., 2000; Parvatker and Eckelman, 2020). Afterwards, this LCI is translated into LCIA by multiplying the energy demand with the corresponding characterisation factors.

- c) Predictive LCA approaches dealing with the disposal of molecules have not been published so far. However, proxies from LCI databases for generic wastewater treatment or waste incineration can be used (Canals et al., 2011).

Methods for predictive LCA of each life cycle stage are combined with a CAMPD method in the proposed COSMO-susCAMPD framework (Figure 3.1) to yield a fully predictive framework with cradle-to-grave environmental assessment. COSMO-susCAMPD is introduced in the following Section 3.1.1.

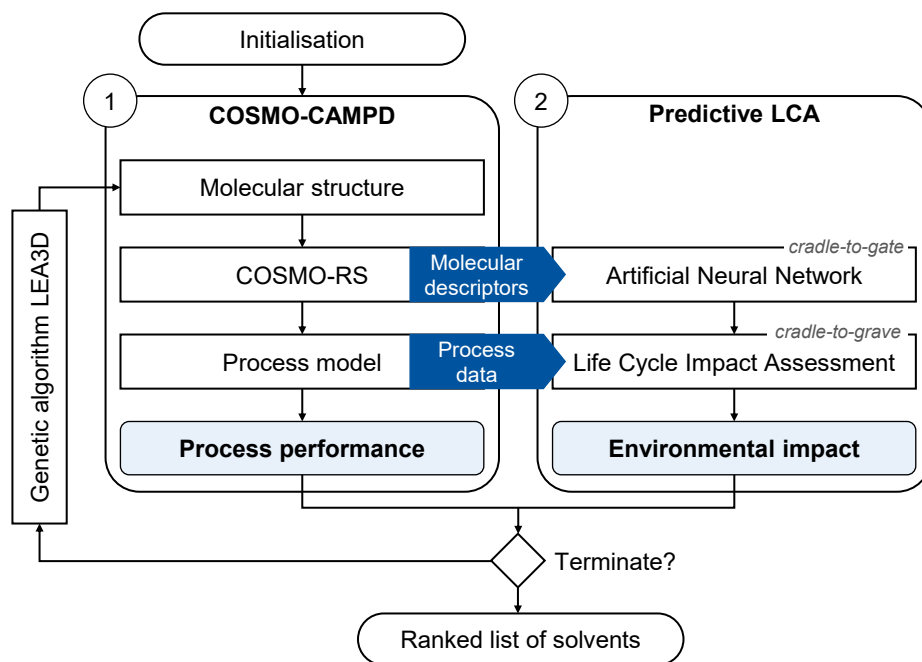


Figure 3.1: COSMO-susCAMPD: Fully automated framework to design environmentally beneficial solvents by combining COSMO-CAMPD with predictive LCA.

3.1.1 Implementation of the COSMO-susCAMPD framework

The COSMO-susCAMPD framework expands COSMO-CAMPD by a predictive LCA method as follows:

1. The basis for COSMO-susCAMPD is the COSMO-CAMPD method initially developed to design solvents for optimum process performance and economic objectives (Scheffczyk et al., 2018). The CAMPD method involves three steps in each iteration of the optimisation procedure:

- a) Generation of a molecular structure: The generation of candidate solvents is part of the molecular optimisation using the genetic algorithm LEA3D (Douguet et al., 2005). LEA3D builds molecules from 3D-molecular fragments. The fragments are specified in the initialisation of the algorithm via a fragment library. The fragment library is created by the users to reflect their preferences. The algorithm starts by randomly combining fragments for the first generation of molecules. After these molecules have been evaluated by the constraints and the objective function (Steps 1b and 1c), LEA3D alters the population of molecules for each following generation using genetic operations on the candidate molecules, i.e. crossover and mutation. Thereby, LEA3D explores the vast molecular design space towards an objective function. After a predefined number of generations is reached, the molecular optimisation stops.

Already during the generation of the molecular structures, LEA3D ensures the chemical feasibility of the molecules, e.g. all candidate molecules fulfil the octet rule. Moreover, the 3D structure of the candidate molecules allows evaluating constraints on the molecular size or functional groups. If such constraints exist, undesirable candidate molecules can already be discarded before the time-consuming computational steps.

- b) Prediction of thermodynamic properties: For each candidate solvent of each generation, thermodynamic properties are obtained using the predictive thermodynamic model COSMO-RS (Klamt et al., 2010). COSMO-RS uses surface charge interactions from quantum chemical Density Functional Theory (DFT; Kohn and Sham, 1965). By applying statistical thermodynamics to the interactions between the surface charges, COSMO-RS can then predict many thermodynamic properties of pure components and mixtures, such as activity coefficients, Liquid-Liquid Equilibria (LLE) or vapour pressures with low computational effort.

In COSMO-susCAMPD, COSMO-RS is used on the TZVP-MF level of theory for each molecule, i.e. full geometry optimisation and determination of the screening charge density (σ -surface) using the DFT functional BP86 and a TZVP basis set performed on a semiempirical conformer generation (Klamt et al., 2010). TZVP-MF exhibits a good balance between computational cost and accuracy for application in CAMPD (Scheffczyk et al., 2017b). The optimised geometries and σ -surfaces are computed in parallel for pure components based on the 3D-molecular structure and are stored in a local database for reuse. Thus, the time-consuming DFT calculations are only performed once for each candidate solvent.

The thermodynamic properties are used to evaluate constraints, e.g. the existence of LLE or an appropriate boiling point. These constraints on thermodynamic properties can reduce the search space and help to identify feasible solvents. In addition, in COSMO-susCAMPD, the thermodynamic properties serve as an input for the ANN to predict cradle-to-gate impacts.

- c) Process model evaluation: For each candidate solvent of each generation that fulfils property constraints, a process flowsheet is evaluated. The process is modelled using pinch-based process models for each unit operation (Bausa et al., 1998; Redepenning et al., 2017). Pinch-based process models are reduced-order models that provide an accurate and efficient calculation of process units assuming minimum thermodynamic driving force. By this assumption, computationally demanding tray-by-tray calculations can be omitted, but no simplifications of thermodynamic modelling are required. In literature, it has been shown that the pinch-based process models agree well with results from rigorous tray-by-tray models for operation near the thermodynamic minimum (Scheffczyk et al., 2018; Redepenning et al., 2017). As a result, the pinch-based process models yield a maximum achievable process performance for each solvent considering full equilibrium thermodynamics. Due to the computational efficiency of the process evaluation, the process flowsheet can be optimised for each solvent.

In COSMO-susCAMPD, the process model not only evaluates process performance but also provides process data as LCI of the use phase for gate-to-gate LCIA.

2. To enable an environmental objective in COSMO-susCAMPD, a predictive LCA is added for every candidate solvent to COSMO-CAMPD. For this purpose, the life cycle of the candidate solvents is divided into three stages (Figure 3.2): (a) solvent production (cradle-to-gate), (b) solvent use in the process (gate-to-gate) and (c) solvent disposal (gate-to-grave).

- a) Solvent production: The environmental impacts from solvent production (cradle-to-gate system boundary) are estimated using an ANN. As shown by Wernet et al. (2008), ANNs outperform other regression methods such as multi-linear regression in LCA applications. The ANN uses molecular and thermodynamic solvent properties as input as already proposed in the literature (Song et al., 2017; Calvo-Serrano and Guillén-Gosálbez, 2018; Papadopoulos et al., 2020). In particular, thermodynamic properties of the candidate solvents from COSMO-RS are included. Properties calculated from COSMO-RS have already been proven to be suitable molecular descriptors by Calvo-Serrano et al. (2018). In COSMO-susCAMPD, molecular descriptors from COSMO-RS provide the additional advantage that a consistent set of descriptors is used for both the LCA and the techno-economic assessment. More details on the training and set-up of ANN are given in Section 3.1.2.
- b) Solvent use: Impacts related to the solvent use in the process (gate-to-gate system boundary) are calculated from the Life Cycle Inventories provided by the process model. Process evaluation solves the mass and energy balances

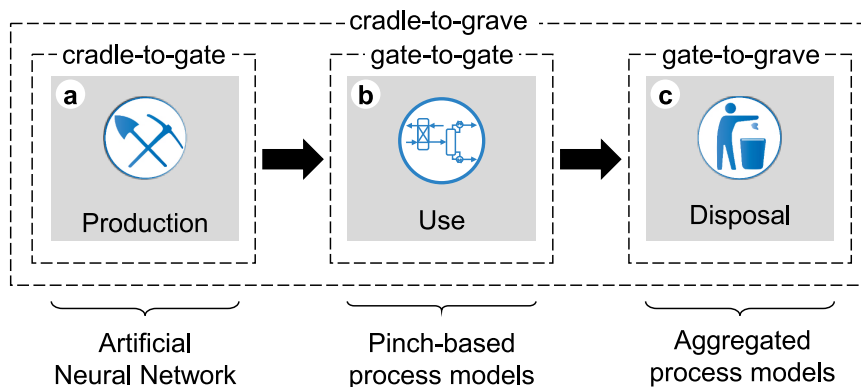


Figure 3.2: Life cycle stages of a solvent and frequently used system boundaries in environmental assessment. In COSMO-susCAMPD, cradle-to-grave system boundaries are enabled by combining an artificial neural network with pinch-based and aggregated process models.

providing all required LCI information for LCIA. In particular, the minimum amount of solvent used in the process is determined accurately by the pinch-based process models. Knowledge of the amount of solvent used allows for a comparison of candidate solvents in process-specific objectives rather than a specific comparison per kilogram of solvent. The LCIA is performed by multiplying the LCI with specific environmental impacts from LCA databases or the ANN prediction. For example, the process heat demand is converted into emissions using the specific impact for natural gas combustion per megajoule heat obtained from the GaBi database (Thinkstep AG, 2017). Additional emissions, such as fugitive emissions, are not considered.

- c) Solvent disposal: For the disposal of solvents (gate-to-grave system boundary), aggregated process models are known in the literature. Here, the solvent disposal is modelled by LCIA for wastewater treatment based on the mass of wastewater including solvent contamination. The literature model yields a specific impact per kilogram of wastewater (Ruiz, 2019). The gate-to-grave LCIA is completed by multiplying the specific impact with the flow rate of wastewater. Both the flow rate and the contamination of wastewater with the solvent result from the process model evaluation.

By combining COSMO-CAMPD and predictive LCA as described, COSMO-susCAMPD yields a fully automated and predictive framework for solvent design. As an objective for the design, process performance, environmental impacts from cradle-to-grave as well as combined objective functions are possible. Alternatively, the predictive LCA can serve as a constraint. More details on used soft- and hardware can be found in Appendix A.1.

3.1.2 Set-up and accuracy of the predictive LCA method

The predictive LCA method uses an Artificial Neural Network (ANN) as a regression model, which is trained on known environmental impacts of solvents from databases or literature. After training, the ANN is capable of predicting environmental impacts for candidate solvents similar to the solvents from the training data. Here, consistent cradle-to-gate LCA data from the GaBi Database (Thinkstep AG, 2017) on 73 solvents is used for training purpose. While the data set is small, it is important to use consistent, high-quality data and to avoid data based on generic heuristics. Thus, the present data set is the largest high-quality data set available to the author. To facilitate the set-up of the ANN, an automated framework is used in four steps (Figure 3.3), as already outlined by Kleinekorte et al. (2019):

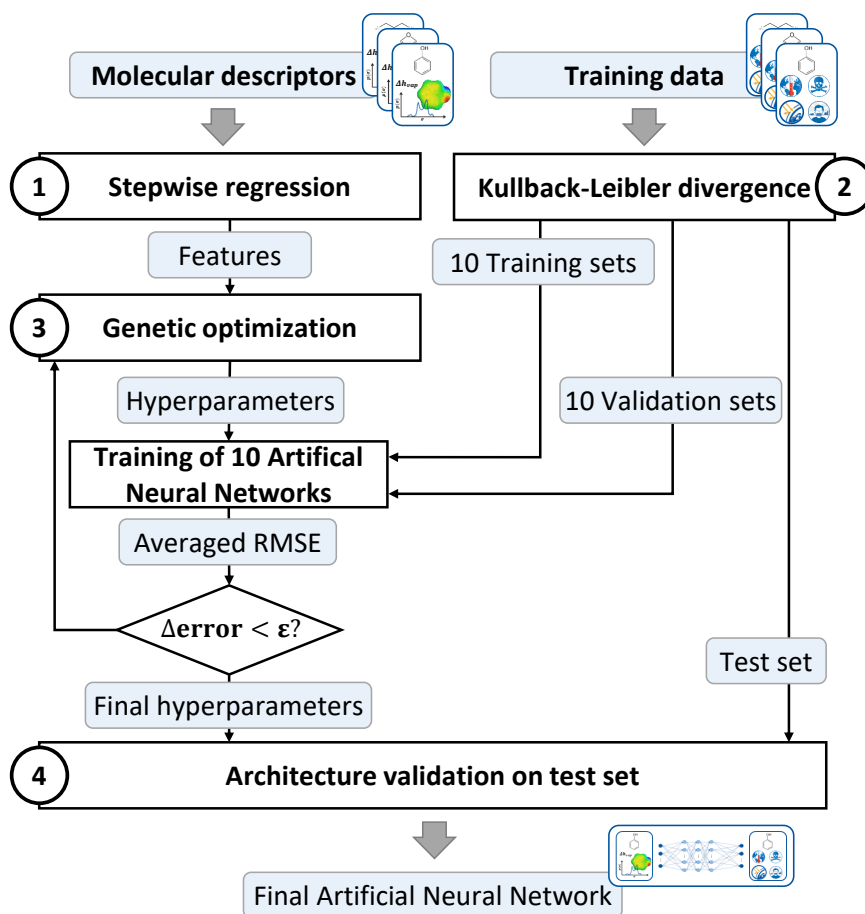


Figure 3.3: Flow diagram for the automated set-up of the artificial neural network.

1. First, suitable features for the ANN are selected from various molecular descriptors using linear stepwise regression as a feature selection method (Draper and Smith, 1998; Lindsey and Sheather, 2010). For all molecules in the training data, various molecular descriptors are calculated as prospective features, e.g. information on the molecular structure, such as the number of carbon or oxygen atoms, or thermodynamic properties from COSMO-RS, such as the normal boiling point. The molecular descriptors which show the highest correlation with the environmental impacts are selected as features (see Appendix A.2 for details).
2. Secondly, the training data is split into three sets to allow for training, validation and testing of the ANN (Goodfellow et al., 2016). At first, a test set is separated from the training data for the final accuracy evaluation of the ANN. The test set contains approximately 10 % of the training data and is not used within the

training or validation of the ANN to obtain a final accuracy value on unseen data. Extreme points at the edge of the data set are not selected for the test set due to the low extrapolation capability of the ANN beyond the training set. Afterwards, the remaining training data is split ten times into a validation (10 % of the remaining training data) and a training set (90 % of the remaining training data) to increase the generalisability of the final architecture.

All sets are chosen so that the statistical distribution of the test, training and validation sets are similar (Goodfellow et al., 2016). Therefore, various test, training and validation sets are randomly generated first. For each random set, the Kullback-Leibler divergence is calculated based on the features as a measure of statistical distribution for data sets (Kullback and Leibler, 1951). A low Kullback-Leibler divergence indicates similar and uniform statistical distribution between the data sets, which is a requirement for the training and application of ANN. Therefore, the test set with the lowest Kullback-Leibler divergence is chosen for the final accuracy evaluation. For training and validation sets, the ten splits with the lowest Kullback-Leibler divergence are chosen for the training of the ANN.

3. Thirdly, the hyperparameters of the ANN, e.g. the number of layers or the number of neurons per layer, are selected. Setting the hyperparameters is not trivial and has a considerable influence on the accuracy of the ANN. Therefore, a genetic algorithm (GA) (The MathWorks Inc., 2018a) is used to find optimal hyperparameters. The objective of the GA is the minimisation of the average root-mean-square error (RMSE) of the ANN predictions on the validation sets:

$$\min \sum_{i=1}^{10} \frac{RMSE_i^{\text{val}}}{n} \quad (3.1)$$

For each instance of the GA, 10 ANNs are trained with the same hyperparameters using the 10 training sets. Afterwards, each ANN is used to predict the corresponding validation set, and the RMSE of the prediction is calculated. By averaging the RMSE over the 10 sets, extreme prediction errors due to the small set sizes are flattened and bootstrapping and accuracy evaluation are enabled (Carney et al., 1999).

To avoid local optima due to the statistical optimisation, 100 runs of the GA are performed from random starting points by varying the initial hyperparameters.

4. Finally, the ANN is trained with the optimised architecture on the combined training and validation set to perform an accuracy evaluation by predicting the

test set. The test set has neither been used to train the ANN nor optimise the hyperparameters. Therefore, the ANN predicts the unseen test set with similar accuracy as the molecules designed within the COSMO-susCAMPD framework.

After applying the described set-up, one trained ANN is obtained with optimal hyperparameters and an estimation of its accuracy for one impact category. The ANN can directly be integrated to predict cradle-to-gate impacts for candidate solvents.

In the following, the accuracy of the ANN predictions is investigated, and COSMO-susCAMPD is subsequently applied to a case study. Using the described framework, one ANN is set up for each of the 17 midpoint impact categories from the ReCiPe method (Goedkoop et al., 2009). In the main text, only the two LCA impact categories are discussed, for which the most reliable LCIA methods are available: Climate Change (CC) and Ozone Depletion (OD) (European Commission-Joint Research Centre, 2011). Details on all 17 impact categories can be found in Appendix A.3. The accuracy of predictions is measured with the coefficient of determination (R^2) and the normalised RMSE (nRMSE).

The coefficient of determination R^2 indicates the trend-capturing correlation between the ANN predictions and the database values (Alexander et al., 2015). The nRMSE indicates how much the predictions deviate on average from the database values (Otto et al., 2018). The normalised RMSE, which is normalised by the range of the database values, is reported to make all impact categories comparable.

Currently, the availability of LCA data on solvents is limited for the training of an ANN. Our training data contains only 73 solvents, a comparably small number for machine learning approaches (Alwosheel et al., 2018). Therefore, the accuracy of the ANN predictions is currently limited (c.f. Table 3.1). On average, the ANN achieves an already acceptable nRMSE of 10%, but the average R^2 is low with a value of only 0.43. The low R^2 can be explained by the small training data set: If very few data points are used, the R^2 value is highly sensitive. Due to the small set sizes, inaccurate predictions for a few solvents decrease the R^2 already significantly despite otherwise acceptable predictions. Therefore, it is important to focus not only on the R^2 but also consider the (n)RMSE. For Ozone Depletion, for example, the nRMSE of the validation and of the test set match very well, indicating acceptable predictions despite substantial differences in the R^2 . In particular, the predictions deviate significantly from the database values for areas of sparse training data. In these areas, a high variance between the 10 ANN predictions and generally large deviations from the database values can be observed, indicating high sensitivity on the training set due to limited data. For example, for the impact on Climate Change (CC), the predictions deviate from the database values by up to 5 kg CO₂-eq. kg_{chem.}⁻¹ for solvents in the

sparse data region greater than 6 kg CO₂-eq. kg_{chem.}⁻¹ (see Figure 3.4 A). Therefore, some solvents with extreme impacts on CC at the edges of the training data are currently predicted inaccurately and need future improvement. Similarly, the ANN yields a few physically not meaningful results, i.e. negative values for some impact categories, which are removed when applying the ANN in COSMO-susCAMPD.

However, an already acceptable accuracy of prediction is achieved for most solvents and, in particular, in ranges with sufficient data (Figure 3.4). Generally, the predictions meet the database values with acceptable confidence except for a few strong outliers in sparse regions. The accuracy is comparable to the state-of-the-art in literature: For example, the estimation of CC had a coefficient of determination R² of 0.41 in work by Wernet et al. (2009) or a coefficient of determination R² of 0.48 in work by Song et al. (2017). Future improvement in accuracy is expected with more data available. For the design of solvents, ultimately, the uncertainty of the final cradle-to-grave environmental impact is most relevant. Therefore, the propagation of the uncertainty caused by the ANN prediction to the cradle-to-grave impact is investigated in Section 3.2.2.

Table 3.1: Prediction accuracy of the artificial neural network for the impact categories Climate Change and Ozone Depletion, as well as an average of all 17 regarded impact categories in terms of coefficient of determination (R²) and normalised root-mean-square error (nRMSE).

Data set	Climate Change		Ozone Depletion		Average of all impact categories	
	R ²	nRMSE	R ²	nRMSE	R ²	nRMSE
Training set	0.44	17%	0.81	8%	0.57	12%
Validation set	0.56	14%	0.76	16%	0.56	14%
Test set	0.51	9%	0.08	15%	0.43	10%

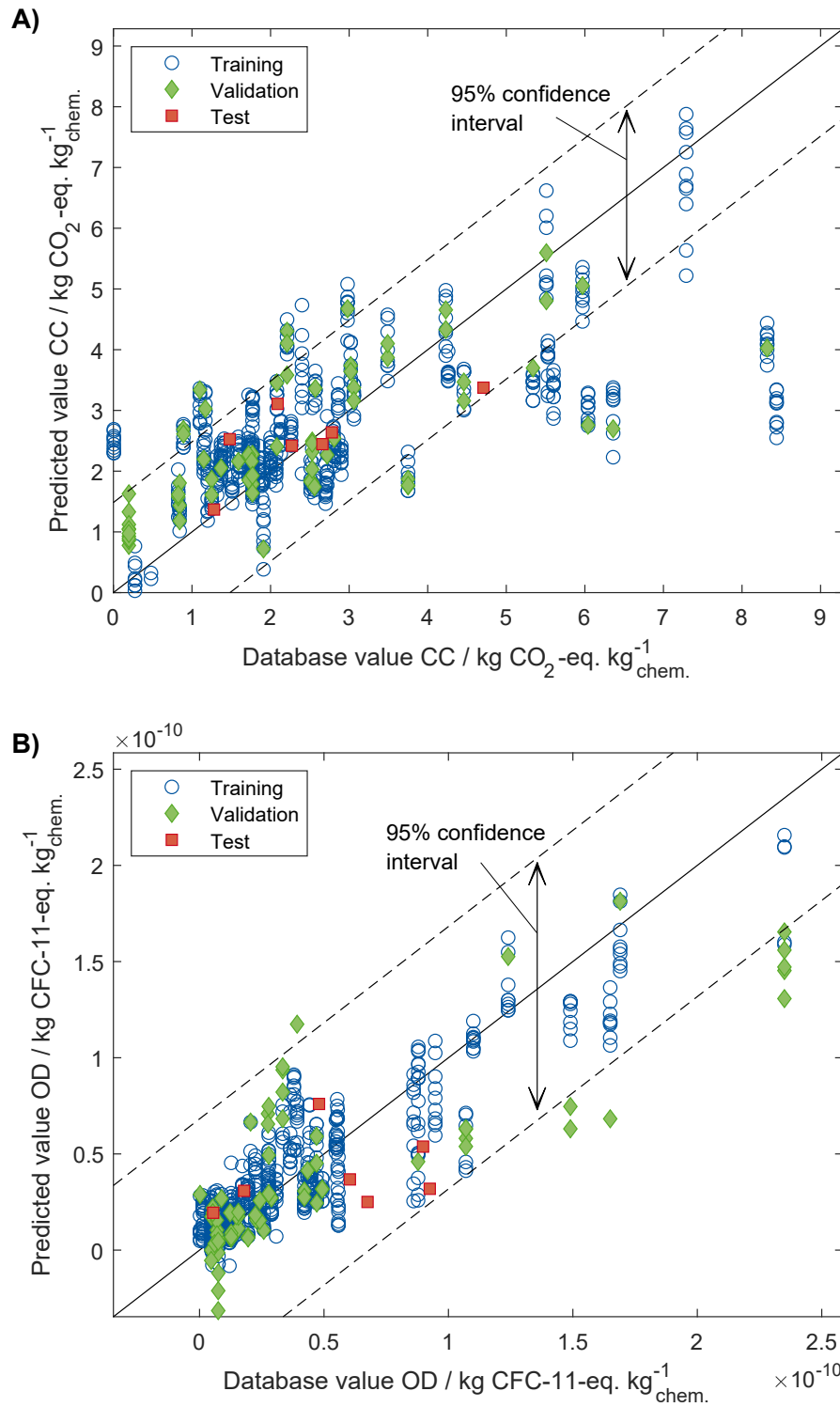


Figure 3.4: Accuracy of the ANN predictions for the LCA impact categories Climate Change (CC) and Ozone Depletion (OD). The confidence interval is calculated from the standard deviation of the predictions on the test set.

3.2 Case study and results: Design of benign solvents for the hybrid extraction-distillation of γ -valerolactone

To demonstrate the application of COSMO-susCAMPD, the hybrid extraction-distillation of γ -valerolactone (GVL) is investigated, as proposed by Murat Sen et al. (2012). Recently, GVL has attracted attention as a bio-derived platform chemical, a green solvent or a renewable fuel (Zhang, 2016). A promising pathway to GVL is the production from lignocellulosic biomass and purification from aqueous solution using hybrid extraction-distillation. As an extraction solvent, n-butyl acetate has been suggested in the literature (Murat Sen et al., 2012). Therefore, n-butyl acetate serves as a benchmark for the solvent design with COSMO-susCAMPD.

3.2.1 Problem specification

As a case study, the process of GVL purification consists of an extraction column, a distillation column and a decanter (Figure 3.5). A mixture of GVL and water containing 5 mol-% GVL is fed to the extraction column, where the solvent extracts the GVL entirely into the extract stream. The resulting extract is split in the distillation column into pure GVL at the bottom and a water-solvent stream at the top of the distillation column. The water-solvent stream is recycled to the extraction column. If the water-solvent stream splits into two liquid phases, the aqueous phase is separated from the organic phase in a decanter, and only the organic phase is fed back into the extraction column. Both the raffinate and the aqueous phase from the decanter, if present, are sent to wastewater treatment.

Candidate solvents are considered for property prediction and process evaluation if they are expected to be stable within the extraction process based on their functional groups and if they are smaller than 13 non-hydrogen atoms. The process specifications further constrain suitable candidate solvents based on their physical properties: Suitable candidate solvents must have a liquid-liquid equilibrium with water. Furthermore, the candidate solvents must not exceed the boiling point for GVL to allow for separation of GVL at the bottom of the distillation column. For simple distillation, candidate solvents also must not form an azeotrope with GVL. The constraints on the molecular properties are evaluated for each candidate solvent in each generation of the genetic algorithm with the thermodynamic properties predicted by COSMO-RS (Step 1b of the COSMO-susCAMPD framework). Candidate solvents that do not fulfil these requirements are discarded and not considered suitable candidate solvent for subsequent process optimisation and environmental assessment.

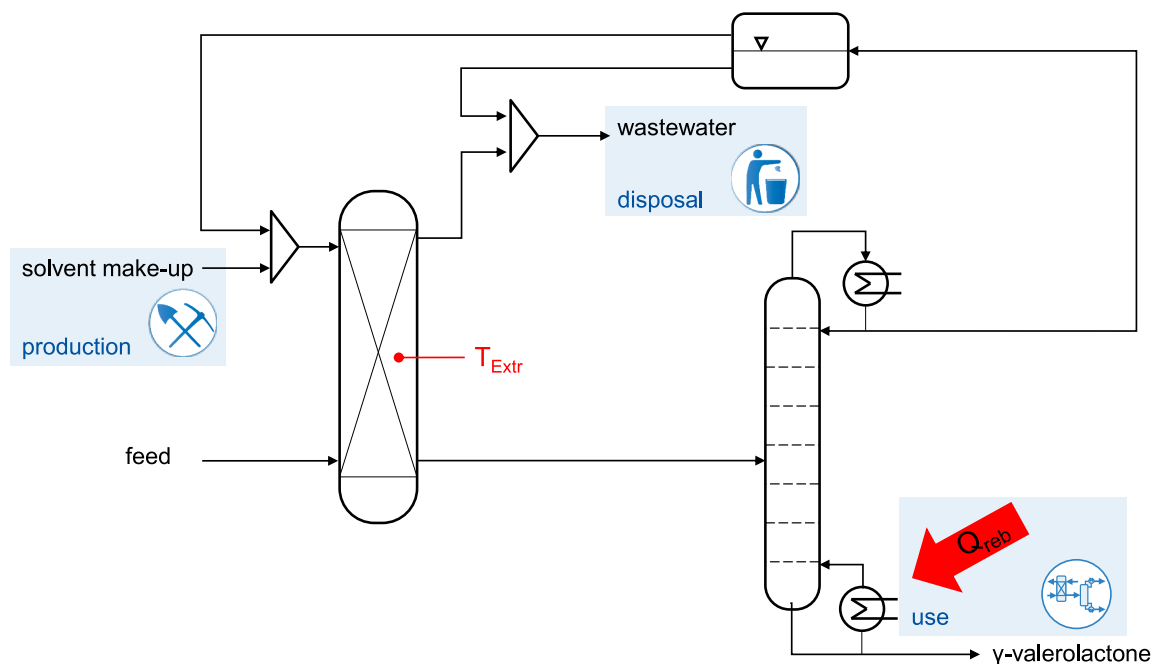


Figure 3.5: Flowsheet and contributions to environmental impacts of the hybrid extraction-distillation process for γ -valerolactone purification. The extraction column temperature T_{Ext_r} is the process design degree of freedom in this case study.

For each suitable candidate solvent, the process settings, i.e. the temperature of the extraction, are optimised to obtain the minimum energy demand for distillation (Step 1c). For the environmental assessment (Step 2), three types of emissions are considered: The emissions from solvent production due to solvent make-up, the emissions from solvent use due to energy consumption in the distillation reboiler, and the emissions from solvent disposal in wastewater treatment. For each candidate solvent, the cradle-to-gate impacts of all 17 LCA impact categories are predicted using the ANNs (Step 2a), and the full cradle-to-grave LCA is conducted by exploiting the LCI from the process evaluation for use phase (Step 2b) and solvent disposal (Step 2c). All emissions are calculated for the functional unit of 1 kmol of GVL produced in this process.

In total, four optimisation runs of the genetic algorithm LEA3D are performed to find an optimal solvent for the GVL purification. For molecular design, all functional groups are included that were in the training set of the ANN, e.g. alkane-, benzene-, amine-, sulfone- keto- or hydroxyl-fragments (see Appendix A.4 for details). Thus, all molecules that are designed should be predictable by the ANN without forcing the ANN to extrapolate. For all optimisation runs, the objective is to minimise the

cradle-to-grave impact on Climate Change ($CC_{\text{cradle-to-grave}}$) by summing the impacts on Climate Change of the three life cycle stages of this process: solvent production ($CC_{\text{Production}}$), solvent use in the process (CC_{Process}) and solvent disposal (CC_{Disposal}):

$$\min CC_{\text{cradle-to-grave}} = CC_{\text{Production}} + CC_{\text{Process}} + CC_{\text{Disposal}} \quad (3.2)$$

The impact on Climate Change of the process (CC_{Process}) is linearly proportional to the energy demand. The energy demand in the distillation column captures the operating cost of the process. Thus, economically attractive solvents have a low impact from the use phase. Therefore, the optimisation of the cradle-to-grave impact on Climate Change yields solvents with a balanced contribution from all life cycle phases and low operational cost. If desired, multi-objective optimisation could be employed to optimise cost and impact on Climate Change explicitly.

3.2.2 Results and discussion

In total, the optimisation generates more than 1600 unique solvents, which are evaluated in the 4 design runs in about 5 days (121 hours). From all candidate solvents, 703 solvents fulfil the property constraints and are suitable for the process. Therefore, a ranking of 703 solvents according to their cradle-to-grave impact on Climate Change is obtained as a result (Figure 3.6).

The solvent with the highest reduction in the impact on Climate Change is 2,3,3,5-tetramethyl-hexane, with a cradle-to-grave impact on Climate Change of about 4.4 kg CO₂-eq. kmol_{GVL}⁻¹. Compared to the benchmark n-butyl acetate (Murat Sen et al., 2012), 2,3,3,5-tetramethylhexane reduces the impact on Climate Change by about 68%. More generally, 291 of the 703 candidate solvents have a lower impact on Climate Change than the benchmark, and 169 solvents outperform the benchmark in terms of Climate Change and process energy demand Q_{reb} . COSMO-susCAMPD thus designs successfully many suitable alternatives. For the top 15 candidates, very similar solvents are found: The top 15 solvents are all alkanes and alkenes, most of which are highly branched and therefore not yet commercially available. The highest-ranking commercially available bulk chemical is n-octane on rank 8. N-octane reduces the impact on Climate Change by about 67.5% compared to the benchmark solvent, which is very close to the impact reduction of the optimal solvent.

To challenge the use of the cradle-to-grave impact as an objective function, the cradle-to-grave impact on Climate Change is compared with the gate-to-gate impact from process energy demand during solvent use. The impact on Climate Change from

process energy depends linearly on the process energy demand Q_{reb} (black line in Figure 3.6) and thus represents the result of an economic optimisation for minimum process energy demand as typically used in CAMPD. Intuitively, one might expect that energy demand in the use phase captures the cradle-to-grave impact on Climate Change already well. However, in this case study, the cradle-to-grave impact of 187 candidate solvents deviates by more than 50% from the impact of process energy (Figure 3.6). The deviation from the impact caused by the process energy is due to the production and disposal of the candidate solvents. This deviation highlights the importance of the cradle-to-grave system boundary. Still, for this case study, the top 15 solvents with the lowest impact on Climate Change equal the top 15 solvents with the lowest process energy demand Q_{reb} . For these solvents, the process requires 60.6 - 62.4 MJ kmol_{GVL}⁻¹ energy for distillation, corresponding to a reduction of about 46 - 48% compared to the benchmark n-butyl acetate.

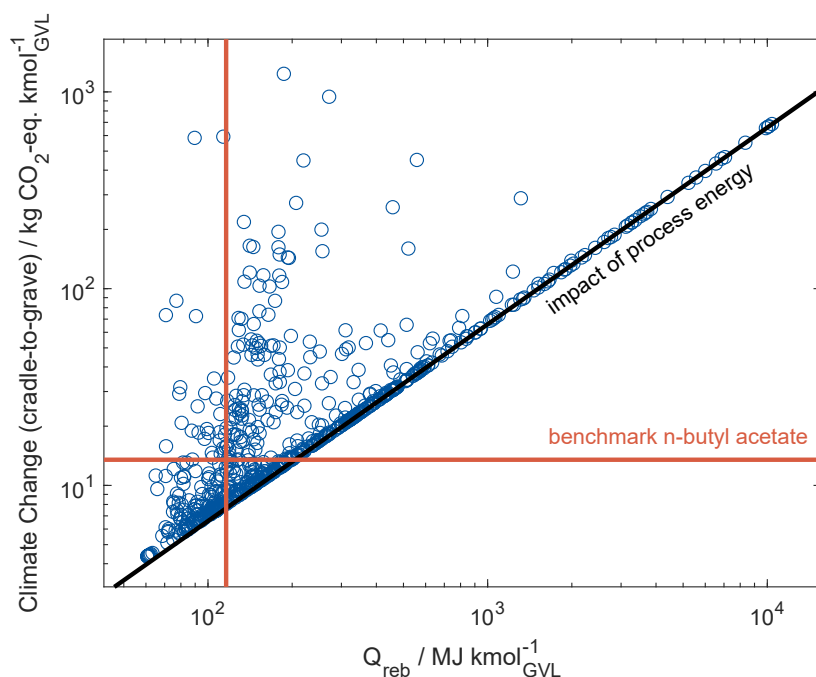


Figure 3.6: Cradle-to-grave impacts on Climate Change (CC) of all solvents designed versus corresponding process energy demand Q_{reb} . Each blue circle represents one candidate solvent. The black line is the impact resulting from the process energy demand; the red lines represent the impact on Climate Change and process energy demand of the benchmark solvent n-butyl acetate.

Moreover, the importance of the cradle-to-grave system boundary is shown by comparison to the ranking by cradle-to-gate LCA: A cradle-to-gate LCA based on the specific impacts of solvent production yields a very different ranking (Figure 3.7 A). The solvent with the lowest cradle-to-gate impact on Climate Change per kilogram solvent is divinyl ether with about $2.1 \text{ kg CO}_2\text{-eq. kg}_{\text{solvent}}^{-1}$. However, divinyl ether has a cradle-to-grave impact on Climate Change of about $7.3 \text{ kg CO}_2\text{-eq. kmol}_{\text{GVL}}^{-1}$ ranking only 75th in cradle-to-grave impact. 2,3,3,5-tetramethyl-hexane, the solvent with the lowest cradle-to-grave impact, ranks only 139th with a higher cradle-to-gate impact on Climate Change of about $2.5 \text{ kg CO}_2\text{-eq. kg}_{\text{solvent}}^{-1}$. Therefore, concentrating only on the specific cradle-to-gate LCA of the solvent production proves to be a misleading objective. Specific assessment of molecular properties is not sufficient. Instead, the amount of solvent used in the process needs to be considered for solvent selection with an environmental objective. In particular, the specific cradle-to-gate impacts are quite similar for all solvents (x-axis of Figure 3.7 A) in this case study. In contrast, the cradle-to-grave impact spans multiple orders of magnitude (y-axis of Figure 3.7 A) yielding a more selective objective.

The differences in the ranking between cradle-to-gate and cradle-to-grave LCA can be explained by the neglect of the solvent use phase: For solvents with a high cradle-to-grave impact, a high amount of solvent is lost in the wastewater stream (Table 3.2). A high solvent loss to wastewater causes a high make-up demand to run the process in steady-state. Therefore, a high amount of solvent needs to be produced for make-up, causing high absolute impacts from solvent production regarding the functional unit of 1 kmol GVL. Conversely, a low impact of solvent production is only achieved with a small make-up demand of solvent, in particular as the specific cradle-to-gate impacts are within the same order of magnitude for all candidate solvents.

If the solvent loss is small, the use phase impact due to process energy dominates the cradle-to-grave LCA. Furthermore, a low solvent loss reduces uncertainty propagation of the ANN predictions. As a result, the uncertainties of the cradle-to-grave impact decrease (Table 3.2). Therefore, accurate LCI of the use phase and thus accurate process modelling and precise property data are crucial. As an indicator of the accuracy, the predicted solubilities of solvent in water from COSMO-RS are compared with experimental data from the literature. The solubilities of the solvents in water are crucial for the LCA because they determine the solvent loss and make-up and, consequently, the environmental impact of the solvent production. For the benchmark n-butyl acetate, solubilities of 6.7 - 8.3 g/l at 25 °C have been determined experimentally (Yalkowsky and He, 2003) compared to 6.1 g/l from the COSMO-RS predictions. For n-octane, COSMO-RS predicts a solubility at 25 °C of 3 mg/l compared to 0.4 - 0.9 mg/l

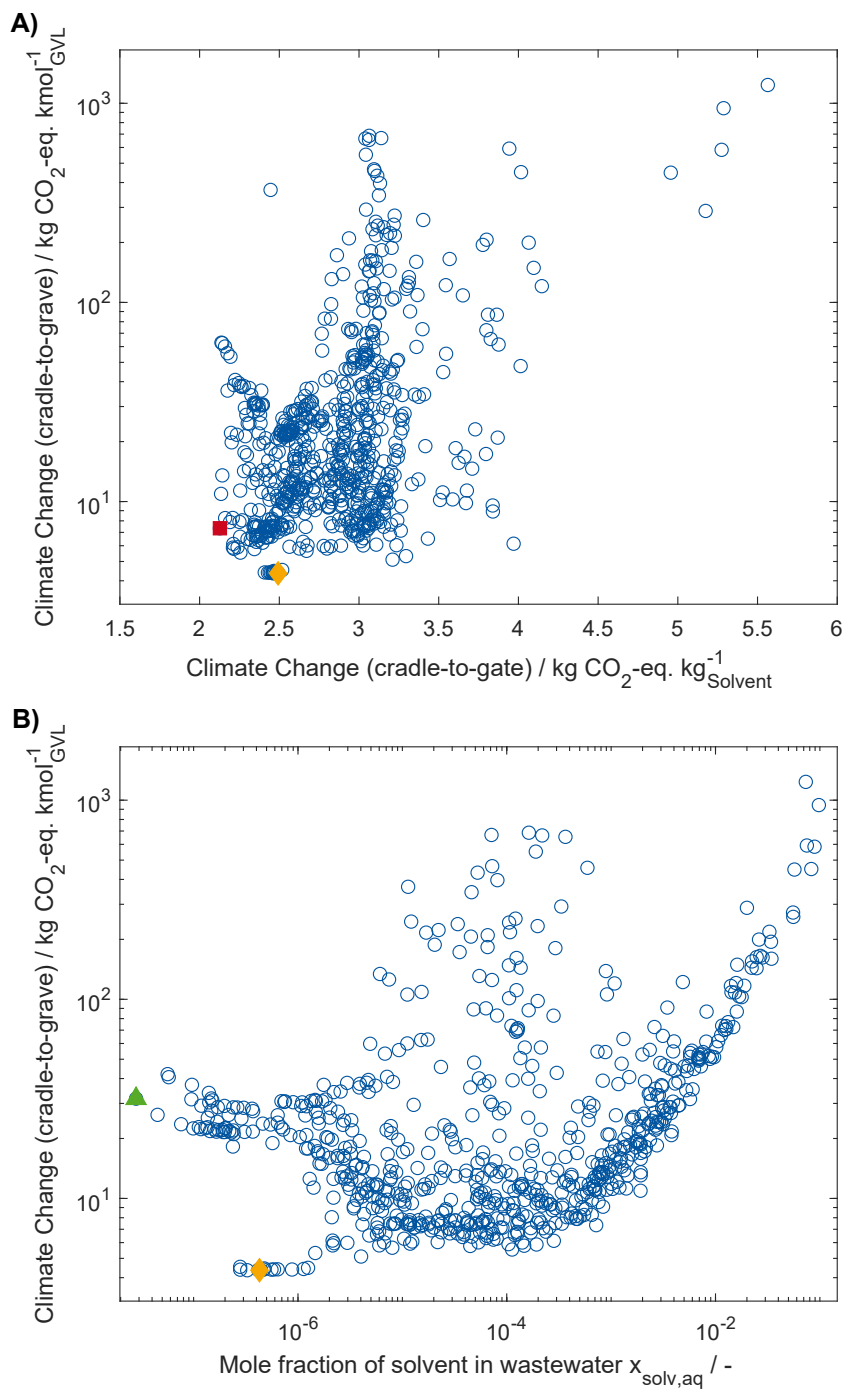


Figure 3.7: (A) Cradle-to-grave impact on Climate Change per kmol GVL versus specific impact from solvent production (cradle-to-grave system boundary) per kilogram solvent. (B) Cradle-to-grave impact on Climate Change versus mole fraction of solvent in the wastewater stream. The red square indicates the solvent with the lowest cradle-to-grave impact on Climate Change, the yellow diamond the solvent with the lowest cradle-to-grave impact on Climate Change, and the green triangle the solvent with the lowest solvent loss to wastewater.

3.2 Case study: Benign solvents for the hybrid extraction-distillation of γ -valerolactone

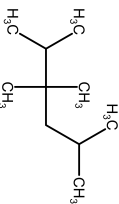

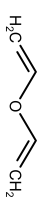
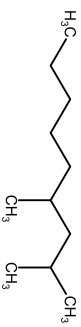
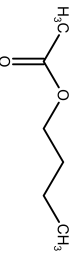
experimentally (Yalkowsky and He, 2003). Considering the broad range of solubilities over multiple orders of magnitude, the experimental measures are both in good agreement with the COSMO-RS prediction. Thus, in conclusion, COSMO-RS can be used for property prediction to generate accurate LCI for this process, even for the challenging hydrocarbon-water interactions (Klamt, 2003).

Still, the solvent loss alone is also not sufficient as an objective for molecular design (Figure 3.7 B). The solvent with the lowest solvent loss, 2,4-dimethyl-nonane, ranks only 600th in process energy demand Q_{reb} and 524th in cradle-to-grave impact on Climate Change. The advantageous low solvent loss does not guarantee a low cradle-to-grave impact on CC, as low solvent loss and low energy demand for separation do not correlate. The high energy demand in distillation outweighs the favourable low solvent loss and make-up. Therefore, top solvents balance solvent loss as well as specific production impact and process energy demand. To include all these relevant factors, the cradle-to-grave LCA is required as objective function.

Besides the impact on Climate Change, the other 16 ReCiPe midpoint impact categories (Goedkoop et al., 2009) are evaluated as well for every candidate solvent in COSMO-susCAMPD. Generally, solvents ranked well in the impact on Climate Change and process energy demand show also a balanced performance in most of the other impact categories. For example, the top solvent in cradle-to-grave impact on Climate Change is also among the top 10 solvents in 15 of the other 16 impact categories. As for the impact on Climate Change, the low solvent loss in the process combined with low energy demand in separation yields a low cradle-to-grave LCA impact.

For this case study, only the impact category Ozone Depletion (OD) differs from the trend of all other impact categories (Figure 3.8). For Ozone Depletion, a strong trade-off between the cradle-to-grave impact on Ozone Depletion and the process energy demand for most solvents can be identified. As a result, the solvent ranking differs substantially for Ozone Depletion. For example, 5 of the top 10 solvents in Ozone Depletion occupy ranks 400 and higher in Climate Change or ranks 500 and higher in process energy demand. The change in ranking for Ozone Depletion is due to the fact that the impacts due to solvent production and solvent loss dominate the impact on Ozone Depletion. This outcome is reasonable since process energy is supplied as heat from natural gas combustion with no substantial impact on Ozone Depletion. Thus, the presented method also captures the variable weighting of and trade-offs between the life cycle stages depending on the impact category considered.

Table 3.2: Comparison of the candidate solvents with the lowest cradle-to-grave and the lowest cradle-to-gate impact on Climate Change, as well as the solvent with the lowest solvent loss and the benchmark solvent. The comparison includes the absolute values for the cradle-to-grave impact on Climate Change ($CC_{\text{cradle-to-grave}}$, the ranking in cradle-to-grave and cradle-to-gate impact on Climate Change (Rank $CC_{\text{cradle-to-grave}}$ and Rank $CC_{\text{cradle-to-gate}}$), the ranking in process energy demand Q_{reb} (Rank Q_{reb}), as well as the molar fraction of solvent in wastewater (Predicted $x_{\text{soliv,aq}}$) predicted by COSMO-RS. The absolute values for $CC_{\text{cradle-to-grave}}$ also include the 95% confidence interval from the uncertainty propagation of the ANN.

Solvent	Molecular structure	$CC_{\text{cradle-to-grave}} / \text{kg CO}_2\text{-eq, kmol}_{\text{GVL}}^{-1}$	Rank $CC_{\text{cradle-to-grave}}$	Rank $CC_{\text{cradle-to-gate}}$	Rank Q_{reb}	Predicted $x_{\text{soliv,aq}}$
lowest Climate Change (cradle-to-grave)		4.36 ± 0.0017	1	139	1	4.3×10^{-7}
- commercially avail.		4.40 ± 0.0015	8	97	9	4.7×10^{-7}
lowest Climate Change (cradle-to-gate)		7.34 ± 1.4	75	1	27	7.2×10^{-4}
lowest solvent loss		31.7 ± 0.00012	524	281	600	2.8×10^{-8}
benchmark (n-butyl acetate)		13.5 ± 3.1	292	247	198	9.4×10^{-4}

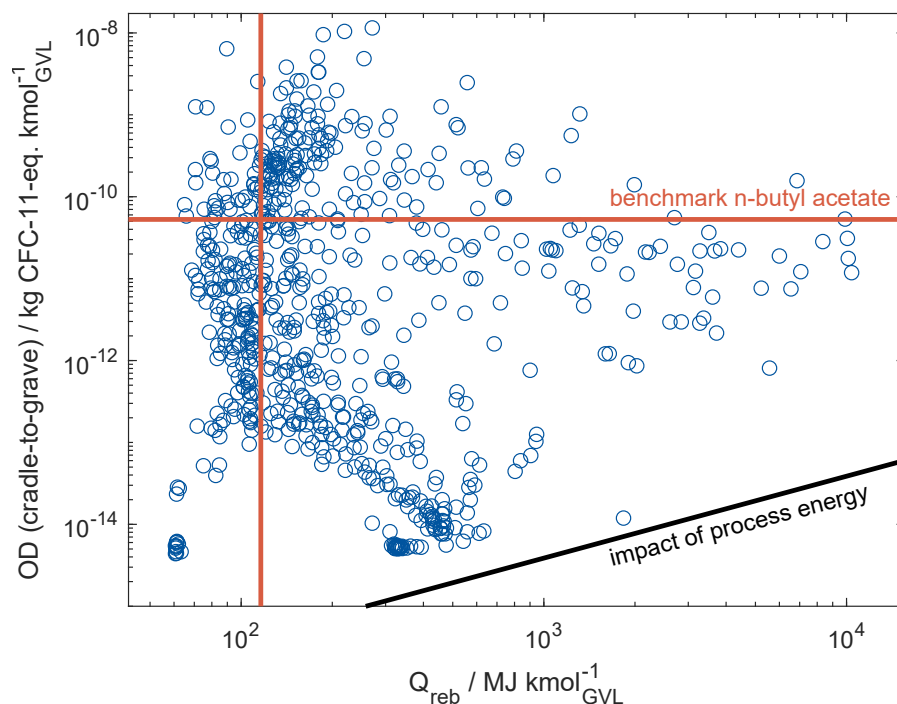


Figure 3.8: Cradle-to-grave impacts on Ozone Depletion (OD) of all solvents designed depending on the corresponding process energy demand Q_{reb} . Each blue circle represents one candidate solvent. The black line is the impact resulting from the process energy demand; the red lines stand for the impact on Ozone Depletion and process energy demand of the benchmark solvent n-butyl acetate.

3.3 Conclusion

In this chapter, a framework for the design of solvents and processes with an environmental objective is presented: COSMO-susCAMPD. The COSMO-susCAMPD framework extends state-of-the-art Computer-Aided Molecular and Process Design (CAMPD) by integrating predictive Life Cycle Assessment (LCA) with a cradle-to-grave system boundary. Cradle-to-grave LCA is achieved by the combination of (1) an Artificial Neural Network (ANN) predicting cradle-to-gate impacts with (2) process optimisation using pinch-based process models providing life cycle inventory for solvent use and disposal. Both the ANN and the process models use molecular and thermodynamic properties calculated from the predictive thermodynamic model COSMO-RS. Therefore, the assessment of environmental impacts and process performance is based on one consistent set of descriptors. For simultaneous molecular and process design, the

predictive LCA and the process optimisation are combined with the genetic algorithm LEA3D, which optimises 3D-molecular structures based on the results from LCA and process optimisation.

As an application for COSMO-susCAMPD, the purification of the bio-based platform chemical γ -valerolactone from aqueous solution by hybrid extraction-distillation is investigated. The process is optimised for minimum environmental impact by exploiting the degrees of freedom from molecular and process design. As a result, promising candidate solvents are identified from a vast design space outperforming the literature benchmark n-butyl acetate by reducing the impact on Climate Change by about 68%. The candidate solvents identified exhibit both a high process performance, i.e. a low process energy demand, as well as a low cradle-to-grave environmental impact in various ReCiPe midpoint impact categories.

The results show that a cradle-to-grave assessment at the system level is necessary for the design of environmentally beneficial solvents. Simplified objectives, such as cradle-to-gate LCA or solely economic evaluation, lead to suboptimal solutions. Only the cradle-to-grave LCA balances conflicting molecular properties for an optimal result.

The COSMO-susCAMPD framework now provides a method for CAMPD based on process evaluation and environmental assessment using LCA. The results of COSMO-susCAMPD serve as an input for further validation by refined process simulations, life cycle assessment and experiments. In particular, the LCA could be refined to include other emissions, such as fugitive emissions of the process (Smith et al., 2017). Further work is required to extend the LCA data for training the ANN. Currently, training data for the ANN is rare, leading to limited accuracy of the ANN predictions. An improvement in the prediction quality of the ANN is expected if more consistent LCA data on solvents is available. Importantly, any additional training data needs to be obtained from process data by consistent allocation and with consistent background data. However, for the given case study, the prediction of accurate process data outweighs the influence of inaccuracies of the ANN.

In conclusion, the presented COSMO-susCAMPD framework extends the environmental assessment of state-of-the-art molecular design by predictive cradle-to-grave life cycle assessment. Thus, the COSMO-susCAMPD framework enables the computer-aided design of sustainable solvents and processes by evaluating environmental impacts at the system level. In the next chapter, the COSMO-susCAMPD framework is advanced by comprehensive process modelling at the application level including additional unit operations to apply the framework to reaction-separation processes beyond hybrid extraction-distillation.

Computer-aided design of solvents and processes using quantum chemistry

Chapter 3 shows the importance of considering the full life cycle of solvents for CAMPD with environmental objective at the system level. This chapter focuses on a comprehensive consideration of the application level in CAMPD. Comprehensive modelling and design of the application level are achieved by increasing the process design scope and the resolution of chemical process modelling. For this purpose, the computer-aided molecular design of solvents is integrated with the design of heat-integrated processes to consider the interactions between molecular properties and process performance for minimum utility demand or minimum environmental impact. The method is based on the COSMO-(sus)CAMPD method for the integrated design of molecules and processes using COSMO-RS described in the previous Chapter 3.

To explain and demonstrate the extended COSMO-(sus)CAMPD method, this chapter is structured as follows: In Section 4.1, the integrated CAMPD problem is formulated as an optimisation problem, and the solution algorithm is explained. The integrated models and the methods used are described in detail. In Section 4.2, the extended COSMO-(sus)CAMPD method is applied to two case studies: A hybrid extraction-distillation process (Section 4.2.1) and an integrated carbon capture and utilisation process (Section 4.2.2). The extended COSMO-(sus)CAMPD method is compared with the state of the art, and the new capabilities of the method are highlighted before conclusions are drawn in Section 4.3.

Major parts of this chapter are reproduced by permission of John Wiley & Sons, Inc., from:

Fleitmann, L.; Gertig, C.; Scheffczyk, J.; Schilling, J.; Leonhard, K. and Bardow, A. (2023). From molecules to heat-integrated processes: Computer-aided design of solvents and processes using quantum chemistry. *Chemie Ingenieur Technik*, 95(3), 368–380.

The author of this thesis contributed to the methodology and the implementation of the CAMPD framework. The author investigated, validated and visualised the results of the molecular and process design and wrote the first draft as the principal author. The conceptualisation and methodology were jointly developed in discussion with the co-authors.

4.1 COSMO-(sus)CAMPD for heat-integrated processes

The extended COSMO-(sus)CAMPD method integrates molecular and process design using quantum chemistry-based property prediction and pinch-based process models for unit operations and heat integration. The method can be formulated as an optimisation problem for maximum process performance:

$$\begin{array}{ll}
 \min_{x,y} & f(x, y, \chi, \psi, \theta) \quad \text{Objective function} \\
 \text{s.t.} & \chi = j(y, \psi) \quad \text{Heat integration model} \\
 & \psi = h(y, \theta) \quad \text{Process model} \\
 & \theta = g(x, y) \quad \text{Model for predicting thermodynamic properties} \\
 & 0 = m(x) \quad \text{Representation of molecules} \\
 & k_1(x) \leq 0 \quad \text{Molecular constraints} \\
 & k_2(\theta) \leq 0 \quad \text{Thermodynamic constraints} \\
 & k_3(y, \theta) \leq 0 \quad \text{Process constraints} \\
 & x \in \mathbb{X} \quad \text{Molecular structure} \\
 & y \in \mathbb{Y} \quad \text{Process variables}
 \end{array} \tag{4.1}$$

In this optimisation problem, the molecular structure x and the process variables y are optimised for a process design objective function $f(x, y, \chi, \psi, \theta)$, e.g. the environmental impacts measured by an LCA impact category, resource consumption measured by the exergy demand, or the operating cost of the process. The objective function f may depend on the molecular structure x , process variables y , targets for heat integration χ , process model results ψ , and thermodynamic properties θ . To calculate the objective, models for heat integration, process, thermodynamic properties, and molecular structure are integrated:

- The **heat integration model** $j(y, \psi)$ calculates maximum feasible heat integration and corresponding minimum demands of utilities χ based on the process variables y (e.g. optimised temperatures in unit operations), and process model results ψ from the process model h (e.g. heat and mass flows).
- The **process model** $h(y, \theta)$ combines individual models of unit operations and depends on process variables y (e.g. temperature or pressure settings in unit operations) and thermodynamic properties θ (e.g. activity coefficients, heat capacities or enthalpies of vaporisation).

- The predictive **thermodynamic models** $g(x, y)$ calculate thermodynamic properties for process model evaluation depending on the molecular structure x and process variables y .
- The **molecular structure model** $m(x)$ represents the candidate molecules using 3D-fragments and ensures chemical feasibility, e.g. the octet rule.

The individual models are combined in one method for molecular and process design (Figure 4.1). The optimisation of the integrated design problem is subject to constraints at the molecular level $k_1(x)$ (e.g. the maximum number of non-hydrogen atoms), thermodynamic constraints $k_2(\theta)$ (e.g. a minimum boiling point of the solvent), and process constraints $k_3(y, \theta)$ (e.g. limits on process variables). The constraints k_1 , k_2 and k_3 tighten the optimisation problem and thus increase the computational efficiency of the algorithm.

Solution algorithm: Property prediction using quantum chemistry and process optimisation are integrated into a molecular design algorithm (Douguet et al., 2005) resulting in an evolutionary optimisation procedure of four steps per iteration (Figure 4.1): (1) generation of candidate solvents, (2) prediction of thermodynamic properties, (3) process optimisation and (4) ranking of candidate solvents.

1. Generation of candidate solvents: As the first step of each iteration, candidate solvents are generated using the genetic algorithm LEA3D (Douguet et al., 2005), as described in Section 3.1.1. The genetic algorithm forms the outer loop of the optimisation procedure and runs property prediction (Step 2) and process optimisation (Step 3) for each candidate solvent (Figure 4.1). LEA3D optimises the molecular structure based on the results of process optimisation for each solvent of each generation, i.e. optimum process settings y^* , optimum process model results ψ^* , and optimum heat integration χ^* :

$$\begin{aligned}
 \min_x \quad & f(x, y^*, \chi^*, \psi^*, \theta) \\
 \text{s.t.} \quad & \chi^* = j(y^*, \psi^*) \\
 & \psi^* = h(y^*, \theta) \\
 & \theta = g(x, y^*) \\
 & 0 = m(x) \\
 & k_1(x) \leq 0 \\
 & k_2(\theta) \leq 0 \\
 & x \in \mathbb{X}
 \end{aligned} \tag{4.2}$$

At any time, the candidate solvents from LEA3D satisfy chemical feasibility expressed through the equality constraints $m(x)$. LEA3D also checks the molecular constraints $k_1(x)$ for every candidate solvent. Candidate solvents violating the molecular constraints are discarded from further evaluation.

2. Prediction of thermodynamic properties: For each candidate solvent, thermodynamic properties θ are predicted based on quantum chemistry. Evaluating a process for operating cost or thermodynamic performance requires equilibrium and thermochemical properties. For this purpose, two quantum chemistry-based methods are employed: COSMO-RS (Klamt et al., 2010) for equilibrium properties (Paragraph 2a) and automated thermochemistry calculations for ideal gas properties (Paragraph 2b).
 - a) COSMO-RS predicts equilibrium properties of pure components and mixtures, allowing to compute liquid phase properties as well as transitions between gas and liquid phase, e.g. activity or Henry coefficients or enthalpies of vaporisation (cf. Section 3.1.1). Based on the thermodynamic equilibrium properties, property constraints $k_2(\theta)$ on thermodynamic requirements are evaluated, e.g. limits for boiling points or the existence of azeotropes (Skiborowski et al., 2016). Candidates are only further considered for thermochemical calculations and process optimisation if they are considered thermodynamically suitable by fulfilling these constraints.
 - b) Thermochemistry is used to calculate ideal gas properties, i.e. ideal gas heat capacities. Based on the pre-optimised geometries from BP86/TZVP calculations, the molecular geometries are optimised, and vibrational frequencies are computed using the DFT functional B3LYP (Becke, 1993; Stephens et al., 1994) with TZVP basis set assuming the rigid rotor harmonic oscillator (RRHO) approximation (Atkins and Friedman, 2011). B3LYP is more accurate than BP86 for geometry optimisation and vibrational frequencies and known for a good balance between computational cost and accuracy (Zheng et al., 2009; Gottschalk et al., 2018). Based on the optimised geometries and the vibrational frequencies, frequency analysis is performed using the TAMkin package (Ghysels et al., 2010) to yield the thermochemical properties.

Details on the software used for the quantum chemistry calculations and a brief comparison of the property prediction accuracy with experimental data can be found in Appendix B.1 and Appendix B.2.

3. Process optimisation: With the thermodynamic properties from Step 2, the entire process flowsheet is modelled using pinch-based process models for individual process units (Paragraph 3a) and the transshipment model for heat integration (Paragraph 3b). The used models are computationally efficient and converge more robustly than rigorous models. Therefore, the process degrees of freedom, e.g. operating temperatures or pressures, considering subsequent heat integration can be optimised for each candidate solvent.

The process optimisation problem for each solvent is thus a subproblem of Problem 4.1 optimising the process settings y with fixed molecular structure x' and corresponding fixed thermodynamic properties θ' :

$$\begin{aligned}
 \min_y \quad & f(x', y, \chi, \psi, \theta') \\
 \text{s.t.} \quad & \chi = j(y, \psi) \\
 & \psi = h(y, \theta') \\
 & k_3(y, \theta') \leq 0 \\
 & y \in \mathbb{Y}
 \end{aligned} \tag{4.3}$$

Process optimisation yields optimum process settings y^* for each solvent resulting in optimum process model results ψ^* and optimum heat integration χ^* , which are used by LEA3D to optimise the molecular degrees of freedom x (see Step 1 and Problem 4.2).

- a) Process units are modelled using equilibrium- and pinch-based process models. Pinch-based process models are available for the most common separation unit operations: absorption (Redepenning et al., 2017), extraction (Redepenning et al., 2017), and distillation (Bausa et al., 1998). An equilibrium-based multiphase reactor from Scheffczyk et al. (2017a) is available using the homotopy continuation algorithm by Bausa and Marquardt (2000) for phase equilibrium calculations. These equilibrium- and pinch-based process models allow modelling of entire process flowsheets for many processes.

The pinch-based process models are well suited for CAMPD since they consider non-ideal thermodynamics without simplifications to heuristic performance indicators (see Appendix B.3 for a brief comparison to rigorous process models). Nevertheless, the calculation of process units is efficient and robust, e.g. by avoiding tray-to-tray calculations. Instead, pinch-based process models calculate the minimum operating point of a column by assuming vanishing thermodynamic driving forces. This assumption

corresponds to columns with an infinite number of trays. Thus, in the context of balancing operating and investment expenditures, the results of the pinch-based models represent the limiting case of minimal operating expenditures without considering investment cost. CAMPD using pinch-based process models is suited to optimise thermodynamic quantities, e.g. exergy loss, energy and solvent demand, or operating cost, e.g. cost of utility consumption.

- b) Heat integration is performed by pinch analysis using the transshipment model of Papoulias and Grossmann (1983). The transshipment model yields the maximum heat integration of a thermodynamically optimal heat recovery network. Similar to the pinch-based process models for the unit operations, no investment costs for the heat recovery network but targets for minimum utility consumption are calculated based on heat and mass flows. Using the transshipment model allows formulating the heat integration problem model as a linear program and solving it computationally efficiently. Thus, in combination with the pinch-based process models, maximum heat-integrated process performance is evaluated for each solvent based on a process-level objective function.

During process optimisation, process models and heat integration are solved iteratively for each candidate solvent. Based on the gradient of the predefined objective function, the process degrees of freedom are optimised using a numerical optimisation method from multiple starting points (see Appendix B.1 for details).

4. Ranking of molecules and next generation: Based on each candidate's optimised process performance, the candidate solvents are scored and ranked according to the predefined objective function. The objective function value is used as a fitness score for the genetic algorithm, which applies genetic operations to generate a new set of molecules (Step 1).

The four steps of the method are repeated until a predefined number of generations is met or a desired improvement is achieved, and the algorithm terminates. The result is a ranked list of molecules and corresponding optimal process settings that can be further refined by additional design criteria and validation. In this chapter, all candidate solvents of the ranked list are checked for commercial availability or synthesisability. Commercial availability is verified by searching online databases. If a candidate solvent is not commercially available, we check synthesisability using a retrosynthesis method based on the attention-based molecular transformer model (Schwaller et al., 2019, 2020). A candidate solvent is considered synthesisable if at most three subsequent reactions are required to form the candidate solvent from commercially available reactants with a confidence of the retrosynthesis algorithm greater than 50%.

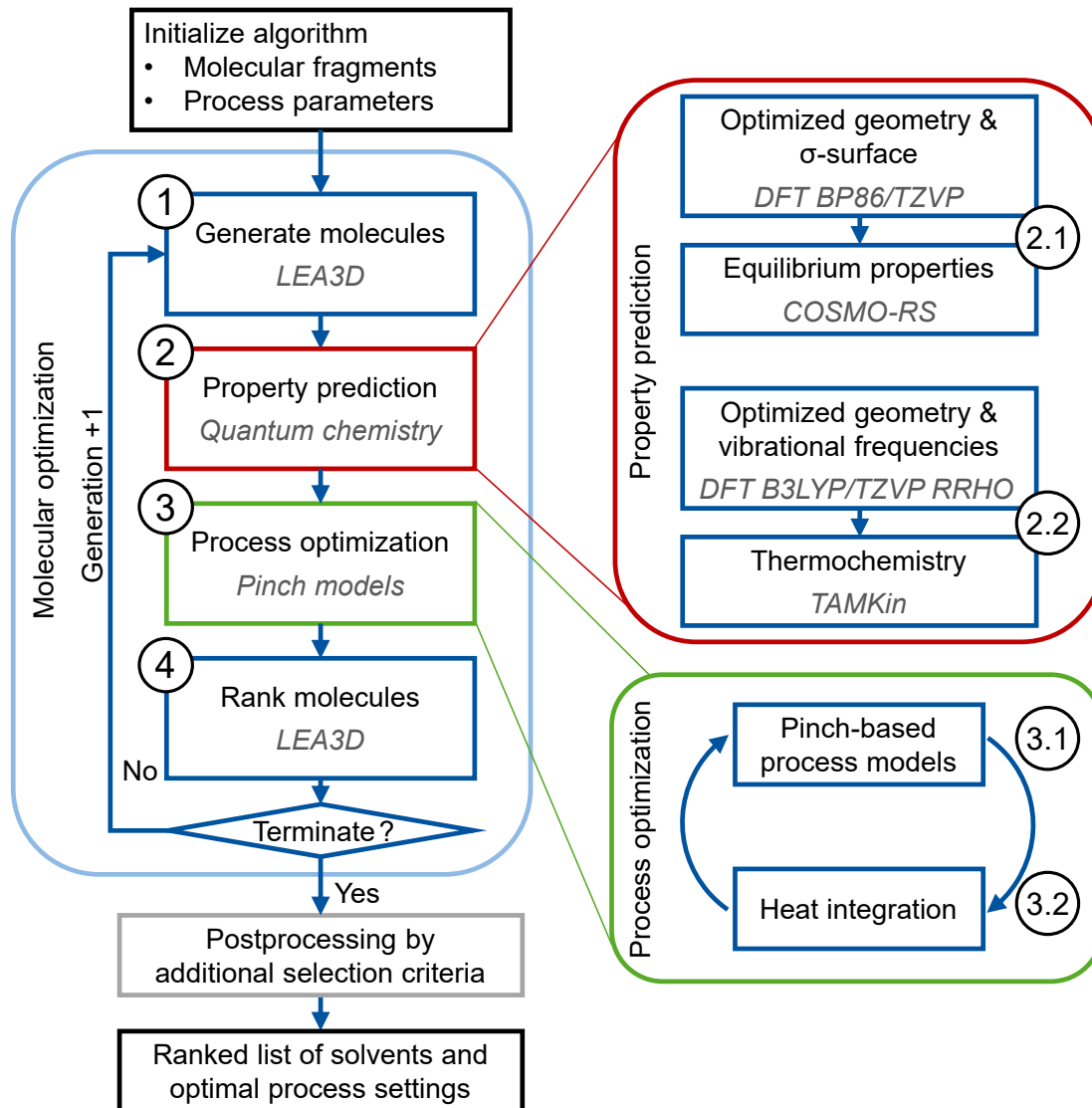


Figure 4.1: COSMO-(sus)CAMPD method for the design of heat-integrated processes by including thermochemistry from quantum chemistry calculations and process optimisation including heat integration. The methods used are given in italics.

4.2 Case studies

The extended COSMO-(sus)CAMPD method is applied to two case studies: (1) a hybrid extraction-distillation process for the purification of γ -valerolactone and (2) an Integrated Carbon Capture and Utilisation (ICCU) process to produce carbon monoxide. The case studies demonstrate the capabilities of COSMO-(sus)CAMPD for integrated design of solvents and heat-integrated processes. In particular, the results from COSMO-(sus)CAMPD with extended property prediction using thermochemistry and heat integration are compared to state-of-the-art methods.

In addition to the environmental assessment, an analysis of process exergy demand can yield additional information on minimum required work and overall efficiency of the process. Therefore, the design for minimum environmental impact in this chapter is complemented by an analysis of process exergy demand in Appendix B.5.

4.2.1 Hybrid extraction-distillation of γ -valerolactone

As a first case study, the hybrid extraction-distillation of γ -valerolactone (GVL) is investigated, as introduced in Section 3.2. The objective of the integrated solvent and process design is to minimise the cradle-to-grave environmental impact on Climate Change ($CC_{c2grave}^{tot}$) of the solvent after heat integration by choosing an optimal solvent with the corresponding optimal process settings. In contrast to Section 3.2, the process modelling considers the reboiler and condenser duties in the distillation column, as well as the sensible heats for heating and cooling of various flows (Figure 4.2). In this case study, the degrees of freedom of the process are the extraction and decanter temperatures T_{Extr} and T_{Dec} and the pressure in the distillation column p_{Dist} . Heat is supplied by low-pressure (3 bar) and high-pressure (70 bar) steam at 410 K and 558.15 K, and cooling is provided by cooling water at 283 K. The heat recovery approach temperature (HRAT) is assumed to be 10 K. More details on the process specifications can also be found in Appendix B.6.

For the LEA3D algorithm, the 3D molecular fragment library is limited to fragments containing carbon, hydrogen, and oxygen to design potentially green and bio-based solvents (for details on the molecular fragments, see Appendix B.7.1).

In total, COSMO-(sus)CAMPD designs 715 unique candidate solvents in approximately 3.5 days. Of the 715 candidate solvents, 348 candidate solvents fulfil the property constraints and are feasible as solvents for the process. 40 candidate solvents are neither commercially available nor predicted to be synthesisable as revealed by database search and retrosynthesis and are thus discarded after the design. The solvent with

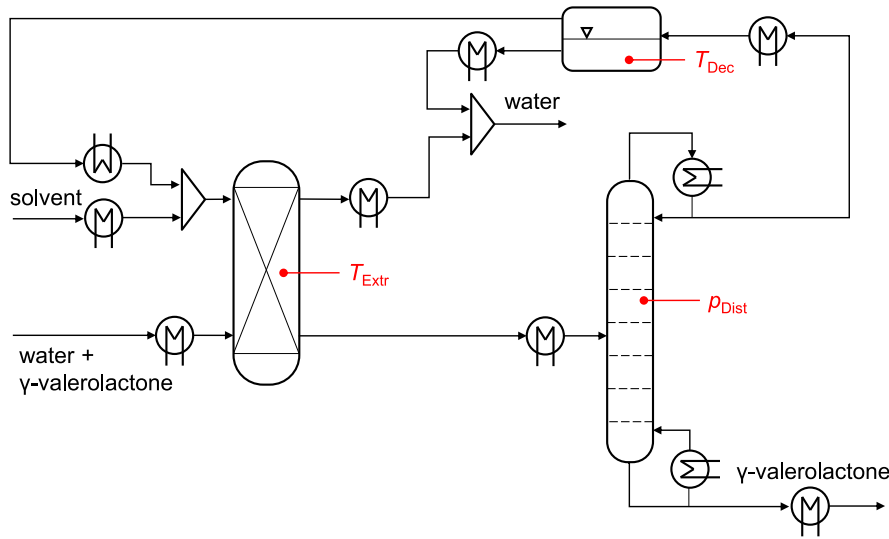


Figure 4.2: Flowsheet and degrees of freedom for the hybrid extraction-distillation of γ -valerolactone with process degrees of freedom highlighted for the process units.

minimal $CC_{c2grave}^{tot}$ is 6-methyl-1,3-heptadiene (cf. Table 4.1). 6-methyl-1,3-heptadiene yields $CC_{c2grave}^{tot} = 6.00 \text{ kg CO}_2\text{-eq. kmol}_{GVL}^{-1}$ for the hybrid extraction-distillation process corresponding to a reduction of 56% compared to the benchmark n-butyl acetate with $CC_{c2grave}^{tot} = 13.8 \text{ kg CO}_2\text{-eq. kmol}_{GVL}^{-1}$. The highest-ranking candidate solvent that is commercially available is 4,6-dimethyl-1-heptene on rank 2. The impact on Climate Change resulting from using 4,6-dimethyl-1-heptene as a solvent is approximately the same as for 6-methyl-1,3-heptadiene, totalling $CC_{c2grave}^{tot} = 6.01 \text{ kg CO}_2\text{-eq. kmol}_{GVL}^{-1}$. Besides the top 2 solvents, the method designs 140 additional candidate solvents with a lower impact on Climate Change than n-butyl acetate highlighting the systematic generation of promising alternatives. Of the top 50 candidate solvents, 42 candidate solvents contain the vinyl group or the furan group, which are thus identified as promising by the method. However, since molecules with vinyl groups tend to polymerise (Tobita, 2000) and furanic compounds are suspected to be toxic and carcinogenic (Bakhiya and Appel, 2010; Moro et al., 2012), these candidates need further assessment. The most promising candidate solvent without a vinyl or furan group is 2,3,3,5-tetramethyl-hexane with an impact on Climate Change of $7.5 \text{ kg CO}_2\text{-eq. kmol}_{GVL}^{-1}$, corresponding to a reduction by 45% compared to n-butyl acetate. 2,3,3,5-tetramethyl-hexane has already been identified as a promising candidate solvent in Chapter 3.

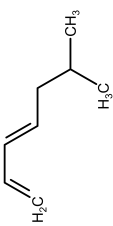
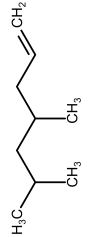
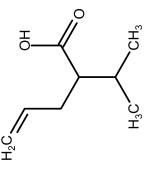

Solvents for hybrid extraction-distillation processes are commonly selected using heuristic selection rules (Gertig et al., 2020b; Adjiman et al., 2021; Chai et al., 2022).

Thus, the solvents identified using the impact on Climate Change of the process as a selection criterion are compared to solvents identified using standard selection rules from the literature. Commonly, heuristic selection rules for an extraction solvent focus only on the solvent's performance in the extraction column (Gertig et al., 2020b; Chai et al., 2022), e.g. the minimum solvent demand for extraction (S_{\min}). From this analysis, S_{\min} moderately correlates with $CC_{c2grave}^{\text{tot}}$ (Pearson's Correlation Coefficient $\rho = 0.53$, Figure 4.3). Thus, the heuristic can be confirmed that promising extraction solvents usually exhibit a low solvent demand for extraction. However, this correlation does not apply among the high-ranking solvents. For example, the correlation between S_{\min} and $CC_{c2grave}^{\text{tot}}$ is weak for the top 100 ranking solvents (Pearson's Correlation Coefficient $\rho = 0.04$), and the solvent with the lowest solvent demand for extraction ranks only 186nd in impact on Climate Change (Table 4.1). Therefore, the minimum solvent demand is not sufficient as an objective to yield a low impact on Climate Change. The entire process needs to be considered for selecting an optimal extraction solvent.

The hybrid extraction-distillation process was already analysed in Section 3.2.2 using only the distillation reboiler energy demand to model the use phase environmental impact ($CC_{c2grave}^{\text{reb}}$). Since the thermochemical estimation of heat capacities was not included for each candidate solvent, COSMO-(sus)CAMPD as presented in Chapter 3 did not consider sensible heat and heat integration. However, ranking the solvents designed for minimum heat-integrated impact on Climate Change by $CC_{c2grave}^{\text{reb}}$ as used in Chapter 3 reveals only minor changes in solvent selection: The rankings are very similar, as indicated by a Spearman's rank correlation coefficient of $\rho_{\text{rank}} = 0.99$. Although heat integration reduces the emissions of the total process energy demand on average by 30 %, heat integration and comprehensive modelling, including sensible heats, have a negligible effect on solvent design for the considered process flowsheet (Figure 4.3). The reboiler duty, already calculated in Chapter 3, represents the main energy demand of the process before and after heat integration since it cannot be heat integrated with this process. Therefore, designing solvents considering only the reboiler duty is sufficient for the process in this case study.

However, neglecting the sensible heat by using only the reboiler energy demand underestimates the heat-integrated impact on Climate Change on average by 19 %. In particular, the reboiler energy demand is underestimated on average by 21 % because the additional heat demand from the temperature increase within the distillation column is not considered. Thus, accurate quantitative results require comprehensive process modelling including sensible heat and heat integration.

Table 4.1: Highest ranking solvents for the hybrid extraction-distillation of γ -valerolactone based on the corresponding objective functions: Minimising the heat-integrated cradle-to-grave impact on Climate Change considering the total process heat ($CC_{c2grave}^{tot}$), minimising the cradle-to-grave impact on Climate Change neglecting sensible heat considering the reboiler heat only ($CC_{c2grave}^{reb}$), or minimising the solvent demand for extraction (S_{min}).

Solvent	Molecular structure	Climate Change per kmol _{GVL}				Solvent demand S_{min}	
		total process heat $CC_{c2grave}^{tot}$ /kg CO ₂ -eq.	Rank	reboiler heat only $CC_{c2grave}^{reb}$ /kg CO ₂ -eq.	Rank	/mol mol ⁻¹	Rank
impact on Climate Change		6.00	1	4.34	1	0.032	30
- commercially avail.		6.01	2	4.34	2	0.026	12
solvent demand for extraction		18.1	186	17.0	197	0.017	1
benchmark (n-butyl acetate)		13.8	143	12.2	162	0.050	100

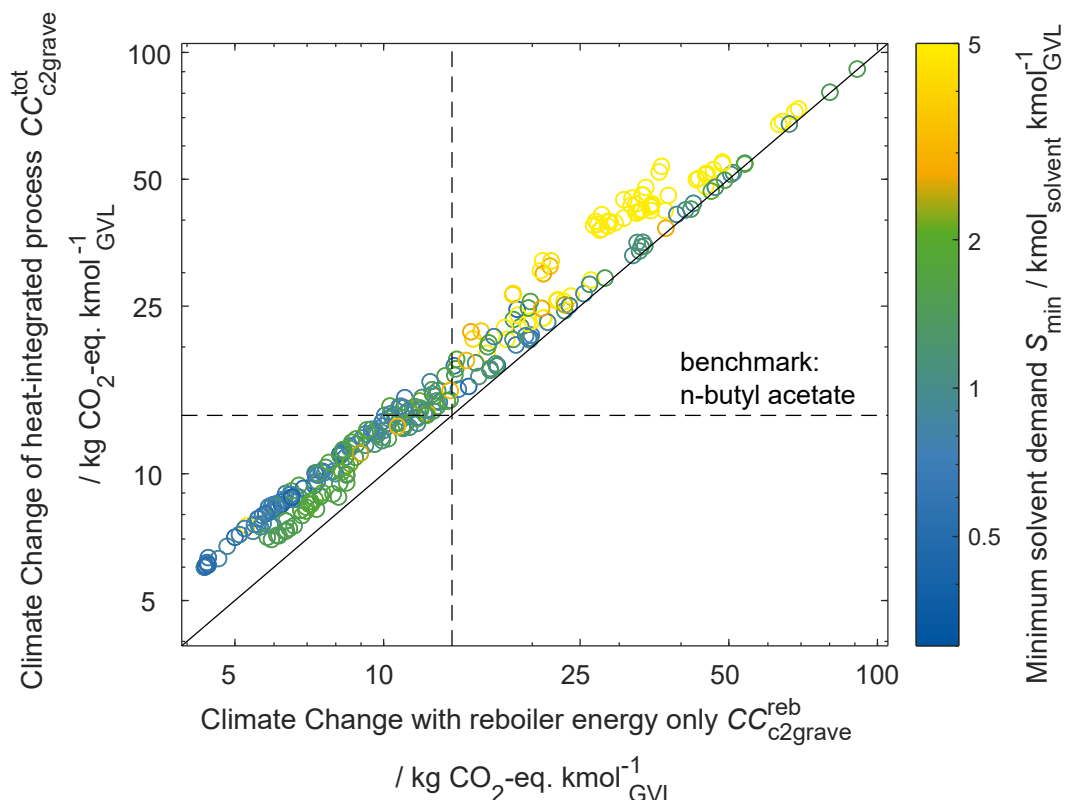
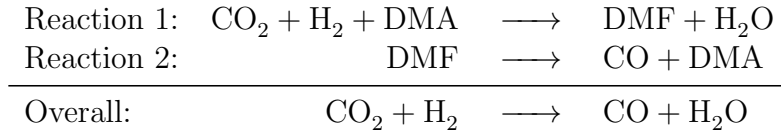


Figure 4.3: Comparison of the impact on Climate Change of the heat-integrated process ($CC_{c2grave}^{tot}$) with the impact on Climate Change neglecting sensible heat and considering only the reboiler energy demand ($CC_{c2grave}^{reb}$). The colour code indicates the heuristic selection criterion of minimum solvent demand for extraction (S_{min}).

4.2.2 Integrated carbon capture and utilisation for the production of carbon monoxide

Integrating carbon capture and utilisation (ICCU) into one process can yield efficient process concepts to capture and utilise carbon dioxide (CO_2) as a feedstock for the chemical industry (Jens et al., 2019). A promising CCU concept is the conversion of CO_2 with hydrogen (H_2) from fluctuating renewable energy to produce carbon monoxide (CO) via a storage molecule that compensates for the fluctuations in electricity supply as a liquid energy carrier (Behr et al., 2004; Supronowicz et al., 2015; Jens et al., 2016). In previous studies by Jens et al. (2016) and Scheffczyk et al. (2017a), the most efficient process was achieved using dimethylformamide (DMF) as a storage molecule.

DMF is produced in the synthesis reaction from CO_2 , H_2 and dimethylamine (DMA, Reaction 1) and subsequently reacted to form CO and dimethylamine in the reforming step (Reaction 2):



So far, the described process has been investigated as a Carbon Capture and Utilisation (CCU) process with captured CO_2 as a pure feedstock for utilisation. The potential of integrating the carbon capture from a CO_2 point source, e.g. CO_2 -rich natural gas, by physical absorption has not yet been evaluated for this ICCU process. An ICCU process omits the energy-intensive CO_2 desorption step from the solvent by converting the CO_2 to a valuable product directly within the solvent (Jens et al., 2016). However, instead, the final product needs to be separated from the solvent.

As a result, process performance and environmental impact of this ICCU process are substantially impacted by solvent and process design, as unit operations for physical absorption, reaction, and distillation are included in the process flowsheet and influenced by the employed solvent (Figure 4.4). Therefore, the optimal solvent needs to balance various properties: (1) high absorption capacity and selectivity, (2) shift of the reaction equilibrium and phase split with water to allow catalyst recovery, as well as (3) low energy demand for heating, cooling, and separation in distillation.

The process design is tailored for each candidate solvent by optimising the reactor pressure p_{Rx} , the pressures in the distillation columns p_{Dist1} and p_{Dist2} and the molar flow of water to the reactor $\dot{n}_{\text{H}_2\text{O}}$ to ensure phase separation. The CAMPD optimisation objective is minimising the cradle-to-grave environmental impact on Climate Change $CC_{\text{c2grave}}^{\text{tot}}$ of the solvent employed in the heat-integrated process. As the CO_2 point source, CO_2 -rich natural gas with 30 mol-% CO_2 and 70 mol-% methane is assumed (Jens et al., 2019). Due to the multifunctionality of the ICCU process, producing purified methane and carbon monoxide, the functional unit (FU) of the process is defined as 2.33 kmol s^{-1} methane and 1 kmol s^{-1} carbon monoxide (cf. Appendix B.4).

The optimised ICCU process is compared with an optimised separated CCU process from the literature, going forward called the benchmark process. The CCU process uses the solvent 3,5-dimethylpiperidine (Jens et al., 2016). For separated carbon capture, conventional chemical absorption using monoethanolamine is assumed (Lee et al., 2016). As a second benchmark, the sum of impacts of conventional methane and carbon monoxide production is considered, going forward called the conventional production of the FU. More details on the LCA, including system boundary and

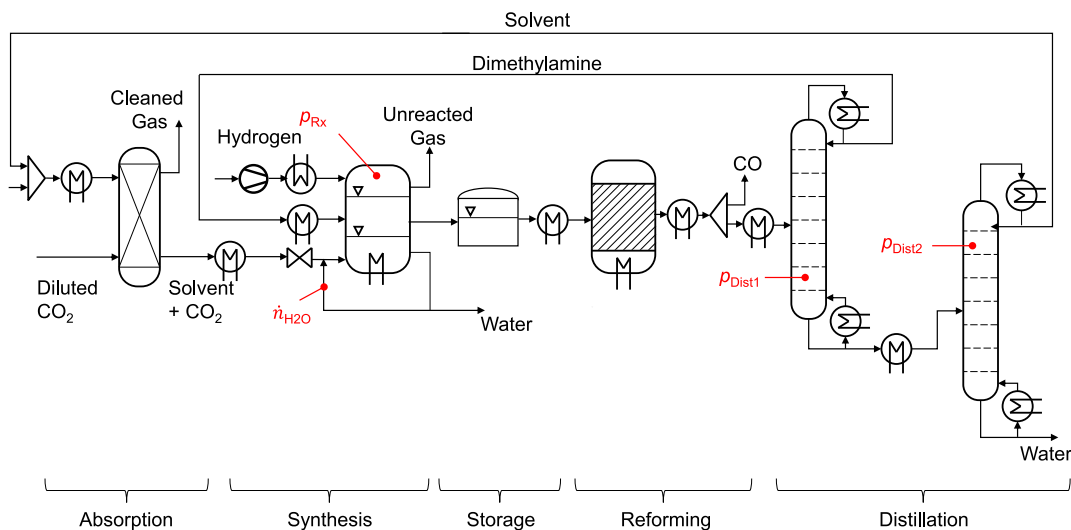


Figure 4.4: Flowsheet for the integrated carbon capture and utilisation process producing carbon monoxide via the liquid energy carrier dimethylformamide. Dimethylformamide is produced in the synthesis reaction, stored, and reformed to carbon monoxide in the reforming step.

functional unit of the investigated ICCU process and the benchmark process, are available in Appendix B.4.

In contrast to the first case study (Section 4.2.1), in this design, halogens, sulphur, and tertiary amines are allowed as building blocks for the LEA3D algorithm (see Appendix B.7.2). These groups are expected to be inert. The choice of amines is limited to tertiary amines, which are not reactive during dry CO_2 capture (Vaidya and Kenig, 2007; Behrens et al., 2017), as considered here. However, tertiary amines can catalyse the formation of bicarbonates from CO_2 in the presence of water (Vaidya and Kenig, 2007; Behrens et al., 2017). The effect of bicarbonate formation on CO_2 capture due to water impurities in the solvent or feed streams is subject to refined evaluation and not considered in the present study. Fragments with non-aromatic carbon double bonds are removed, as these would be hydrogenated in the reactor. Thus, other reactions than Reactions 1 and 2 are not assumed to occur.

Since the training of the ANN in Section 3.1.2 did not include all of these building blocks in the training data, e.g. halogens, the solvent-specific ANN prediction is replaced by a constant value of $3 \text{ kg CO}_2\text{-eq. kg}_{\text{solvent}}^{-1}$. A value of $3 \text{ kg CO}_2\text{-eq. kg}_{\text{solvent}}^{-1}$ corresponds approximately to the average of the ANN predictions from the case study in Section 3.2.

Similar to the first case study (Section 4.2.1), heat that is not provided by heat integration is supplied by external utilities. Here, low-pressure steam (3 bar) at 410 K and furnace heat at 750 K is assumed, as well as cooling water at 283 K and refrigeration at 233 K. The heat recovery approach temperature (HRAT) equals 10 K. More details on the process specifications can also be found in Appendix B.6.2.

The COSMO-(sus)CAMPD method generates 1162 unique candidate solvents in approximately 9.5 days. Of these candidate solvents, 390 solvents are feasible for the ICCU process, and 330 candidate solvents are additionally commercially available or synthesisable as determined by database search or retrosynthesis. As the optimal solvent, the method discovers 2-dimethylamino-ethanethiol, also known as captamine, with an impact on Climate Change of $CC_{c2grave}^{tot} = 0.30 \text{ kg CO}_2\text{-eq. kg}_{FU}^{-1}$. Thus, the optimal solvent for the ICCU process reduces the impact on Climate Change by 64 % compared to the benchmark separated CCU process with an impact on Climate Change for this case study of $CC_{c2grave}^{tot} = 0.84 \text{ kg CO}_2\text{-eq. kg}_{FU}^{-1}$ (Figure 4.5).

Since captamine is a hydrolysis product of a chemical warfare agent (Glasco and Bell, 2021), its use might be prohibited. The second best commercially available solvent is 4-bromo-dimethylbutan-1-amine, ranking second with $CC_{c2grave}^{tot} = 0.32 \text{ kg CO}_2\text{-eq. kg}_{FU}^{-1}$, which is 6 % higher than for the optimal solvent captamine. In total, the method designs 204 candidate solvents with a lower $CC_{c2grave}^{tot}$ than the benchmark process that are commercially available or synthesisable. Therefore, with an optimal combination of process and solvent, the ICCU process concept is an efficient alternative to the separated process and advantageous in terms of Climate Change. However, the ICCU process concept alone is not a guarantee for a lower environmental impact in general but requires careful and integrated solvent and process design as a key design decision: About 40 % of all evaluated solvents cause a higher $CC_{c2grave}^{tot}$ for the ICCU process than the benchmark separated process, which is in line with the conclusion by Jens et al. (2019) for the feed specifications of the ICCU process.

In the considered scenario, the impact on Climate Change of conventional production equals $CC_{c2grave}^{tot} = 0.52 \text{ kg CO}_2\text{-eq. kg}_{FU}^{-1}$ and is thus lower than the impact of the benchmark separated CCU process. Thus, the CCU process is not environmentally beneficial in terms of Climate Change to conventional production, emphasising the challenge of finding CCU processes mitigating climate change. For the ICCU process, 41 candidate solvents yield a lower impact on Climate Change than the conventional production. Therefore, with an optimal combination of process and solvent, Climate Change mitigation seems possible with the ICCU process but are subject to further, more detailed investigation.

To challenge the integrated design for minimum $CC_{c2grave}^{tot}$, the solvent design based on the heat-integrated impact on Climate Change is compared to solvent rankings considering process subsystems only, i.e. unit operations. Generally, a higher absorption selectivity of the solvent for CO_2 leads to a higher yield in the reactor and thus to a lower impact on Climate Change (cf. Figure 4.5). This trade-off is confirmed by a Pearson correlation coefficient between the absorption selectivity and the impact on Climate Change of $\rho = -0.65$. Similarly, a high equilibrium yield of dimethylformamide in the organic reactor outlet leads to a low impact on Climate Change ($\rho = -0.62$). Importantly, equilibrium conversion does not correlate with impact on Climate Change ($\rho = -0.12$) since only the product concentration in the organic phase at the reactor outlet impacts the separation effort. However, product concentration is not only determined by the equilibrium conversion but also by the phase equilibrium.

Despite the correlation, choosing absorption selectivity or equilibrium yield as the design objective lead to suboptimal solvent selection: The solvent with the highest absorption selectivity is 2-phenylethanol (triangle in Figure 4.5) and the solvent with the highest equilibrium yield is thiooxalane (diamond in Figure 4.5). 2-phenylethanol and thiooxalane only rank 44th and 73rd in $CC_{c2grave}^{tot}$ with impacts 74 % and 90 % higher than the optimal solvent captamine (Table 4.2). Therefore, considering targets for single unit operations is not sufficient to select the optimal solvent for the overall process. Only an objective function based on the entire process successfully captures all process-relevant trade-offs within the molecular properties.

Heat integration strongly affects the environmental impact on Climate Change of the process for every solvent. On average, heat integration reduces the impact on Climate Change by 49 % compared to the impact on Climate Change before heat integration, with a maximum of 70 % for the top solvent captamine (Figure 4.6). Therefore, a quantitative estimation of the impact on Climate Change of the process requires the consideration of heat integration within the integrated design of process and solvent.

Heat integration also influences solvent ranking since the solvent properties impact the amount of heat that can be integrated. However, considering the impact on Climate Change without heat integration ($CC_{c2grave}^{tot,woHI}$) still enables differentiation between high- and low-ranking solvents. The correlation coefficient between solvent ranking with and without heat integration is $\rho_{rank} = 0.92$. Remarkably, for this case study, the gate-to-gate impact on Climate Change from distillation only (CC_{g2g}^{reb}) is a good estimator for $CC_{c2grave}^{tot}$ ($\rho = 0.66$). In distillation, the candidate solvents exhibit substantial differences in energy demand and thus impact on Climate Change. Thus, ranking according to distillation effort is effective, although the impact of distillation accounts for only 21 % of the total process impact on Climate Change without heat integration on average.

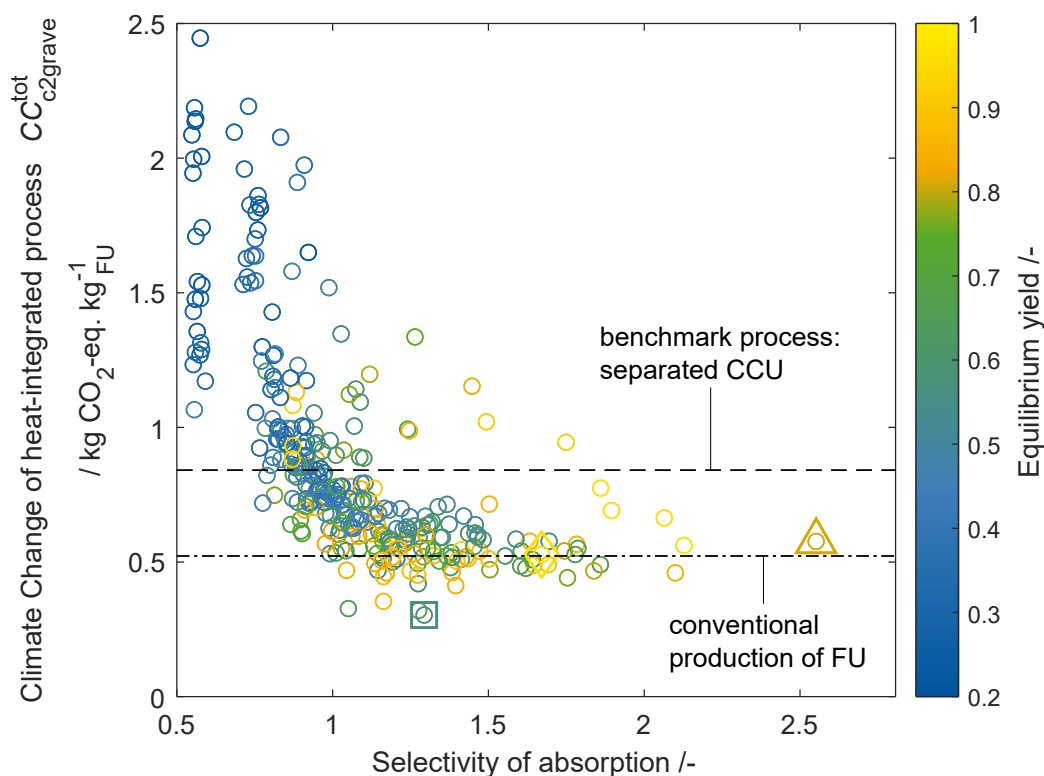
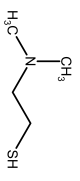

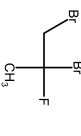
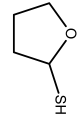
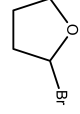


Figure 4.5: Results of the integrated molecular and process design of the heat-integrated ICCU process: Each circle represents a molecular candidate with its corresponding optimised process. The candidate with the lowest impact on Climate Change is additionally marked with a square, the candidate with the highest selectivity of absorption is additionally marked with a triangle, and the candidate with the highest equilibrium yield is additionally marked with a diamond. The dashed and the dashed-dotted line represent the impact on Climate Change of the benchmark process and conventional production, respectively.

In contrast to the low impact on the overall ranking, the heat integration potential significantly impacts solvent ranking among the top solvents. For example, when heat integration is not considered, only 31 candidates of the top 50 candidate solvents continue to be included in the revised top 50 candidate list. The correlation coefficient between solvent ranking with and without consideration of heat integration among the top 50 solvents is only $\rho_{\text{rank}} = 0.38$, indicating a weak correlation between the two rankings. Similarly, the rank correlation coefficient between $CC_{\text{c2grave}}^{\text{tot}}$ and $CC_{\text{g2g}}^{\text{reb}}$ reduces to $\rho = -0.18$ for the top 50 candidate solvents.

Table 4.2: Ranking of candidate solvents based on the chosen objective function and impact on Climate Change of the corresponding process. The impact on Climate Change is given in kg CO₂-eq. per kg of the functional unit and normalised by the impact on Climate Change of the solvent with the lowest total heat-integrated impact on Climate Change $CC_{c2grave}^{tot,1}$. The list contains the candidate solvents with the lowest and the second lowest total heat-integrated impact on Climate Change $CC_{c2grave}^{tot}$; the solvent with the lowest impact on Climate Change without heat integration $CC_{c2grave}^{tot,woHI}$, the solvent with the highest absorption selectivity and the solvent with the highest reactor equilibrium yield.

Solvent with the ...	Molecular structure	Rank		Climate Change $CC_{c2grave}^{tot,i}$			
		Climate Change with heat integration	Climate Change without heat integration	equilibrium yield	absorption selectivity	kg CO ₂ -eq. kg ⁻¹ FU	/ $CC_{c2grave}^{tot,1}$
lowest Climate Change $CC_{c2grave}^{tot}$		1	26	148	65	0.30	1
- alternative to Rank 1		2	9	93	71	0.32	1.06
lowest Climate Change $CC_{c2grave}^{tot,woHI}$		18	1	96	21	0.48	1.57
highest reactor equilibrium yield		44	12	1	17	0.53	1.74
highest absorption selectivity		73	53	56	1	0.58	1.90

When heat integration is considered in the design, the optimal solvent can successfully exploit increasing pressure in the distillation columns to 2.7 bar and 4.1 bar for optimal heat integration. Thus, for this solvent, the impact on Climate Change decreases by 70 % from 1.00 kg CO₂-eq. kg_{FU}⁻¹ to 0.30 kg CO₂-eq. kg_{FU}⁻¹. In contrast, the optimal solvent without heat integration saves only 38 % of the impact on Climate Change by heat integration. As a result, the impact on Climate Change after heat integration is 57 % higher than for the optimal solvent. Therefore, identifying the top solvents requires comprehensive modelling and consideration of heat integration besides a sound cradle-to-grave environmental objective.

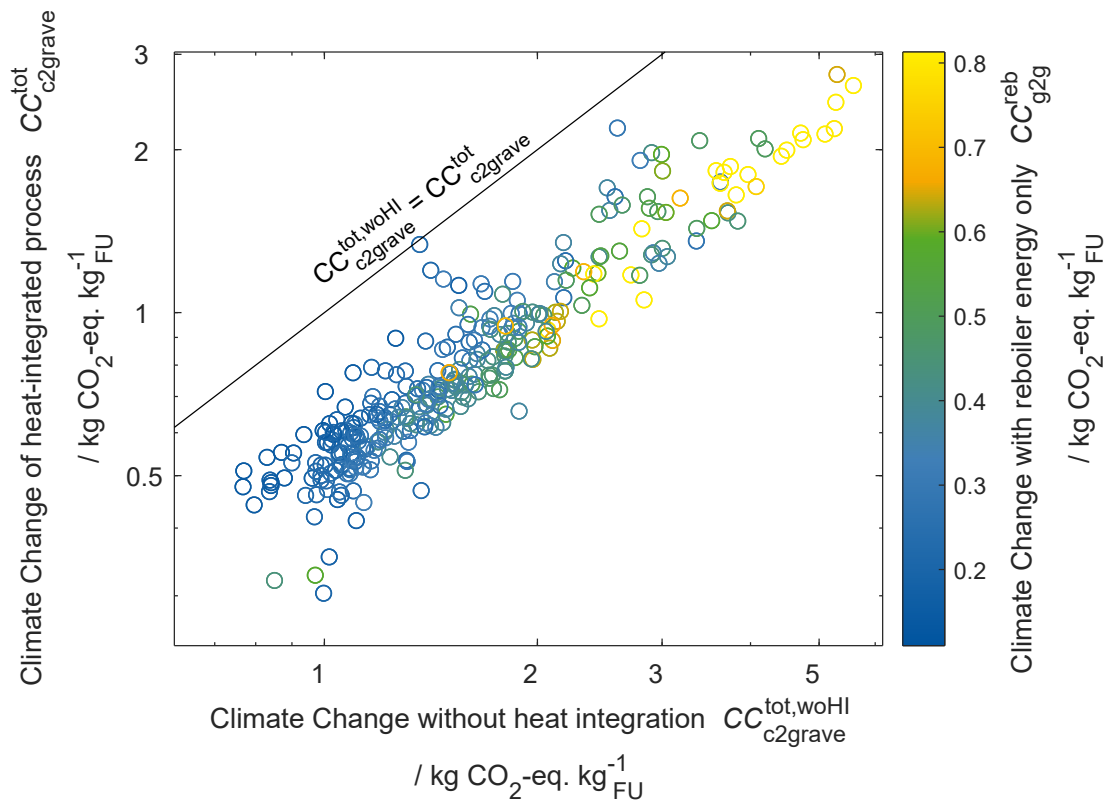


Figure 4.6: Environmental impact on Climate Change depending on process design scope: Considering process optimisation with heat integration ($CC_{c2grave}^{tot}$) compared to a process optimisation without considering heat integration potential ($CC_{c2grave}^{tot,woHI}$) and consideration of the reboiler energy demand only (CC_{g2g}^{reb}). All results include feedstock and electricity emissions and only differ in process design scope.

4.3 Conclusion

This chapter presents the COSMO-(sus)CAMPD method for computer-aided, integrated molecular and process design for heat-integrated chemical processes. Similar to the original COSMO-(sus)CAMPD method (Chapter 3), the method is based on a genetic algorithm that optimises molecules evaluated by property prediction and process optimisation. For all candidate solvents, properties used in process optimisation are predicted by quantum chemistry. Quantum chemistry allows the calculation of thermodynamic properties from a large molecular design space independent from the availability of parametrised functional groups. For computationally efficient and accurate process design, pinch-based process models are used for extraction, absorption and distillation columns, and multiphase reaction. In addition, heat integration is considered for each candidate solvent within the process optimisation. Therefore, process modelling and optimisation overcome the limitations of state-of-the-art simplified process performance indicators often used in CAMPD.

The method is applied to two case studies of (1) hybrid extraction-distillation and (2) integrated carbon capture and utilisation. In both case studies, promising candidate solvents are designed that are commercially available or synthesisable and reduce the environmental impact on Climate Change by up to 56 % and 64 %, compared to literature benchmark processes, respectively. Furthermore, the case studies reveal mutual dependencies of optimal solvents and processes. For optimal process performance, CAMPD requires a process-level objective that captures overall process performance, e.g. the total heat-integrated cradle-to-grave environmental impact. Separate consideration of individual unit operations or performance targets of process subsystems is insufficient to design optimal solvents for the entire process, as evident by low correlation coefficients between the objective function values of the heuristics and the entire process.

The case studies show that heat integration significantly impacts quantitative estimates of, e.g. environmental impact on Climate Change. Heat integration reduces the impact on Climate Change in the case studies on average by 19 % and 49 %. Due to the large savings that can be achieved by heat integration depending on the candidate solvent, comprehensive modelling considering heat integration is crucial for selecting solvents for large process flowsheets with various unit operations. For the considered case studies, an accurate ranking of promising candidate solvents cannot be achieved by simplified process representations. However, minimising the main process energy drivers also provides a suitable selection criterion for generating promising candidates that should then be analysed in subsequent detailed investigations.

The main focus of this chapter was the comprehensive application modelling and its impact on molecular design in chemical processes. To exploit the degrees of freedom in process design resulting from comprehensive modelling, heat integration and process optimisation are included for each candidate solvent using established methods. However, recently, advanced optimisation frameworks for simultaneous design of process and heat integration have been developed, e.g. by Kong et al. (2016) and Liesche et al. (2019). Future work in CAMPD should build on these frameworks to include heat exchanger network design and process structure optimisation, or to improve solution quality to global optima, provided that heat integration and process optimisation are crucial for the considered processes.

The predictive methods still contain uncertainties that propagate through the presented CAMPD method. Therefore, valuable future work could quantify uncertainty in detail and explore potential improvements for CAMPD. e.g. from experimental data. The experiments required in this context can and should ideally be tailored to the process under consideration for minimum experimental effort. Thus, the tailored design of experiments for accuracy increase of application- and system-level simulations is presented in the next chapter.

Optimal experimental design of physical property measurements for optimal chemical process simulations

Validating predictions as well as training and parametrising models in CAMPD requires experiments that are optimally tailored to the purpose of experimentation. With regard to CAMPD with environmental objective, optimal experiments maximise the accuracy of predictions on process performance and environmental impacts. Therefore, in this chapter, optimal experimental design (OED) of physical property measurements is investigated that results in the most accurate chemical flowsheet simulations: *c*-optimal experimental design (*c*-OED). *c*-OED links decisions at the property level with the modelling at the application and the assessment at the system level.

In Section 5.1, the fundamentals and the implementation of *c*-OED for chemical engineering problems are described. The benefit of *c*-OED is demonstrated in Section 5.2 for three case studies of equilibrium- and rate-based extraction and hybrid extraction-distillation processes. *c*-OED is used to design liquid-liquid equilibrium (LLE) and diffusion experiments minimising the uncertainty of thermodynamic, economic or environmental performance metrics of the solvent-based processes. The *c*-optimal experimental design is compared with the state-of-the-art OED in chemical engineering for parameter accuracy and conventionally used experimental designs without OED. The results from OED theory are challenged by a Monte Carlo analysis of the designed experiments (Section 5.3). The limits of *c*-OED for highly non-linear process models are investigated before this chapter is concluded in Section 5.4.

Major parts of this chapter are reproduced by permission of Elsevier from:

Fleitmann, L.; Pyschik, J.; Wolff, L.; Schilling, J. and Bardow, A. (2022). Optimal experimental design of physical property measurements for optimal chemical process simulations. *Fluid Phase Equilibria*, 557, 113420.

The author of this thesis contributed to the methodology and the implementation of the research topic. The author investigated, validated and visualised the results of the experimental design and wrote the first draft as the principal author. The conceptualisation was jointly developed in discussion with the co-authors.

5.1 Optimal experimental design using the c-optimality criterion

In this section, the state-of-the-art theory and fundamentals of OED are briefly explained with a focus on c-OED, which is adapted to the estimation of thermodynamic properties for chemical process simulations. First, the c-optimal objective function is derived (Section 5.1.1). Subsequently, solution approaches to OED problems are explained (Section 5.1.2). In Section 5.1.3, quality measurement criteria are introduced to compare and validate the results in Section 5.2.

5.1.1 Derivation of the c-optimal objective function

In general, the goal of statistical OED is to minimise parameter uncertainty (cf. Section 2.4). For this purpose, the objective is defined as a measure of the variance-covariance matrix of parameters \mathbf{V}_θ . The parameter variance-covariance matrix \mathbf{V}_θ can be approximated by the product of the Fisher-Information-Matrix $\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi})$ and the number of experiments N_{exp} (Bard, 1974):

$$\mathbf{V}_\theta \approx [N_{\text{exp}} \mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi})]^{-1} \quad (5.1)$$

The Fisher-Information-Matrix $\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi})$ depends on the chosen experimental design $\boldsymbol{\xi}$ and an initial parameter guess $\hat{\boldsymbol{\theta}}$ if the model is not linear in the parameters (Bard, 1974). For example, for OED of phase equilibria measurements to parametrise the NRTL-model (Renon and Prausnitz, 1968), an initial set of NRTL-parameters has to be provided.

As the parameter variance-covariance matrix \mathbf{V}_θ is proportional to the inverse of $\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi})$, OED usually focuses on optimising the Fisher-Information-Matrix $\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi})$ by selecting an optimal design $\boldsymbol{\xi}^*$ that contains the distribution of experiments independent from the total number of experiments N_{exp} .

Every experimental design $\boldsymbol{\xi}$ is represented by a design vector of N distinct experimental settings \mathbf{z}_i , e.g. temperature and pressure of each experiment, and corresponding N normalised weights v_i , which indicate the share of the total experimental effort (Fedorov and Leonov, 2014):

$$\boldsymbol{\xi} = \left\{ \begin{array}{cccc} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_N \\ v_1 & v_2 & \cdots & v_N \end{array} \right\} \quad \text{with} \quad \sum_{i=1}^N v_i = 1 \quad (5.2)$$

The number of distinct experimental settings N is usually not known a priori and a result of OED besides the specification of the experimental settings.

The Fisher-Information-Matrix $\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi})$ is calculated from the underlying model of the experiment $\mathbf{g}(\mathbf{z}, \boldsymbol{\omega}, \boldsymbol{\theta})$ (Bard, 1974). The model $\mathbf{g}(\mathbf{z}, \boldsymbol{\omega}, \boldsymbol{\theta})$ describes the experiments by relating the parameters $\boldsymbol{\theta}$ for given experimental settings \mathbf{z} to the experimental measurement results $\boldsymbol{\omega}$. For flowsheet simulation of chemical processes, for example, the model $\mathbf{g}(\mathbf{z}, \boldsymbol{\omega}, \boldsymbol{\theta})$ describes the experiments to measure liquid-liquid equilibria or diffusion coefficients. The experiments are characterised by experimental settings \mathbf{z} given as input from the experimental design $\boldsymbol{\xi}$, e.g. temperatures and concentrations. The experimental measurements $\boldsymbol{\omega}$ are, for example, measured phase compositions.

The Fisher-Information-Matrix $\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi})$ for a given experimental design $\boldsymbol{\xi}$ is calculated by multiplying the variance-covariance matrix of the experimental measurements \mathbf{V}_ω by the model sensitivity to experimental measurements \mathbf{A}_i and the model sensitivity to parameters \mathbf{B}_i for each distinct experimental setting i of the experimental design $\boldsymbol{\xi}$ (Bard, 1974):

$$\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi}) = \sum_{i=1}^N v_i \mathbf{B}_i^T (\mathbf{A}_i \mathbf{V}_\omega \mathbf{A}_i^T)^{-1} \mathbf{B}_i + \boldsymbol{\Sigma}_0^{-1} \quad (5.3)$$

with the local model sensitivity to experimental measurements: $\mathbf{A}_i = \left. \frac{\partial \mathbf{g}}{\partial \boldsymbol{\omega}} \right|_{\boldsymbol{\omega}_i, \hat{\boldsymbol{\theta}}}$

and the local model sensitivity to parameters: $\mathbf{B}_i = \left. \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\omega}_i, \hat{\boldsymbol{\theta}}}$

The variance-covariance matrix of the experimental measurements \mathbf{V}_ω is a key input parameter, which needs to be specified a priori from uncertainty measurements, highlighting the need for uncertainty reporting as part of good reporting practice for property measurements (Bazyleva et al., 2021). Already available information on the parameter variance-covariance matrix, e.g. from previously performed experiments or the literature, can be included for the design of further experiments in the Fisher-Information-Matrix $\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi})$ through $\boldsymbol{\Sigma}_0$ since $\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi})$ is additive (Fedorov and Leonov, 2014). In this work, no previously performed experiments or predictions are assumed; thus, $\boldsymbol{\Sigma}_0^{-1}$ is not further considered.

To account for the parameter use, the c -optimal design objective is to minimise a linear combination of the parameter variances, which is calculated by the product of a vector $\mathbf{c}(\hat{\boldsymbol{\theta}})$ and the inverse of the Fisher-Information-Matrix $[\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi})]^{-1}$ (Atkinson

et al., 2006). The c-optimal experimental design ξ_c^* is the solution to this optimisation problem:

$$\xi_c^* = \arg \min_{\xi} \mathbf{c}(\hat{\boldsymbol{\theta}})^T [\mathbf{F}(\hat{\boldsymbol{\theta}}, \xi)]^{-1} \mathbf{c}(\hat{\boldsymbol{\theta}}) \quad (5.4)$$

For considering the property parameter use in a process simulation, $\mathbf{c}(\hat{\boldsymbol{\theta}})$ should reflect the linearised variance propagation of the property parameter uncertainties through the process model. Therefore, the first-order sensitivities of a scalar simulation output to property parameters are chosen as weights of the linear combination. Thereby, the property parameters are weighted by their impact on the process model, e.g. the sensitivity of total process energy demand with respect to NRTL-parameters. The variance of a simulation output is thus obtained as c-OED objective. The vector $\mathbf{c}(\hat{\boldsymbol{\theta}})$ is calculated from the sensitivities of a scalar result $\psi(\boldsymbol{\theta}, \mathbf{y})$ of the process model h to property parameters $\boldsymbol{\theta}$:

$$\mathbf{c}(\hat{\boldsymbol{\theta}}) = \left. \frac{\partial \psi(\boldsymbol{\theta}, \mathbf{y})}{\partial \boldsymbol{\theta}} \right|_{\hat{\boldsymbol{\theta}}, \hat{\mathbf{y}}} \quad (5.5)$$

As both the model sensitivities for the Fisher-Information-Matrix $\mathbf{F}(\hat{\boldsymbol{\theta}}, \xi)$ and the process model sensitivities for the vector $\mathbf{c}(\hat{\boldsymbol{\theta}})$ are calculated for given initial parameters $\hat{\boldsymbol{\theta}}$, the resulting optimal experimental design is locally optimal for the given initial parameters $\hat{\boldsymbol{\theta}}$. The vector $\mathbf{c}(\hat{\boldsymbol{\theta}})$ can also depend on further parameters, e.g. specifications of the process model \mathbf{y} , resulting in an additional dependence of the optimal experimental design on these parameters. These additional specifications, such as operation settings, must be known a priori, e.g. from experience, known operation of similar systems or process design.

In contrast to c-OED, state-of-the-art OED criteria do not consider the process sensitivity to the property parameters expressed by $\mathbf{c}(\hat{\boldsymbol{\theta}})$. For example, the commonly used D-optimal experimental design yields the most accurate parameters using only the Fisher-Information-Matrix. A D-optimal experimental design ξ_D^* minimises the uncertainty of all parameters by maximising the determinant of the Fisher-Information-Matrix (Bard, 1974):

$$\xi_D^* = \arg \max_{\xi} \log [\det (\mathbf{F}(\hat{\boldsymbol{\theta}}, \xi))] \quad (5.6)$$

Generally, statistical OED as presented here requires several assumptions on the model and the errors that have been summarised, e.g. in Bard (1974) or Dechambre et al. (2014b): (1) The model parameters $\boldsymbol{\theta}$ need to be identifiable, and the true values for

the measurements need to lead to the true parameter values, i.e. no model bias is assumed. (2) No errors are assumed in the independent variables, i.e. the experimental settings \mathbf{z} , and no systematic errors are assumed in the measured variables $\boldsymbol{\omega}$. (3) Errors in different experiments are independent of each other and normally distributed with the same covariance matrix \mathbf{V}_ω .

Importantly, the thermodynamic model \mathbf{g} and the experimental measurements $\boldsymbol{\omega}$ need to be carefully selected since multiple options exist for the model and the measured quantities. For example, van Ness and coworkers showed that isobaric vapour-liquid equilibrium (VLE) measurements usually lead to large uncertainties and model errors in contrast to isotherm VLE experiments (van Ness, 1995; Gmehling and Kleiber, 2014). Isobaric VLEs rely on vapour pressure equations used as input. If inadequately parametrised, this input can cause a model bias in the temperature dependence of the vapour pressure. The intrinsic model bias then leads to incorrect parameters – independent of the experimental design.

5.1.2 Solving OED problems

The computation of the OED objectives requires sensitivities. The model sensitivities of the thermodynamic model to experimental measurements and property parameters as well as the sensitivities of the process model to property parameters are calculated by first-order numerical differentiation using central differences. To ensure stable numerical differentiations, a parameter study was performed. The step size was chosen to 1×10^{-7} for the calculation of the Fisher-Information-Matrix using complex-step differentiation (Squire and Trapp, 1998) and to 1×10^{-4} for the sensitivities of the process model.

To solve optimal experimental design problems, several general-purpose algorithms have been proposed in the literature (García-Ródenas et al., 2020). In this work, the general algorithm for computing optimal designs with monotonic convergence by Yu (2010) is used.

The algorithm yields optimal experimental designs with a continuous distribution of experimental effort, also called continuous designs (Atkinson et al., 2006). A continuous design quantifies which share of the total experimental effort should be spent on which measurements. Continuous designs suit as targets for experiments in the laboratory, as these designs specify only relative experimental effort for an infinite number of experiments. In practice, only a limited number of experiments can be performed. Therefore, implementable experimental designs for the laboratory, so-called exact designs, can be calculated for a predefined number of experiments, e.g. by rounding the

continuous designs (Atkinson et al., 2006). However, as rounding does not guarantee close approximation of continuous designs (Atkinson et al., 2006), various algorithms for the calculation of exact designs have been developed, e.g. non-sequential algorithms (Wynn, 1972) or exchange methods (Nguyen and Miller, 1992). In the validation section of this chapter, a non-sequential algorithm by Wynn (1972) is used for exact optimal designs due to its simple implementation (Section 5.3). Exact designs can also be calculated considering previous experiments or literature data (cf. Equation 5.3), as frequently required in practice. The exact design then yields the optimal subsequent experiments, as demonstrated by Duarte et al. (2021).

5.1.3 Comparison of experimental designs

Experimental designs can be compared by OED-efficiencies, which measure the effectiveness of an experimental design $\boldsymbol{\xi}$ compared to an optimal design $\boldsymbol{\xi}^*$. In this work, the focus lies on c-efficiency as a measure of process simulation accuracy and D-efficiency as a measure of parameter accuracy. The efficiencies are defined based on the c-optimal design $\boldsymbol{\xi}_c^*$ or D-optimal design $\boldsymbol{\xi}_D^*$ as (Atkinson et al., 2006):

$$\begin{aligned}
 - \text{c-efficiency: } \zeta_c(\boldsymbol{\xi}) &= \frac{\mathbf{c}(\hat{\boldsymbol{\theta}})^T \mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi}_c^*)^{-1} \mathbf{c}(\hat{\boldsymbol{\theta}})}{\mathbf{c}(\hat{\boldsymbol{\theta}})^T \mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi})^{-1} \mathbf{c}(\hat{\boldsymbol{\theta}})} \\
 - \text{D-efficiency: } \zeta_D(\boldsymbol{\xi}) &= \left[\frac{\det(\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi}))}{\det(\mathbf{F}(\hat{\boldsymbol{\theta}}, \boldsymbol{\xi}_D^*))} \right] \frac{1}{n_{\text{parameter}}}
 \end{aligned}$$

with $n_{\text{parameter}}$ for the number of estimated model parameters. The efficiencies are valuable metrics since they allow to determine the number of experiments to achieve a particular accuracy. The inverse of the c- or D-efficiency describes how many additional experiments are required for the same accuracy compared to the optimal design of the respective criterion. For example, a design with a c-efficiency $\zeta_c = 0.25$ needs 4 times as many experiments for the same process simulation accuracy as a c-optimal design.

5.2 Application of c-OED for extraction and hybrid extraction-distillation processes

The c-OED is applied by computing continuous c-optimal experimental designs for liquid-liquid equilibrium and diffusion experiments for two process models of solvent-based processes as an example:

1. Pinch-based process models for extraction and distillation (Bausa et al., 1998; Redepenning et al., 2017)
2. Countercurrent rate-based extraction model with the HTU-NTU method for sizing (Chilton and Colburn, 1935; Sattler and Feindt, 1995)

In all case studies, the thermodynamic model for liquid-liquid equilibrium measurements is taken from Dechambre et al. (2014b) and uses the NRTL activity coefficient model (Renon and Prausnitz, 1968). For the HTU-NTU sizing of the extraction column, diffusion coefficients are additionally required and assumed to be measured using a closed cell with fixed geometries as the experimental setup (Wolff et al., 2016).

In this chapter, c-OED is exemplified for the ternary system water-acetone-toluene in both case studies. The study is limited to ternary systems for ease of interpretation and visualisation. However, the method of c-OED is not limited to ternary systems but is applicable for multi-component systems with more than three components.

The chemical system water-acetone-toluene is a model system of great interest in research and industry since it is applicable for studying various processes such as extraction and distillation (Enders et al., 2007). The components represent a variety of chemical interactions: Water and toluene are almost completely immiscible since water is a highly polar molecule, whereas toluene is highly unpolar. Acetone is mildly polar and, thus, soluble in both water and toluene. Toluene is consequently a suitable solvent for the extraction of acetone from water since only acetone is attracted to the extract toluene phase leading to a selective separation. Therefore, the system is well suited to study the estimation of binary interaction parameters for extraction and hybrid extraction-distillation processes.

The continuous c-optimal experimental designs are compared as targets for maximum experimental efficiency with state-of-the-art OED for maximum parameter precision (D-optimal experimental design) and a conventional experimental design without OED, which equally distributes the experimental effort over the design space. Numerical details on the experimental designs and the initial property parameters for each design can be found in Appendix C.1 and C.2.

5.2.1 OED for estimating isothermal NRTL- τ -parameters for a pinch-based extraction model

As the first case study, the extraction of acetone from aqueous solution is investigated at 25 °C using toluene as a solvent. The extraction column is modelled using a pinch-based process model, taking NRTL-parameters as input (Redepenning et al., 2017). Pinch-based process models assume infinite columns operating at vanishing thermodynamic

driving force but consider the full non-ideal thermodynamics. Therefore, the model yields the minimum solvent demand S_{\min} required for this separation. The minimum solvent demand characterises the extraction process as the performance metric ψ for the selection of an extraction solvent (cf. Section 4.2.1).

The aim is to determine the minimum solvent demand S_{\min} for extraction as accurately as possible. For this purpose, OED is used to decide which LLE experiments should be performed to estimate the six NRTL- τ -parameters for the binary interactions. LLE experiments are typically performed by equilibrating a liquid mixture with known overall composition and miscibility gap in an equilibrium cell. After equilibration, samples are drawn from each liquid phase and the molar composition of each phase is measured. Today, mixing, equilibration, sample drawing and measuring are preferably integrated into an automated set-up (Kuzmanović et al., 2003; Dechambre et al., 2014a; Thien et al., 2020).

For simplicity, it is assumed that LLE experiments are performed with the overall composition of the components that corresponds to the centre of the tie lines. The overall compositions of all experiments lie on a line running from the centre of the miscibility gap of the binary subsystem to the critical point (cf. Figure 5.1A). Each position on this line is labelled by the scalar quantity α defined linearly from the beginning in the binary subsystem ($\alpha = 0$) to the end at the critical point ($\alpha = 1$), which thus defines each experiment exactly (Dechambre et al., 2014b). As a result, the three-dimensional representation of the overall composition of the experiment is exactly described by the parameter α , without simplifying the problem. Measurements are challenging close to the critical point. In addition, the NRTL model is known to describe the phase equilibrium poorly close to the critical point, leading to model bias. To ensure experimental feasibility and applicability of the NRTL-model, α is limited to a maximum of 0.9. Since ternary mixtures are investigated in this work, two molar fractions are measured for each experiment and each liquid phase (cf. thermodynamic model of LLE experiments by Dechambre et al. (2014b)). For each measurement, the same constant absolute standard deviation in measured mole fractions $\sigma_w = 0.005$ is assumed (cf. Section 5.3).

The c-optimal experimental design selects three distinct locations for measurements (Figure 5.1B): About 80 % of the experimental effort is placed near the operating range of the extraction column at $\alpha = 0.22$. However, no experiments are performed in the actual operating range of the extraction column. Instead, 20 % of the experimental effort is placed in the high-curvature region of the binodal curve at $\alpha > 0.65$, in particular, 6 % at the design space boundary at $\alpha = 0.9$. Thus, the c-optimal experimental design provides another argument to support the previous conclusions that the process operation settings should not be mimicked or copied for physical property experiments (van Ness, 1995).

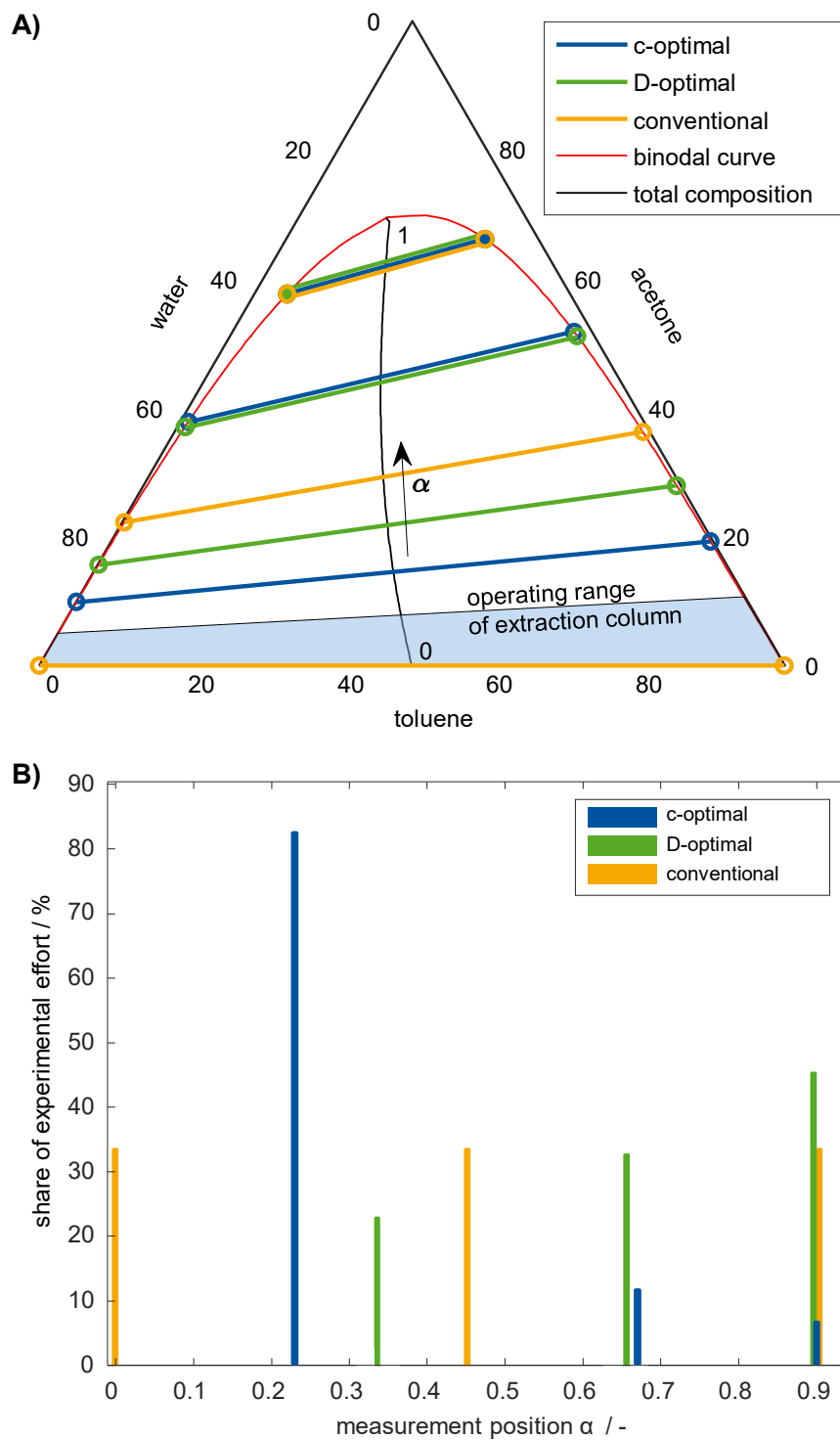


Figure 5.1: Experimental designs for LLE experiments for the extraction process: (A) Location of LLE experiments. (B) Share of experimental effort. Part A of the figure also shows the scalar quantity α , which characterises the centre of the tie lines and the total composition of an experiment.

The D-optimal experimental design similarly focuses on three distinct locations for measurements. However, in contrast to the c-optimal design, 65 % of the experiments are placed in the high-curvature region of the binodal curve, as already discovered by Dechambre et al. (2014b). In the high-curvature region of the binodal curve, the phase equilibrium model is highly sensitive to the property parameters. Thus, placing experiments in the high-curvature region leads to low uncertainty in the parameter estimation. However, the c-efficiency of the D-optimal design is only $\zeta_c^D = 0.36$ (Table 5.1). Consequently, about three times more D-optimal than c-optimal experiments are required to achieve the same accuracy in the process simulation.

For an equidistantly distributed conventional experimental design, three experimental settings are specified since the c- and D-optimal designs yielded three distinct settings. The experimental effort is equally distributed across all experiments. The conventional design yields a low c-efficiency of only $\zeta_c^{\text{con}} = 0.10$ despite placing experimental effort within the operating range of the extraction column in the solvent-carrier binary subsystem. As a result, the conventional design requires about ten times more experiments for the same process simulation accuracy as the c-optimal design. Thus, the c-optimal experimental design promises to significantly reduce experimental effort.

In terms of parameter precision, the c-optimal design scores a D-efficiency of $\zeta_D^c = 0.44$. In contrast, the conventional design yields a D-efficiency of $\zeta_D^{\text{con}} = 0.56$ and, thus, returns more accurate parameter values than the c-optimal design. The low D-efficiency of the c-optimal design illustrates the varying influence of each property parameter on simulation accuracy. For the extraction process, not all parameters of the thermodynamic model are equally important. For example, the binary interactions between solvent and solute as well as carrier and solute are of major importance. However, the simulation results are much less sensitive to the solvent-carrier interaction parameters, although low miscibility between solvent and carrier resulting from the solvent-carrier interactions is key to the extraction process. Nevertheless, highly accurate estimation of the solvent-carrier interaction is not required for accurate process simulations since small inaccuracies in the solvent-carrier interaction parameters still lead to low miscibility between solvent and carrier. Therefore, spending additional experimental effort on increasing the parameter precision of these less important parameters for the simulation reduces c-efficiency. Instead, the experimental effort is more efficiently spent on experiments targeting the more influential property parameters of the simulation. For example, near the operating range of the extraction column, the thermodynamic model of the LLE experiments is most sensitive to solvent-solute interaction, while the sensitivity to solvent-carrier interactions is low. Therefore, the c-optimal design places the majority of the experimental effort on the experimental settings near the operating range. However, exclusive focus on the property parameters

most important for the chemical process simulation neglects the accuracy increase resulting from experiments for overall high parameter precision. Therefore, the c-optimal design also includes experiments for overall parameter accuracy, such as experiments in the high-curvature region of the binodal curve at $\alpha = 0.9$.

Table 5.1: c- and D-efficiencies ζ_c and ζ_D of the c-optimal ξ_c^* , D-optimal ξ_D^* , and equidistantly distributed conventional ξ_{con} experimental designs for the estimation of isothermal NRTL- τ -parameters and use in the pinch-based extraction process model.

Design ξ	c-efficiency ζ_c	D-efficiency ζ_D
c-optimal ξ_c^*	1	0.44
D-optimal ξ_D^*	0.36	1
conventional ξ_{con}	0.10	0.56

5.2.2 OED for estimating isothermal NRTL- τ - and diffusion parameters for a countercurrent rate-based extraction column with HTU-NTU sizing

In the second case study, the same extraction of acetone from aqueous solution is considered. However, in contrast to Case Study 1 (Section 5.2.1), a countercurrent rate-based extraction model with HTU-NTU sizing based on the PhD thesis by Wolff (2021) is used instead of a pinch-based process model. The countercurrent extraction column assumes mass transfer of the solute only, following two-film theory with a constant mass transfer coefficient and thermodynamic equilibrium at the interface. In contrast to the pinch-based process models, sizing using the HTU-NTU method (Chilton and Colburn, 1935; Sattler and Feindt, 1995) and costing (Biegler et al., 1997) are included (see Appendix C.2.4 for detailed equations). Therefore, the final model result is the total annualised cost, which should be determined as accurately as possible using c-OED.

Consequently, the model needs both isothermal NRTL- τ - and diffusion parameters as property data. Therefore, the OED is extended to the selection of experiments for several thermodynamic properties. The optimal experimental design not only yields which LLE and diffusion experiments to perform but also balances the experimental effort between LLE and diffusion experiments. Thus, the design vector of experiments ξ includes the scalar measure α of LLE experiments and additionally the effort on experiments for the diffusion coefficients of acetone in water D_W and acetone in toluene D_T .

The diffusion experiments are assumed to be performed in a closed diffusion cell filled with equal volumes of two substances. The diffusion coefficients are derived from concentration measurements using Fick's second law. Here, measurements are assumed at one position δ in the closed cell and at one dimensionless measurement time given by the Fourier number Fo (Wolff et al., 2016). Therefore, the OED methodology determines the optimal Fourier number Fo^* and the optimal measurement position δ^* . For the calculations, a standard deviation of $\sigma_w = 0.5\%$ is assumed in measuring the phase compositions in the LLE experiments and the concentrations in the diffusion experiments.

For the LLE experiments, c-optimal design selects the same three measurements with the same relative distribution of experimental effort among the LLE experiments as the OED for the pinch-based process model for extraction in Section 5.2.1 (cf. Figure 5.1). Therefore, the dominating interactions for describing the minimum solvent demand using the pinch-based process model are also most important for the countercurrent rate-based extraction model. The selection of the same experimental settings is reasonable since both process models use the same thermodynamic model describing the liquid-liquid equilibrium as a basis for the solvent demand and cost calculations. Therefore, a precise description of the extraction process is a prerequisite for accurate cost calculation. Naturally, the D-optimal design equals the D-optimal design in Section 5.2.1 if focusing only on the LLE experiments since the same thermodynamic model describing the experiments is considered.

The diffusion experiments have the same optimal design using either c- or D-OED: The most accurate estimation of diffusion coefficients is achieved at a Fourier number $Fo^* = 0.1$ and the measurement position $\delta^* = 0$, i.e. at the wall of the closed diffusion cell. Moreover, the experimental effort is equally distributed between the diffusion coefficients of acetone in water D_W and acetone in toluene D_T for both designs (cf. Figure 5.2).

However, the designs differ strongly in the distribution of experimental effort between LLE and diffusion experiments: The c-optimal design focuses 96% of the total experimental effort on LLE experiments and only 4% on diffusion experiments. Therefore, the property parameters describing the phase behaviour are 24 times more important for accurate process simulation and costing than the parameters describing the diffusion. In contrast, the D-optimal design places 25% of the experimental effort on diffusion experiments and yields a c-efficiency $\zeta_c^D = 0.46$ (Table 5.2). Compared to D-optimal design in Case Study 1, the c-efficiency of the D-optimal design improves since the D-optimal design selects the same optimal diffusion experiments as the c-optimal design.

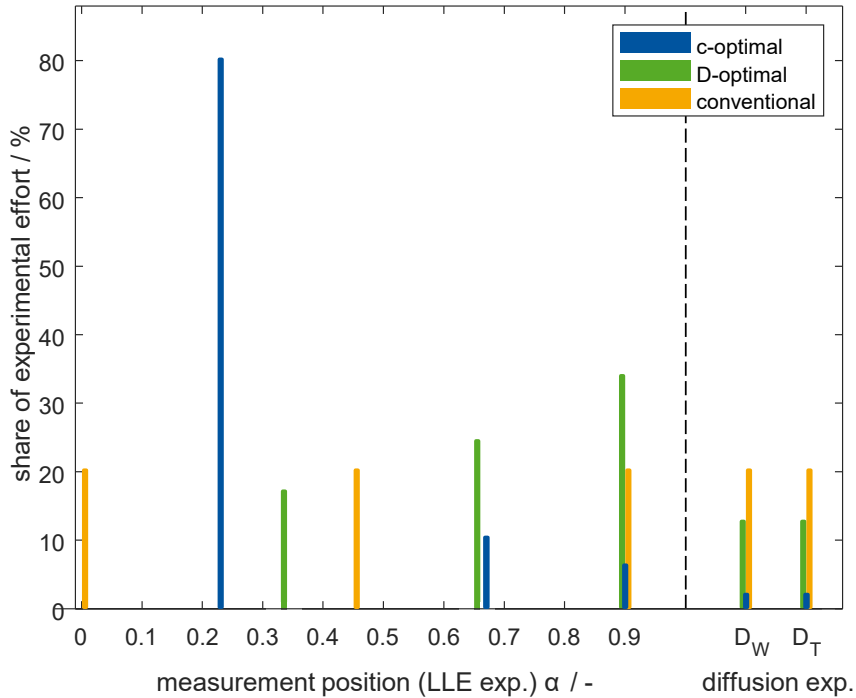


Figure 5.2: Experimental designs of LLE and diffusion experiments for the extraction model using the HTU-NTU method.

The process simulation accuracy of the D-optimal experimental design increases more from the optimal diffusion experiments than from the D-optimal LLE experiments that are suboptimal in terms of process model accuracy. Still, the D-optimal design doubles the experimental effort compared to the c-optimal design.

In terms of parameter accuracy, the limited experimental effort on diffusion experiments in the c-optimal design results in low expected accuracy of the diffusion coefficient estimation. As a result, the D-efficiency of the c-optimal design is only $\zeta_D^c = 0.39$.

For the conventional experimental design, 60 % of the experimental effort is manually allocated to three equally weighted and distributed LLE experiments as in Section 5.2.1. The remaining 40 % of the experimental effort is equally distributed between the two diffusion experiments. The optimal settings with $Fo^* = 0.1$ and $\delta^* = 0$ are assumed to be selected for each diffusion experiment in the conventional design, as these settings have been disclosed in previous work (Wolff et al., 2016). As a result, the conventional design yields c- and D-efficiencies of $\zeta_c^{\text{con}} = 0.11$ and $\zeta_D^{\text{con}} = 0.62$. Similarly to Case Study 1, the conventional design has a substantially lower c-efficiency

than the D-optimal design but a comparable D-efficiency to the c-optimal design. Both the c- and D-efficiencies of the conventional design increase by 10% compared to Case Study 1. Similar to the D-optimal design, the conventional design benefits from the optimal diffusion experiments, which increases both the overall accuracy of parameter estimation and process simulation results.

Table 5.2: c- and D-efficiencies ζ_c and ζ_D of the c-optimal ξ_c^* , D-optimal ξ_D^* , and equidistantly distributed conventional ξ_{con} experimental designs estimation of isothermal NRTL- τ -and diffusion parameters and use in the countercurrent rate-based extraction process model.

Design ξ	c-efficiency ζ_c	D-efficiency ζ_D
c-optimal ξ_c^*	1	0.39
D-optimal ξ_D^*	0.46	1
conventional ξ_{con}	0.11	0.62

5.2.3 OED for estimating temperature-dependent NRTL- τ -parameters for a hybrid extraction-distillation process

As the third case study, the extraction process previously investigated in Section 5.2.1 and Section 5.2.2 is extended by a distillation column. First, acetone is extracted from the aqueous solution using toluene before acetone is separated from the extract using distillation. Both the extraction and distillation columns are modelled using a pinch-based process model (Bausa et al., 1998; Redepinning et al., 2017). For this case study, the key performance metric ψ is the cradle-to-grave environmental impact on Climate Change (CC) of the process that should be estimated as accurately as possible. The environmental impact is calculated from the three life cycle stages for the solvent toluene as described in Chapter 3. A c-optimal experimental design that minimises only the uncertainty of the reboiler energy demand of the distillation column Q_{min} is additionally provided with corresponding OED efficiencies in Appendix C.4.

Two isothermal and two temperature-dependent NRTL- τ -parameters are estimated for each binary interaction pair because of the temperature profile in the distillation column. Therefore, in this case study, c-OED is used to determine at which temperatures and concentrations LLE experiments should be performed to calculate the environmental impact on Climate Change as accurately as possible. For the demonstration of the c-OED, the designed experiments are limited to LLE experiments. In practice, however, the NRTL parameters should be estimated through liquid-liquid- and vapour-liquid equilibrium experiments for higher accuracy (Forte et al., 2020). As

in Section 5.2.1, the design space is limited to concentrations corresponding to $\alpha < 0.9$ to ensure experimental feasibility and applicability of the NRTL-model, and now also consider a temperature range of 10–80 °C.

The experimental designs consist of a non-trivial combination of nine experimental settings for the *c*-optimal and eight for the D-optimal design across the entire design space (Figure 5.3). Both the *c*- and D-optimal experimental designs mainly focus on the boundaries of the design space but avoid experiments at and near the binary subsystem of solvent and carrier with $\alpha < 0.3$ (Figure 5.3). At the boundaries, where the temperature is high or near the critical point at high α , the thermodynamic model is particularly sensitive to the property parameters reducing parameter uncertainty more than in the centre of the design space. The binary subsystem is avoided for all temperatures since both isothermal and temperature-dependent parameters are more accurately estimated in the high-curvature region of the binodal curve (cf. Section 5.2.1).

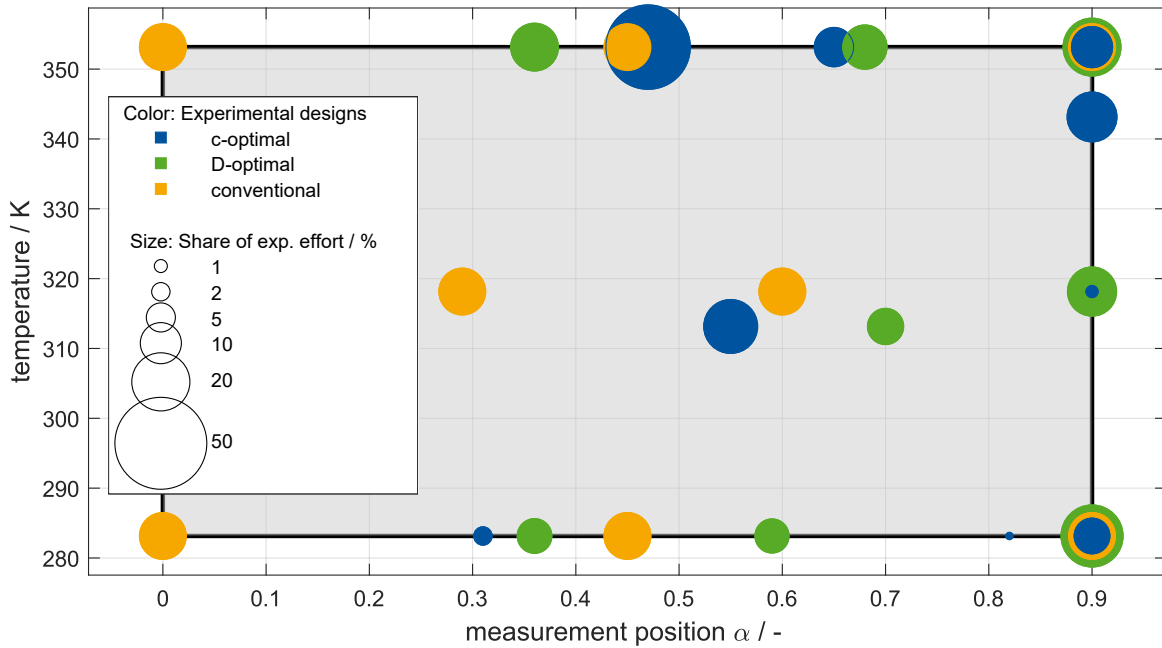


Figure 5.3: Experimental designs of LLE experiments for the hybrid extraction-distillation process. The size of the circles corresponds to the share of the experimental effort. The grey box indicates the design space.

The c-optimal design favours experiments at lower α than the D-optimal design as already discovered in the isothermal case study (Section 5.2.1), e.g. the total experimental effort spent in the c-optimal design for $\alpha < 0.6$ equals 56 %. At low α , the binary interactions between solvent and solute as well as carrier and solute can be more accurately determined, which is crucial for a high process simulation accuracy. In contrast, the D-optimal design places only 27 % of the total experimental effort on measurements with $\alpha < 0.6$. Similarly to the estimation of isothermal NRTL-parameters, experiments at higher α are important for higher parameter precision.

To capture the temperature dependency accurately, the c-optimal design places a large share of the experimental effort on experiments with higher temperatures: 73 % of the total experimental effort is spent for temperatures higher than 60 °C. The focus on higher temperatures in the c-optimal design can be explained by the temperature glide from the condenser ($T_{\text{cond}} = 74$ °C) to the reboiler ($T_{\text{reb}} = 110$ °C) in the distillation column. For an accurate description of distillation and corresponding environmental impacts, capturing the temperature dependence of the NRTL parameters is important, and at higher temperatures, the temperature-dependent parameters are more sensitive to the measurements.

The D-optimal design only allocates 43 % of the experimental effort for temperatures higher than 60 °C to balance the estimation of isothermal and temperature-dependent NRTL-parameters resulting in a c-efficiency of $\zeta_c^D = 0.63$ (Table 5.3). Compared to the D-optimal design for isothermal NRTL-parameters in Section 5.2.1, the c-efficiency of the D-optimal design increases by 75 %. Therefore, for the hybrid extraction-distillation process, parameter precision is generally more important for accurate process simulation than for the extraction only. As a result, parameter precision as measured by the D-efficiency also increases to $\zeta_D^c = 0.65$ for the c-optimal design by about 48 % compared to the c-optimal design for isothermal NRTL-parameters.

For the conventional design, equally distributed experimental effort on eight experimental settings across the design space is assumed (Figure 5.3). As in the first case study (Section 5.2.1), the optimal experimental designs significantly outperform the conventional design in both process simulation and parameter accuracy ($\zeta_c^{\text{con}} = 0.32$ and $\zeta_D^{\text{con}} = 0.55$, Table 5.3) emphasising the benefits of OED. In particular, the c-optimal experimental design is predicted to reduce experimental effort by 68 % compared to the conventional design for the same process simulation accuracy. Therefore, manually distributing the experimental effort across a large design space is particularly inefficient for subsequent use of the property parameters in a process simulation.

Table 5.3: c- and D-efficiencies ζ_c and ζ_D of the c-optimal ξ_c^* , D-optimal ξ_D^* , and equidistantly distributed conventional ξ_{con} experimental designs for the estimation of temperature-dependent NRTL- τ -parameters and use in the pinch-based hybrid extraction-distillation process model.

Design ξ	c-efficiency ζ_c	D-efficiency ζ_D
c-optimal ξ_c^*	1	0.65
D-optimal ξ_D^*	0.63	1
conventional ξ_{con}	0.32	0.55

5.3 Discussion: Uncertainties resulting from the experimental designs

In this section, the predictions from OED theory are challenged by Monte Carlo analysis for the pinch-based process models of extraction and hybrid extraction-distillation since these two case studies exhibit the minimum and maximum difference in c-efficiencies reported in this chapter. In c-OED theory, the standard deviation of the process simulation result is predicted assuming linear variance propagation. However, process models are usually highly non-linear. Therefore, assuming linear variance propagation from experiments through parameter estimation and process model only approximates the actual variance propagation. Here, the uncertainty from linear variance propagation in c-OED is compared with propagation from uncertain experimental measurements using a Monte Carlo approach. For this purpose, LLE experiments are simulated by calculating phase compositions using the initial property parameters (see Appendix C.2) and adding normally distributed noise to account for measurement errors. Afterwards, property parameters are estimated from the simulated experiments and run the process simulation using the estimated parameters to obtain the actual process model uncertainty. In detail, the following five-step procedure is applied (Figure 5.4):

1. Design optimal experiments: First, exact c- and D-optimal designs are calculated for a predefined number of experiments using a non-sequential algorithm (Wynn, 1972). Since the resulting exact designs depend on the initialisation of the algorithm, the algorithm is run repeatedly from random starting points to aim for a globally optimal solution. Additionally, a conventional design is created, which equidistantly distributes the same number of experiments across the design space.

2. Simulate experiments: From the initial property parameters and the thermodynamic model of the LLE experiments, the phase compositions are calculated that result from the experimental designs of Step 1. For this purpose, it is assumed that the initial property parameters lead to the true phase compositions. Subsequently, measurement errors are added to the true phase compositions by sampling from a Gauss distribution with a mean of zero and a standard deviation corresponding to typical uncertainty for phase compositions in LLE measurements published in the literature. The typical standard deviation for measuring molar fractions σ_w ranges between 0.001 (Nagata, 1984, 1987) and 0.005 (Dechambre et al., 2014a; Thien et al., 2017, 2020), depending on the measurement method. In this work, a standard deviation $\sigma_w = 0.005$ is chosen, as higher uncertainties are more challenging for the experimental design methodology because of the assumption of locally optimal designs. For comparison, Monte Carlo analyses were also performed for $\sigma_w = 0.001$ and $\sigma_w = 0.01$, which can be found in Appendix C.3.
3. Estimate property parameters: From the simulated experiments, the property parameters are estimated by fitting the thermodynamic model of the LLE experiments. For this purpose, the MATLAB solver `lsqcurvefit` (The MathWorks Inc., 2019) is used, considering 10 starting points for each fit to aim for a globally optimal solution. The direct use of global optimisation methods as proposed by Mitsos et al. (2009) would be a promising extension for future work.
4. Calculate process simulation: The estimated property parameters are used as input for the process simulation to obtain the actual propagation of the estimated property parameters on the process simulation result.
5. Calculate uncertainty of process simulation: Steps 2-4 are repeated until 1000 process simulation results are obtained for each experimental design. From the 1000 simulation results, the root-mean-square error (RMSE) is calculated between the simulation results of the estimated parameters from the Monte Carlo analysis and the simulation result of initial parameters for each design. The RMSE of the Monte Carlo samples is compared to the expected uncertainty from linear error propagation given by the standard deviation of OED theory. Both the RMSE and the standard deviation are normalised by the actual value of the process simulation result to allow for relative comparisons.

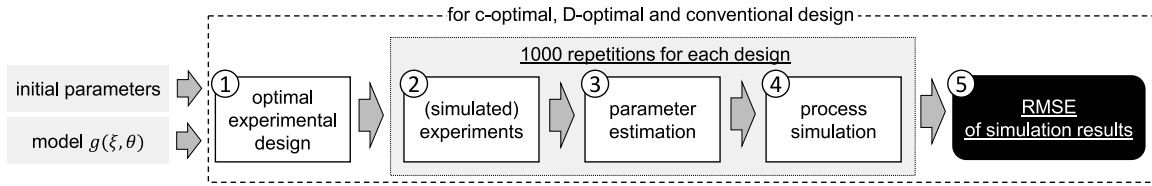


Figure 5.4: Procedure to determine the uncertainties of the process simulation results for each experimental design.

5.3.1 Accuracy of the extraction process simulation

The accuracy of the extraction process simulation depending on the experimental design is investigated by estimating the six isothermal NRTL- τ -Parameters of the ternary system for 5, 7, 10, 15 and 20 LLE experiments as an example. The uncertainty of the process simulation is measured by computing the relative RMSE of the minimum solvent demand resulting from the pinch-based process model. Generally, the uncertainty of the process simulation results is low, with a relative RMSE of 2–6 % (Figure 5.5). Therefore, a small number of experiments, e.g. 5 to 10, is already sufficient for an accurate description of the extraction process.

The predictions from linear variance propagation using the c-optimal objective function (hatched bars) and Monte Carlo analysis (full bars) agree well for each design. The c-optimal objective function successfully predicts qualitatively and quantitatively the uncertainties of the simulation results: For the investigated numbers of experiments, the c-optimal design yields the lowest uncertainty in the Monte Carlo analysis as predicted, followed by the conventional and the D-optimal design. For each experimental design, the relative RMSE decreases monotonically with an increasing number of experiments, as expected from OED theory (cf. Equation 5.1). Thus, for the simulation of the extraction process, Monte Carlo analysis confirms the benefits promised by c-OED theory on simulation accuracy.

Notably, the exact conventional designs with 5–20 experiments yield c-efficiencies between 0.43 and 0.46 and thus, exceed the c-efficiencies of the continuous conventional designs with only three distinct experimental settings (cf. Section 5.2.1). Therefore, the exact conventional designs outperform the exact D-optimal designs in simulation accuracy for this example. The differences in c-efficiency compared to Section 5.2.1 result from the differences between continuous and exact conventional designs and are correctly reflected by the c-optimal objective function.

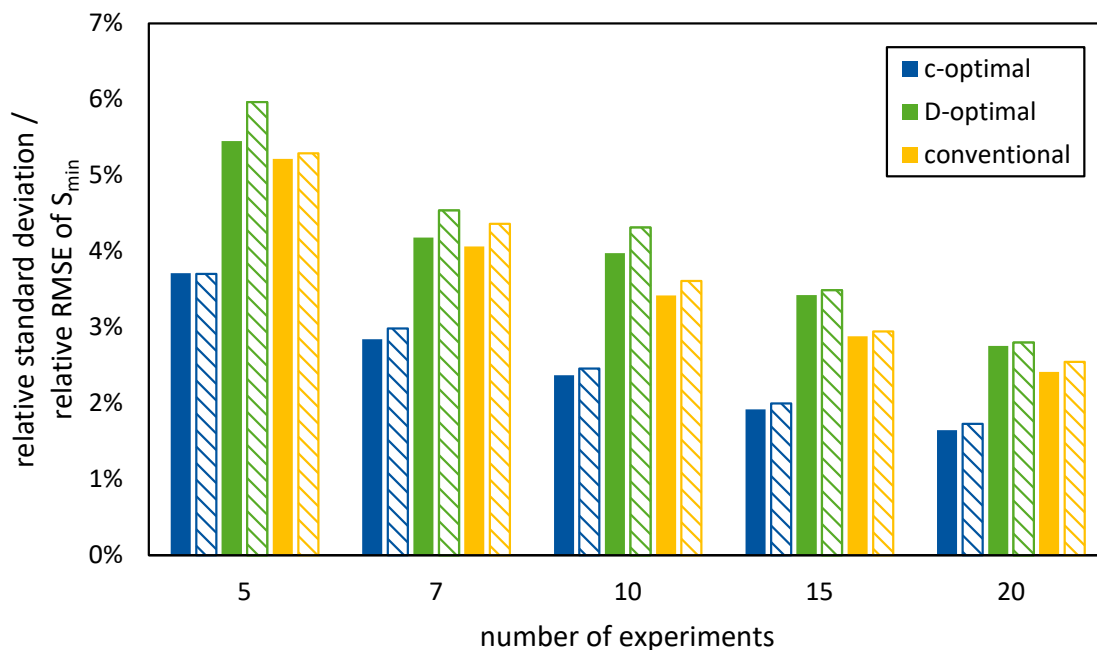


Figure 5.5: Uncertainties of the solvent demand in the extraction process for c-optimal, D-optimal, and equidistantly distributed conventional experimental designs. The full bars are the relative RMSE from the Monte Carlo sampling; the hatched bars are the expected relative standard deviation from OED theory.

5.3.2 Accuracy of the hybrid extraction-distillation process simulation

For the hybrid extraction-distillation process, the six isothermal and six temperature-dependent NRTL- τ -parameters of the ternary system are estimated by performing 20, 25, 30, 40 or 50 experiments. More experiments are chosen than for the extraction process to capture the temperature dependence with additional parameters. The uncertainty of the process model is measured by the relative RMSE of the cradle-to-grave environmental impact on Climate Change (CC) resulting from the process.

The relative RMSE of the simulation result range between 12–19% for the Monte Carlo analysis (full bars, Figure 5.6) and are thus about one order of magnitude higher for the hybrid extraction-distillation with temperature-dependent NRTL-parameters compared to the extraction with isothermal NRTL-parameter. Qualitatively, the results from the Monte Carlo analysis agree with the ranking obtained from OED theory: The c-optimal design provides the lowest uncertainty in process simulation results, followed by the D-optimal and the conventional design. However, the results from the Monte Carlo analysis deviate quantitatively from OED theory: For the

investigated c-optimal designs, the OED theory underestimates the actual uncertainty of process simulation results by up to 50 % of the predicted uncertainty. In contrast, for the investigated conventional designs, the OED theory overestimates the actual uncertainty of the process simulation results by up to 58 % of the actual uncertainty. The predictions for the D-optimal designs match well with the Monte Carlo analysis for 20 and 25 experiments. For 30, 40 and 50 experiments, however, the accuracy for the D-optimal designs is increasingly overestimated by up to 27 % of the predicted uncertainty. The results indicate that the assumption of linear error propagation is limited. The improvements predicted by linear variance propagation for c-OED cannot always be achieved. However, c-OED still proves to provide the most accurate simulation results.

In contrast to the c-optimal design, the relative RMSE of the D-optimal design from the Monte Carlo sampling is not always underestimated by OED theory. For fewer experiments, i.e. 20 experiments, the uncertainty in the Monte Carlo analysis resulting from the D-optimal design is predicted by OED theory. Therefore, the

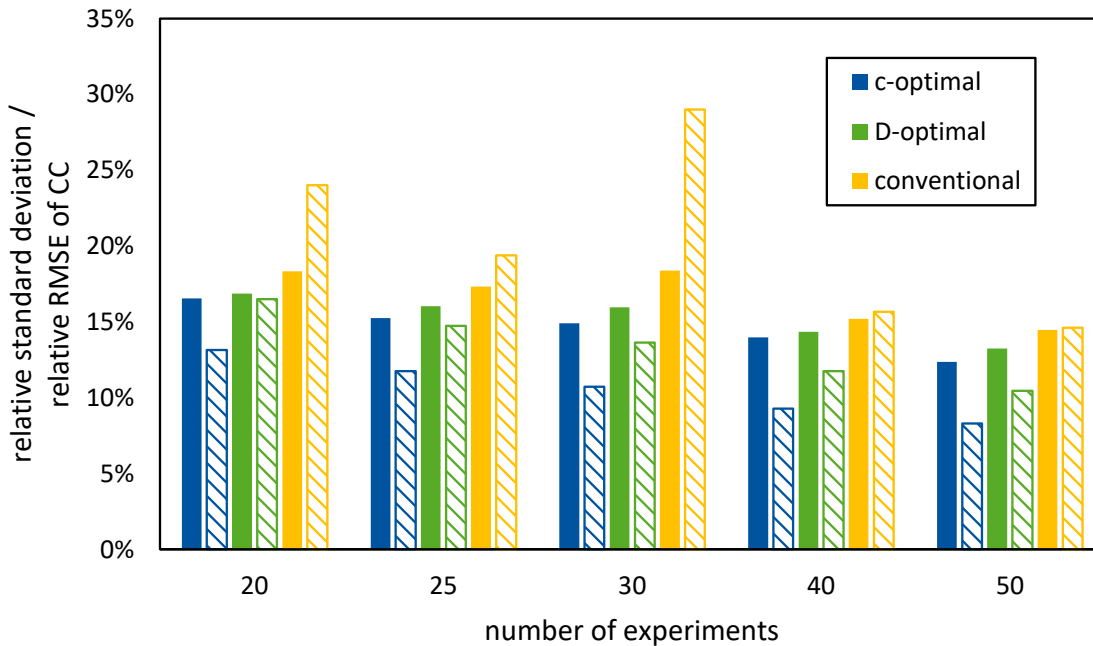


Figure 5.6: Uncertainties of the environmental impact on Climate Change in the hybrid extraction-distillation process for c-optimal, D-optimal, and equidistantly distributed conventional experimental designs. The full bars are the relative RMSE from the Monte Carlo sampling; the hatched bars are the expected relative standard deviation from OED theory.

accuracy improvement of the c-optimal design decreases compared to the D-optimal design and eventually disappears for the hybrid extraction-distillation process if only a minimum number of experiments is performed (cf. analysis in Appendix C.3.4). Thus, for a small number of experiments, an improvement in simulation accuracy cannot be guaranteed by c-OED.

For these experiments, the accuracy increase in the process simulation through c-OED is counterbalanced by the impact of inaccurate property parameters. The property parameter accuracy of c-OED is lower since overall parameter precision is not the goal of c-OED. Thus, the parameters from c-optimal experiments are more prone to measurement uncertainties and more strongly affected by inaccurate measurements for a small number of experiments.

However, if the number of experiments is increased beyond the minimal number, c-optimal design outperforms the D-optimal and conventional designs. The c-optimal design monotonically decreases simulation uncertainty, which is not guaranteed for the D-optimal and conventional designs. For higher uncertainties in mole fraction measurements, e.g. $\sigma_w = 0.01$, the same qualitative trend can be observed (see analysis in Appendix C.3.4).

In conclusion, the property parameters should be tailored for use in a process simulation, but the overall parameter accuracy cannot always be ignored to obtain accurate and robust simulation results. Since the parameters estimated from c-OED are tailored to a specific process, these parameters are not optimal for every purpose. If the parameters are not only used for process simulation but also for, e.g. gaining thermodynamic insights, the OED objective needs to be adapted.

A single OED optimality criterion rarely leads to optimal parameters for all purposes since the individual OED objectives conflict with each other. For the hybrid extraction-distillation process, e.g. c- and D-efficiency form a well-defined Pareto frontier (Figure 5.7). Neither c-efficiency nor D-efficiency of an optimal multi-objective design can be improved without deteriorating the efficiency of the other objective. Therefore, a multi-objective design needs a carefully balanced optimality criterion. However, the Pareto frontier shows good trade-off solutions: c-efficiency for accurate process simulations can be substantially increased with small losses to the D-efficiency representing parameter accuracy. E.g. a trade-off point minimising the distance to the utopia point ($\zeta_D^{\text{uto}} = \zeta_c^{\text{uto}} = 1$) retains a D-efficiency of $\zeta_D^{\text{to}} = 0.90$ while increasing c-efficiency from $\zeta_c^{\text{D}} = 0.6$ to $\zeta_c^{\text{to}} = 0.91$. The study of such trade-offs could be a valuable use of the introduced c-efficiency concept.

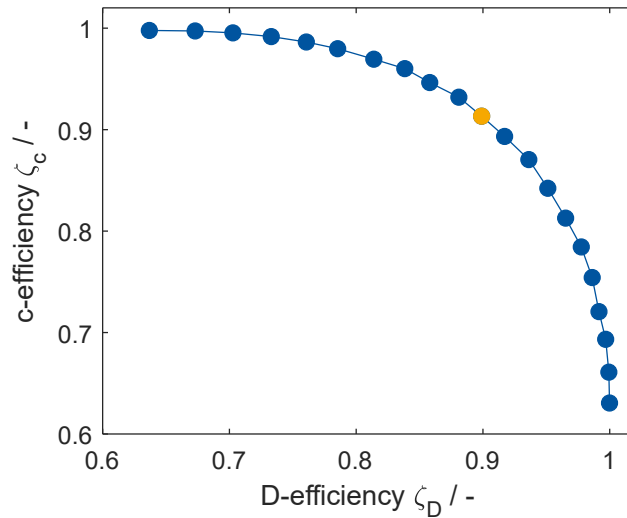


Figure 5.7: Pareto frontier between c- and D-efficiency for parameter estimation of the hybrid extraction-distillation process. The orange circle marks the trade-off solution with minimum distance from the utopia point ($\zeta_D^{\text{uto}} = \zeta_c^{\text{uto}} = 1$).

In future work, the parameter precision of the c-optimal design can further be considered towards more robust designs for parameter estimation, e.g. by compound design (Fedorov and Leonov, 2014) such as combined c- and D-optimal design (Atkinson and Bogacka, 2002), or by introducing a minimum D-efficiency as a constraint within the optimisation (Holland-Letz et al., 2018; Holland-Letz and Kopp-Schneider, 2018). Methods from the area of robust experimental design can also be explored to strengthen the c-optimal design for reliable improvement of simulation accuracy, e.g. by considering the most inaccurate process simulation as objective for OED (Rojas et al., 2007).

5.4 Conclusion

In this chapter, c-optimal experimental design (c-OED) was introduced as a method of optimal experimental design for chemical engineering problems. c-OED minimises the uncertainty of the process simulation result instead of parameter precision as the design objective. Thus, c-OED considers the application of estimated property parameters in a process simulation already during the design of experiments and links the experiments at the property level with the application and system level. To estimate the uncertainty of the process simulation results, c-OED uses linear variance propagation from uncertain property parameters through the process model.

c-OED is demonstrated for estimating isothermal and temperature-dependent NRTL-parameters from liquid-liquid equilibrium experiments for an extraction column and a hybrid extraction-distillation process modelled by pinch-based process models. The LLE experiments are designed to minimise the uncertainty of the main thermodynamic and environmental performance measures: the minimum solvent demand of the extraction and the minimum environmental impact on Climate Change of the hybrid extraction-distillation process. Moreover, for a rate-based extraction column sized by the HTU-NTU method, liquid-liquid equilibrium and closed-cell diffusion experiments are simultaneously designed to minimise the uncertainty of the total annualised cost of the extraction column.

The application of c-OED for chemical processes shows that considering the sensitivity of the process within OED highly impacts the selection of experiments for property parameter estimation. c-OED yields non-trivial experimental designs that outperform state-of-the-art OED in accuracy of process simulation results. The c-optimal experiments focus on the accurate estimation of parameters most relevant for accurate process simulations. The prioritisation of experiments for specific parameters is particularly evident in the simultaneous design of LLE and diffusion experiments: The major experimental effort of the c-optimal design for the rate-based extraction column is spent on LLE instead of diffusion experiments.

Compared to state-of-the-art OED, c-OED reduces the experimental effort by up to 64 % for the same accuracy in the case studies. Conventionally designed experiments without using OED would increase the experimental effort compared to c-OED by up to a factor of 10, highlighting the need for (c)-OED.

The predictions on accuracy from c-OED theory are examined by Monte Carlo Analysis to challenge the linear approximation of variance propagation. Generally, the OED predictions agree well with the results from Monte Carlo Analysis, and thus, the assumption of linear variance propagation is a good approximation of the actual variance propagation. In the case studies, process simulation accuracy significantly increases through c-OED. The uncertainty of process model results decreases by 30–40 % for an extraction process and by 2–20 % for a hybrid extraction-distillation process compared to conventional experimental designs and state-of-the-art OED that does not consider the process. Therefore, c-OED increases accuracy even for highly non-linear process models and is thus successfully shown to tailor experiments for thermodynamic properties to process simulations.

For future work, the focus should be directed to strengthening the robustness of the c-optimal design, e.g. by compound design (Fedorov and Leonov, 2014). The predictions from c-OED theory can fail due to overall inaccurate property parameters

if too few experiments are considered. The c-optimal experimental designs increase the accuracy of process simulations at the expense of other OED efficiencies, e.g. D-efficiency for parameter accuracy. However, efficient trade-off solutions can be identified balancing process simulation and parameter accuracy. Balanced compound or multi-objective designs allow identifying such trade-off solutions.

Considering multiple operating points of the process simulation instead of only one operating point, e.g. by L- or D_k -optimality (Atkinson et al., 2006; Holland-Letz and Kopp-Schneider, 2018), could extend the accuracy increase by c-OED for a broader simulation range of the process model. For an extension from process simulation to process optimisation and design, the sensitivities of the optimal process variables to the uncertain property parameters need to be considered. For example, the first-order derivatives of optimised process simulation outputs with respect to property parameters could be integrated into c-OED. Ultimately, this approach would formally transform the idea of c-OED into the method of weighted A-optimality presented by Houska et al. (2015).

Moreover, in practice, the initial parameter guesses rarely match the optimal parameters. Thus, an iterative procedure is usually required (Mukkula et al., 2021) that involves not only OED but also parameter fitting, validation and consistency tests. Therefore, future work should investigate the influence of initial property parameters on the benefits of c-optimal design.

In the broader context of CAMPD, c-OED can provide an efficient way to generate experimental data for validation and accuracy increase, which is discussed in more detail in Chapter 7.

Molecular design of spark-ignition fuels for maximum engine efficiency

Molecules are not only key for a sustainable chemical industry due to their influence on process design as processing chemicals but also as chemical products. In contrast to the computer-aided design of processing chemicals, computer-aided product design (CAPD) is still usually performed by screening the physico-chemical properties of candidate products for property targets instead of directly assessing a performance metric of the application. To demonstrate targeted model-based design of chemical products through a model of the application, this chapter presents a method for the design of fuels for spark-ignition engines that incorporates an empirical model of engine efficiency as the objective function.

The chapter begins with a brief introduction to state-of-the-art fuel design as a special case of product design (Section 6.1). In Section 6.2, the details of the fuel design method are explained. The models and methods are outlined for the prediction of properties describing combustion (Section 6.2.1), thermodynamics (Section 6.2.2), and environmental impacts and synthesizability (Section 6.2.3). The merit function calculating the expected engine efficiency increase is described in Section 6.2.4. Section 6.2.5 explains the constraints applied to the fuel design for spark-ignition engines. In Section 6.3, the fuel design method is applied to the design of pure-component fuels (Section 6.3.1) and the design of binary blends with ethanol (Section 6.3.2). The chapter is summarised and concluded with a brief outlook on future research in the area of fuel design in Section 6.4.

Major parts of this chapter are reproduced by permission of the American Chemical Society from:

Fleitmann, L.; Ackermann, P.; Schilling, J.; Kleinekorte, J.; Rittig, J.G.; vom Lehn, F.; Schweidtmann, A.M.; Pitsch, H.; Leonhard, K.; Mitsos, A.; Bardow, A. and Dahmen, M. (2023). Molecular design of spark-ignition fuels for maximum engine efficiency by combining predictive thermodynamics and machine learning. *Energy & Fuels*, 37(3), 2213–2229.

The author of this thesis developed the fuel design method and integrated the individual property prediction methods. The author investigated and validated the design results with respect to the CAPD methodology. The author wrote the first draft of the paper jointly with PA.

6.1 Fuel design as a special case of product design

To date, fuels are usually designed for physico-chemical property targets as surrogate measures rather than the expected engine efficiency itself (König et al., 2020b). Some studies on fuel design rely on database screenings using experimental data rather than models for property prediction: McCormick et al. (2017) screened a database of approximately 500 potential biomass-based blendstocks and blends to identify feasible gasoline blends. To assess the candidates, experimental data was collected from various databases for physico-chemical properties, environment, health and safety indicators, and corrosivity. Similarly, Fioroni et al. (2019) screened a database for potential diesel blendstocks based on thermodynamic properties and cetane numbers. Using the database created by Fioroni et al. (2019), Huo et al. (2019) and Huq et al. (2019) evaluated chemo-catalytic conversion pathways from potential bio-based platform chemicals to hydrocarbons targeting physico-chemical and combustion properties. Recently, Kuzhagaliyeva et al. (2022) published a data-driven framework to design gasoline blends with tailored properties from a database of fuel molecules.

To expand the molecular design space beyond molecules contained in databases, generate-and-test approaches have been developed. The idea of generate-and-test CAPD is to create candidate structures first and assess their fuel properties subsequently by predictive models. Hechinger (2014) employed the structure generator MOLGEN (Gugisch et al., 2015) and dedicated Quantitative Structure-Property Relation (QSPR) models to predict physico-chemical properties. The combustion performance was not included in the screening but was assessed manually a posteriori because of limited training data (Hechinger et al., 2012). Dahmen and Marquardt (2016) later extended the generate-and-test approach by a group contribution method for the derived cetane number (Dahmen and Marquardt, 2015) to include combustion behaviour. The authors tailored the structure generation to model catalytic refunctionalisation of platform chemicals derived from lignocellulosic biomass. This approach yields a list of candidate fuels that meet a range of fuel properties associated with high engine efficiency and high synthesisability (Dahmen and Marquardt, 2016). Recently, Rittig et al. (2022) employed generative graph machine learning models to design molecules with maximum research octane number and octane sensitivity.

Since generate-and-test approaches enumerate candidate fuels, numerous evaluations of candidate fuels are required. For a more targeted exploration of the molecular design space, optimisation-based fuel design has been developed based on the general mathematical product design problem by Gani (2004). The optimisation-based approach was first applied to determine the composition of biofuel blends of pre-selected blend

components respecting fuel standards (Kashinath et al., 2012; Hashim et al., 2017) and was later extended to arbitrary components using a decomposed optimisation strategy (Yunus et al., 2014). The integrated design of the molecular structures and their optimal composition in a blend was finally achieved by formulating and solving the blend design problem as a mixed integer non-linear programme based on functional groups as molecular building blocks (Zhang et al., 2018a; Kalakul et al., 2018; Liu et al., 2019c).

Optimisation-based fuel design has attracted particular attention in combination with the selection of optimal conversion routes. Marvin et al. (2013) used a rule-based reaction network generator to generate possible gasoline fuel components and their production pathways. Based on the reaction network, gasoline blends are optimised with respect to production process performance constrained by fuel blend properties. In contrast, Ng et al. (2015a) designed bio-based fuel blends by first optimising the properties of a blend and subsequently solving a superstructure optimisation problem for an integrated biorefinery. Dahmen and Marquardt (2017) combined blend design with mass-based screening of processing pathways using experimental yields to obtain renewable fuel blends maximising resource efficiency. The method was extended by early-stage process design using process network flux analysis, allowing the minimisation of production cost and global warming impact of the designed fuel blend (König et al., 2020a). Subsequent engine testing of selected blends (Dahmen and Marquardt, 2017; König et al., 2021) has confirmed the superior engine performance compared to fossil gasoline (Burkardt et al., 2021; Ackermann et al., 2021).

The studies mentioned above successfully identified promising molecules based on a list of target properties. However, such a fuel design does not consider the combined effect that these properties exert on engine performance. To date, no method has been proposed that explicitly designs fuels for maximum engine efficiency.

To consider engine efficiency as an explicit design objective, an engine model is required that predicts engine efficiency based on the fuel's physico-chemical properties. Recently, two models for spark-ignition (SI) engines were presented in the literature: (1) a zero-dimensional engine model (Gschwend et al., 2017) and (2) the engine efficiency merit function (Farrell et al., 2018; Szybist et al., 2021).

The zero-dimensional engine model was developed to calculate a fuel's maximum engine efficiency considering knock limitation and has already been coupled with reaction network analysis to find the optimal upgrading of lignin pyrolysis oil (Gschwend et al., 2018). Moreover, the model has been applied to a detailed performance evaluation of fifty pre-selected fuel candidates (Gschwend et al., 2019).

The engine efficiency merit function is a correlation that predicts the relative engine efficiency increase compared to a base fuel, i.e. RON95 gasoline, based on fuel properties, such as research octane number, octane sensitivity, and heat of vaporisation. vom Lehn et al. (2021b) and Li et al. (2022) have used the merit function to rank candidate fuels within database screenings. However, the screening studies are limited to the existing database molecules and cannot discover novel molecular structures.

6.2 Fuel design method

The fuel design method uses molecular optimisation to maximise the predicted achievable engine efficiency increase of a fuel combusted in dedicated spark-ignition engines. Various constraints are imposed to design efficient and safe fuels. Specifically, thermodynamic and combustion properties are predicted as well as environmental, health, and safety (EHS) indicators and synthesisability.

For thermodynamic properties, established models from the literature use group contribution (GC; Gani, 2019) or quantum chemistry-based methods (Gertig et al., 2020a), e.g. COSMO-RS (Klamt et al., 2010). GC methods have also been applied to predict EHS indicators of chemicals (Hukkerikar et al., 2012a). Recently, advanced machine learning-based methods, e.g. deep learning and Bayesian regression, have progressed rapidly in the field of molecular property prediction (Walters and Barzilay, 2021). Therefore, the fuel design method is based on a hybrid approach for property prediction: Thermodynamic properties are predicted using quantum chemistry-based COSMO-RS, and combustion and EHS properties as well as synthesisability are predicted using machine learning-based models.

Based on the predicted properties, each candidate fuel is evaluated with the engine efficiency merit function. Both the property prediction and the evaluation of the objective function and constraints are integrated into a molecular optimisation framework that is based on COSMO-CAMD (Scheffczyk et al., 2017b; cf. Section 3.1.1).

The fuel design method involves five steps for property prediction and candidate evaluation in each generation of the genetic optimisation (Figure 6.1):

1. Prediction of combustion properties
2. Prediction of thermodynamic properties
3. Prediction of EHS indicators and synthesisability
4. Objective function evaluation
5. Constraints evaluation

Based on the fitness values of the current generation of molecules, a next generation is created through the genetic operations crossover and mutation. The method proceeds to systematically explore the molecular design space until a pre-defined maximum number of generations is met.

In the following subsections, each step in the fuel design method is briefly explained. Details on the used soft- and hardware are included in Appendix D.1. Appendix D.2 contains the molecular fragments, which are specified as building blocks for the genetic algorithm. In this work, fragments are included to design oxygenated and non-oxygenated hydrocarbons.

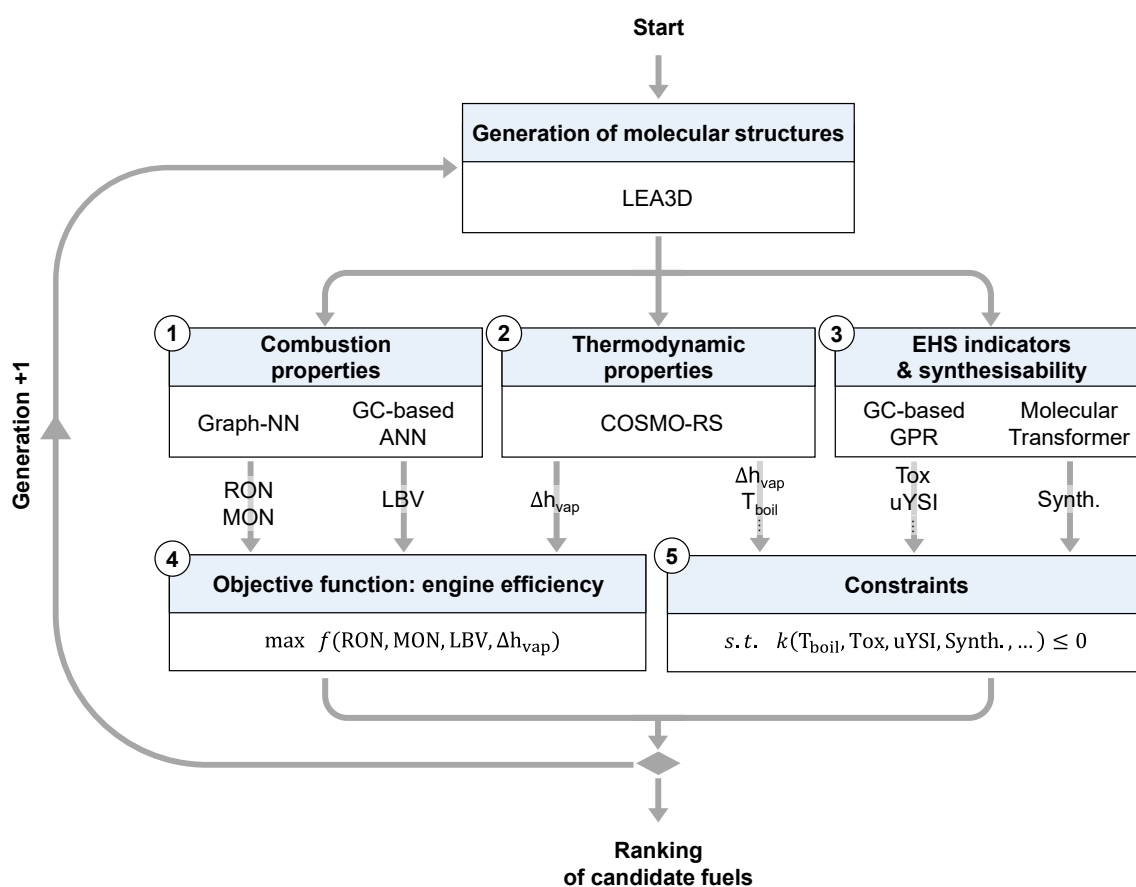


Figure 6.1: Fuel design method for maximum engine efficiency considering constraints on environment, health, and safety hazards, as well as synthesisability. For molecular optimisation of candidate fuels, property prediction and performance assessment are integrated into the genetic algorithm LEA3D in five steps (1-5): prediction of combustion (1) and thermodynamic properties (2), and environment, health, and safety indicators including synthesisability (3), evaluation of objective function f (4) and constraints k (5).

6.2.1 Combustion properties

Combustion properties that substantially influence SI engine efficiency are the Research and the Motor Octane Numbers (RON and MON) as well as the Laminar Burning Velocity (LBV) (Szybist et al., 2021). RON and MON are predicted with the Graph Neural Network (GNN) developed by Schweidtmann et al. (2020) that was trained by simultaneous, so-called multi-task learning on training data of RON, MON, and derived cetane number (DCN) of oxygenated and non-oxygenated hydrocarbons. By multi-task learning, a higher prediction accuracy is achieved compared to the accuracy of single-task learning of the individual properties. Moreover, the GNN directly uses the molecular graph as feature and thus eliminates the need for manual feature selection.

The LBV is predicted with a Group Contribution (GC) based Artificial Neural Network (ANN) by vom Lehn et al. (2021a). Following the idea of group contributions (Gani, 2019), the GC-based ANN uses the number of structural groups in a molecule as input features. For this purpose, molecules are divided into the structural groups originally proposed by Joback and Reid (1987). The LBV not only depends on the molecular structure but also on combustion parameters, i.e. temperature, pressure, and the fuel-air equivalence ratio, which are therefore additional inputs to the ANN. As suggested by Farrell et al. (2018), the LBV is evaluated for a stoichiometric mixture at ambient pressure and 358 K. It should be noted that these conditions differ from typical engine conditions. However, as Szybist et al. (2021) point out, LBV measurements at engine relevant conditions are associated with high uncertainties.

6.2.2 Thermodynamic properties

Thermodynamic properties are predicted with COSMO-RS on TZVP-MF level (Klamt et al., 2010; cf. Section 3.1.1). For the fuel design, COSMO-RS predicts boiling/bubble points and enthalpies of vaporisation of pure components and mixtures. Furthermore, melting points of pure components are available via a random forest-based QSPR model trained on structural molecular information and the σ -moment descriptors from COSMO-RS by Loschen and Klamt (2016). For fuel blends, COSMO-RS additionally calculates liquid-liquid-equilibria to estimate immiscibility and phase segregation.

6.2.3 Environment, health, and safety indicators & synthesisability

In addition to technical performance and technical feasibility, the candidate fuel needs to be assessed with regard to practical constraints as an optimally designed fuel should allow for safe handling and minimum hazards to environment and health (Ackermann et al., 2021). Furthermore, some in-silico designed molecules are challenging or impossible to synthesise in practice (Gao and Coley, 2020). Therefore, the prediction of Environment, Health, and Safety (EHS) indicators and synthesisability is included to design non-hazardous and attainable fuels.

EHS properties: Alshehri et al. (2021) recently presented models for predicting various EHS indicators of pure components using Group Contribution (GC) based Gaussian Process Regression (GPR). A similar approach was presented by Li et al. (2021) for the prediction of sooting tendencies. Similar to Alshehri et al. (2021) and Li et al. (2021), the following EHS indicators are considered for the fuel design through GC-based GPR prediction as constraints:

- Autoignition Temperature (AiT; American Society for Testing and Materials, 2000)
- Bioconcentration Factor (BCF; Arnot and Gobas, 2006)
- aqueous toxicity as Lethal Concentration for Fathead Minnow fish ($LC_{50}(\text{FM})$; Ankley and Villeneuve, 2006)
- oral toxicity as Lethal Dose for rats (LD_{50} ; Walum, 1998)
- Permissible Exposure Limit using the OSHA time-weighted average ($PEL_{\text{OSHA-TWA}}$; Spear and Selvin, 1989)
- chemical tendency to form soot expressed through the unified yield sooting index (uYSI; Das et al., 2018)

For integration in the fuel design method, the models by Alshehri et al. (2021) for AiT, BCF, $LC_{50}(\text{FM})$, LD_{50} , and $PEL_{\text{OSHA-TWA}}$ are re-trained using UNIFAC groups as descriptors and the training and test data from Alshehri et al. (2021). The uYSI model is developed using the data from McEnally et al. (2017). Noteworthy, the uYSI does not predict engine-out soot emissions but rather the chemical tendency of a fuel to form soot. A more practical measure for engine-out soot emissions would be the Particulate Matter Index (PMI) Aikawa et al. (2010), where the number of double-bond equivalents as a proxy for the chemical tendency to form soot is divided by the vapour pressure as a measure for in-cylinder mixture formation quality. However, the different oxygenate functionalities of alternative fuels are differently effective in reducing soot formation (Westbrook et al., 2006). Accordingly, comparing different soot indices,

Leach et al. (2021) found that the correlation between the number of double-bond equivalents and (u)YSI is stronger for hydrocarbon fuels than for oxygenated fuels. Unfortunately, no model is available for predicting nano soot number density, which is becoming more important in regulations (Samaras et al., 2020).

For the set-up and accuracy assessment of the models, the training data for each model is split into a set for training and testing (Goodfellow et al., 2016). The test set contains approximately 10 % of the training data and is not used within the training to assess the accuracy of the model on unseen data. The data is split so that the statistical distribution of the features in the test and training sets are similar (Goodfellow et al., 2016). A test set with a statistical distribution similar to the training set represents the model domain well and thus reflects model performance across the whole domain rather than just in a particular region. For this purpose, 10000 random splits are performed and the split with the lowest Kullback-Leibler divergence (Kullback and Leibler, 1951) is chosen, indicating the most similar and uniform statistical distribution between the training and test sets.

The models are set up by fragmenting the molecules contained in the training data into UNIFAC groups using the automated fragmentation tool by Müller (2019). The kernels for the GPR of each model are selected by employing the automated kernel-search algorithm developed by Duvenaud et al. (2013), Duvenaud (2014), and Lloyd et al. (2014).

The models achieve an accuracy comparable to the models in the literature (Hukkerikar et al., 2012a; Alshehri et al., 2021; see Table 6.1). The accuracy of the predictions on the test sets measured by the coefficient of determination (R^2) equals on average $R^2 = 0.73$. The corresponding Root-Mean-Square Error normalised by the range of values (nRMSE) is $nRMSE = 8.5\%$. Parity plots of predicted and target values of the test sets visualising prediction accuracy can be found in Appendix D.3.

Synthesisability: By assembling molecular fragments, the LEA3D algorithm generates molecules that always satisfy chemical feasibility, i.e. the octet rule. However, a chemically feasible molecule is not necessarily similar to known molecules and may therefore be technologically challenging to obtain, i.e. hardly synthesisable or only synthesisable with considerable effort via numerous synthesis steps.

The synthesisability of candidate molecules can be assessed through retrosynthesis algorithms (Gao and Coley, 2020). Retrosynthesis algorithms generate synthesis routes for a given product and thus allow an investigation into whether and how a chemical can be synthesised. Various retrosynthesis models have been developed in the past (Sun and Sahinidis, 2022). For the fuel design algorithm, the graph exploration algorithm

developed by Schwaller et al. (2020) is employed for retrosynthesis. The algorithm is based on the molecular transformer, a multi-head attention-based neural network model for forward synthesis prediction with high accuracy (Schwaller et al., 2019).

In this work, a fuel is considered synthesisable if a maximum of three subsequent reactions are required from commercially available reactants to the desired fuel. Otherwise, the synthesis route is deemed too costly to be viable. Moreover, the confidence of the retrosynthesis algorithm in the synthesis route has to be greater than 50 %.

Notably, this definition of synthesisability discards candidate fuels for which efficient synthesis routes are likely to be currently unknown and thus can be classified as hypothetical or technologically unattainable molecules. This assessment of synthesisability does not ensure that the synthesisable molecules can be produced in large quantities, at low cost, or from renewable resources, which should also be a design target for a novel fuel but cannot yet be predicted.

Table 6.1: Data set sizes N_{Train} and N_{Test} and prediction accuracies on the test sets of the EHS indicators using group contribution-based GPR models for the categories autoignition temperature (AiT), bioconcentration factor (BCF), aqueous toxicity of fathead minnow fish ($LC_{50}(\text{FM})$), oral rat toxicity (LD_{50}), permissible exposure limit ($PEL_{\text{OSHA-TWA}}$), and unified yield sooting index (uYSI). The accuracy is measured by the coefficient of determination (R^2), the root-mean-square error (RMSE), and the root-mean-square error normalised by the range of values (nRMSE).

EHS indicator	N_{Train}	N_{Test}	R^2_{test}	$\text{RMSE}_{\text{test}}$	$\text{nRMSE}_{\text{test}}$
<i>AiT</i>	487	54	0.78	58 K	6.4 %
$\log(\text{BCF})$	366	41	0.78	0.69	11 %
$-\log(\text{LC}_{50}(\text{FM}))$	490	54	0.66	0.75	8.6 %
$-\log(\text{LD}_{50})$	2157	240	0.58	0.40	9.1 %
$-\log(\text{PEL}_{\text{OSHA-TWA}})$	346	38	0.60	1.1	13 %
<i>uYSI</i>	397	44	0.99	39	2.9 %
average performance	-	-	0.73	-	8.5 %

6.2.4 Engine efficiency merit function as objective function

Various fuel properties can have a positive impact on achievable SI engine efficiency, including RON, MON, Δh_{vap} , and LBV (Farrell et al., 2018; Szybist et al., 2021). Alternative fuels can exhibit favourable values in one or more of these properties. To evaluate the potential of an alternative fuel candidate for use in advanced highly-boosted SI engines, the impact of these fuel properties on engine efficiency has been empirically quantified through experimental sensitivity analyses under stoichiometric boosted combustion conditions, resulting in the so-called engine efficiency merit function (Farrell et al., 2018; Szybist et al., 2021). The merit function linearly correlates the fuel properties with the achievable improvement in maximum brake thermal engine efficiency (η) compared to a reference fuel, e.g. RON95 gasoline. For example, assuming an engine efficiency of 30 % with RON95 gasoline, a merit value of 16 % leads to an absolute engine efficiency increase of 4.8 %. Efficiency increases in this range have already been achieved in single-cylinder research engines (Hoppe et al., 2016b; Burkardt et al., 2021; Ackermann et al., 2021).

Using the merit function, the expected relative engine efficiency increase of a candidate fuel is calculated based on RON, MON, Δh_{vap} , and LBV compared to RON95 gasoline:

$$\begin{aligned}
 \frac{\textit{merit}}{100\%} &= \frac{\eta - \eta_{\text{ref}}}{\eta_{\text{ref}}} \\
 &= \underbrace{\frac{RON - RON_{\text{ref}}}{1.6}}_{\text{octane number}} - K \cdot \underbrace{\frac{RON - MON - (RON_{\text{ref}} - MON_{\text{ref}})}{1.6}}_{\text{octane sensitivity}} \\
 &\quad + 0.0085 \cdot \underbrace{\frac{\frac{\Delta h_{\text{vap}}}{AFR + 1} - \frac{\Delta h_{\text{vap,ref}}}{AFR_{\text{ref}} + 1}}{1.6}}_{\text{effective octane rating}} + \underbrace{\frac{\frac{\Delta h_{\text{vap}}}{AFR + 1} - \frac{\Delta h_{\text{vap,ref}}}{AFR_{\text{ref}} + 1}}{15.2}}_{\text{charge cooling}} \\
 &\quad + \underbrace{\frac{LBV - LBV_{\text{ref}}}{5.4}}_{\text{laminar burning velocity}}
 \end{aligned} \tag{6.1}$$

Each term in the merit function reflects an empirically found linear influence on maximum engine efficiency (Farrell et al., 2018; Szybist et al., 2021): the influence of octane number, octane sensitivity (defined as $RON - MON$), effective octane rating, charge cooling, and laminar burning velocity. The reference values for RON, MON, and Δh_{vap} are taken from Leitner et al. (2017) (Table 6.2). Lacking a value for LBV from Leitner et al. (2017), the LBV of a commercial RON95 measured by Dirrenberger

et al. (2014) is used as a reference value (Table 6.2). The stoichiometric air-to-fuel ratio of each candidate fuel is denoted by AFR, and the parameter K is a normalised value describing the engine’s operating conditions relative to those of the RON and MON tests, i.e. $K_{\text{RON}} = 0$ and $K_{\text{MON}} = 1$, respectively (Kalghatgi, 2001).

Since the K -value depends on engine operating conditions, selecting a single, representative value is difficult. For modern downsized, turbocharged SI engines, the K parameter is usually negative (Remmert et al., 2014; Kassai et al., 2019; Singh et al., 2021). Kassai et al. (2019) determined values in a single-cylinder research engine with a moderate compression ratio of 10.5 between -0.1 and -1.9 , with K being the lowest at high intake pressure and low engine speed. Since this design aims at fuels for engines with higher compression ratios (e.g. a compression ratio of 16.4 as in a recent work by Ackermann et al. (2021)), K is pragmatically chosen to $K = -1.5$. Appendix D.4 contains an example analysis on how the choice of the K -value affects the predicted efficiency gains.

Table 6.2: Reference values for RON95 gasoline used in the engine efficiency merit function.

Property	Reference value	Reference
RON_{ref}	96	Leitner et al. (2017)
MON_{ref}	85	Leitner et al. (2017)
$\Delta h_{\text{vap,ref}}$	$350 \text{ kJ kg}_{\text{air}}^{-1}$	Leitner et al. (2017)
AFR_{ref}	14	Leitner et al. (2017)
LBV_{ref}	48 cm s^{-1}	Dirrenberger et al. (2014)

6.2.5 Property constraints

The candidate fuels need to meet several thermodynamic, environmental, and practical requirements that are formulated as design constraints (Table 6.3). The normal boiling and melting points of the candidate fuels are constrained to ensure that the fuel is liquid at ambient conditions. Additional constraints on the maximum boiling point and the maximum enthalpy of vaporisation ensure sufficient volatility and, thus, proper in-cylinder mixture formation under cold conditions. If the boiling point or the enthalpy of vaporisation of a candidate fuel is too high, the candidate fuel may not completely evaporate under cold conditions but dissolves in the engine oil potentially causing engine failure due to oil dilution (Larsen et al., 2009; Thewes et al., 2011; Hoppe et al., 2016a). The constraints on boiling and melting points and enthalpy

of vaporisation are taken from a previous design study by Dahmen and Marquardt (2016). Noteworthy, today’s fossil fuel standards use the Reid vapour pressure and characteristic points on the distillation curve to address cold start issues and neglect the enthalpy of vaporization (American Society for Testing and Materials, 2021; DIN German Institute for Standardization, 2017). However, studies on pure alcohol fuels such as ethanol, 1-butanol, and 2-butanol have linked high enthalpies of vaporization to higher pollutant formation (Chen and Stone, 2011; Thewes et al., 2012) and excessive oil dilution (Hoppe et al., 2016a). Further research is needed to better define appropriate upper limits on boiling point and enthalpy of vaporization in case of pure-component alternative fuels.

The candidate fuels’ EHS indicators are constrained to ensure that the candidate fuels are less hazardous than RON95 gasoline with respect to AiT, LC₅₀(FM), LD₅₀, and PEL_{OSHA-TWA}. For BCF, candidate fuels must not be bioaccumulative, i.e. $\log(BCF) < 3.3$ according to Arnot and Gobas (2006). The sooting tendency expressed through the uYSI is not restricted by regulations or policy but should be as small as possible for clean and efficient combustion (Szybist et al., 2021). A strict upper bound is difficult to define. In this work, the uYSI value of n-hexane is considered acceptable and therefore chosen as an upper bound.

Since the EHS indicators are predicted with Gaussian Process Regression (GPR), predicted values are provided with uncertainty quantification. This prediction uncertainty is considered in chance constraints to minimise the number of incorrectly discarded fuel candidates. Candidate fuels are only discarded if a property’s 95% confidence interval violates a constraint, i.e. if a constraint’s lower bound $lb > \Omega_{\max} = \Omega + 1.96 \cdot \sigma_{\Omega}$, or if a constraint’s upper bound $ub < \Omega_{\min} = \Omega - 1.96 \cdot \sigma_{\Omega}$, where Ω is the considered property and σ_{Ω} the property’s prediction uncertainty.

6.2.6 Evaluation of mixture properties

For the design of binary blends with ethanol (cf. Section 6.3.2), mixture properties have to be predicted. Mixture bubble points and mixture enthalpies of vaporisation are calculated considering full non-ideal thermodynamic behaviour using COSMO-RS. The prediction of non-ideal behaviour is a particular strength of COSMO-RS. Pragmatically, the bubble point temperature of the mixture is constrained to the same value (120 °C) as the normal boiling point used in the pure-component design. Noteworthy, mixtures of components with strongly different evaporation characteristics may lead to in-cylinder mixture inhomogeneity, potentially causing wall wetting and oil dilution (Itani et al., 2015; Bardi et al., 2019; Kranz and Kaiser, 2019). Consideration

Table 6.3: Property constraints for the fuel design used in this work. T_{melt} : melting point temperature; $T_{\text{boil/bubble}}$: boiling / bubble point temperature; Δh_{vap} : enthalpy of vaporisation; AiT : autoignition temperature; BCF : bioconcentration factor; $LC_{50}(\text{FM})$: aqueous toxicity of fathead minnow fish; LD_{50} : oral rat toxicity; $PEL_{\text{OSHA-TWA}}$: permissible exposure limit; $uYSI$: unified yield sooting index.

Property	Constraint	Reason	Reference
T_{melt}	$T_{\text{melt}} \leq -20 \text{ }^\circ\text{C}$	liquid fuel	Dahmen and Marquardt (2016)
$T_{\text{boil/bubble}}$	$T_{\text{boil/bubble}} \geq 60 \text{ }^\circ\text{C}$	liquid fuel	Dahmen and Marquardt (2016)
Δh_{vap}	$T_{\text{boil/bubble}} \leq 120 \text{ }^\circ\text{C}$	in-cylinder mixture formation	Dahmen and Marquardt (2016)
AiT	$\Delta h_{\text{vap}} \leq 60 \text{ kJ kg}_{\text{air}}^{-1}$	in-cylinder mixture formation	Dahmen and Marquardt (2016)
BCF	$AiT > 553.15 \text{ }^\circ\text{C}$	RON95 gasoline value	BP Europa (2021)
$LC_{50}(\text{FM})$	$\log(BCF) < 3.3$	REACH administrative	Arnot and Gobas (2006)
LD_{50}	$-\log(LC_{50}(\text{FM})) < 4$	RON95 gasoline value	BP Europa (2021)
$PEL_{\text{OSHA-TWA}}$	$-\log(LD_{50}) < 1.3$	RON95 gasoline value	BP Europa (2021)
$uYSI$	$-\log(PEL_{\text{OSHA-TWA}}) < 1.96$	RON95 gasoline value	BP Europa (2021)
	$uYSI < 30$	n-hexane value	Das et al. (2018)

of such non-ideal evaporation effects is, however, beyond the scope of this work. With the given data-driven models, the non-ideal mixture behaviour for the combustion properties and EHS indicators cannot be predicted. In the absence of more accurate, non-linear models, these mixture properties are approximated using a linear molar mixing rule:

$$\Omega_{\text{blend}} = \sum_{i=1}^n x_i \Omega_i \quad (6.2)$$

In this equation, Ω_i stands for the predicted pure-component properties, and x_i is the mole fraction of component i in the blend. The mixture property is denoted by Ω_{blend} .

For combustion properties, the linear molar mixing rule approximates non-ideal behaviour more accurately than, e.g. a linear liquid volume-based mixing rule, in particular for blends with ethanol (vom Lehn et al., 2021b). For the EHS indicators, the linear molar mixing rule is in line with previous blend design studies (Yunus et al., 2014; Zhang et al., 2018a), following the concept of dose addition for toxicity (LC₅₀(FM), LD₅₀; Altenburger et al., 2003) and exposure hazards (PEL_{OSHA-TWA}; Craig et al., 1999).

6.3 Design of pure fuels and fuel blends for spark-ignition engines

The fuel design method is applied to design (1) pure-component fuels and (2) blend components for binary fuel blends with ethanol.

6.3.1 Pure component fuel design

The fuel design method is started twice with 50 generations and 40 candidate molecules per generation to accommodate for the stochastic nature of the approach. In total, the method investigates 1033 unique molecules in approximately 3 days and 9 hours in parallel on 24 computer cores (see Appendix D.1 for details on the hardware). From these 1033 unique molecules, 22 are candidate fuels that fulfil all constraints (Figure 6.2). 11 candidate fuels outperform the benchmark RON95 gasoline in predicted engine efficiency (Table 6.4). As the optimal fuel, tert-butyl formate is identified with a predicted increase in engine efficiency of approximately 7.9%, followed by ethyl acetate (3.8%), isopropyl formate (3.7%), and vinyl propionate (3.6%). The remaining 6 candidate fuels increasing engine efficiency achieve only minor improvements in predicted engine efficiency between 1% and 3.3%.

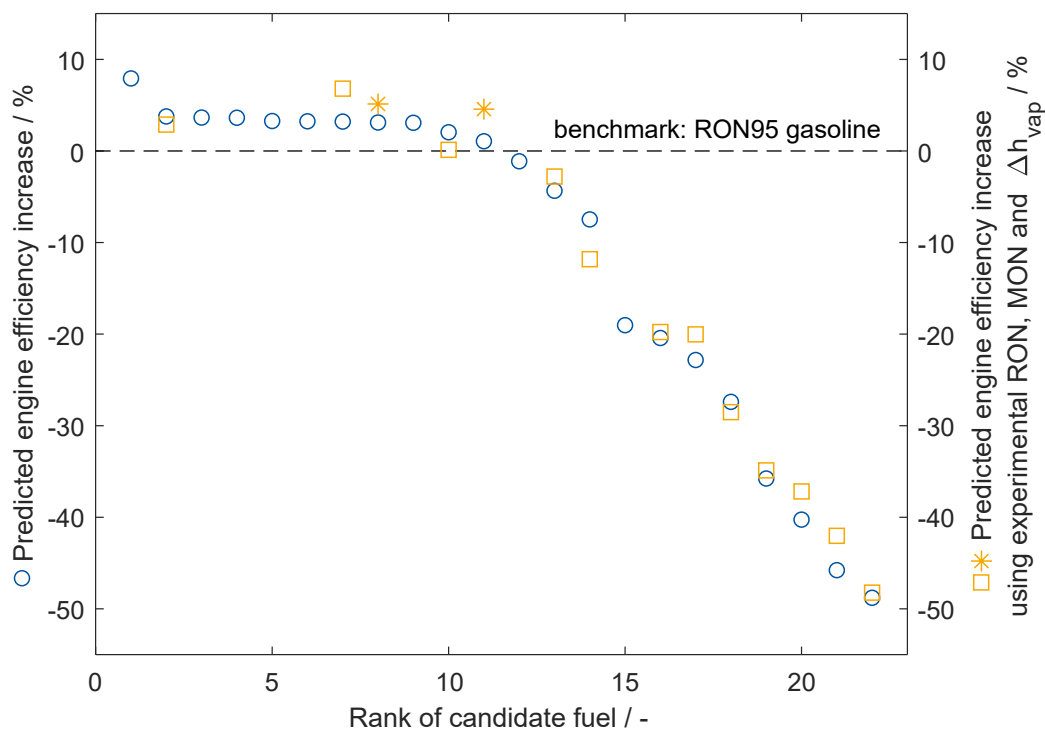


Figure 6.2: Predicted engine efficiency increase for the pure-component design. The blue circles represent the model predictions for the 22 candidate fuels. The orange squares and asterisks are calculated using experimental values for RON, MON, and Δh_{vap} . The experimental values of the molecules marked with asterisks were not used for training of the GNN.

From the 22 identified candidate fuels, 14 candidate fuels have also been considered in the database screening by vom Lehn et al. (2021b) that relied on experimental property data. This experimental data is used to re-calculate the engine efficiency increase and compare the results with the predicted engine efficiency increase. The mean absolute error (MAE) of the predicted engine efficiency increase is only 2.2 %, indicating an accurate assessment by the fuel design method (cf. Figure 6.2). However, 12 of these 14 candidate fuels were also included in the training data set of the GNN that contributes RON and MON values to the engine efficiency assessment (Schweidtmann et al., 2020), leading to a high prediction accuracy of these candidates. Nevertheless, the MAE for 3-methyl-2-butanone and isopropyl acetate, which were not included in the training data set, is comparable with $MAE = 2.8\%$ to the MAE of the training data ($MAE_{\text{train}} = 2.1\%$), indicating generalisability beyond the training data of the GNN.

Using the experimental data, 1-butanol (rank 7, $merit_{exp} = 6.8\%$), 3-methyl-2-butanone (rank 8, $merit_{exp} = 5.1\%$), and isopropyl acetate (rank 11, $merit_{exp} = 4.6\%$) are highly promising candidate fuels achieving higher engine efficiency according to the engine efficiency merit function than predicted during the design. Thus, the fuel design method suggests promising candidate fuels. Still, the final ranking requires subsequent experimental verification.

In contrast to the present study, vom Lehn et al. (2021b) identified methanol, methyl formate, and ethanol as the highest-ranking candidate fuels. However, under the constraints of the present study, methanol and ethanol exceed the maximum permissible heat of vaporisation with predicted values of $214 \text{ kJ kg}_{air}^{-1}$ (experimental $183 \text{ kJ kg}_{air}^{-1}$ (Majer and Svoboda, 1985)) and $106 \text{ kJ kg}_{air}^{-1}$ (experimental $93 \text{ kJ kg}_{air}^{-1}$ (Majer and Svoboda, 1985)), respectively. Methyl formate violates the lower bound on the boiling point with a predicted boiling temperature of 38°C (experimental 32°C (Majer and Svoboda, 1985)). Formates with longer alkyl chains and higher boiling points have been identified as promising pure-component fuels in the present study as well as other esters. In particular, ethyl acetate and isopropyl acetate are suggested as pure-component fuels by the present study, vom Lehn et al. (2021b), and Dahmen and Marquardt (2016). Moreover, 3-methyl-2-pentanone and 2-pentanone have been proposed by Dahmen and Marquardt (2016) as blend candidates with moderate knock resistance, which is confirmed by the predicted moderate pure-component engine efficiency increase found in the present study (3.2% and 2.0% , respectively). 1-butanol and its isomers have also been studied in the literature as 1-butanol has similar knock resistance to RON95 and is known to fulfil the property constraints for an SI fuel (Dahmen and Marquardt, 2016). The fuel design in this work is thus confirmed by findings from the literature but also yields additional candidate fuels that have not yet been investigated.

Influence of property constraints on optimal fuel candidates

In the design of pure-component fuels for SI engines, the tight property constraints drastically limit the number of candidate molecules, as also discovered by Dahmen and Marquardt (2016). In the present study, 1011 of 1033 candidates are already discarded before the engine efficiency evaluation due to constraint violation (Figure 6.3). The most selective constraint is the constraint on the boiling point, which is also enforced first: 960 of 1011 excluded candidate fuels violate the constraint, with 929 of 960 candidates exceeding the maximum boiling point. The subsequently applied constraints on melting point, enthalpy of vaporisation, EHS indicators, and synthesisability discard

Table 6.4: Details on the pure-component fuel design for the 11 candidate fuels predicted to exceed the engine efficiency of RON95 gasoline.

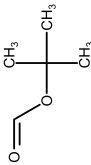
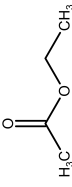
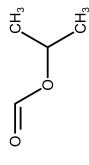
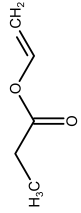
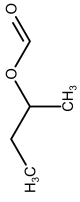
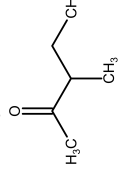

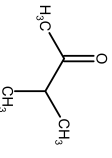
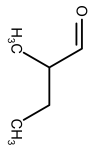
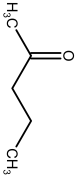
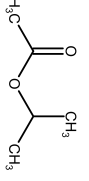
Molecule	Pure component properties					
	$T_{\text{boil}} / \text{K}$	RON / K	MON / -	$\Delta h_{\text{vap}} / \text{kJ kg}_{\text{air}}^{-1}$	$\text{LBV} / \text{cm s}^{-1}$	Merit value
	371.6	115.5	108.6	37.5	38.6	7.92 %
tert-butyl formate						
	372.0	116.7	116.7	49.1	41.2	3.77 %
ethyl acetate						
	357.7	113.1	110.9	47.3	42.3	3.65 %
isopropyl formate						
	379.9	111.3	108.3	42.7	46.9	3.63 %
vinyl propionate						
	378.9	108.6	103.1	39.1	43.0	3.27 %
sec-butyl formate						
	390.5	104.6	96.4	30.7	47.8	3.24 %
3-methyl-2-pentanone						

Table 6.5: Details on the pure-component fuel design for the 11 candidate fuels predicted to exceed the engine efficiency of RON95 gasoline (continued).

Molecule	T_{boil} / K	RON / -	MION / -	Δh_{vap} / kJ kg ⁻¹	$\text{kg}_{\text{air}}^{-1}$	LBV / cm s ⁻¹	Merit value
 1-butanol	387.7	97.8	88.5	54.9	50.1	3.21 %	
 3-methyl-2-butanone	368.9	104.7	96.9	34.9	46.5	3.11 %	
 2-methylbutyraldehyde	351.3	103.3	95.2	33.3	50.2	3.08 %	
 2-pentanone	372.2	106.7	102.2	35.5	49.5	2.04 %	
 isopropyl acetate	391.8	115.1	115.5	40.2	38.6	1.06 %	

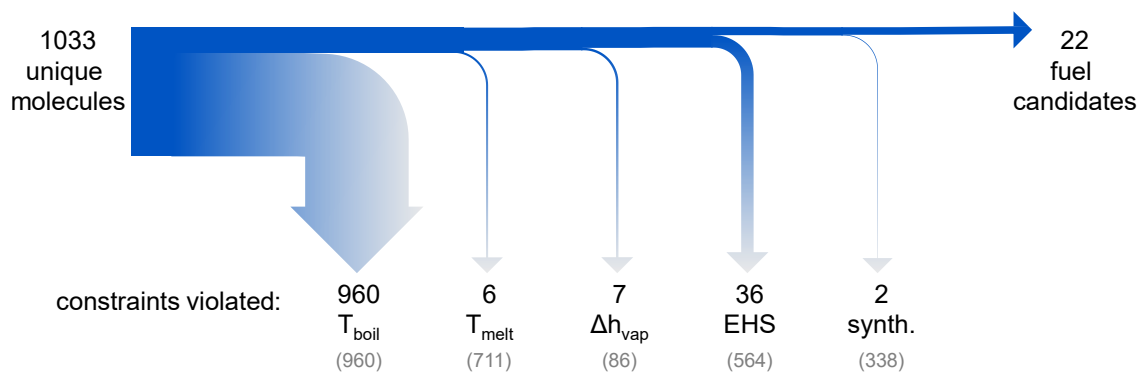


Figure 6.3: Influence of property constraints on the number of candidate fuels in the pure-component fuel design. The grey numbers in brackets are the total number of candidate fuels violating the corresponding constraint.

additional 6, 7, 36, and 2 molecules, respectively. The majority of candidate molecules does not only violate the boiling point constraint but also the melting point, EHS, and synthesizability constraints (grey numbers in Figure 6.3).

The constraints on boiling point and enthalpy of vaporisation are relaxed to investigate whether technical advances in engine design could substantially increase the number of feasible candidate fuels. A relaxation of the constraints on T_{boil} and Δh_{vap} by 10 K and $10 \text{ kJ kg}_{\text{air}}^{-1}$ compared to the original values (cf. Table 6.3) yields three additional fuel candidates: propen-2-ol (engine efficiency increase 12 %, $\Delta h_{\text{vap}} = 69 \text{ kJ kg}_{\text{air}}^{-1}$), 2-propanol (engine efficiency increase 10 %, $\Delta h_{\text{vap}} = 69 \text{ kJ kg}_{\text{air}}^{-1}$), and 2-hydroxy-2-methylpropanal (engine efficiency increase 8.5 %, $T_{\text{boil}} = 120.2 \text{ }^\circ\text{C}$). Further relaxation of the constraints by 10 K and $10 \text{ kJ kg}_{\text{air}}^{-1}$ to $140 \text{ }^\circ\text{C}$ and $80 \text{ kJ kg}_{\text{air}}^{-1}$, respectively, additionally yields 1-propanol (engine efficiency increase 12 %, $\Delta h_{\text{vap}} = 72 \text{ kJ kg}_{\text{air}}^{-1}$).

Moreover, various highly-branched alkenes and alkanes are designed under the relaxed property constraints, e.g. 3,3,4-trimethyl-1-pentene, 4,4,5-trimethyl-1-hexene, or 2,3,3,4-tetramethylpentane. The designed highly-branched alkenes and alkanes are predicted to exhibit high RONs (104 – 110) and high OSs (12-16), leading to predicted engine efficiency increases between 8.5 % and 11 %. However, high uYSI values (74 – 84) are predicted for these molecules, which are only considered fuel candidates by the algorithm as their 95 % confidence intervals reach below the threshold value of $uYSI = 30$. Since highly branched alkenes and alkanes are known to cause soot formation, the actual values of the corresponding candidate fuels are likely in the area of $uYSI = 70 - 80$, as also suggested by the online uYSI estimator by St. John et al. (2022).

In conclusion, pure-component fuel design for high-efficiency SI engines is extremely challenging as it yields only a limited number of candidates. Moreover, relaxations in property constraints only result in a few additional candidate fuels.

6.3.2 Design of binary blends with ethanol

Unfavourable properties of individual molecules can often be balanced by blending (Dahmen and Marquardt, 2017; König et al., 2020a, 2021). Therefore, in fuel design, blending constitutes an additional degree of freedom that can be used to broaden the number of candidate fuels. In this section, the fuel design method is applied to binary blends with ethanol.

Ethanol is an established blend component for commercial fuels (Leitner et al., 2017). It is known for its excellent knock resistance (high RON) and its high enthalpy of vaporisation that provides charge cooling. Both properties increase engine efficiency (Shirazi et al., 2020), leading to a predicted engine efficiency increase by the merit function of 25%. However, as a pure-component fuel, ethanol is troublesome as its high enthalpy of vaporisation (cf. Section 6.3.1) can lead to oil dilution under cold conditions (Larsen et al., 2009; Thewes et al., 2011; Hoppe et al., 2016a). Therefore, ethanol is a suitable base component for a blend due to its favourable combustion properties but must be balanced with a tailor-made secondary component.

For the purpose of blend design, both the molecular structure and the molar fraction of the candidate blend component in a binary blend with ethanol are optimised. The fuel design method is run twice for 50 generations with 40 candidate molecules per generation. In total, the method investigates 1310 unique blends in 9 days and 20 hours in parallel on 24 computer cores (see Appendix D.1 for details on the hardware). Of these unique blends, 226 fulfil the property constraints and are thus candidate blends (see Appendix D.5 for details). 184 candidate fuel blends lead to a positive merit function value indicating an engine efficiency increase compared to RON95 gasoline (Figure 6.4). Compared to the optimal pure-component fuel tert-butyl formate, 136 blends yield a higher engine efficiency. Therefore, as Figures 6.3 and 6.4 demonstrate, the design space for fuel blends is much larger than the design space for pure-component fuels and offers many more possibilities for increasing engine efficiency.

The optimal blend with ethanol is obtained with 3,4-dimethyl-3-propan-2-yl-1-pentene, a highly-branched alkene. The engine efficiency increase of this optimal blend is predicted to be 19.5% (Table 6.6). The majority of the blend is composed of ethanol (83 mol-%) and only 17 mol-% of 3,4-dimethyl-3-propan-2-yl-1-pentene. The blends with the second and third highest engine efficiency increases are formed with 22 mol-% 3,3,4-trimethyl-1-pentene, and with 16 mol-% 4,5-dimethyl-4-propan-2-yl-1-hexene,

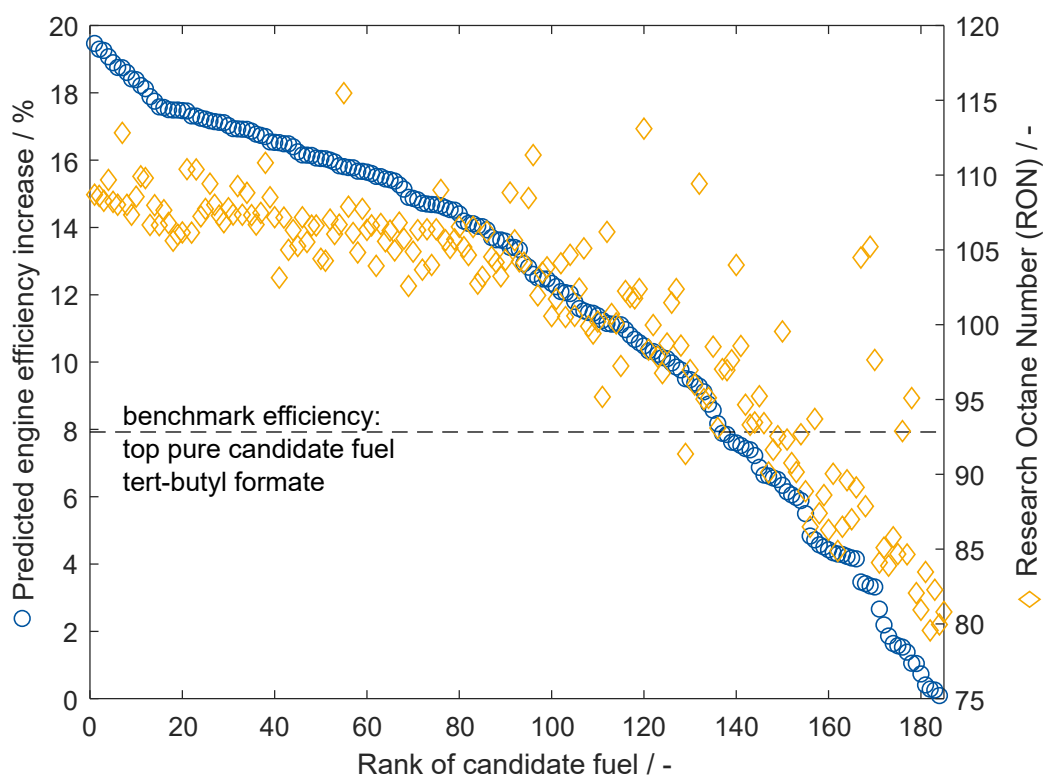


Figure 6.4: Predicted engine efficiency increase for the 184 binary blends with ethanol. The blue circles represent the predictions from the predictive models. The orange diamonds are the predicted blend RONs.

again highly-branched alkenes. The engine efficiencies are predicted to increase by approximately 19.3%.

The top candidates combine a high RON (104 – 107) and a high octane sensitivity (15 – 16), as this combination is known to enable high engine efficiency in modern highly-boosted engines (Kassai et al., 2019). Due to these favourable properties, alkenes are generally high-ranking blend components: 18 of the top 50 candidate blends are formed with an alkene and 1 with an alkadiene. The positive effect of the vinyl groups on engine efficiency is known from the literature (vom Lehn et al., 2021b).

The top 3 blend components are not suitable as pure-component fuels, e.g. due to constraint violations on boiling point, toxicity, or soot formation. Since the top 3 candidates are large alkenes (C8-C11), they have high boiling points of 129 °C, 91 °C, and 145 °C, and high sooting tendencies as evident by predicted uYSI values of between 75 and 107. Moreover, the components are highly toxic to aqueous organisms as indicated by a predicted $-\log(LC_{50}(FM)) = 4.0 - 4.8$ and potentially have a low

exposure limit (predicted $-\log(PEL_{\text{OSHA-TWA}}) = 3.3 - 3.5$). In a blend with ethanol, ethanol is predicted to compensate for these properties while the low enthalpies of vaporisation of the top blend components ($18 - 21 \text{ kJ kg}_{\text{air}}^{-1}$) help to mitigate ethanol's high enthalpy of vaporisation.

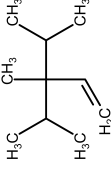
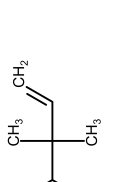
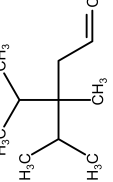
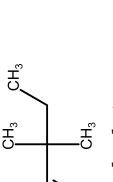
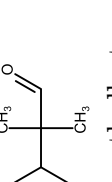
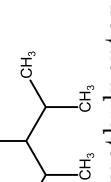
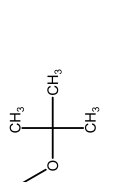
The top 3 blend components are synthesisable as indicated by retrosynthesis through 1 or 2 reactions from commercially available components. However, they are currently not commercially available themselves as determined by database searches. The highest-ranking commercially available component is 3,3-dimethyl-1-pentene on rank 8. An optimised blend of 26 mol-% 3,3-dimethyl-1-pentene and 74 mol-% ethanol increases engine efficiency by approximately 18.6 % compared to RON95. Similar to the top 3 blend components, 3,3-dimethyl-1-pentene has a high RON of 105 and an octane sensitivity of 15. The predicted EHS indicators suggest similar toxicity ($-\log(LC_{50}(\text{FM})) = 3.8$, $-\log(LD_{50}) = 1.6$) and permissible exposure limit ($-\log(PEL_{\text{OSHA-TWA}}) = 3.8$), but lower soot formation ($uYSI = 57$).

A commercially available alternative to 3,3-dimethyl-1-pentene is 2,2,3-trimethylbutanal on rank 12. A blend of 28 mol-% 2,2,3-trimethylbutanal and 72 mol-% ethanol is predicted to increase engine efficiency by 18 %. Compared to 3,3-dimethyl-1-pentene, 2,2,3-trimethylbutanal has the advantage of not containing a vinyl group but instead contains an aldehyde group. However, molecules with vinyl or aldehyde groups can both age the fuel blend and reduce its stability due to polymerisation (Pereira and Pasa, 2006) and high reactivity (Baehr et al., 2022). The highest-ranking commercially available candidate blend component without a vinyl or aldehyde group is 2,3,4-trimethylpentane on rank 25. A blend of 21 mol-% 2,3,4-trimethylpentane and 79 mol-% ethanol is predicted to increase engine efficiency by 17 %. In comparison to the top 3 and the high-ranking commercially available blend components, the predicted EHS indicators of 2,3,4-trimethylpentane suggest safer handling and use with a higher permissible exposure limit ($-\log(PEL_{\text{OSHA-TWA}}) = 1.9$) and lower toxicity ($-\log(LC_{50}(\text{FM})) = 3.6$, $-\log(LD_{50}) = 1.4$). Therefore, 2,3,4-trimethylpentane could be the most promising commercially available blend component in this design, highlighting the additional requirements for the practical selection of a promising fuel besides engine efficiency.

Noteworthy, aromatic molecules are not suggested as promising blend components although aromatic molecules are known to have high RONs (vom Lehn et al., 2021b) and today's RON95 fuel contains up to 35 vol-% aromatic components (DIN German Institute for Standardization, 2017). In this blend design, aromatic molecules are discarded because they fail to meet the melting point constraint and/or violate the acceptable uYSI value of the mixture when added in an amount that sufficiently reduces the high enthalpy of vaporization of ethanol.

6.3 Design of pure fuels and fuel blends for spark-ignition engines

Table 6.6: Blend design results: the three highest-ranking candidate blend components, the highest-ranking commercially available blend components, and the optimal pure-component fuel tert-butyl formate. The predicted pure-component engine efficiency increase is calculated to show the effect of blend design. Note that except for tert-butyl formate, the components do not satisfy all property constraints as pure components.

Molecular structure	Blend properties		Engine efficiency increase pure comp.	Engine efficiency increase blend	Molar fraction in blend x_{blend}	Rank in design eng. eff.	RON _{blend}
	RON _{blend}	$T_{\text{bubble,blend}}$					
 3,4-dimethyl-3-propan-2-yl-1-pentene	108.7	352.7 K	10.2%	19.5%	0.175	1	17
 3,3,4-trimethyl-1-pentene	108.6	346.2 K	10.9%	19.3%	0.222	2	18
 4,5-dimethyl-4-propan-2-yl-1-hexene	108.3	353.8 K	8.28%	19.3%	0.159	3	22
 3,3-dimethyl-1-pentene	108.0	337.5 K	9.18%	18.6%	0.261	8	25
 2,2,3-trimethylbutanal	109.2	355.4 K	8.38%	18.1%	0.275	12	9
 2,3,4-trimethylpentane	107.7	344.2 K	6.42%	17.2%	0.214	25	31
 tert-butyl formate	112.0	354.3 K	7.92%	15.8%	0.453	55	1

Comparison between ranking by RON and by engine efficiency increase

The blend design maximising the merit function of engine efficiency increase (blue circles, Figure 6.4) is challenged by a blend design maximising an individual blend property, here, the RON of the blend (orange diamonds, Figure 6.4). In general, the correlation between the predicted engine efficiency increase of the blend and the RON of the blend is strong, as evident by a Pearson correlation coefficient of $\rho = 0.96$. The strong correlation is not surprising since the RON is a key property determining engine efficiency with the largest impact on the engine efficiency increase. Consequently, a blend design maximising RON would be sufficient to identify many high-ranking candidate blends and could discard low-ranking candidate blends.

However, the correlation becomes weak among the high-ranking candidates in engine efficiency, e.g. for the highest-ranking 50 candidate blends, the Pearson correlation coefficient equals only $\rho = 0.49$. Only 2 of the top 10 blend components in the RON maximisation are among the top 10 blend components in the engine efficiency maximisation. 3 of the top 10 blend components in the RON maximization are not even among the top 50 blend components in the engine efficiency maximization.

The blend component maximising RON is the best pure-component fuel tert-butyl formate. The pure-component RON of tert-butyl formate ($RON = 116$) is higher than the RON of ethanol ($RON = 109$). Since tert-butyl formate already meets the property constraints as a pure component, ethanol neither increases the blending RON nor contributes to fulfilling constraints. Therefore, the “blend” maximising RON does not contain any ethanol and is equal to pure tert-butyl formate. Apart from tert-butyl formate, 8 more “blends” do not contain ethanol or reduce the ethanol fraction if optimised for RON.

Thus, focusing the optimisation on a single characteristic property, e.g. the blending RON, is insufficient for a ranking which accurately reflects engine efficiency. The most promising candidate blend components yield not only a high RON of the blend but also balance octane sensitivity, heat of vaporisation, and laminar burning velocity. Therefore, CAPD and fuel design should use an application-based objective function that combines the effects of the individual properties.

Comparison between pure-component and blend results

Of the 226 candidate blend components, 26 fulfil the pure-component constraints. Of these pure fuel candidates, 10 candidates would yield positive engine efficiency increases as pure-component fuels and thus are predicted to increase engine efficiency compared

to RON95 gasoline. Six of these candidates were not discovered in the pure-component design. Particularly promising are 2,2-dimethylbutyraldehyde, isobutyraldehyde, and isovaleraldehyde with predicted engine efficiency increases of 6.3 %, 3.9 %, and 3.6 %, respectively. The additionally identified pure-component candidates indicate that the pure-component molecular design study is not exhaustive. For a more comprehensive list of candidates, the method needs to be re-run, e.g. for more generations, or the general solution strategy needs to be improved.

To compare the predicted engine efficiency increase of the blends to that of the pure components, the pure-component efficiency increase are calculated for all blend components ignoring property constraint violations (Figure 6.5). As evident by a Pearson correlation coefficient of $\rho = 0.69$, the pure-component engine efficiency increase is a good indicator of the engine efficiency of the blend. High-ranking pure-component fuels usually lead to high-ranking blends; e.g. 18 of the 25 components with the highest pure efficiency increase are among the top 50 blend components. However, not all high-ranking blend components necessarily have a high pure-component efficiency increase: 19 of the top 50 blend components are not among the top 50 pure components. Thus, the effects of blending go beyond simple weighting of the objective function.

Generally, all candidate fuel blends are predicted to achieve a higher engine efficiency than the pure blend components if optimisation targets maximum engine efficiency increase. In particular, even the optimal pure fuel tert-butyl formate benefits from a binary blend with ethanol: Engine efficiency is predicted to increase from 7.9 % to 16 % in a blend with 55 mol-% ethanol. The optimal balance of molecular properties also enables high-ranking blends for molecules with low pure-component efficiency increases. For example, 4-ethyl-4,5-dimethyl-1-hexene ranks 79th in the pure-component ranking and improves to rank 17 in a blend with 82 mol-% ethanol through an increase in predicted engine efficiency from -0.9% as a pure component to 18 % in the fuel blend.

The increased engine efficiency of the blends compared to the pure-component fuels is primarily attributed to the favourable combustion properties of ethanol. Ethanol has the highest predicted engine efficiency increase of the pure components (25 %). Therefore, only 18 of the 184 blend components with a positive efficiency increase are blended with ethanol at more than 50 mol-%. The majority of blend components act as enablers for ethanol since pure ethanol violates the constraint on the heat of vaporisation. For 179 of the 184 blends, the blend's heat of vaporisation matches the constraint limit of $60 \text{ kJ kg}_{\text{air}}^{-1}$. Conversely, 144 of the 184 candidate blend components violate the maximum boiling temperature as pure components but are feasible in the blend through a reduction of the bubble point by ethanol. Thus, blending allows for meeting the strict property constraints and significantly enlarges the molecular design space. To fulfil fuel property constraints for SI engines, blend design is therefore key.

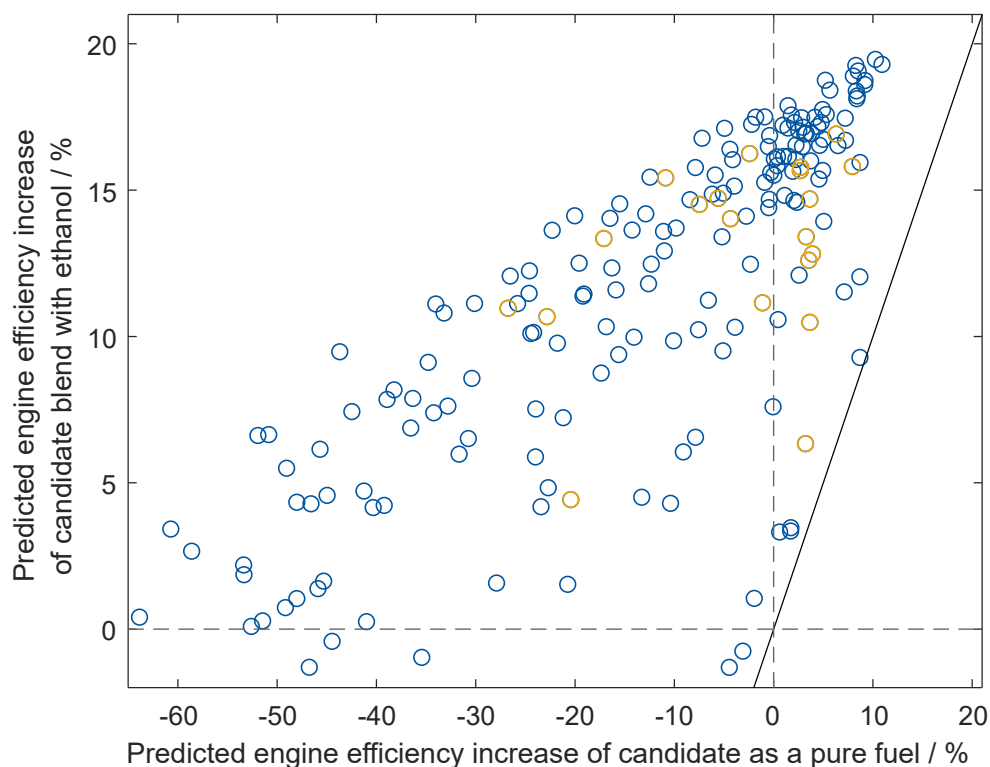


Figure 6.5: Parity plot comparing pure-component and blend merit function value of the candidate fuels. Each circle stands for one blend component. The orange circles represent blend components that fulfil the property constraints of the pure-component design as well. The black line indicates equal efficiency increase as a pure substance and as a blend component. The dashed grey lines represent the benchmark engine efficiency of RON95.

Comparison between engine efficiency increase and direct CO₂ emissions caused as alternative objective

An optimal renewable fuel should not only lead to high engine efficiency but ultimately must enable sustainable mobility with low environmental impact, e.g. with low CO₂ emissions. Therefore, for each candidate blend, the direct CO₂ emissions generated by driving 100 km are calculated based on the maximum engine efficiency. The CO₂ emissions of the candidate blends are compared to the CO₂ emissions of RON95 gasoline, assuming a fuel consumption of RON95 of 7 L/100 km. Furthermore, constant engine operation is assumed at the optimal operating point achieving maximum engine efficiency. Although this assumption limits the significance for practical implementation

because no full driving cycle is considered, the analysis still provides insight into whether the maximisation of engine efficiency minimises direct CO₂ emissions, which are additionally influenced by the heating value and the amount of carbon of each fuel.

To calculate the CO₂ emissions of the blends (\dot{m}_{CO_2}), the thermal engine efficiency increase of a fuel is assumed to reduce the energy demand for the same engine power P_{engine} as with RON95:

$$\begin{aligned}\dot{m}_{\text{CO}_2} &= \frac{M_{\text{CO}_2}}{M_{\text{blend}}} \cdot N_{\text{C,blend}} \cdot \dot{m}_{\text{blend}} \\ &= \frac{M_{\text{CO}_2}}{M_{\text{blend}}} \cdot N_{\text{C,blend}} \cdot \frac{\dot{m}_{\text{RON95}} \cdot LHV_{\text{RON95}}}{\frac{\text{merit}}{100\%} + 1} \cdot \frac{1}{LHV_{\text{blend}}}\end{aligned}\quad (6.3)$$

In this equation, the mass-based fuel consumption of the candidate blend and RON95 are denoted by \dot{m} . The molar mass of the candidate blend and CO₂ are denoted by M , and $N_{\text{C,blend}}$ stands for the average number of carbon atoms per molecule in the blend. The lower heating value of a fuel blend (LHV_{blend}) is calculated from linear mixing of the pure-component heating values (Equation 6.2). The pure components' lower heating values are calculated from the stoichiometric combustion reaction using standard enthalpies of formation. The gas-phase standard enthalpy of formation for each candidate fuel is predicted from a GC-based GPR similar to the EHS hazards (see Section 6.2.3, $N_{\text{Train}} = 697$, $N_{\text{Test}} = 78$, Accuracy: $R_{\text{test}}^2 = 1.00$, $RMSE_{\text{test}} = 16 \text{ kJ mol}^{-1}$, $nRMSE_{\text{test}} = 1.5\%$). More details and property data used for RON95 can be found in Appendix D.6.

In total, 177 candidate blends with an increased engine efficiency compared to RON95 (Figure 6.6) reduce the direct CO₂ emissions compared to RON95. The blend with the lowest CO₂ emissions (red upward-pointing triangle in Figure 6.6) contains 19% 2,3,3,4-tetramethyl-pentane and 81% ethanol and reduces the exhaust gas CO₂ emissions compared to RON95 by 19% to 12.9 kg CO₂ / 100 km. 2,3,3,4-tetramethyl-pentane also achieves one of the highest engine efficiency increases with 19%, ranking fourth in engine efficiency increase. The blend with the highest efficiency increase (orange downward-pointing triangle in Figure 6.6) also reduces direct CO₂ emissions by approximately 19% to 13.0 kg CO₂ / 100 km and ranks third in CO₂ emissions reduction. Generally, high engine efficiency is closely aligned with lower direct CO₂ emissions ($\rho = -0.88$). However, the energy and carbon content of the fuel also significantly affect CO₂ emissions. Thus, a fuel design maximising engine efficiency is not sufficient to guarantee minimum direct CO₂ emissions.

Besides the direct exhaust emissions of a vehicle, emissions during the production of the blend are highly relevant for environmental assessment (Deutz et al., 2018). Therefore, a meaningful fuel design with an environmental objective must include a "well-to-wheel" assessment.

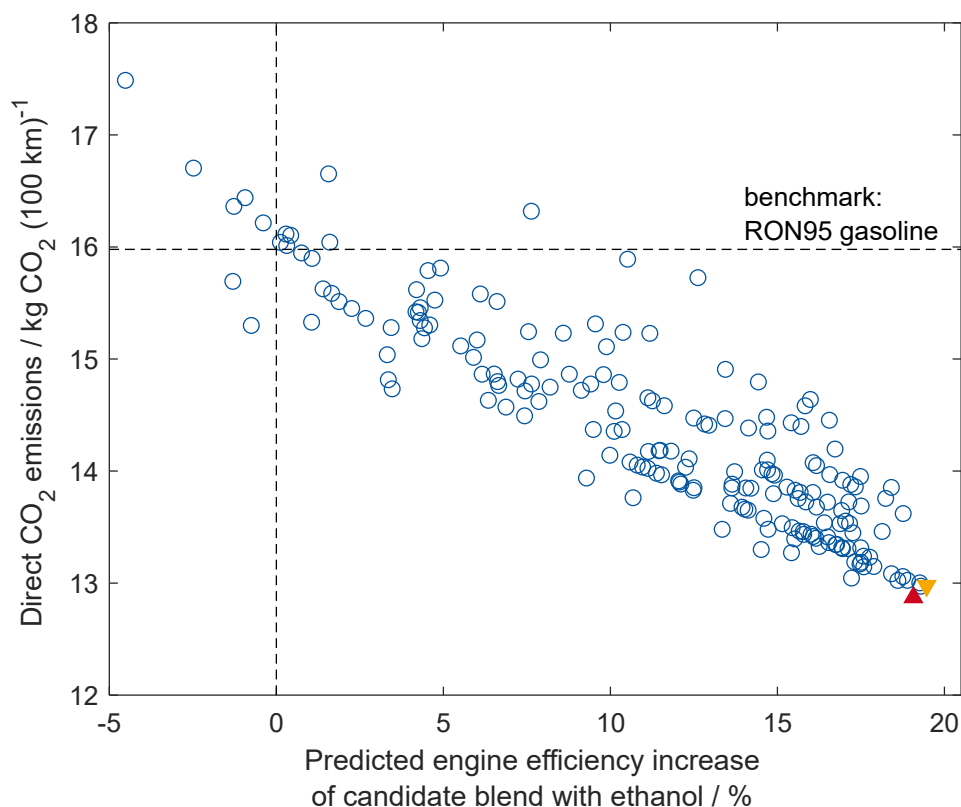


Figure 6.6: Direct CO₂ emissions from fuel combustion compared to predicted engine efficiency increase for each candidate blend. The red upward-pointing triangle is the candidate blend with the lowest direct CO₂ emissions. The orange downward-pointing triangle is the candidate blend with the highest predicted engine efficiency increase. The dashed grey lines represent the benchmark engine efficiency and emissions of RON95.

6.4 Conclusion

This chapter presents a method for the molecular design of fuels for future, dedicated spark-ignition engines with high compression ratios that uses an empirical model of engine efficiency as the objective function. The method is based on a genetic algorithm

for molecular optimisation and integrates the prediction of various properties to evaluate the feasibility and expected engine efficiency of each candidate fuel. Thermodynamic properties are calculated using COSMO-RS. Combustion properties are predicted by a graph neural network and a group contribution-based artificial neural network from the literature. Constraints on environmental, health, and safety indicators as well as synthesisability are assessed using Gaussian process regression and a retrosynthesis method.

The method is applied to design candidate fuels with high spark-ignition engine efficiency in two case studies: (1) the design of pure-component fuels and (2) the design of binary blends with ethanol. The application highlights the challenge of finding suitable pure-component fuels that meet all property constraints. In the design, only 11 of the 1033 investigated molecules fulfil the property constraints and increase predicted engine efficiency compared to RON95 gasoline. As an optimal pure-component fuel, tert-butyl formate is identified with a predicted engine efficiency increase of 7.9%.

The molecular design space is extended by designing a two-component fuel blend with ethanol. 184 blend components are identified that exceed the engine efficiency of RON95. 136 of the 184 candidate blends also exceed the predicted engine efficiency of the best pure-component fuel tert-butyl formate, highlighting the significant potential of blend design. As optimal blend components, highly-branched alkenes are identified increasing engine efficiency by up to 19%. In most cases, the designed blend component represents the minor constituent of the blend but satisfies the property constraints to enable the inclusion of a substantial amount of ethanol.

The fuel design method has successfully designed candidate fuels that are known to increase engine efficiency from experiments in the literature. However, for the less well-known candidate fuels proposed by the method, the predicted properties need to be confirmed. Furthermore, experimental engine tests are required to confirm the predicted efficiency increases.

In future work, the accuracy of the fuel design method should be improved, since the uncertainties in the predictions are still large for some properties. To increase the reliability and the significance of the predictions, the prediction uncertainties have already been considered for constraint evaluation and could be extended to design under uncertainty. Moreover, accuracy can be improved by integrating additional models accounting for non-ideal mixture behaviour. The mixture property models except for COSMO-RS are currently linear molar mixing rules, e.g. for toxicity and RON. However, toxicity and RON are known to frequently exhibit non-ideal mixing behaviour (vom Lehn et al., 2021b; Smith et al., 2013). Moreover, for the design of renewable fuels, e.g. produced from biomass and carbon dioxide, the assessment of synthesisability

should be adapted to constrain the molecular design space to appropriate processing pathways. To lower the market-entry barriers of alternative fuels, blending with fossil gasoline to comply with existing fuel standards could be considered already in the design phase.

Future work could also extend the design method to a higher number of blend components or even optimise the number of blend components as a design degree of freedom, since this work showed the opportunities of blend design for spark-ignition engine fuels.

Finally, to improve the engine efficiency assessment, a more detailed engine model is desirable that considers a typical driving cycle instead of a correlation of potential engine efficiency. Ultimately, the model for fuel assessment should consider not only the combustion of the fuel but the emissions of the entire life cycle of the fuel aiming at "well-to-wheel" optimisation.

The fuel design method presented in this chapter is a special case of CAPD optimising the product's performance in application. Therefore, the progress in designing fuels for maximum engine efficiency and minimum direct CO₂ emissions for mobility is an important demonstration towards the life cycle optimisation of products.

Summary, conclusions, and future perspectives

This thesis investigates the computer-aided design of molecules, experiments, and processes for a sustainable chemical industry with tailored products and processes. In this chapter, the thesis is summarised, and conclusions are drawn (Section 7.1). Perspectives for further research are outlined in Section 7.2.

7.1 Summary and conclusion

The tailored design of sustainable processes and products in the chemical industry is a key contribution to reducing the industry's environmental impact. In this thesis, design methods for molecules, experiments, and processes are developed to incorporate new design metrics beyond feasibility, technical function, or cost. Present methods frequently use simplified metrics as objectives resulting in non-optimal and rather general designs.

The design for minimum environmental impact needs to consider and address multiple levels: The environmental impacts of chemical products and processes are influenced from the molecular level via the physical properties and applications to the environmental impacts at the system level. This thesis advances each level and the links between the individual levels towards life cycle design.

At the system level, this thesis overcomes the limitations of computer-aided molecular and process design (CAMPD), which has previously only considered environmental impact potentials and cradle-to-gate system boundaries. Cradle-to-grave Life Cycle Assessment (LCA) is integrated with CAMPD of solvents to optimise solvents and processes for minimum life cycle environmental impacts. The integration is achieved by combining predictive LCA with the COSMO-CAMPD framework for solvent and process design, forming the COSMO-susCAMPD framework. The predictive LCA uses an Artificial Neural Network (ANN) to estimate the cradle-to-gate impacts of candidate

solvents. COSMO-CAMPD is based on COSMO-RS and pinch-based process models and, thus, provides gate-to-grave life cycle inventory data from the process models. Since the ANN is trained on molecular descriptors available in COSMO-CAMPD, the framework is still fully predictive. As a result, the environmental impacts of the complete life cycle of candidate solvents can be used as objective for CAMPD. The COSMO-susCAMPD framework is applied to design solvents in a hybrid extraction-distillation process. The results highlight the need for cradle-to-grave LCA as objective function: Heuristics, economics, or cradle-to-gate LCA lead to suboptimal solvent choices and are outperformed by the COSMO-susCAMPD solutions.

At the application level, the effects of comprehensive process modelling and optimisation on the molecular design are investigated. Computer-aided molecular design of solvents is extended for the design of heat-integrated processes for minimum utility demand to consider the interactions between molecular properties and process performance. The entire process flowsheet is modelled accurately by pinch-based process models for the most common unit operations: extraction, distillation, absorption, multiphase reaction, and heat integration. Process settings are optimised considering heat integration for each candidate solvent; thus, the solvents are designed based on their optimised process performance. In two case studies, solvents are designed for minimum cradle-to-grave environmental impacts in a hybrid extraction-distillation process and a process for integrated carbon capture and utilisation. In both, designed solvents outperform literature benchmarks on process- and system-level design objectives. The results confirm existing heuristics for solvent selection but also highlight the importance of integrating molecular and process design to achieve maximum process performance and quantitative process- and system-level estimates.

At the property level, the integration of various properties into the COSMO-CAM(P)D method is demonstrated using a hybrid approach. Thermodynamic properties of pure components and mixtures are predicted using quantum chemistry-based models, i.e. COSMO-RS and thermochemistry. Various other molecular properties, i.e. combustion properties, environmental impacts and hazards, and synthesizability, are estimated by recent models using deep learning and Bayesian regression. The extension of property prediction available in CAM(P)D and CAPD is key to substantially expanding the accessible design objectives, e.g. for minimising environmental impacts.

For validating predictions and collecting experimental data for model training and parametrisation, experiments are inevitable. Thus, experimentation and the design of experiments is an important part of product and process design at the property level. To reduce the experimental effort required for accurate modelling of applications, *c*-optimal experimental design (*c*-OED) is presented. *c*-OED aims to minimise the

uncertainty of the simulation results as design objective, which is estimated by linear uncertainty propagation from uncertain property parameters through the application model, e.g. a chemical process model or the LCA of a process. c-OED is applied to design liquid-liquid equilibrium and diffusion experiments minimising the uncertainty of thermodynamic, economic, and environmental performance metrics of three solvent-based processes. In all three case studies, the c-optimal design can substantially reduce the experimental effort for the same simulation accuracy compared to state-of-the-art OED that neglects the parameter application. The findings are confirmed by a Monte Carlo analysis of the designed experiments but also discover the limits of c-OED for highly non-linear process models.

At the molecular level, the design of fuels as chemical products is investigated with the expertise in designing processing chemicals. Based on the advances in CAMPD on property prediction and candidate assessment, an optimisation-based fuel design method is developed incorporating an empirical model of spark-ignition engine efficiency as objective function. Engine efficiency increase compared to conventional RON95 gasoline is calculated for each candidate fuel as a function of the fuel's properties. The method designs pure-component fuels and blend components in a binary blend with ethanol predicted to substantially exceed the engine efficiency of conventional RON95 gasoline. The results show that optimising a key property determining engine efficiency is sufficient to differentiate between high- and low-ranking candidate fuels. However, accurate ranking is only possible with a targeted application-based objective function, i.e. an engine model combining the effects of various properties. Thus, the presented method successfully designs promising candidate fuels for spark-ignition engines and demonstrates product design based on an application-based objective function. Similar to CAM(P)D, the results highlight the importance of accurately formulating the objective function in computer-aided product design (CAPD), which is key for future life cycle design of chemical products.

In conclusion, the computer-aided design of molecules, experiments, and processes is advanced in this thesis by systematically integrating sustainability assessment and targeting towards application. The presented methods successfully link environmental impacts with molecular, experimental, and process degrees of freedom and enable application modelling beyond mere technical feasibility. Thus, the methods presented in this thesis demonstrate possibilities for shaping a sustainable chemical industry.

7.2 Future perspectives

The integrated design methods presented in this thesis successfully design solvents, fuels, experiments, and processes. The methods provide a strong basis for further extensions in the domains of property prediction, solution algorithm, integration of experimental data, and towards life cycle design.

Accelerating and extending property prediction

Property prediction forms the core of molecular design and determines which applications can be designed. The molecular design in this thesis uses quantum chemistry and COSMO-RS for thermodynamic property prediction. The framework also allows for flexible integration of, e.g. machine learning-based methods to predict properties not related to thermodynamics. These predictive methods can be extended and improved for more accurate or faster property prediction.

1. Extended property prediction: In this work, various properties are predicted for molecular design and assessment. However, the list of predicted properties is not exhaustive, e.g. transport properties for sizing and costing of equipment are lacking. Transport properties can be calculated via entropy scaling using the PCP-SAFT equation of state as demonstrated by Hopp and Gross (2017, 2019) and already integrated in CAMPD by Schilling et al. (2017, 2020). The prediction of properties that quantify reactivity and stability of molecules could provide valuable information for practical and reliable application. Recently, a GC method for inertness prediction was presented (Liu et al., 2019c). In the future, inertness evaluation could be included more rigorously through reaction network prediction, e.g. through the Chemtrayzer model (Döntgen et al., 2015, 2018; Krep et al., 2022). The design of reaction networks could be a long-term goal.

2. More accurate property prediction: The combination of methods for property prediction could not only extend the considered properties but also increase the accuracy of the predictions. Kaminski et al. (2017) demonstrated that combining PCP-SAFT with COSMO-RS increases prediction accuracy. However, additional quantum chemistry calculations have to be performed to remain predictive (Kaminski and Leonhard, 2020). The integration of quantum chemistry-based methods to predict charged species thermodynamics, e.g. through Cluster Continuum models (Kröger et al., 2020), can be explored to accurately design charged species in applications such as battery and electrolyte design. A recent combination of density functional theory (DFT) and machine learning (ML) based methods showed higher accuracy through a

neural network-enhanced DFT functional compared to conventional state-of-the-art functionals (Kirkpatrick et al., 2021).

3. Faster property prediction: While the quantum chemistry-based methods provide high accuracy and have been developed to account for non-ideal mixture behaviour, e.g. through COSMO-RS, the underlying DFT calculations are computationally demanding and represent the main computational cost of the integrated algorithm. Since the DFT calculations are performed for pure-component properties, e.g. optimised molecular geometries or screening charge profiles (σ -profiles), future effort should be directed to increasing the speed of the DFT calculations or even replacing them by, e.g. ML-based methods. For the generation of σ -profiles, promising ML-based methods have recently been proposed in the literature (Chen et al., 2021; Liu et al., 2021). Faber et al. (2017) directly predicted properties of quantum chemistry calculations such as (free) energies, heat capacities, or dipole moments. However, since ML is not the universal solution to all problems (Walters and Barzilay, 2021), hybrid approaches should be explored, e.g. towards combining ML and physical-based models (Zhou et al., 2021; Jirasek and Hasse, 2021; Sharma and Liu, 2022). Hybrid approaches seem particularly promising for mixture properties, which are currently rarely predicted or estimated with simple linear mixing rules.

Accelerating problem solution

The fragment-based genetic algorithm LEA3D (Douguet et al., 2005) forms the basis of the presented molecular design algorithm and has proven to be very versatile and applicable for many problems in CAMD and CAPD. The molecular design can easily be adapted to include extended property prediction and other application models as objective, particularly, since no derivatives are required.

Genetic algorithms are frequently used in molecular design as stochastic solution methods (Sun et al., 2020). However, in comparison to other derivative-free optimisation methods, genetic algorithms need a large number of evaluations and do not provide a termination criterion leading to long solution times and high computational cost (Sun et al., 2020). In the literature, various other derivative-free optimisation methods were developed for continuous and mixed-integer problems with high solution performance (Rios and Sahinidis, 2013; Ploskas and Sahinidis, 2021). Particularly promising optimisation algorithms optimise a surrogate model of the problem to reduce function evaluations (Müller, 2016). In CAMPD, fewer function evaluations could significantly decrease the solution time by reducing the computational demand for process optimisation and property prediction by quantum chemistry methods.

The surrogate-based optimisation of molecules and materials has recently been demonstrated successfully using a continuous representation of molecules (Gómez-Bombarelli et al., 2018). Within this approach, molecules are optimised directly in a continuous, so-called latent space before the result is decoded to an interpretable molecular structure. Since the latent space optimisation is a continuous non-linear optimisation problem, it is more efficient to solve than the direct optimisation problem of the 3D molecular structure. Currently, methods frequently focus on individual molecular properties (Sousa et al., 2021) but could potentially be extended to optimise state-of-the-art CAMPD models. In particular, the latent space optimisation is being further developed to consider the integer constraints of molecular structures. Approaches include modifying the underlying covariance function of the surrogate model (Garrido-Merchán and Hernández-Lobato, 2020; Häse et al., 2021) or imposing constraints on the acquisition function (Müller, 2016) or in the latent space (Griffiths and Hernández-Lobato, 2020).

Moreover, a new solution algorithm could substantially benefit from using a molecular representation that does not depend on 3D-molecular fragments. Recently, the SELFIES notation for molecules was presented (Krenn et al., 2020) and used for molecular design in a genetic algorithm (Nigam et al., 2019, 2021). The SELFIES representation allows for arbitrary string manipulations on SELFIES that still generate chemically feasible candidates, making the fragment-based representation obsolete.

Combining property prediction with experimental data

The presented framework uses predictive methods for property estimation, i.e. quantum chemistry and machine learning. These purely predictive methods can be applied when no experimental data is available. However, the accuracy of property prediction and modelling can substantially improve by integrating experimental data, e.g. from optimally designed experiments. Future work could benefit from directly integrating experimental data into the design framework. Closing the loop from performing experiments for validation to improving the predictive methods will enhance the accuracy and reliability of the molecular and process design.

For this purpose, property prediction methods need to be developed that allow incorporating experimental data. In the literature, the COSMO-UNIFAC method was proposed (Dong et al., 2018; Zhu et al., 2020). Here, the UNIFAC method, which relies on parameters fitted to experimental data, is used when available. If no parameters for the UNIFAC model are known, properties are estimated by COSMO-RS. Any available experimental data could thus directly be integrated into the molecular design.

Machine learning models can also easily include new experimental data. Recently, first models for the prediction of thermodynamic mixture properties, e.g. activity coefficients at infinite dilution, were developed (Jirasek et al., 2020; Sanchez Medina et al., 2022). Jirasek et al. (2020) estimated thermodynamic properties using a data-driven matrix completion method, and Sanchez Medina et al. (2022) successfully used a Graph Neural Network. Hybrid approaches based on QM and ML could be promising since QM guarantees a prediction even outside the applicability domain of the ML model. For example, COSMO-RS could be extended by a correction term fitted from experimental data. Winter et al. (2022) and Winter et al. (2023) have successfully shown how to use COSMO-RS as a basis for the ML model that is refined with experimental data.

In the future, the experimental data could be provided by automated experiments. Robotic platforms have been presented for automated experimentation (Steiner et al., 2019; Coley et al., 2019; Burger et al., 2020). The experiments can be planned using optimal experimental design, as for example presented in this thesis, possibly in combination with self-learning algorithms (Schweidtmann et al., 2018; Häse et al., 2019; Clayton et al., 2020). An integrated workflow that refines property models would significantly contribute to achieving higher prediction accuracy and validated design.

Designing the life cycle

In this thesis, CAMPD and CAPD are advanced towards the integrated design of molecules and their applications. For the design of solvents, the life cycle environmental impacts from cradle to grave are considered by estimating the environmental impacts of production from the molecular structure and assuming disposal through aggregated process models, e.g. for wastewater treatment. However, as a consequence of the aggregated modelling of production and use phase, the design scope besides the molecular structure is limited to the application, e.g. for solvents to the chemical process. Ultimately, the design scope should also include models for production and fate of the chemicals to make life cycle-optimal decisions (Bakshi, 2019). For example, cross-process life cycle benefits can be achieved by easy re- or upcycling, benign disposal, or production that does not overachieve the requirements of the application (Anastas and Zimmerman, 2003; Zimmerman et al., 2020). In particular, this approach would integrate waste as a resource for further valorisation (Clark, 2017; Zimmerman et al., 2020).

The idea of designing the entire life cycle has been conceptualised in process systems engineering as life cycle optimisation (Guillén-Gosálbez et al., 2019; Kleinekorte et al.,

2020). Zhang et al. (2020b) presented a life cycle design approach by sequentially optimising each life cycle stage for product development. For the design of fuels and their production processes, König et al. (2020b, 2021) solved an integrated optimisation problem considering fuel feasibility requirements and preselected reaction routes for minimum cost and global warming impact. Thus, promising steps are taken towards life cycle design that should be combined with the methods presented and used in this thesis. With the increasing capabilities to predict reaction pathways and to calculate chemical production processes based on predictive methods, integrated life cycle design is the next frontier for process and product design in chemical engineering.

The presented ideas for future research aim to increase the reliability, ease of use, and applicability of the methods for implementation in academic and industrial practice. Undoubtedly, the methods have not yet matured into automated all-purpose tools for every problem in chemical engineering. However, besides technical advances, reducing the environmental impacts of the chemical industry also requires the transformation of the practitioners' mindset to inherently aim for sustainability. Now that the appropriate tools become available, practitioners need to apply these methods embracing life cycle thinking.

Appendices

APPENDIX A

Supporting information to Chapter 3

A.1 Soft- and hardware

The following software is used for COSMO-susCAMPD:

Calculation	Method	Software version
Optimised geometry and σ -surface	DFT BP86 / TZVP (-MF)	COSMOconf16 (4.1) (COSMOlogic GmbH & Co. KG, 2017a) TURBOMOLE [®] 7.2 (COSMOlogic GmbH & Co. KG, 2017c)
Equilibrium properties	BP-TZVP-C30-1701	COSMOtherm17 (C30-1701) (COSMOlogic GmbH & Co. KG, 2017b)
Numerical optimisation	fminbnd	MATLAB [®] R2018b (The MathWorks Inc., 2018b)
Artificial neural network		MATLAB [®] R2018b (The MathWorks Inc., 2018c)
Training data for the ANN	GaBi Database 8007	GaBi 10.6.0.110 (Thinkstep AG, 2017)

The COSMO-susCAMPD method was run on an Intel[®] Xeon[®] E5-1660v3 workstation utilising 32 GB RAM and 8 processor cores with 2 threads each in parallel.

A.2 Molecular descriptors selection for the artificial neural network

In total, 41 molecular properties are calculated from the molecular structure and COSMO-RS as descriptors for each solvent molecule. The 7 molecular descriptors that show the highest correlation with the environmental impacts are selected as features for the ANN using linear stepwise regression as a feature selection method.

- #Oxygens: Number of oxygen atoms
- #Nitrogens: Number of nitrogen atoms
- #Carbons: Number of carbon atoms
- #Halogens: Number of halogen atoms
- #C=C: Number of carbon-carbon double bonds
- #aromatic Rings: Number of aromatic rings
- #Rings: Number of rings
- MW: Molecular weight
- p_{sat} : Saturation pressure at $t = 25\text{ }^{\circ}\text{C}$
- Δh_{vap} : Enthalpy of vaporisation at $t = 25\text{ }^{\circ}\text{C}$
- x^{LLE} : Molar fractions of the liquid-liquid-equilibrium between solvent and water for both the aqueous and organic phase resulting in four descriptors $x_{\text{aq,water}}^{\text{LLE}}$, $x_{\text{aq,solvent}}^{\text{LLE}}$, $x_{\text{org,water}}^{\text{LLE}}$, and $x_{\text{org,solvent}}^{\text{LLE}}$
- $\log(P_{\text{OW}})$: Logarithm of the partition coefficient of the solvent molecule between octanol and water
- T_{boil} : Boiling temperature at $p = 1\text{ bar}$
- Δh^{f} : Standard enthalpy of formation
- E_{COSMO} : COSMO-energy
- σ -Descr. i : 8 σ -descriptors of the σ -Profile from COSMO-RS as adapted from Zhou et al. (2014)
- σ -Mom. i : 7 σ -moments of the σ -Profile from COSMO-RS (Klamt et al., 2010)
- HB_{acc} -Mom. i : 4 hydrogen bond acceptor moments of the σ -Profile from COSMO-RS (Klamt et al., 2010)
- HB_{don} -Mom. i : 4 hydrogen bond donor moments of the σ -Profile from COSMO-RS (Klamt et al., 2010)

Table A.1: Molecular descriptors selected for the artificial neural network

	ALO	CC excl. bio.	CC incl. bio.	FD	FE	FET	HT	IR	ME
#Oxygens		x	x	x					x
#Nitrogens		x	x	x				x	x
#Carbons									
#Halogens	x				x	x	x	x	
#C=C									
#arom. Rings							x		
#Rings									
MW									
p_{sat}		x	x						
Δh_{vap}		x	x		x	x	x		
$x_{\text{aq,water}}^{\text{LLE}}$									
$x_{\text{aq,solvent}}^{\text{LLE}}$	x	x	x	x					
$x_{\text{org,water}}^{\text{LLE}}$	x	x	x						
$x_{\text{org,solvent}}^{\text{LLE}}$								x	
$\log(P_{\text{OW}})$									
T_{boil}									
Δh^{f}				x			x		
E_{COSMO}						x			
σ -Descr. i	2,3			4,7	4,8	3,8	8	6,8	1,4,7
σ -Mom. i		4	4	6					6,7
HB_{acc} -Mom. i	1,2				2,4	1,4	1,4	1,3	
HB_{don} -Mom. i					2				

Table A.2: Molecular descriptors selected for the artificial neural network (continued)

	MET	MRD	OD	PMF	POF	TA	TET	WD
#Oxygens	x			x	x			
#Nitrogens			x		x			
#Carbons	x							
#Halogens		x	x				x	x
#C=C		x						
#arom. Rings	x	x		x				
#Rings	x							
MW						x		
p_{sat}	x				x			
Δh_{vap}						x	x	x
$x_{\text{aq,water}}^{\text{LLE}}$								
$x_{\text{aq,solvent}}^{\text{LLE}}$				x	x	x		x
$x_{\text{org,water}}^{\text{LLE}}$				x		x		x
$x_{\text{org,solvent}}^{\text{LLE}}$			x					
$\log(P_{\text{OW}})$								
T_{boil}			x	x		x		
Δh^{f}								
E_{COSMO}				x		x		x
σ -Descr. i	3,4	4,5,6	6,8		8		4,8	3
σ -Mom. i				4	4,7	1	6	
HB_{acc} -Mom. i			1				1,4	2
HB_{don} -Mom. i		3						

A.3 Accuracy of the artificial neural network predictions for all impact categories

Table A.3: Accuracy of the neural network predictions for all impact categories. The accuracy is measured by the coefficient of determination (R^2) and the root-mean-square error normalised by the range of values (nRMSE).

Impact category	Training Set		Validation Set		Test Set	
	R^2	nRMSE	R^2	nRMSE	R^2	nRMSE
Agricultural Land Occupation (ALO)	0.47	14%	0.5	14%	0.46	11%
Climate Change (CC) excl. biogenic carbon	0.58	14%	0.51	19%	0.79	7%
Climate Change (CC) incl. biogenic carbon	0.44	17%	0.56	14%	0.51	9%
Fossil Depletion (FD)	0.39	17%	0.57	16%	0.18	12%
Freshwater Eutrophication (FE)	0.7	9%	0.62	20%	0.24	5%
Freshwater Ecotoxicity (FET)	0.3	16%	0.48	9%	0.41	9%
Human Toxicity (HT)	0.18	16%	0.38	15%	0.1	13%
Ionising Radiation (IR)	0.35	18%	0.62	10%	0.5	18%
Marine Eutrophication (ME)	0.97	2%	0.64	2%	0.71	2%
Marine Ecotoxicity (MET)	0.8	8%	0.5	15%	0.85	6%
Metal Resource Depletion (MRD)	0.47	9%	0.53	10%	0.33	5%
Ozone Depletion (OD)	0.81	8%	0.76	16%	0.08	15%
Particulate Matter Formation (PMF)	0.67	11%	0.7	18%	0.62	6%
Photochemical Oxidant Formation (POF)	0.77	11%	0.36	16%	0.15	15%
Terrestrial Acidification (TA)	0.46	14%	0.59	13%	0.5	12%
Terrestrial Ecotoxicity (TET)	0.85	6%	0.52	9%	0.46	9%
Water Depletion (WD)	0.52	13%	0.76	13%	0.4	13%
average	0.57	12%	0.56	14%	0.43	10%

Parity plots of predicted versus database value for all impact categories except for Climate Change and Ozone Depletion

The parity plots for Climate Change and Ozone Depletion are shown in Section 3.1.2 as Figure 3.4 A & B.

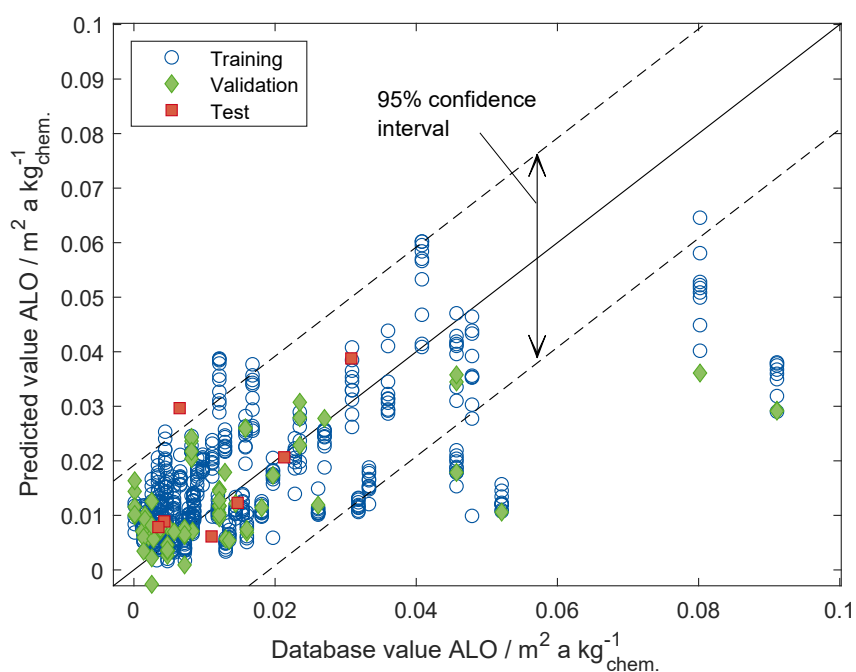


Figure A.1: Parity plot of predicted versus database Agricultural Land Occupation (ALO)

A.3 Accuracy of the artificial neural network predictions for all impact categories

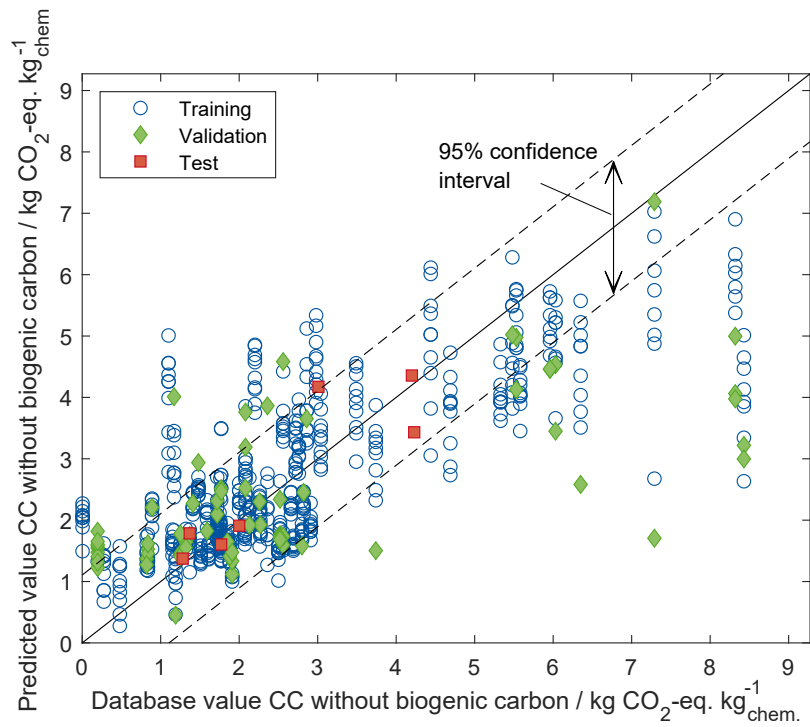


Figure A.2: Parity plot of predicted versus database Climate Change (CC) without biogenic carbon

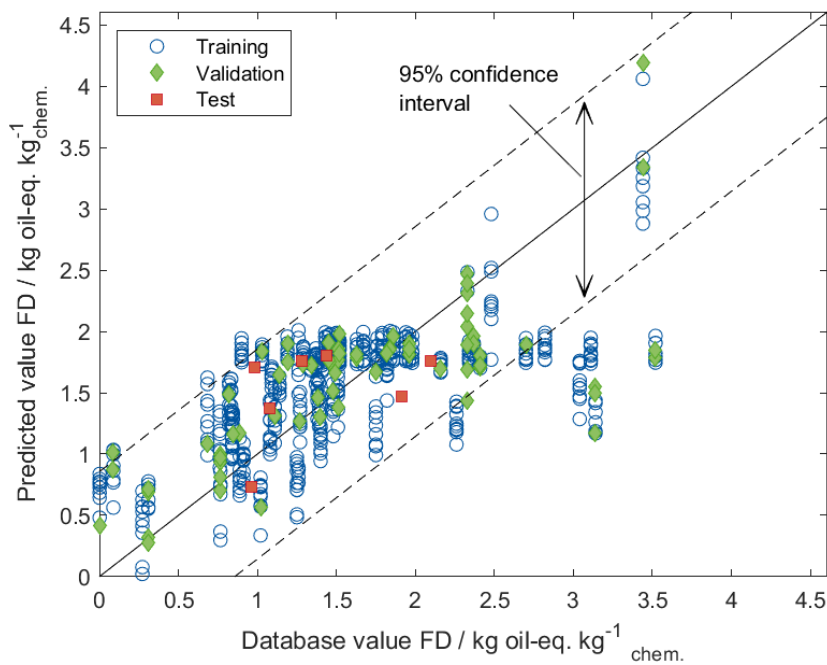


Figure A.3: Parity plot of predicted versus database Fossil Depletion (FD)

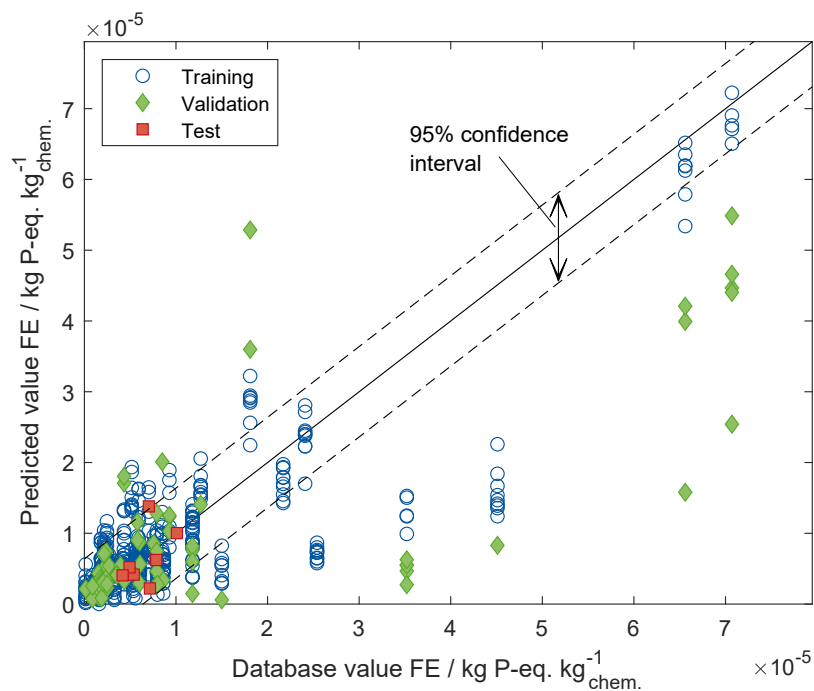


Figure A.4: Parity plot of predicted versus database Freshwater Eutrophication (FE)

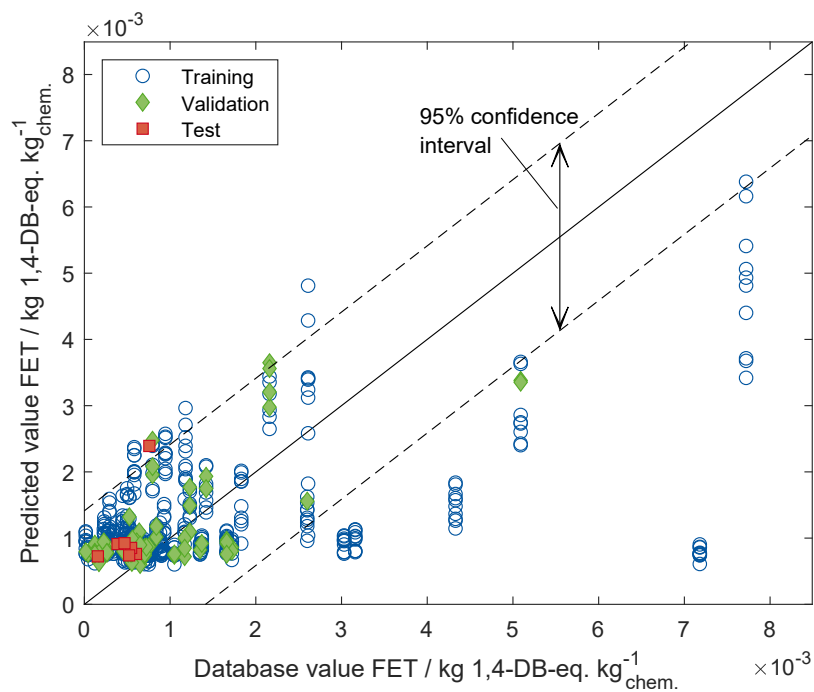


Figure A.5: Parity plot of predicted versus database Freshwater Ecotoxicity (FET)

A.3 Accuracy of the artificial neural network predictions for all impact categories

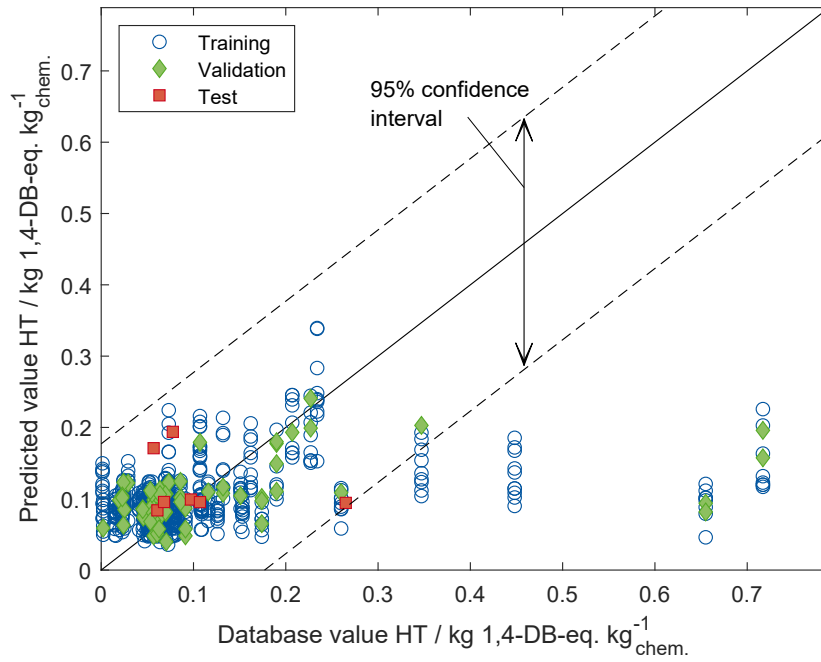


Figure A.6: Parity plot of predicted versus database Human Toxicity (HT)

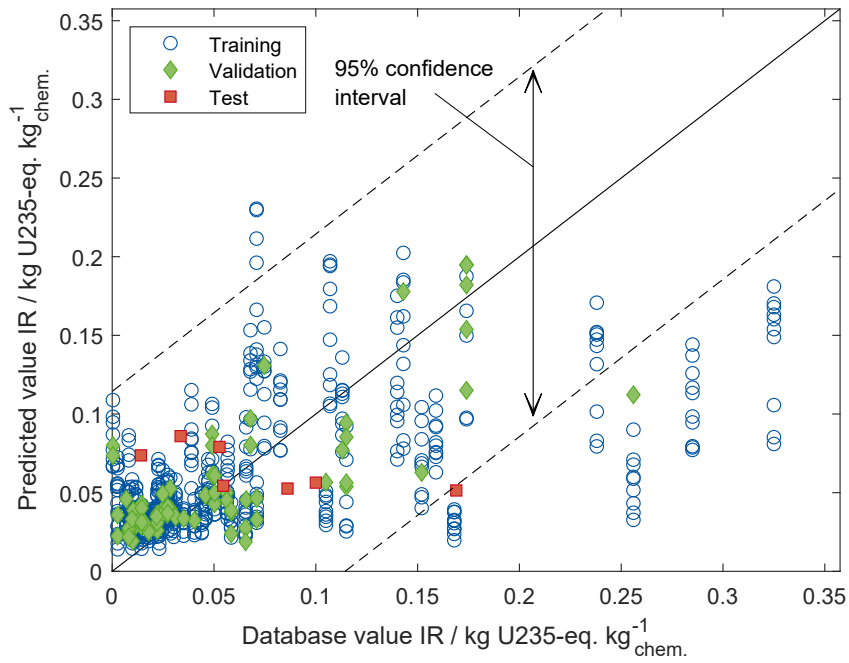


Figure A.7: Parity plot of predicted versus database Ionising Radiation (IR)

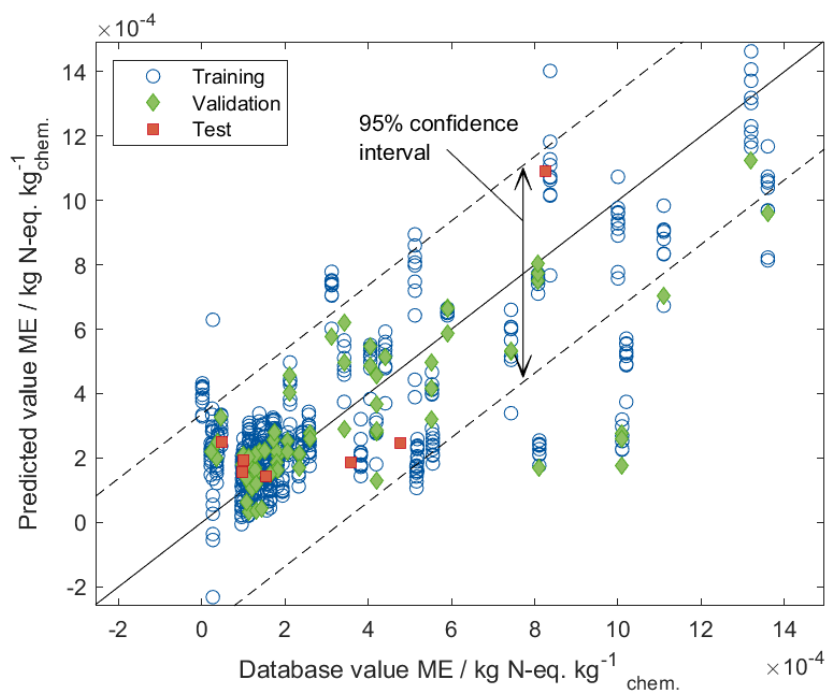


Figure A.8: Parity plot of predicted versus database Marine Eutrophication (ME)

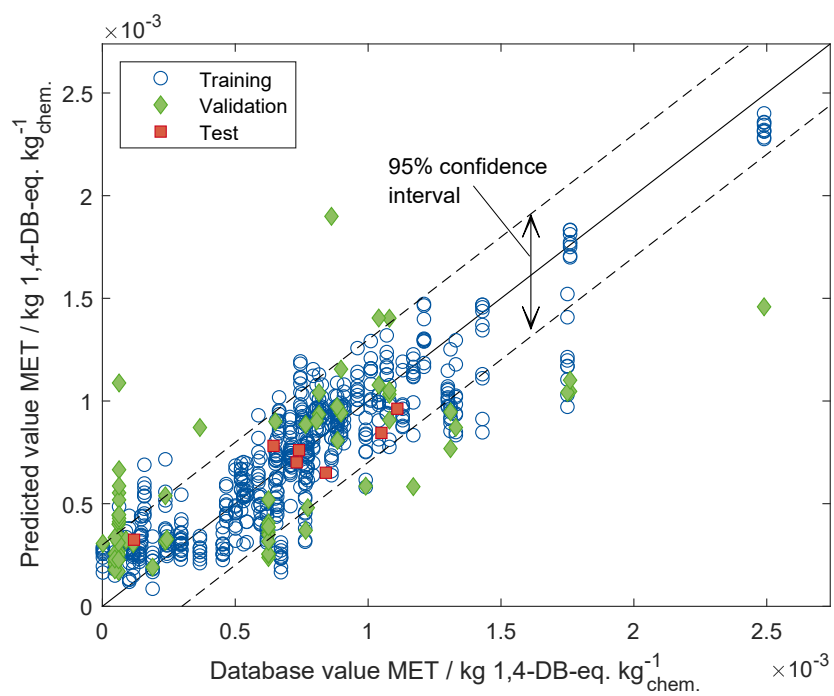


Figure A.9: Parity plot of predicted versus database Marine Ecotoxicity (MET)

A.3 Accuracy of the artificial neural network predictions for all impact categories

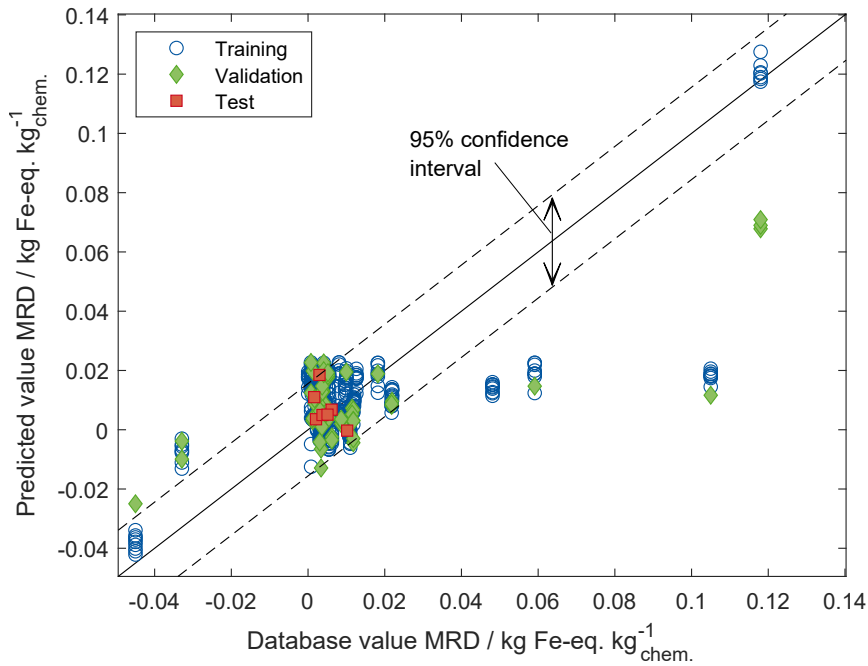


Figure A.10: Parity plot of predicted versus database Metal Resource Depletion (MRD)

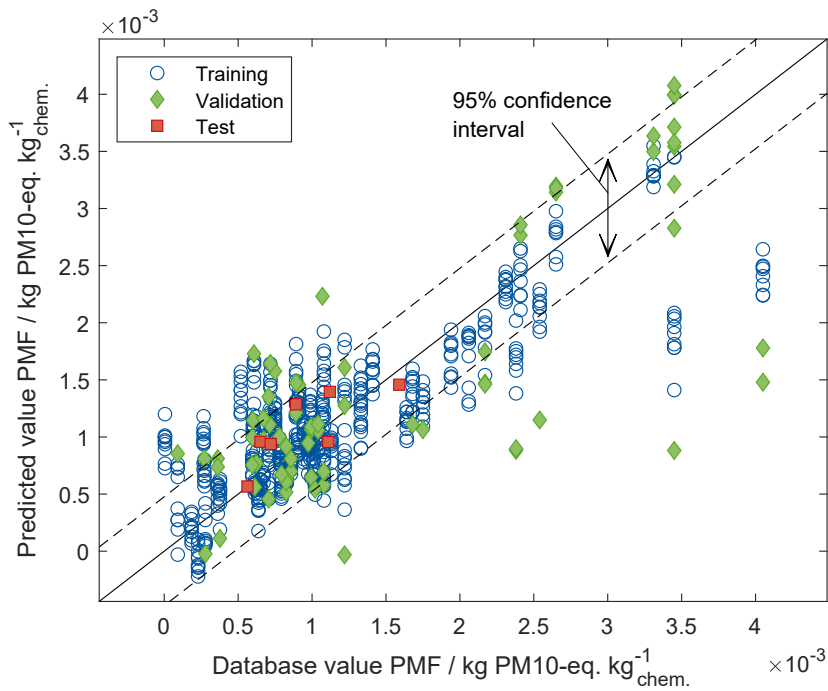


Figure A.11: Parity plot of predicted versus database Particulate Matter Formation (PMF)

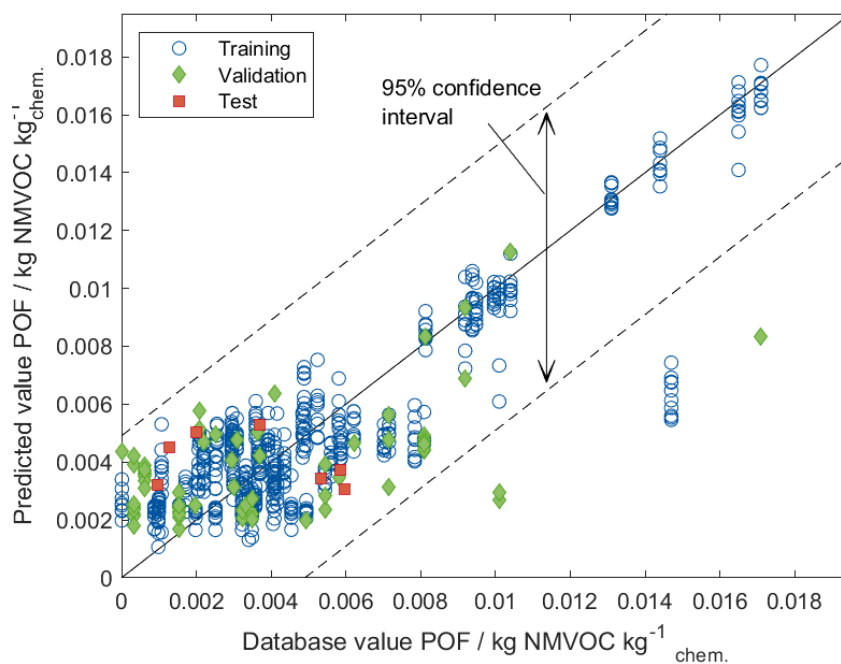


Figure A.12: Parity plot of predicted versus database Photochemical Oxidant Formation (POF)

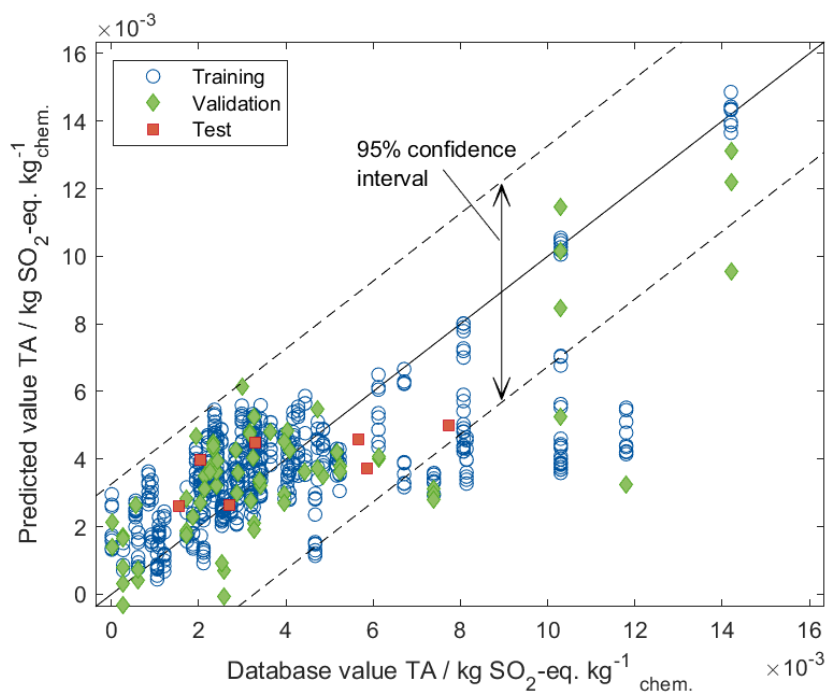


Figure A.13: Parity plot of predicted versus database Terrestrial Acidification (TA)

A.3 Accuracy of the artificial neural network predictions for all impact categories

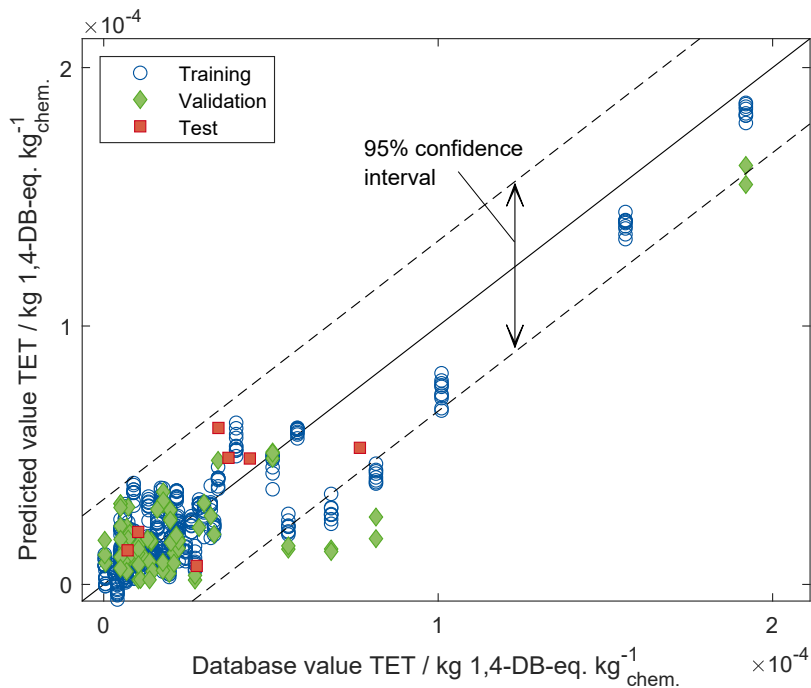


Figure A.14: Parity plot of predicted versus database Terrestrial Ecotoxicity (TET)

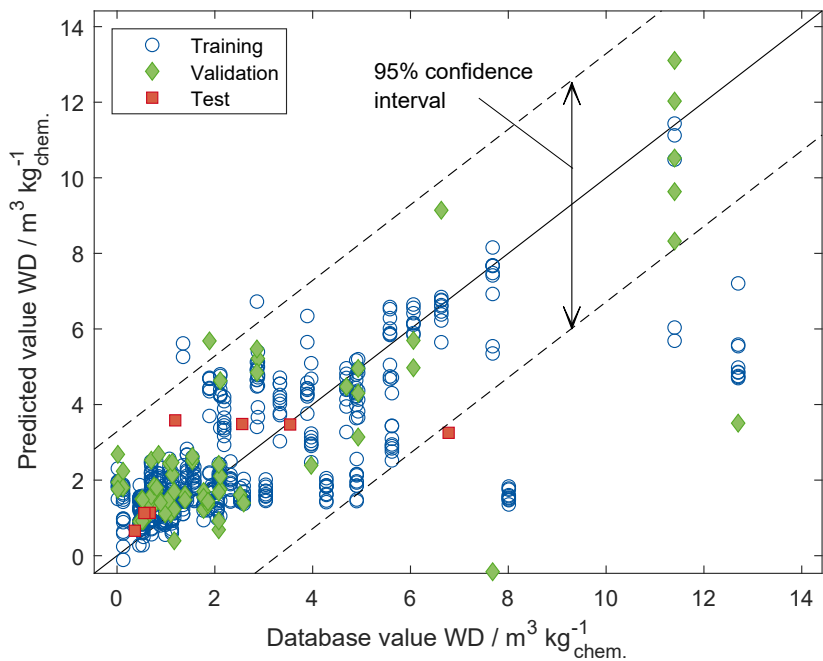


Figure A.15: Parity plot of predicted versus database Water Depletion (WD)

A.4 Molecular fragment library for the genetic algorithm LEA3D

The following 3D-molecular fragments have been specified as the molecular fragment library for the genetic algorithm LEA3D (Douguet et al., 2005). Each “X” marks a connector of the molecular fragment and can be connected to another “X” from another molecular fragment to build a new molecular structure. Unconnected connectors “X” are automatically replaced by hydrogen atoms.

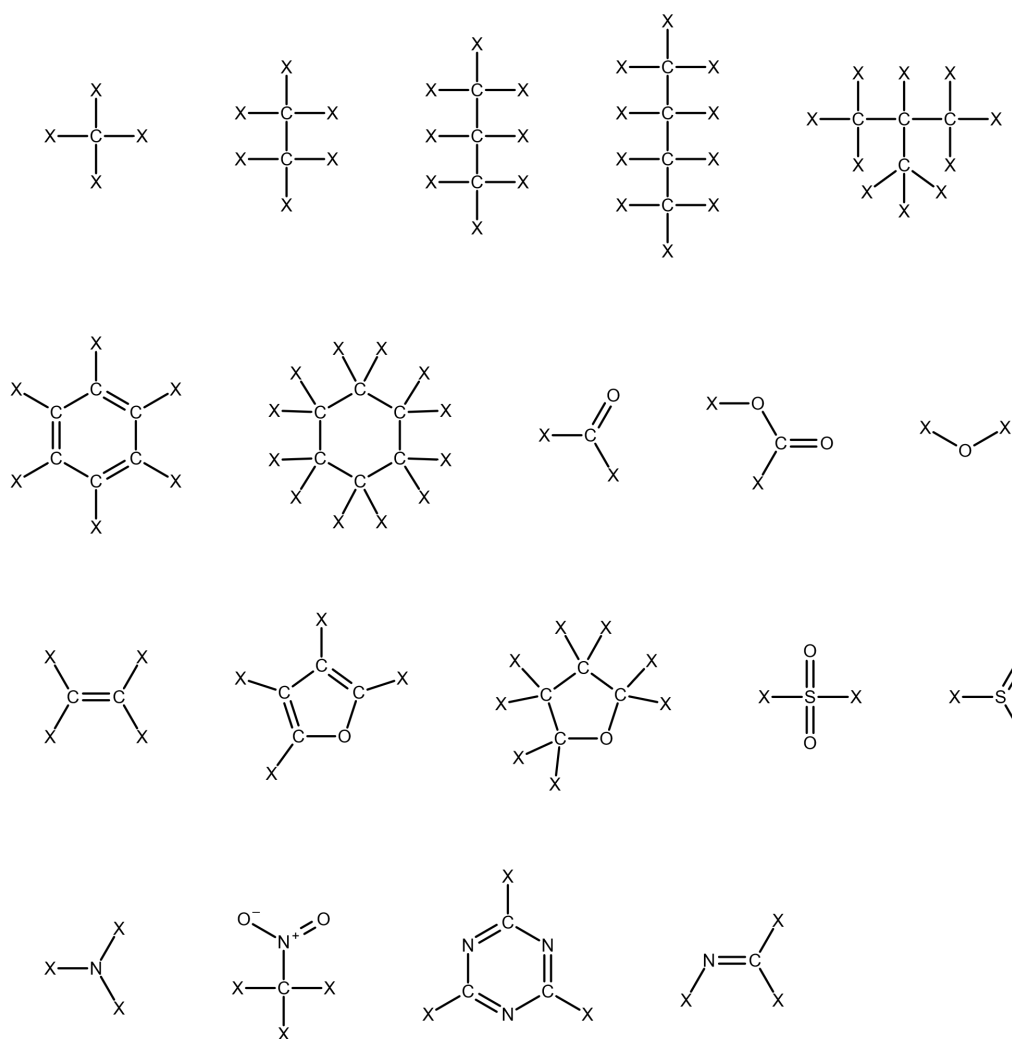


Figure A.16: Molecular fragment library used for the hybrid extraction-distillation of γ -valerolactone.

APPENDIX B

Supporting information to Chapter 4

B.1 Soft- and hardware

The following software is used within COSMO-CAMPD:

Calculation	Method	Software version
Optimised geometry and σ -surface	DFT BP86 / TZVP (-MF)	COSMOconf16 (4.1) (COSMOlogic GmbH & Co. KG, 2017a) TURBOMOLE [®] 7.2 (COSMOlogic GmbH & Co. KG, 2017c)
Equilibrium properties	BP-TZVP-C30-1701	COSMOtherm17 (C30-1701) (COSMOlogic GmbH & Co. KG, 2017b)
Optimised geometry and vibrational freq. Thermochemistry	DFT B3LYP / TZVP RRHO	Gaussian 09.d01 (Frisch et al., 2009) TAMKin 1.2.4 (Ghysels et al., 2010)
Numerical optimisation	fmincon with multistart	MATLAB [®] R2018b (The MathWorks Inc., 2018b)

All calculations were performed on an Intel[®] Xeon[®] E5-1660v3 workstation utilising 32 GB RAM and 8 processor cores with 2 threads each in parallel.

B.2 Accuracy of predicted enthalpies of vaporisation and ideal-gas heat capacities for selected molecules

B.2.1 Enthalpy of vaporisation

Experimental enthalpies of vaporisation are taken for 98 molecules from Chickos and Acree (2003).

Of the 98 considered molecules, 79 (81 %) are within $\pm 10\%$ of the literature values. The mean absolute error (MAE) in the enthalpies of vaporisation between the literature values and the predictions from COSMO-RS is 2.83 kJ mol^{-1} , and the coefficient of determination equals $R^2 = 0.93$.

The molecule with the largest deviation between predicted and literature enthalpy of vaporisation is nitric acid, for which the literature value is 38.6 kJ mol^{-1} and COSMO-RS predicts 18.5 kJ mol^{-1} , followed by formic acid (lit. 36 kJ mol^{-1} , COSMO-RS 50.7 kJ mol^{-1}), methanol (lit. 37.4 kJ mol^{-1} , COSMO-RS 49.1 kJ mol^{-1}), and acetic acid (lit. 43 kJ mol^{-1} , COSMO-RS 53.8 kJ mol^{-1}). Strong polar molecules, e.g. alcohols or acids, are known to have large uncertainties in the predicted enthalpies of vaporisation using COSMO-based models (Lin et al., 2004). In the case studies of this thesis, strong polar molecules such as alcohols and acids are not feasible as solvents due to their miscibility with water. However, the strong deviations between experimental and predicted values highlight the need to investigate the predicted properties of promising solvent candidates including the chosen COSMO-RS parametrization.

The closest predictions are achieved for n-hexane (lit. 31.5 kJ mol^{-1} , COSMO-RS 31.5 kJ mol^{-1}), toluene (lit. 35.7 kJ mol^{-1} , COSMO-RS 35.8 kJ mol^{-1}), and n-butene (lit. 23.3 kJ mol^{-1} , COSMO-RS 23.1 kJ mol^{-1}).

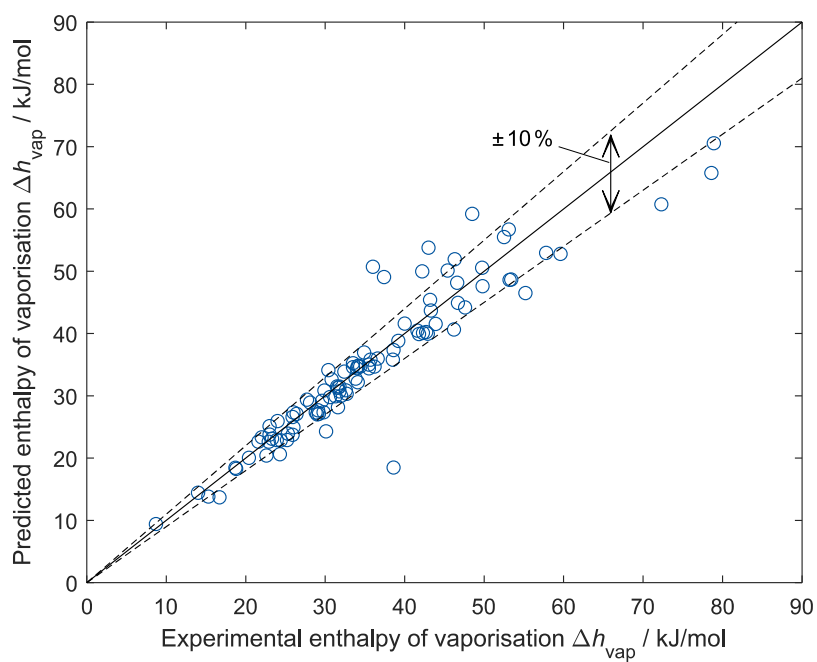


Figure B.1: Comparison of experimental enthalpies of vaporisation with predictions from COSMO-RS on TZVP-MF level.

B.2.2 Ideal-gas heat capacities

Experimental ideal-gas heat capacities are taken for 145 molecules from the NIST Chemistry WebBook (2021).

Of the 145 considered molecules, 140 (97%) are within $\pm 10\%$ of the literature values from the NIST database. The mean absolute error (MAE) in the enthalpies of vaporisation between the literature values and the predictions from quantum chemistry is $3.93 \text{ J mol}^{-1} \text{ K}^{-1}$, and the coefficient of determination equals $R^2 = 0.88$.

In contrast to the enthalpies of vaporisation, no outliers are identified: The molecule with the largest deviation between predicted and literature ideal-gas heat capacity is acetic anhydride, for which the literature value is $100 \text{ J mol}^{-1} \text{ K}^{-1}$ and quantum chemistry predicts $119 \text{ J mol}^{-1} \text{ K}^{-1}$, followed by benzoic acid (lit. $104 \text{ J mol}^{-1} \text{ K}^{-1}$, QM $124 \text{ J mol}^{-1} \text{ K}^{-1}$), and diethyl ether (lit. $120 \text{ J mol}^{-1} \text{ K}^{-1}$, QM $104 \text{ J mol}^{-1} \text{ K}^{-1}$).

Nearly perfect predictions are achieved, e.g. for methylcyclopentane (lit. $110 \text{ J mol}^{-1} \text{ K}^{-1}$, QM $110 \text{ J mol}^{-1} \text{ K}^{-1}$), trioxane (lit. $82.4 \text{ J mol}^{-1} \text{ K}^{-1}$, QM $82.6 \text{ J mol}^{-1} \text{ K}^{-1}$), and acetaldehyde (lit. $55.5 \text{ J mol}^{-1} \text{ K}^{-1}$, QM $55.3 \text{ J mol}^{-1} \text{ K}^{-1}$).

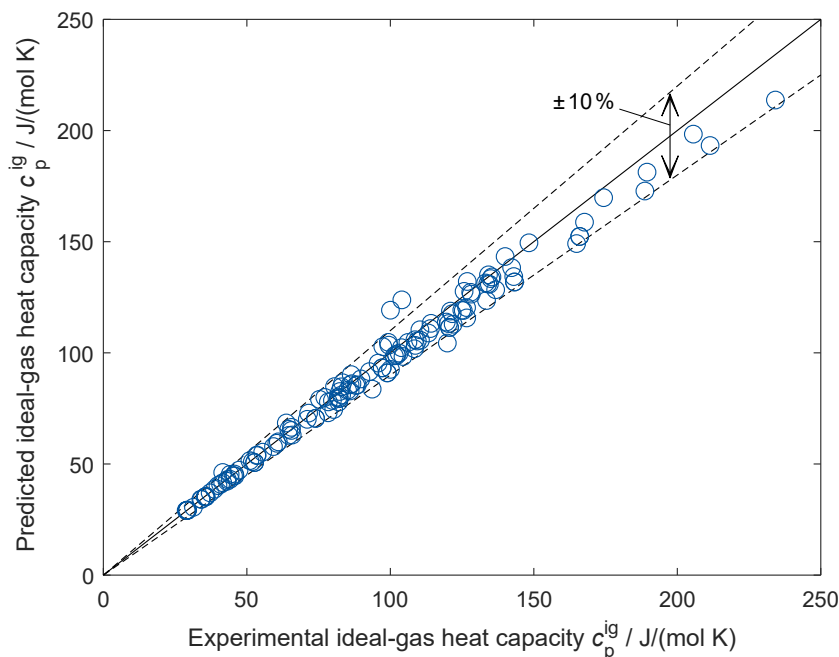


Figure B.2: Comparison of experimental ideal-gas heat capacities with predictions using B3LYP/TZVP with RRHO approximation.

B.3 Accuracy of the pinch-based process model for absorption and the multiphase reactor model

B.3.1 Absorption column model

Comparison between minimum solvent demand S_{\min} calculated by the pinch-based process model (Redepenning et al., 2017) and minimum solvent demand from rigorous simulations in ASPEN Plus (S_{rig}) using 20 stages. The plot shows the results for 1761 solvents that were successfully completed by both models. Both models use the same thermodynamic properties calculated by COSMO-RS.

Overall, the solvent demand calculated by the pinch-based process model agrees well with the results from rigorous simulation in ASPEN Plus, as evident by $MAE = 0.14 \text{ mol mol}_{\text{Feed}}^{-1}$ and $R^2 = 0.95$.

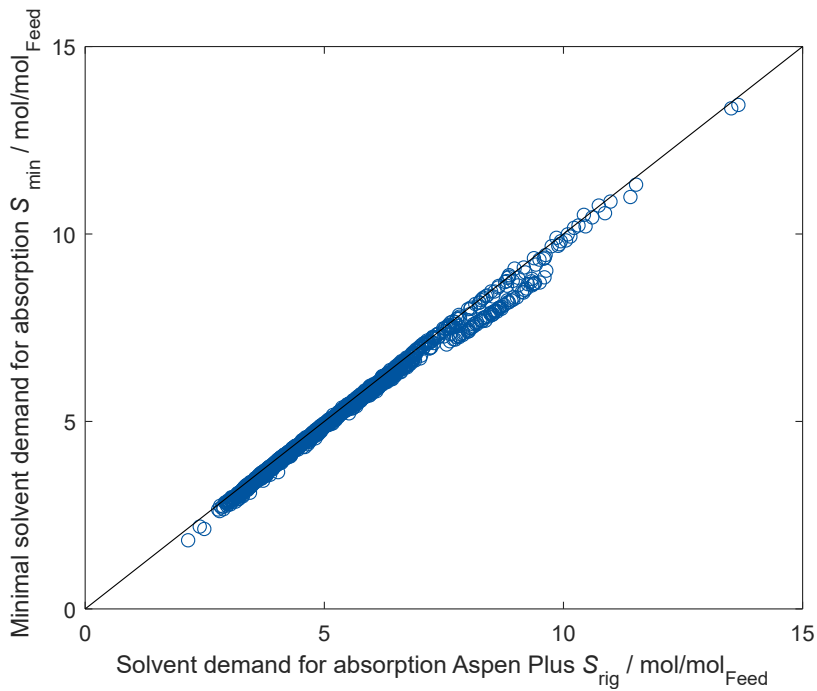


Figure B.3: Comparison of minimal solvent demand calculated by the pinch-based process model with rigorous simulations in ASPEN Plus.

B.3.2 Multiphase reactor model

Comparison of equilibrium conversion $\xi_{\text{eq}}^{\text{SC/rig}}$ and dimethylformamide mole fraction $x_{\text{DMF}}^{\text{SC/rig}}$ of a vapour-liquid-liquid equilibrium reactor calculated by the multiphase reactor based on the homotopy continuation algorithm (Bausa and Marquardt, 2000) and simulations in Aspen Plus. The plots show the results for 2652 solvents that were successfully completed by both models. Both models use the same thermodynamic properties calculated by COSMO-RS.

The results calculated by the pinch-based process model agree very well with the results from rigorous simulation in ASPEN Plus, as evident by $MAE = 2 \times 10^{-5}$ and $R^2 = 1.00$ for the equilibrium conversion and $MAE = 6 \times 10^{-6}$ and $R^2 = 1.00$ for the product mole fraction in the reactor outlet.

Only for one solvent the solutions differ more significantly, i.e. by 0.01 in equilibrium conversion and product mole fraction, potentially due to numerics.

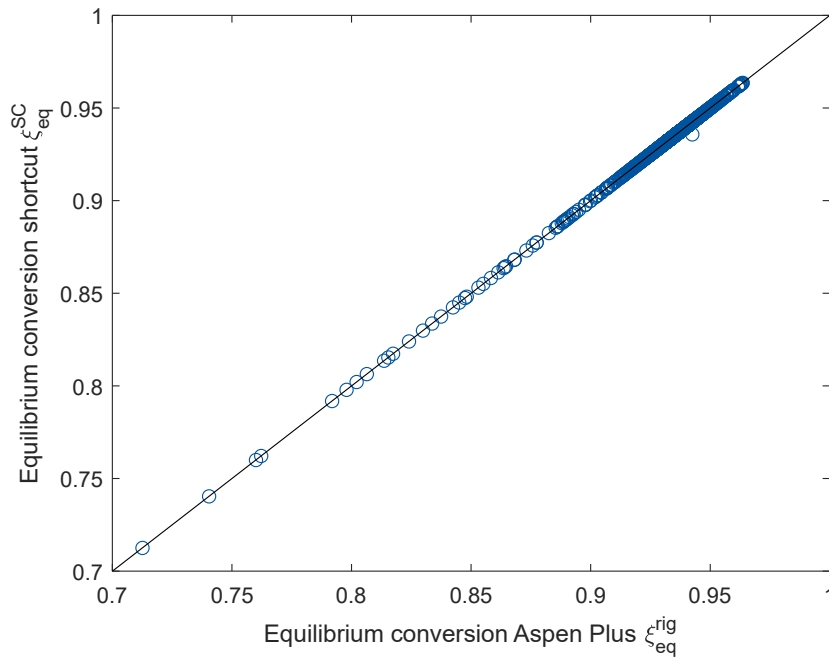


Figure B.4: Comparison of equilibrium conversion ξ_{eq} calculated by the vapour-liquid-liquid equilibrium reactor model used in this work and simulations in ASPEN Plus.

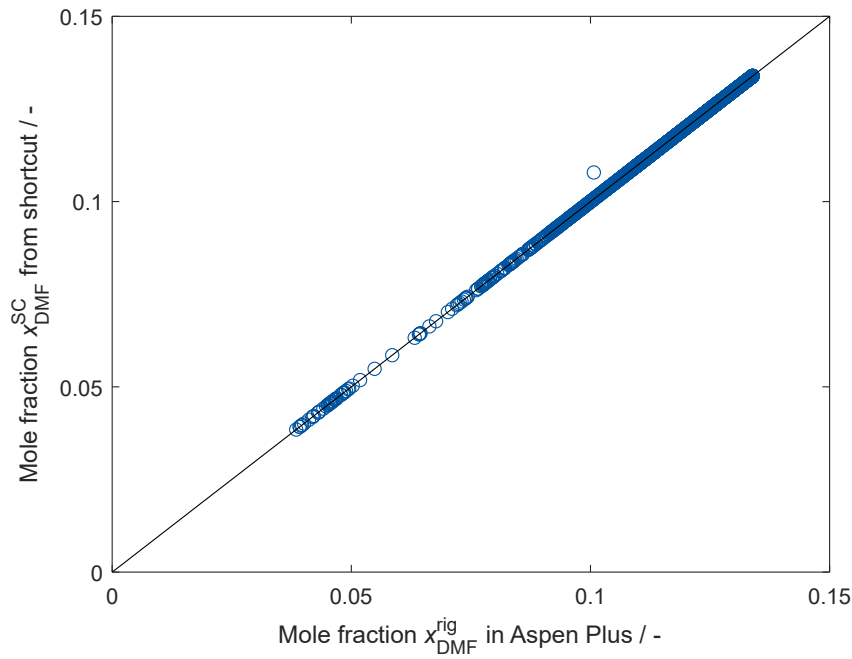
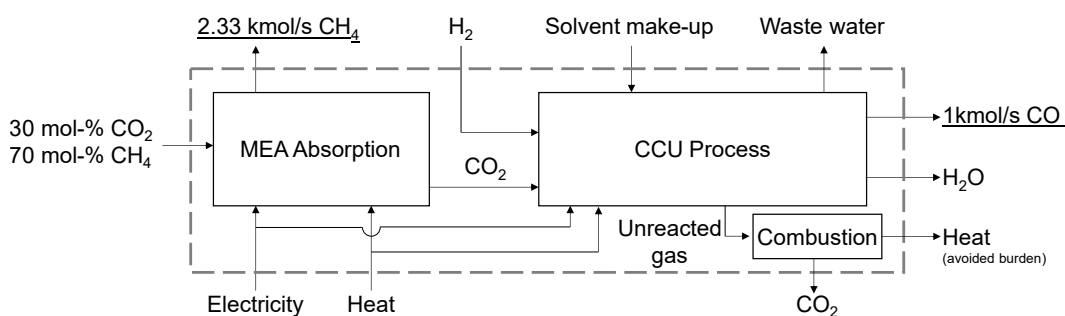


Figure B.5: Comparison of product mole fraction x_{DMF} calculated by the vapour-liquid-liquid equilibrium reactor model used in this work and simulations in ASPEN Plus.

A comparison between minimum energy demand calculated by the pinch-based process models and minimum energy demand from rigorous simulations for a hybrid extraction-distillation process has been presented by Scheffczyk et al. (2018). The pinch-based process models for extraction and distillation were found to be sufficiently accurate for solvent design compared to rigorous ASPEN Plus simulations ($\rho_{rank} = 0.87$), and agreed particularly for promising solvents with low energy demand for hybrid extraction-distillation.

B.4 System boundaries and considered impacts for the LCA of the ICCU process

1. Benchmark process



2. ICCU process

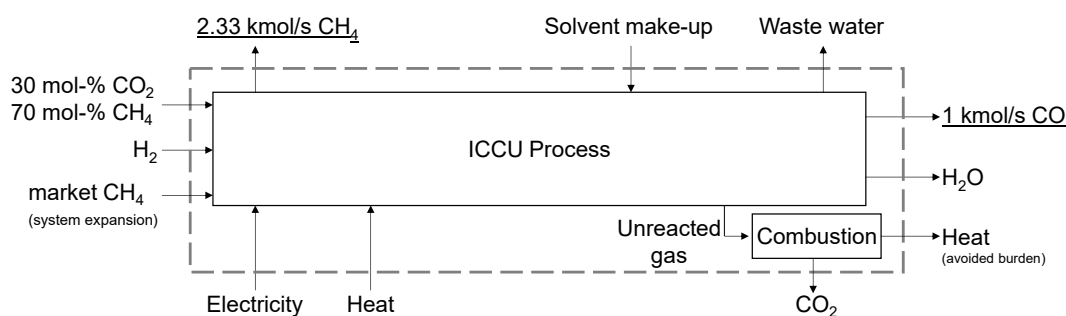


Figure B.6: System boundaries and considered impacts for the LCA of the ICCU process and the benchmark CCU process. The grey dashed box represents the system boundary. The underlined quantities are the functional unit of the LCA. As the avoided burden for heat, heat from low-pressure steam is assumed.

B.4 System boundaries and considered impacts for the LCA of the ICCU process

The impacts are considered according to the following references:

Quantity	Reference
Feedstocks and material flows	
Hydrogen production	DE: Hydrogen (steam reforming heavy fuel oil) ts (Thinkstep AG, 2017)
Methane production	DE: Methane (natural gas purification) ts (Thinkstep AG, 2017)
Natural gas production	Drying, natural gas (Emmenegger, 2000)
Solvent production	constant value: $3 \text{ kg CO}_2\text{-eq. kg}_{\text{solvent}}^{-1}$
Wastewater disposal	treatment of spent solvent mixture (Ruiz, 2019)
Energy flows	
Electricity	Market group for electricity, medium voltage (ecoinvent, 2019b)
Furnace heat	Heavy fuel oil, burned in refinery furnace (Jungbluth, 2019)
Low-pressure steam	DE: Thermal energy from natural gas ts (Thinkstep AG, 2017)
Refrigeration	Market for cooling energy (ecoinvent, 2019a)
Cooling water	$0 \text{ kg CO}_2\text{-eq. MJ}^{-1}$
Conventional production of FU	
Carbon monoxide production	Carbon monoxide (via synthetic gas) ts (Thinkstep AG, 2017)

B.5 Supplementary COSMO-(sus)CAMPD designs for minimum exergy demand

B.5.1 Hybrid extraction-distillation of γ -valerolactone

For this supplementary design, the objective of the integrated solvent and process design is to minimise the total exergy demand (E_{tot}) after heat integration by choosing an optimal solvent with the corresponding optimal process settings. Similar to Section 4.2.1, the degrees of freedom of the process are the extraction and decanter temperatures T_{Extr} and T_{Dec} and the pressure in the distillation column p_{Dist} .

Qualitatively, the same conclusions can be drawn for the design for minimum exergy demand as for the design for minimum environmental impact on Climate Change in Section 4.2.1 (cf. Figure B.7): The heuristic solvent selection criterion of minimum solvent demand for extraction S_{min} correlates well with the total exergy demand of the process E_{tot} (Pearson's Correlation Coefficient $\rho = 0.92$), except for the high-ranking solvents, e.g., ranking higher than the benchmark n-butyl acetate (Pearson's Correlation Coefficient $\rho = 0.07$). Moreover, designing for minimum heat-integrated exergy demand yields only minor changes in solvent selection compared to designing for reboiler exergy demand only (rank correlation coefficient $\rho_{\text{rank}} = 0.98$), similar to the design for minimum environmental impact with and without sensible heat and heat integration in Section 4.2.1.

Noteworthy, the optimal solvent regarding exergy demand of distillation reboiler only is also the solvent minimising total environmental impact on Climate Change, ranking second in total heat-integrated exergy demand (Table B.1). However, the solvent minimising total heat-integrated exergy demand ranks only 27th in total minimum environmental impact on Climate Change in Section 4.2.1, highlighting the important use of a cradle-to-grave environmental objective (cf. Chapter 3).

The optimal solvent with the minimum exergy demand of the heat-integrated process is 3-vinylfuran. 3-vinylfuran leads to a total exergy demand of 43.7 kJ/mol GVL for the extraction-distillation process corresponding to a reduction in the exergy demand by 40% compared to the benchmark n-butyl acetate with a total exergy demand of 72.9 kJ/mol GVL. Besides 3-vinylfuran, the method designs 97 additional candidate solvents with a lower exergy demand than n-butyl acetate highlighting the systematic generation of promising alternatives with higher efficiency than the literature benchmark.

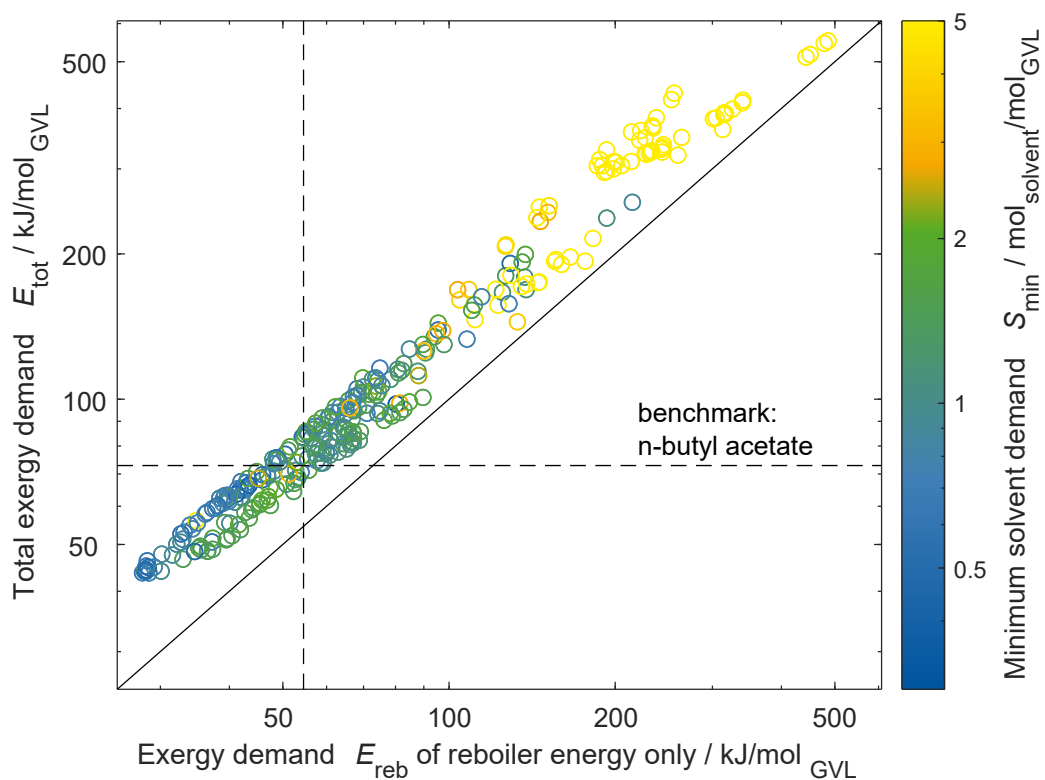
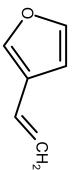
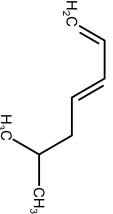
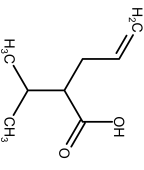
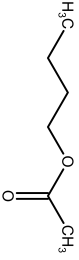


Figure B.7: Comparison of the total exergy demand of the heat-integrated process (E_{tot}) with the exergy demand from reboiler energy neglecting sensible heat (E_{reb}). The colour code indicates the heuristic selection criterion of minimum solvent demand for extraction (S_{min}).

Table B.1: Highest ranking solvents for the hybrid extraction-distillation of γ -valerolactone based on the corresponding objective functions: Minimisation of heat-integrated exergy demand of the entire process (E_{tot}), exergy demand of distillation reboiler neglecting sensible heat (E_{reb}) or solvent demand for extraction (S_{min}). In this case study, the solvent minimising the exergy demand of distillation reboiler only is also the solvent minimising total environmental impact on Climate Change (cf. Chapter 3).

Optimal solvent regarding ...	Molecular structure	Exergy demand		Solvent demand S_{min}			
		process E_{tot} /kJ mol ⁻¹	Rank	reboiler E_{reb} /kJ mol ⁻¹	Rank	mol mol ⁻¹	Rank
total heat-integrated exergy demand		43.7	1	28.6	11	0.036	42
exergy demand of distillation reboiler		43.7	2	27.8	1	0.032	30
solvent demand for extraction		97.7	182	80.2	200	0.017	1
benchmark n-butyl acetate		72.9	98	54.5	109	0.050	100

B.5.2 Integrated carbon capture and utilisation for the production of carbon monoxide

Similar to the design in Section 4.2.2, the process design is tailored for each candidate solvent by optimising the reactor pressure p_{Rx} , the pressures in the distillation columns p_{Dist1} and p_{Dist2} and the molar flow of water to the reactor $\dot{n}_{\text{H}_2\text{O}}$ to ensure phase separation. In contrast to Section 4.2, the optimisation objective is minimising the overall process exergy demand E_{tot} per mole CO normalised by the Gibbs free energy of the overall reaction $\Delta_{\text{R}}G_{\text{CO}}^0$:

$$\min E_{\text{tot}} = \frac{\sum_i \dot{Q}_i \cdot \left(1 - \frac{T_{\text{u}}}{T_i}\right) + P_{\text{comp}}}{\dot{n}_{\text{CO}} \cdot \Delta_{\text{R}}G_{\text{CO}}^0} \quad (\text{B.1})$$

with \dot{Q}_i and T_i representing the heat duties and corresponding temperatures of the utilities, respectively, and P_{comp} representing the power demand for compression of H_2 to reactor pressure p_{Rx} . The overall reaction is the reverse water gas shift reaction, totalling in an overall Gibbs free energy of the reaction of $\Delta_{\text{R}}G_{\text{CO}}^0 = 27.72 \text{ kJ mol}^{-1}$ (Jens et al., 2016).

Qualitatively, the same conclusions can be drawn for the design for minimum exergy demand as for the design for minimum environmental impact on Climate Change in Section 4.2.2: A higher absorption selectivity of the solvent for CO_2 leads to a higher yield in the reactor and thus to a lower total process exergy demand ($\rho = -0.54$; cf. Figure B.8). Similarly, a high equilibrium yield of dimethylformamide in the organic reactor outlet leads to a low exergy demand ($\rho = -0.66$). However, selecting absorption selectivity or equilibrium yield as objective is not sufficient to find the highest ranking candidate solvents in minimum heat-integrated process exergy demand (cf. Table B.2), and a design neglecting heat integration does not yield the same high-ranking candidate solvents as a design considering heat integration for total process exergy demand. The correlation coefficient between solvent rankings with and without consideration of heat integration among the top 50 solvents is only $\rho_{\text{rank}} = 0.40$, indicating a weak correlation between the two rankings. However, considering the total process exergy demand without heat integration still enables differentiation between high- and low-ranking solvents as evident by a correlation coefficient between solvent ranking with and without heat integration of $\rho_{\text{rank}} = 0.93$ (cf. Figure B.9).

Despite the similar trends for exergy demand and environmental impact on Climate Change, the solvent minimising exergy demand (5-fluoro-dimethylpentan-1-amine) is

different from the solvent minimising the heat-integrated environmental impact on Climate Change (2-dimethylamino-ethanethiol; see Table B.2). 5-fluoro-dimethylpentan-1-amine ranks only fourth in heat-integrated environmental impact on Climate Change with an impact 17% higher than the optimal solvent 2-dimethylamino-ethanethiol. The different optimal solvents illustrate the effect of a cradle-to-grave environmental objective compared to an exergy-based assessment, similar to the results of Chapter 3 and Section B.5.1. Thus, choosing an LCA- or exergy-based objective function impacts solvent selection and needs to be carefully addressed.

The optimal solvent for the ICCU process reduces the exergy demand of the process by 38% to 148 kJ/mol CO ($= 5.3 \Delta_{\text{R}}G_{\text{CO}}^0$) compared to the benchmark separated CCU process with an exergy demand for this case study of 238 kJ/mol CO ($= 8.6 \Delta_{\text{R}}G_{\text{CO}}^0$). The solvent minimising the heat-integrated environmental impact on Climate Change yields an exergy demand of 158 kJ/mol CO ($= 5.7 \Delta_{\text{R}}G_{\text{CO}}^0$). Therefore, with an optimal combination of process and solvent, the ICCU process concept is a more efficient alternative to the separated process and can be advantageous both in terms of exergy demand and environmental impact on Climate Change (cf. Section 4.2.2). However, with a minimum exergy demand of $5.3 \Delta_{\text{R}}G_{\text{CO}}^0$, the studied ICCU process is very exergy-intensive even for the best process-solvent-combination.

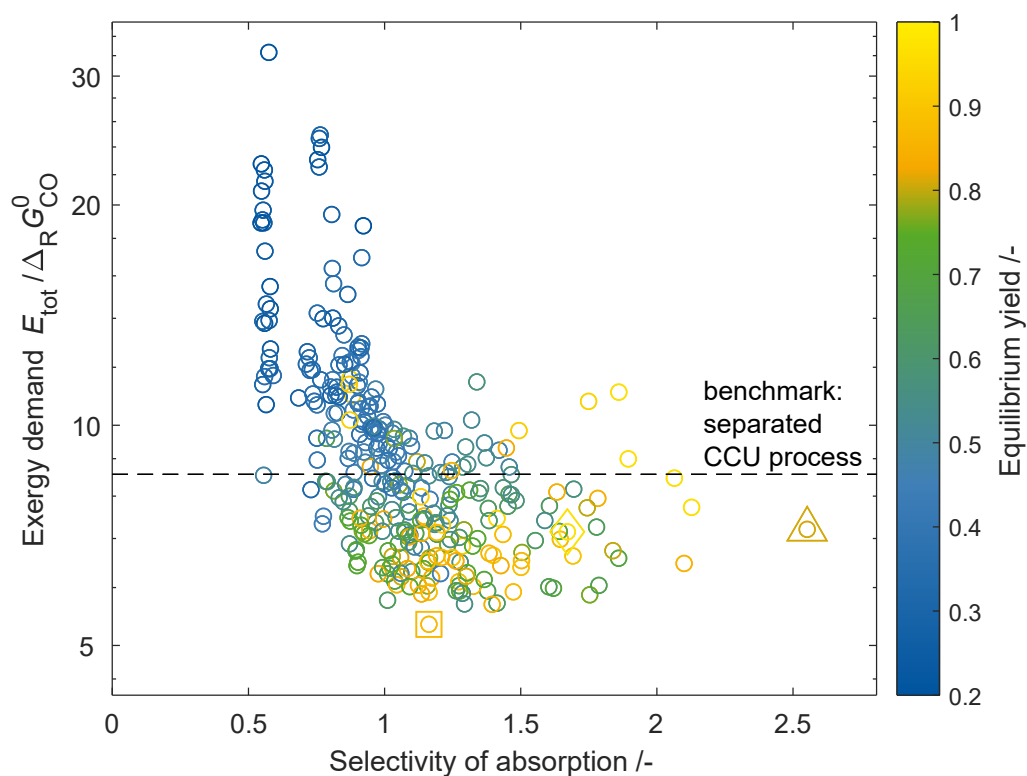

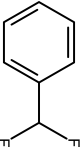
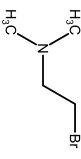
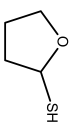
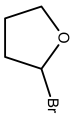


Figure B.8: Results of the integrated molecular and process design of the heat-integrated ICCU process: Each circle represents a molecular candidate with its corresponding optimised process. The candidate with the lowest heat-integrated process exergy demand is marked with a square, the candidate with the highest selectivity of absorption is marked with a triangle, and the candidate with the highest equilibrium yield is marked with a diamond.

Table B.2: Ranking of candidate solvents based on the chosen objective function and exergy demand of the process normalised by the Gibbs free energy of the overall reaction $\Delta_{\text{R}}G_{\text{CO}}^0$ and the exergy demand of the solvent with the lowest total heat-integrated process exergy demand $E_{\text{tot},1}$. The list contains the candidate solvent with the lowest total heat-integrated process exergy demand E_{tot} ; the solvent with the lowest total heat-integrated process exergy demand E_{tot} commercially available; the solvent with the lowest total process exergy demand without heat integration $E_{\text{tot}}^{\text{woHI}}$; the solvent with the highest absorption selectivity and the solvent with the highest reactor equilibrium yield.

Solvent with the ...	Molecular structure	Rank		Exergy demand $E_{\text{tot},i}$			
		exergy demand with heat integration	exergy demand without heat integration	equilibrium yield	absorption selectivity	$/\Delta_{\text{R}}G_{\text{CO}}^0$	$/E_{\text{tot},1}$
lowest exergy demand E_{tot}		1	17	27	106	5.34	1
- commercially available		6	6	68	12	5.86	1.10
lowest exergy demand $E_{\text{tot}}^{\text{woHI}}$		10	1	22	31	5.92	1.11
highest reactor equilibrium yield		89	22	1	17	7.15	1.34
highest absorption selectivity		97	69	56	1	7.21	1.35

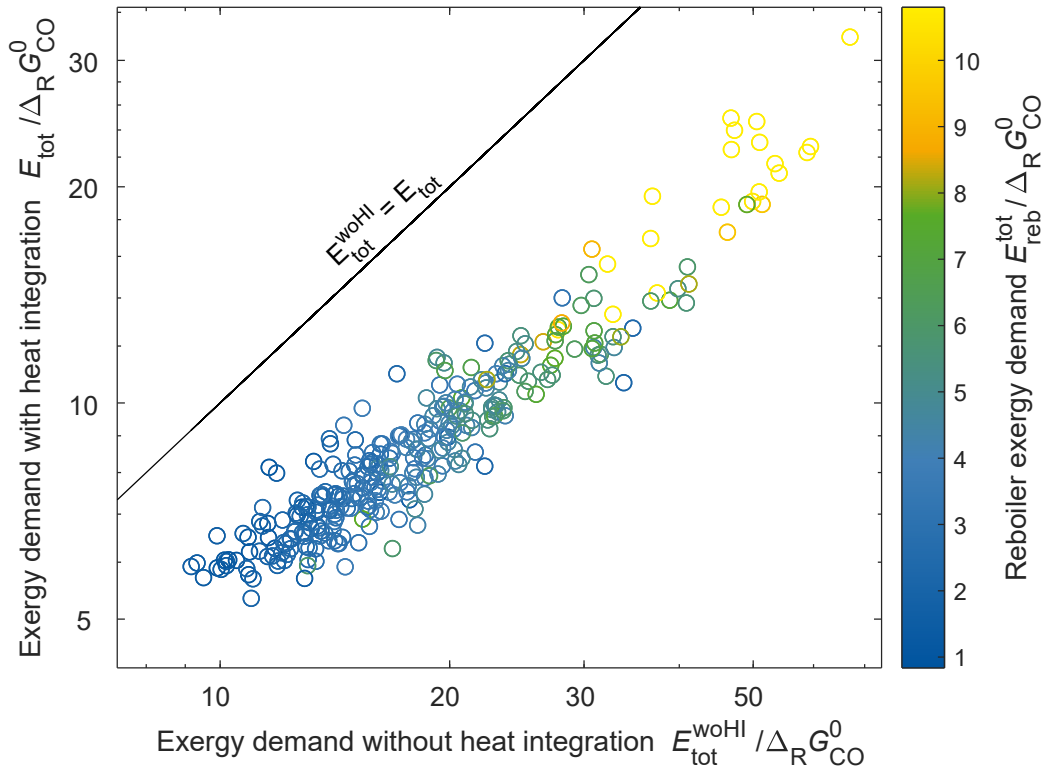


Figure B.9: Comparison of the total process exergy demand considering process optimisation with heat integration (E_{tot}) compared to a process optimisation without considering heat integration potential ($E_{\text{tot}}^{\text{woHI}}$) and the total reboiler exergy demand only (E_{reb}). The black line corresponds to the equality of exergy demand with and without heat integration, demonstrating that heat integration is not possible.

B.6 Specifications of the process models

B.6.1 Hybrid extraction-distillation of γ -valerolactone

Process parameter	Value
Temperature of flows entering the system	25 °C
Temperature of flows leaving the system	25 °C
Pressure in extraction column	1.013 bar
Feed	
- GVL	50 mol s ⁻¹
- Water	950 mol s ⁻¹
Optimisation bounds	
Temperature in the extraction column and the decanter	[25 °C, min (80 °C, boiling point of solvent)]
Pressure in the distillation column	[1 bar, 2 bar]
Utilities	
Hot	
- Low pressure steam (3 bar)	410 K
- High pressure steam (70 bar)	558.15 K
Cold	
- Cooling water	10 °C
Heat Recovery Approach Temperature	10 K

B.6.2 Integrated carbon capture and utilisation for the production of carbon monoxide

Process parameter	Value
Absorption column	
Feed temperature	10 °C
Feed	
- methane	2.33 mol s ⁻¹
- CO ₂	1 mol s ⁻¹
Solvent make-up temperature	-30 °C
Pressure	60 bar
Synthesis reactor	
Temperature	50 °C
$\Delta_R G^0$	12.55 kJ mol ⁻¹
Feed	$\dot{n}_{\text{CO}_2} = \dot{n}_{\text{H}_2} = \dot{n}_{\text{DMA}}$
Bounds for optimisation of pressure	[1 bar, 200 bar]
Bounds for optimisation of water flow	[0.5 mol s ⁻¹ , 5 mol s ⁻¹]
Reforming reactor	
Conversion	1
Temperature	280 °C
Distillation	
Bounds for optimisation of pressure	[1 bar, 5 bar]
Utilities	
Hot	
- Low pressure steam (3 bar)	410 K
- Furnace	750 K
Cold	
- Cooling water	10 °C
- Refrigeration	-40 °C
Heat Recovery Approach Temperature	10 K

B.7 Molecular fragment libraries for the genetic algorithm LEA3D

The following 3D-molecular fragments have been specified as the molecular fragment library for the genetic algorithm LEA3D. Each “X” marks a connector of the molecular fragment and can be connected to another “X” from another molecular fragment to build a new molecular structure. Unconnected connectors “X” are automatically replaced by hydrogen atoms.

B.7.1 Hybrid extraction-distillation of γ -valerolactone

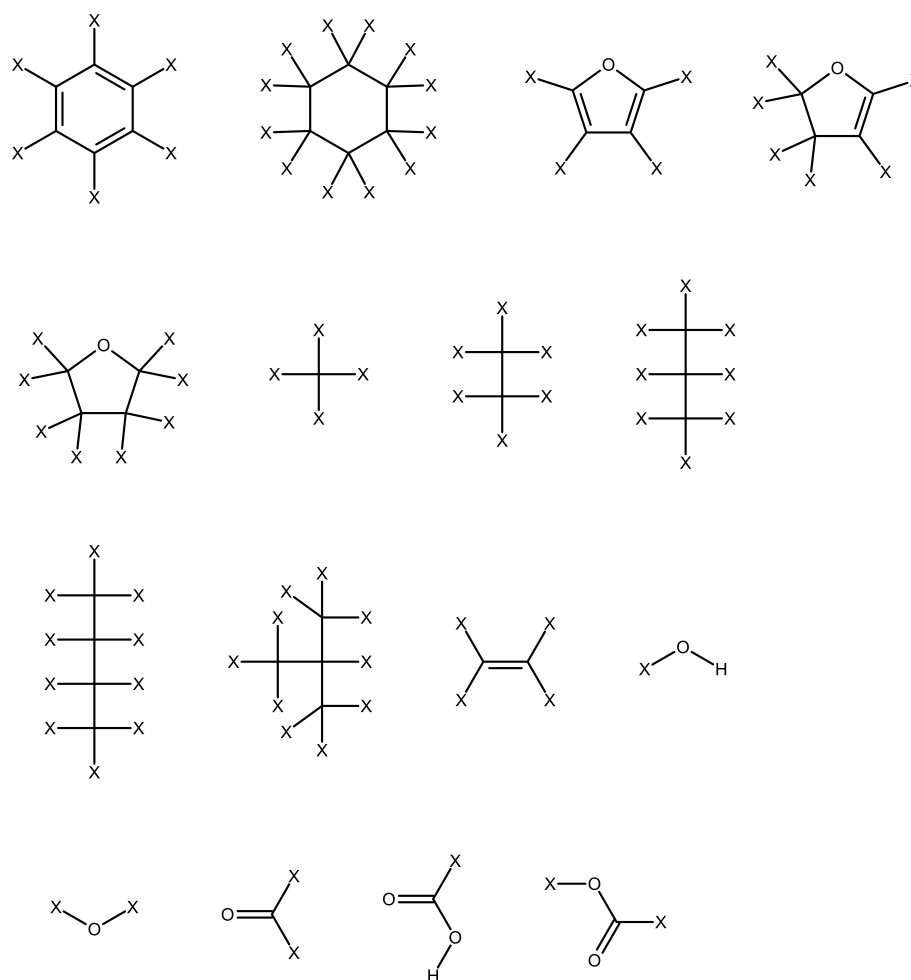


Figure B.10: Molecular fragment library used for the hybrid extraction-distillation case study.

B.7.2 Integrated carbon capture and utilisation via dimethylformamide

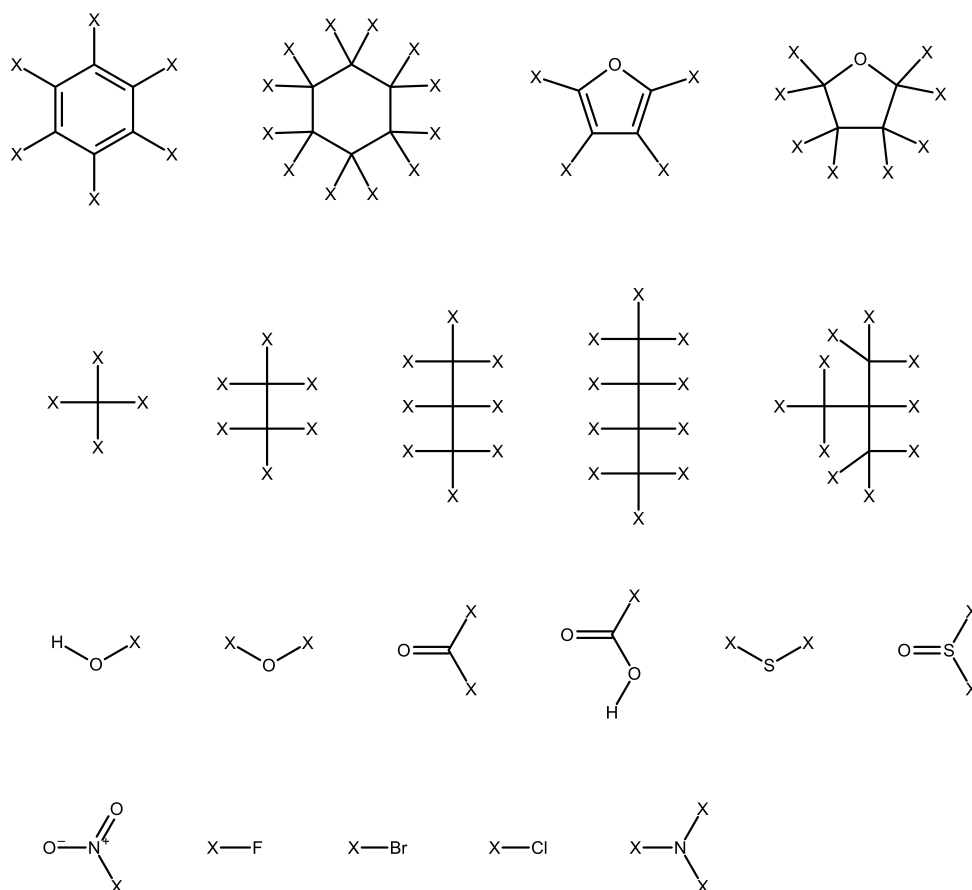


Figure B.11: Molecular fragment library used for the ICCU process case study.

Supporting information to Chapter 5

C.1 Distribution of experimental effort for the experimental designs

Table C.1: Distribution of experimental effort for c-optimal ξ_c^* , D-optimal ξ_D^* and equidistantly distributed conventional ξ_{con} experimental designs for the estimation of isothermal NRTL- τ -parameters for the extraction process

Design ξ	Distribution $\begin{Bmatrix} \alpha_i \\ \nu_i \end{Bmatrix}$
c-optimal ξ_c^*	$\begin{Bmatrix} 0.23 & 0.67 & 0.90 \\ 0.82 & 0.11 & 0.06 \end{Bmatrix}$
D-optimal ξ_D^*	$\begin{Bmatrix} 0.34 & 0.66 & 0.90 \\ 0.23 & 0.32 & 0.45 \end{Bmatrix}$
conventional ξ_{con}	$\begin{Bmatrix} 0.0 & 0.45 & 0.90 \\ 0.33 & 0.33 & 0.33 \end{Bmatrix}$

Table C.2: Distribution of experimental effort for c-optimal ξ_c^* , D-optimal ξ_D^* and equidistantly distributed conventional ξ_{con} experimental designs for the estimation of isothermal NRTL- τ - and diffusion parameters for the extraction process

Design ξ	Distribution $\left(\text{LLE} \begin{Bmatrix} \alpha_i \\ \nu_i \end{Bmatrix} \mid \text{Diffusion} \right)$
c-optimal ξ_c^*	$\left\{ \begin{array}{ccc cc} 0.23 & 0.67 & 0.90 & D_W & D_T \\ 0.81 & 0.10 & 0.06 & 0.02 & 0.02 \end{array} \right\}$
D-optimal ξ_D^*	$\left\{ \begin{array}{ccc cc} 0.34 & 0.66 & 0.90 & D_W & D_T \\ 0.17 & 0.24 & 0.34 & 0.13 & 0.13 \end{array} \right\}$
conventional ξ_{con}	$\left\{ \begin{array}{ccc cc} 0.0 & 0.45 & 0.90 & D_W & D_T \\ 0.20 & 0.20 & 0.20 & 0.20 & 0.20 \end{array} \right\}$

Table C.3: Distribution of experimental effort for c-optimal ξ_c^* , D-optimal ξ_D^* and equidistantly distributed conventional ξ_{con} experimental designs for the estimation of temperature-dependent NRTL- τ -parameters for the hybrid extraction-distillation process

Design ξ	Distribution $\begin{Bmatrix} T_i \\ \alpha_i \\ \nu_i \end{Bmatrix}$
c-optimal ξ_c^*	$\left\{ \begin{array}{ccccccccc} 283.15 & 283.15 & 283.15 & 313.15 & 318.15 & 343.15 & 353.15 & 353.15 & 353.15 \\ 0.31 & 0.82 & 0.90 & 0.55 & 0.90 & 0.90 & 0.47 & 0.65 & 0.90 \\ 0.02 & 0.01 & 0.07 & 0.17 & 0.01 & 0.14 & 0.41 & 0.09 & 0.10 \end{array} \right\}$
D-optimal ξ_D^*	$\left\{ \begin{array}{ccccccc} 283.15 & 283.15 & 283.15 & 313.15 & 318.15 & 353.15 & 353.15 & 353.15 \\ 0.36 & 0.59 & 0.90 & 0.70 & 0.90 & 0.36 & 0.68 & 0.90 \\ 0.07 & 0.07 & 0.22 & 0.07 & 0.14 & 0.13 & 0.11 & 0.19 \end{array} \right\}$
conventional ξ_{con}	$\left\{ \begin{array}{ccccccc} 283.15 & 283.15 & 283.15 & 318.15 & 318.15 & 353.15 & 353.15 & 353.15 \\ 0.00 & 0.45 & 0.90 & 0.30 & 0.60 & 0.00 & 0.45 & 0.90 \\ 0.13 & 0.13 & 0.13 & 0.13 & 0.13 & 0.13 & 0.13 & 0.13 \end{array} \right\}$

C.2 Input parameters for the case studies

C.2.1 Initial property parameters for the NRTL model (Renon and Prausnitz, 1968)

The initial NRTL-parameters for all case studies have been estimated using COSMO-RS (Klamt et al., 2010) on BP-TZVPD-FINE parametrisation (BP-TZVPD-FINE-C30-1701) in the implementation COSMOthermX17 by COSMOlogic GmbH & Co. KG, Leverkusen, Germany.

Isothermal NRTL parameters

$$\alpha = \begin{pmatrix} 0 & 0.7548 & 0.5166 \\ 0.7548 & 0 & 0.1940 \\ 0.5166 & 0.1940 & 0 \end{pmatrix}$$

$$\tau = \begin{pmatrix} 0 & 0.9287 & 0.5751 \\ 0.8724 & 0 & 7.3512 \\ 0.0647 & 5.0405 & 0 \end{pmatrix}$$

Temperature-dependent NRTL parameters

The α -parameters are assumed to be temperature-independent and have the same values as for the isothermal case (see above).

The τ -parameters are assumed to depend on the temperature according to the following equation:

$$\tau_{i,j} = \tau_{i,j}^0 + \frac{\tau_{i,j}^1}{T}$$

$$\tau^0 = \begin{pmatrix} 0 & 1.3516 & 0.5256 \\ 3.5090 & 0 & 2.7044 \\ -0.4226 & -6.9574 & 0 \end{pmatrix}, \quad \tau^1 = \begin{pmatrix} 0 & -126.08 & 14.7637 \\ -786.11 & 0 & 1385.4 \\ 145.29 & 3577.2 & 0 \end{pmatrix}$$

C.2.2 Specifications for the extraction column using a pinch-based process model

Process parameter	Value
bounds for α	[0 0.9]
std. dev. of phase composition measurements σ_w	0.005
feed to extraction column	
-acetone	0.05 mol s ⁻¹
-water	0.95 mol s ⁻¹
solvent composition	100 % toluene
extraction temperature	298.15 K
extraction pressure	1 bar

Process model result	Value
min. solvent demand S_{\min}	0.4207 mol _{solvent} /mol _{feed}

C.2.3 Specifications for the hybrid extraction-distillation process using pinch-based process models

Process parameter	Value
bounds for α	[0 0.9]
std. dev. of phase composition measurements σ_w	0.005
feed to extraction column	
-acetone	0.05 mol s ⁻¹
-water	0.95 mol s ⁻¹
solvent composition	100 % toluene
extraction temperature	298.15 K
extraction pressure	1 bar
distillation pressure	1 bar

Property parameter	Value
Antoine parameter $\left(\frac{p}{\text{mbar}} = \exp\left(A - \frac{B}{T/\text{K} + C}\right)\right)$	
-acetone	A=15.578, B=2574.5, C=-50.400
-water	A=16.836, B=3155.6, C=-73.970
-toluene	A=16.172, B=3102.5, C=-52.552
Enthalpy of vaporisation (DIPPR 106 form in J/mol, $T_r = T/10\,000$ K)	
-acetone	a=36988, b=11.409, c=-475.23, d=12844, e=-99271
-water	a=51367, b=-6.8777, c=384.29, d=-104.56, e=-33539
-toluene	a=41517, b=7.7975, c=-521.85, d=16049, e=-126581

The vapour pressure and the enthalpy of vaporisation have been estimated using COSMO-RS (Klamt et al., 2010) on BP-TZVPD-FINE parametrisation (BP-TZVPD-FINE-C30-1701) in the implementation COSMOthermX17 by COSMOlogic GmbH & Co. KG, Leverkusen, Germany.

The LCIA model for the wastewater treatment is the same as in Chapter A and taken from the Ecoinvent database (Ruiz, 2019). The environmental impacts of toluene are also taken from the Ecoinvent database (Hischier, 2019).

Process model result	Value
min. solvent demand S_{\min}	0.4207 mol _{solvent} /mol _{feed}
min. energy demand Q_{\min}	222 kW / mol _{acetone}
min. impact on Climate Change CC_{\min}	15.2 g CO ₂ / mol _{acetone}

C.2.4 Specifications for the countercurrent rate-based extraction column using the HTU-NTU method for sizing

Process parameter	Value	Reference
Mass flow of raffinate carrier water \dot{m}_1''	5 kg s^{-1}	
Specific mass transfer area of packing a_P	$150.92 \text{ m}^2 \text{ m}^{-3}$	Strigle (1994)
Void fraction area of packing ϵ	0.972	Strigle (1994)
Boundary layer thickness δ'	$1 \times 10^{-4} \text{ m}$	
Diffusion coefficient of acetone in water D_{12}	$0.81 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$	Tyn and Calus (1975)
Diffusion coefficient of acetone in toluene D_{23}	$1.1 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$	Bulicka and Prochazka (1976)
Raffinate inlet loading $X_{2,\text{in}}''$	$0.2 \text{ mol}_{\text{acetone}}/\text{mol}_{\text{water}}$	
Raffinate purification yield $X_{2,\text{out}}''$	$0.05 \text{ mol}_{\text{acetone}}/\text{mol}_{\text{water}}$	
Extract inlet loading $X'_{2,\text{in}}$	$0 \text{ mol}_{\text{acetone}}/\text{mol}_{\text{toluene}}$	
Density of raffinate phase ρ''	997.05 kg m^{-3}	NIST Chemistry WebBook (2021)
Density of extract phase ρ'	862.24 kg m^{-3}	NIST Chemistry WebBook (2021)
Viscosity of raffinate phase μ''	0.00056 Pa s	NIST Chemistry WebBook (2021)
Interfacial tension σ_1	$25.46 \times 10^{-7} \text{ N m}^{-1}$	Enders et al. (2007)
Safety factor S for Crawford-Wilke correlation	5	Strigle (1994)

Cost parameter	Value	Reference
Plant lifetime N_{life}	25 a	
Annual operation hours t_a	8000 h	
Interest rate I	4.18 %	
Disposal costs C_{disp}	1 \$/kg	
Material factor F_m	1	Biegler et al. (1997)
Tray spacing factor F_s	1	Biegler et al. (1997)
Tray type factor F_t	0	Biegler et al. (1997)
Pressure factor F_p	1	Biegler et al. (1997)
Base costs C_0	1000 \$	Biegler et al. (1997)
Reference height of column $h_{K,0}$	4 ft	Biegler et al. (1997)
Reference diameter of column $d_{K,0}$	3 ft	Biegler et al. (1997)
Exponent for column height scaling α_h	0.81	Biegler et al. (1997)
Exponent for column diameter scaling α_d	1.05	Biegler et al. (1997)
Update factor UF	5.353	

Process model result	Value
total annualised cost TAC	624 309 USD/year

Rate-based mass transfer model *

The rate-based model of a counter-current extraction process considers a continuous raffinate phase " and a dispersed extract phase '. The extract and raffinate are assumed to be non-soluble within each other, which is well justified for a system with a large miscibility gap. Therefore, calculations are based on molar loadings X_2 of mol transition component per mol extract carrier and raffinate carrier, respectively:

$$X_2'' = \frac{x_2''}{1 - x_2''} \quad (\text{C.1})$$

$$X_2' = \frac{x_2'}{1 - x_2'} \quad (\text{C.2})$$

* This explanation of the rate-based mass transfer model is reproduced in major parts from the PhD thesis of Ludger W. M. Wolff, RWTH Aachen University, Aachen, Germany, 2021, for transparency and comprehensibility of the results presented in this thesis.

The total mole balance of the extraction process reads as

$$\dot{n}_1''(X_{2,\text{out}}'' - X_{2,\text{in}}'') = \dot{n}_3'(X_{2,\text{in}}' - X_{2,\text{out}}'). \quad (\text{C.3})$$

Here, \dot{n}_1'' is the molar flow of the raffinate carrier component, and $X_{2,\text{out}}''$ and $X_{2,\text{in}}''$ are the outlet and inlet loadings of the raffinate. \dot{n}_3' is the molar flow of the extract carrier component, and $X_{2,\text{in}}'$ and $X_{2,\text{out}}'$ are the inlet and outlet loadings of the extract.

Sizing is performed using the well-known HTU-NTU method (Chilton and Colburn, 1935; Sattler and Feindt, 1995)

$$h_K = \underbrace{\frac{V''\dot{n}_1''}{\beta_{12}a_P \left(\frac{\pi d_K^2}{4}\right)}}_{HTU} \underbrace{\int_{X_{2,\text{out}}''}^{X_{2,\text{in}}''} \frac{dX_2''}{\ln \frac{1+X_2''}{1+X_{2,\text{Ph}}''}}}_{NTU}. \quad (\text{C.4})$$

Here, V'' is the molar volume of the raffinate phase, and β_{12} is the mass transfer coefficient. The mass transfer coefficient β_{12} is assumed to be independent of composition and follows from the diffusion coefficient D_{12} and the thickness of the boundary layer of the raffinate phase, δ'' :

$$\beta_{12} = \frac{D_{12}}{\delta''}. \quad (\text{C.5})$$

Typically, the thickness of the boundary layer is often unknown (Taylor and Krishna, 1993). Therefore, the mass transfer coefficient β_{12} is typically calculated directly from empirical correlations employing the Sherwood number (Taylor and Krishna, 1993). Since the impact of multiple thermodynamic properties on the optimal experimental design is in the focus of this work, the thickness of the boundary layer is assumed to be known and constant.

To calculate the number of transfer units NTU , the loading $X_{2,\text{Ph}}''$ at the interface is calculated assuming thermodynamic equilibrium at the interface.

A relation between bulk phase loadings and loadings at the interface between raffinate and extract is obtained via the molar flux \dot{N}_2 of the transferred component. The molar flux \dot{N}_2 of the transferred component can be calculated from the extract or the raffinate phase via the concentration gradient considering Stefan-correction and has to be equal. Thus, a relation between bulk phase loadings and loadings at the interface between raffinate and extract is obtained:

$$-\frac{V''\beta_{12}}{V'\beta_{23}} = \frac{\ln \frac{1+X_2'}{1+X_{2,\text{Ph}}'}}{\ln \frac{1+X_2''}{1+X_{2,\text{Ph}}''}}. \quad (\text{C.6})$$

Here,

$$\beta_{23} = \frac{D_{23}}{\delta''} \quad (\text{C.7})$$

is the (constant) mass transfer coefficient of acetone in the extract phase.

The column diameter d_K is calculated from the maximum volume flows of the raffinate and extract phase, respectively (Strigle, 1994):

$$\frac{\pi d_K^2}{4} = \left(\underbrace{\frac{\dot{n}_2'' V'' (1 + \max(X''))}{u''}}_{\text{max. volume flow of raffinate}} + \underbrace{\frac{\dot{n}_3' V' (1 + \max(X'))}{u'}}_{\text{max. volume flow of extract}} \right) \cdot S. \quad (\text{C.8})$$

Here, $S = 5$ is a safety factor that ensures sufficient distance of the operating point from flooding (Strigle, 1994).

The maximum volume flows of the raffinate and extract phase depend on the flooding velocities of the raffinate and extract phase, u' and u'' . u' and u'' are calculated from the Crawford-Wilke flooding correlation (Strigle, 1994)

$$\frac{\rho'}{a_P \mu''} (\sqrt{u'} + \sqrt{u''})^2 = f \left(\frac{\mu''}{|\rho'' - \rho'|} \left[\frac{\sigma_I}{\rho''} \right]^{0.2} \left[\frac{a_P}{\epsilon} \right]^{1.5} \right). \quad (\text{C.9})$$

Here, ρ'' and ρ' are the densities of the raffinate and extract phase, μ'' is the viscosity of the continuous phase (here the raffinate phase), σ_I is the interfacial tension, and ϵ is the void fraction of the column packing. The function f is taken from a plot given by Strigle (1994).

Cost estimation model

The total annualised cost TAC result from total annualised operational costs $TAOC$ and the depreciation of the investment costs, the total updated bare module costs $TUBMC$ (Biegler et al., 1997):

$$TAC = TAOC + ccf \cdot TUBMC. \quad (\text{C.10})$$

The factor ccf is the capital charge factor that determines the depreciation. The total annualised operational costs $TAOC$ are approximated by the disposal costs for the extract:

$$TAOC = (\dot{n}_3' (1 + \max(X')))) \cdot t_a \cdot C_{\text{disp}} \quad (\text{C.11})$$

Here, t_a is the annual operational time, and C_{disp} are the specific disposal costs. The total updated bare module costs $TUBMC$ are calculated following the systematic approach of Guthrie (Biegler et al., 1997):

$$TUBMC = UF \cdot (BC (MPF + MF - 1)) \quad (\text{C.12})$$

Optimal experimental design calculation

The aim of c-OED is to determine which diffusion and LLE experiments should be performed to minimise the variance in the total annual costs TAC . Noteworthy, the c-optimal design minimises only the uncertainty σ_{TAC}^2 in the total annual costs and not the total annual costs TAC itself. However, for a reasonable operation of the extraction column, the total annual costs TAC should be low. Therefore, a two-step optimisation procedure is performed: In the first step, the total annual costs TAC are minimised. The total annual costs TAC are determined by the physical properties of the extraction system, i.e. the parameters $\hat{\theta}$, and the extract flow rate \dot{n}'_3 .

In the second optimisation step, the c-optimal design $\mathbf{d}_{c,\text{opt}}$ for the operation point with minimum TAC_{min} is determined. Of course, the c-optimal design could be determined for any other operation point as well.

C.3 Additional results from the Monte Carlo analysis

C.3.1 Accuracy of the extraction process for $\sigma_w = 0.001$ and $\sigma_w = 0.01$

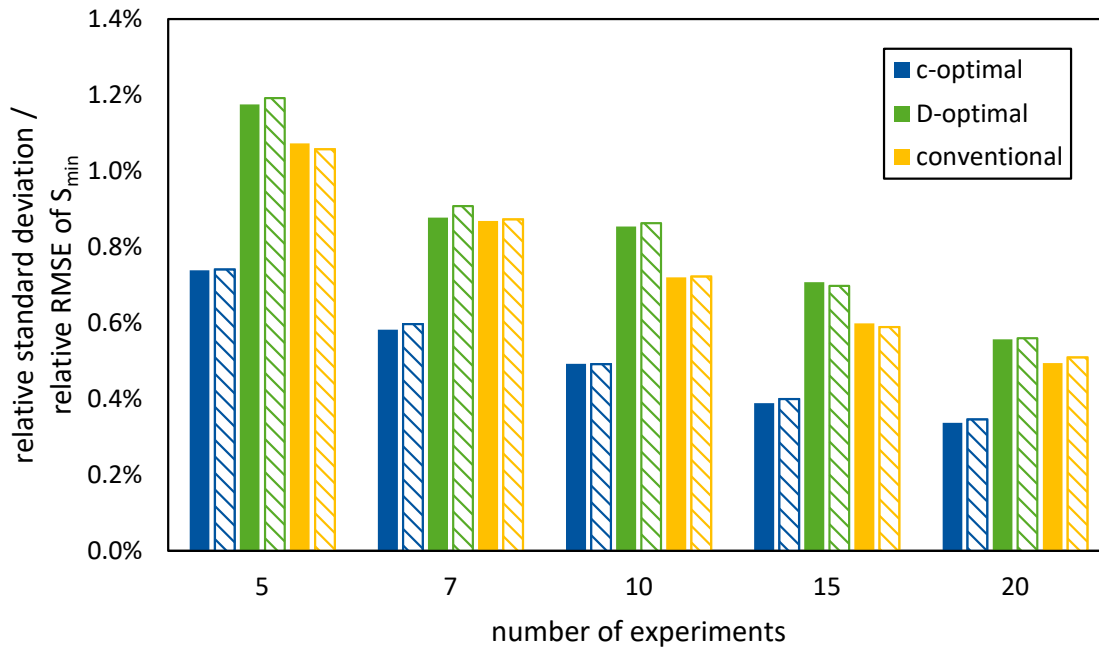


Figure C.1: Uncertainties of the solvent demand in the extraction process for the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for a standard deviation in the measurements of mole fractions of phase equilibrium experiments of $\sigma_w = 0.001$. The full bars are the relative RMSE from the Monte Carlo sampling; the hatched bars are the expected relative standard deviation from OED theory.

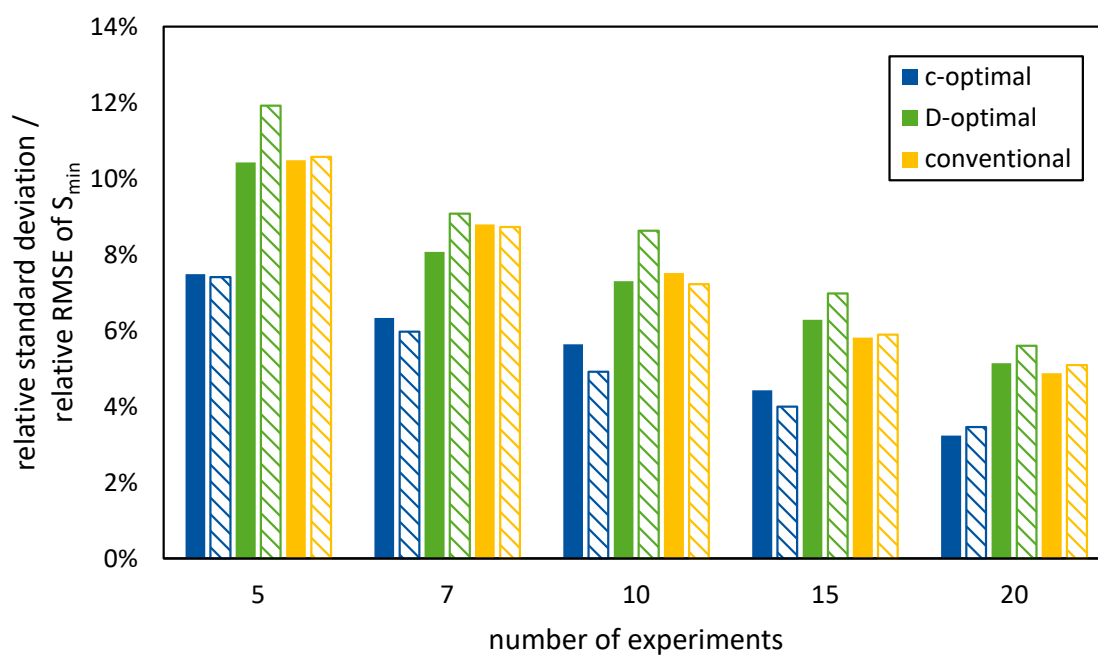


Figure C.2: Uncertainties of the solvent demand in the extraction process for the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for a standard deviation in the measurements of mole fractions of phase equilibrium experiments of $\sigma_w = 0.01$. The full bars are the relative RMSE from the Monte Carlo sampling; the hatched bars are the expected relative standard deviation from OED theory.

C.3.2 NRTL-parameter accuracy for the extraction process for $\sigma_w = 0.005$

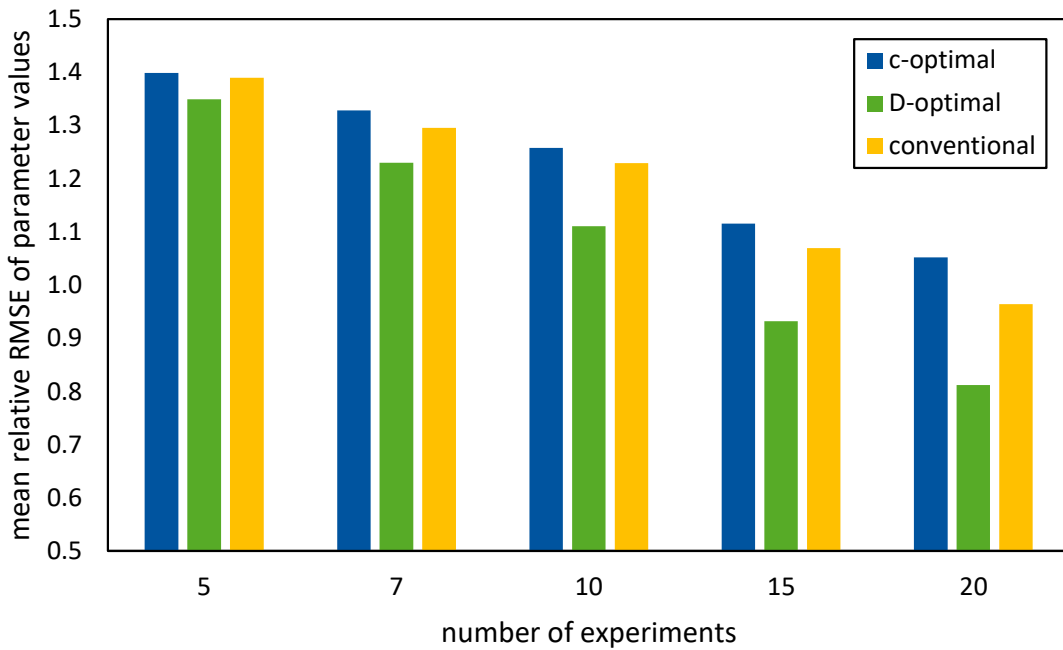


Figure C.3: Parameter accuracy of the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for the extraction process for $\sigma_w = 0.005$. The parameter accuracy is measured using the mean relative root-mean-square error (RMSE) of the property parameters.

C.3.3 Accuracy of the hybrid extraction-distillation process for $\sigma_w = 0.001$ and $\sigma_w = 0.01$

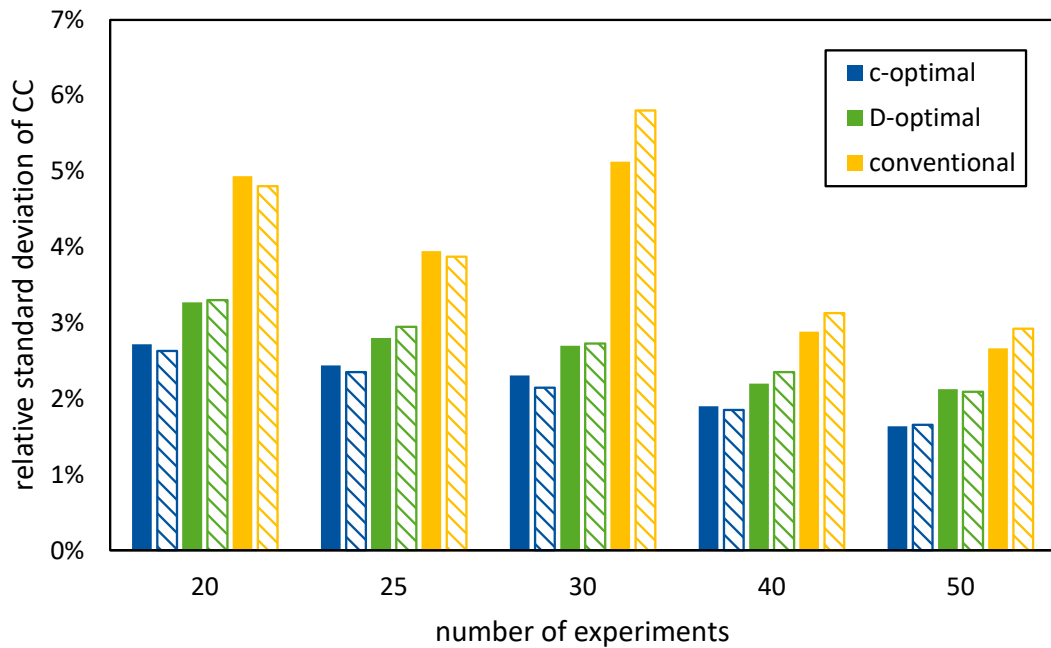


Figure C.4: Uncertainties of the environmental impact on Climate Change in the extraction-distillation process for the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for a standard deviation in the measurements of mole fractions of phase equilibrium experiments of $\sigma_w = 0.001$. The full bars are the relative RMSE from the Monte Carlo sampling; the hatched bars are the expected relative standard deviation from OED theory.

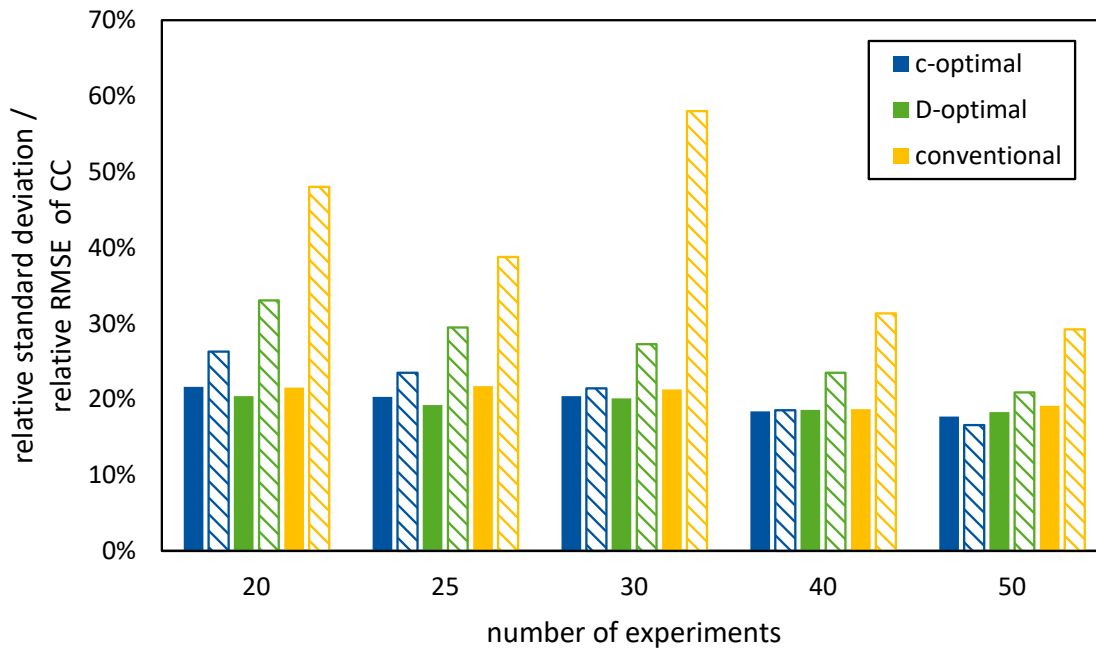


Figure C.5: Uncertainties of the environmental impact on Climate Change in the hybrid extraction-distillation process for the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for a standard deviation in the measurements of mole fractions of phase equilibrium experiments of $\sigma_w = 0.01$. The full bars are the relative RMSE from the Monte Carlo sampling; the hatched bars are the expected relative standard deviation from OED theory.

C.3.4 Influence of the number of experiments for the hybrid extraction-distillation process for $\sigma_w = 0.005$ and $\sigma_w = 0.01$

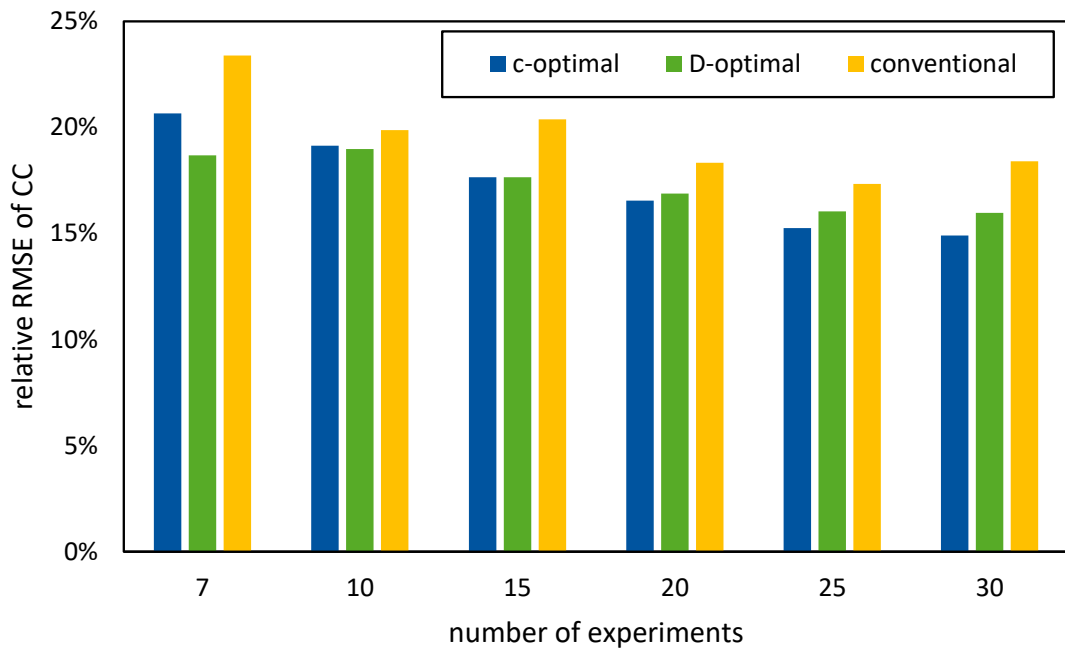


Figure C.6: Effect of number of experiments on the relative RMSE of the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for the hybrid extraction-distillation process for a standard deviation in the measurements of mole fractions of phase equilibrium experiments of $\sigma_w = 0.005$.

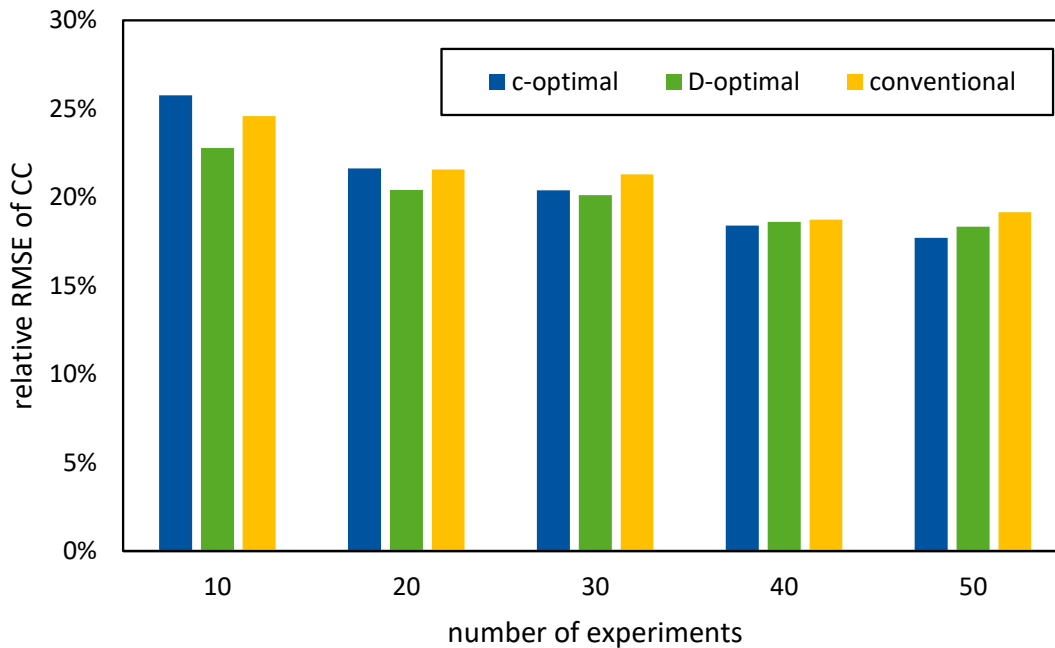


Figure C.7: Effect of number of experiments on the relative RMSE of the c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for the hybrid extraction-distillation process for a standard deviation in the measurements of mole fractions of phase equilibrium experiments of $\sigma_w = 0.01$. Compared to a smaller standard deviation in the measurements of mole fractions of $\sigma_w = 0.005$, the c-optimal design starts to outperform the D-optimal and conventional design for 40 experiments or more.

C.3.5 NRTL-parameter accuracy for the hybrid extraction-distillation process for $\sigma_w = 0.005$

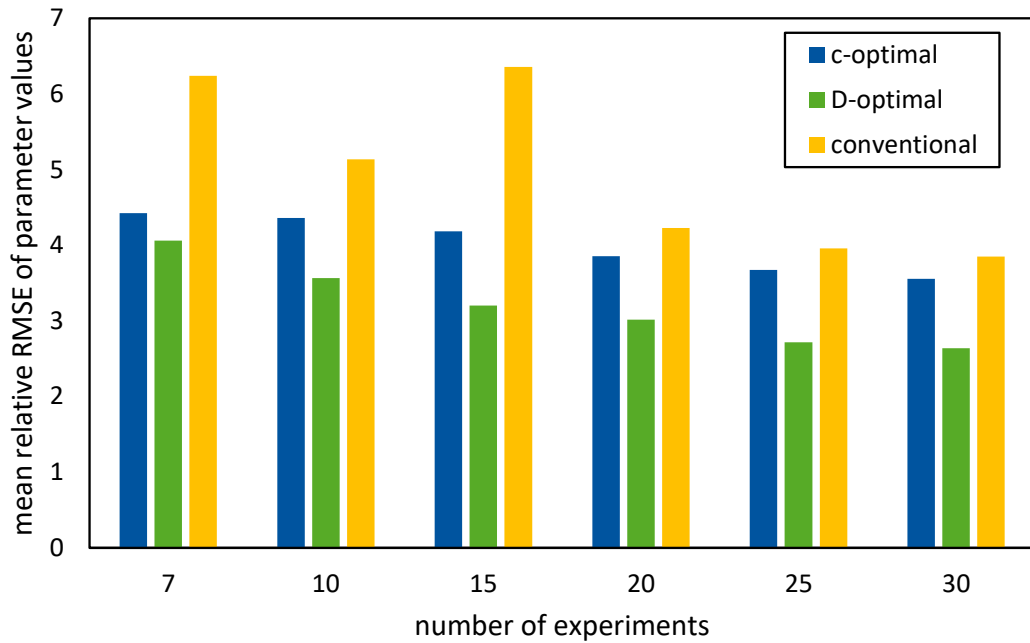


Figure C.8: Property parameter accuracy of c-optimal, D-optimal, and equidistantly distributed conventional experimental designs for the hybrid extraction-distillation process. The parameter accuracy is measured using the mean relative root-mean-square error (RMSE) of the property parameters. As expected from theory, the exact D-optimal designs yield the most accurate parameters for each number of experiments investigated. However, even for the most accurate parameter estimation in this case study for the D-optimal design and 30 experiments, the variance in the property parameter equals a factor of 2.6 on average, with a range between 0.16 and 18.7 for the individual parameters.

C.4 Additional optimal experimental design for the hybrid extraction-distillation process

Figure C.9 shows an additional optimal experimental design for estimating temperature-dependent NRTL- τ -parameters from LLE experiments for the hybrid extraction-distillation process. In contrast to Section 5.2.3, the objective of this c-OED is to minimise the uncertainty of the reboiler energy demand of the distillation column Q_{\min} . Table C.4 contains the corresponding c- and D-efficiencies ζ_c and ζ_D of the c-optimal ξ_c^* , D-optimal ξ_D^* and equidistantly distributed conventional ξ_{con} experimental designs. Generally, the c-optimal experimental design and the corresponding c- and D-efficiencies ($\zeta_c^D = 0.61$ and $\zeta_D^c = 0.62$) are similar to the minimisation of the uncertainty of the environmental impact in Section 5.2.3 ($\zeta_c^D = 0.63$ and $\zeta_D^c = 0.65$).

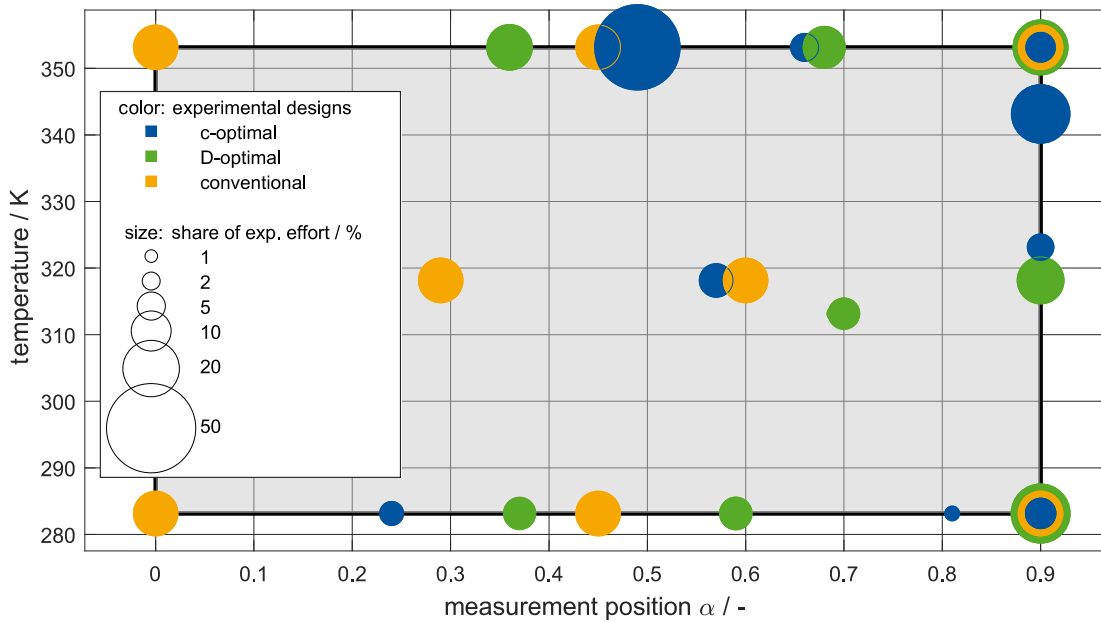


Figure C.9: Experimental designs of LLE experiments for the hybrid extraction-distillation process. The size of the circles corresponds to the share of the experimental effort. The grey box indicates the design space.

The minor differences in the c-optimal design are caused by the reduced parameter requirements of the energy-related objective compared to the LCA objective in Section 5.2.3. In contrast to performing an LCA of the process, describing the energy demand only does not require an accurate description of the aqueous solvent solubility for emissions related to solvent disposal and make-up. The reduced parameter requirements decrease the need for overall accurate parameters and are reflected in a marginally lower D-efficiency of $\zeta_c^D = 0.61$ for the energy-related objective function compared to the LCA objective ($\zeta_c^D = 0.63$).

Table C.4: c- and D-efficiencies ζ_c and ζ_D of the c-optimal ξ_c^* , D-optimal ξ_D^* , and equidistantly distributed conventional ξ_{con} experimental designs for the estimation of temperature-dependent NRTL- τ -parameters and use in the pinch-based hybrid extraction-distillation process model.

Design ξ	c-efficiency ζ_c	D-efficiency ζ_D
c-optimal ξ_c^*	1	0.62
D-optimal ξ_D^*	0.61	1
conventional ξ_{con}	0.24	0.55

APPENDIX D

Supporting information to Chapter 6

D.1 Soft- and hardware

The following software is used within the fuel design method:

Calculation	Method	Software version
Optimised geometry and σ -surface	DFT BP86 / TZVP (-MF)	COSMOconf16 (4.1) (COSMOlogic GmbH & Co. KG, 2017a) TURBOMOLE® 7.2 (COSMOlogic GmbH & Co. KG, 2017c)
Thermodynamic properties	BP-TZVP-C30-1701	COSMOtherm17 (C30-1701) (COSMOlogic GmbH & Co. KG, 2017b)
Fuel ignition quality indicators	GNN	Version v0.2 (Schweidtmann et al., 2020)
Laminar burning velocity	GC-based ANN	Version v2 (vom Lehn et al., 2021a)
Synthesisability	Molecular Transformer	Version 12class-tokens-2021-05-14 (Schwaller et al., 2020)
Numerical optimisation	fmincon	MATLAB® R2018b (The MathWorks Inc., 2018b)

All calculations were performed on one Intel® HNS2600BPB node of the RWTH Compute Cluster CLAIX-2018 employing one 24-core Intel® Xeon® Platinum 8160 Processor with 94 GB RAM.

D.2 Molecular fragment library for the genetic algorithm LEA3D

The following 3D-molecular fragments have been specified as the molecular fragment library for the genetic algorithm LEA3D. Each “X” marks a connector of the molecular fragment and can be connected to another “X” from another molecular fragment to build a new molecular structure. Unconnected connectors “X” are automatically replaced by hydrogen atoms.

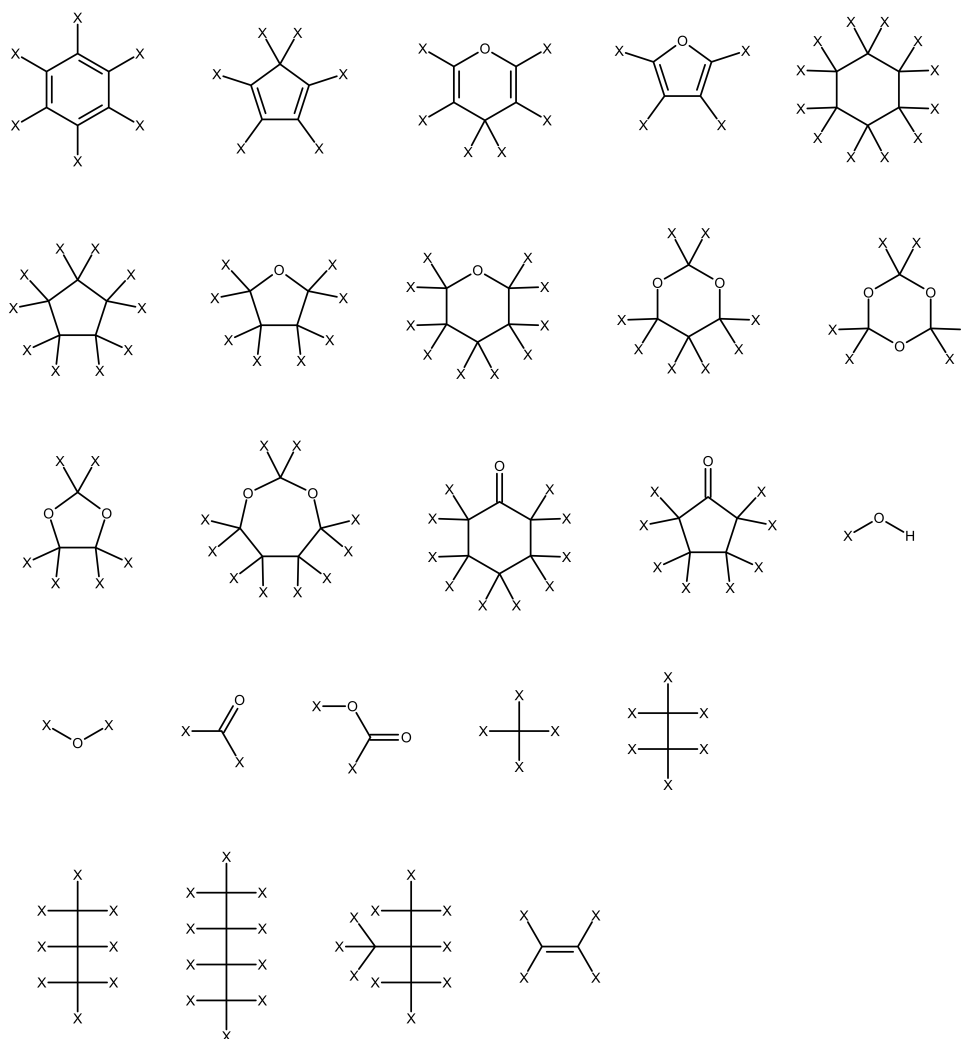


Figure D.1: Molecular fragment library used for the pure-component and blend design.

D.3 Prediction accuracy of environment, health, and safety indicators

The environment, health, and safety indicators are estimated using group contribution-based Gaussian Process Regression (GPR) as proposed by Alshehri et al. (2021). In contrast to Alshehri et al. (2021), the models are trained using UNIFAC groups as features. The fragmentation into UNIFAC groups is performed using the automatic fragmentation algorithm by Müller (2019). For each model, the most accurate kernel configuration is reported. Here, the following kernels were considered:

Kernel	Abbreviation
Linear	Lin
Squared-Exponential	SE
Rational Quadratic	RQ
White Noise	WN

The plots below visualise the prediction accuracy of unseen test data excluded during model training. The training and test data are taken from Alshehri et al. (2021), except for the unified yield sooting index (uYSI). The uYSI data is taken from McEnally et al. (2017).

- Autoignition temperature (AiT; American Society for Testing and Materials, 2000) in Kelvin

Kernel configuration: SE + SE + WN + RQ

Number of data points for training: $N_{\text{Train}} = 487$

Number of data points for testing: $N_{\text{Test}} = 54$

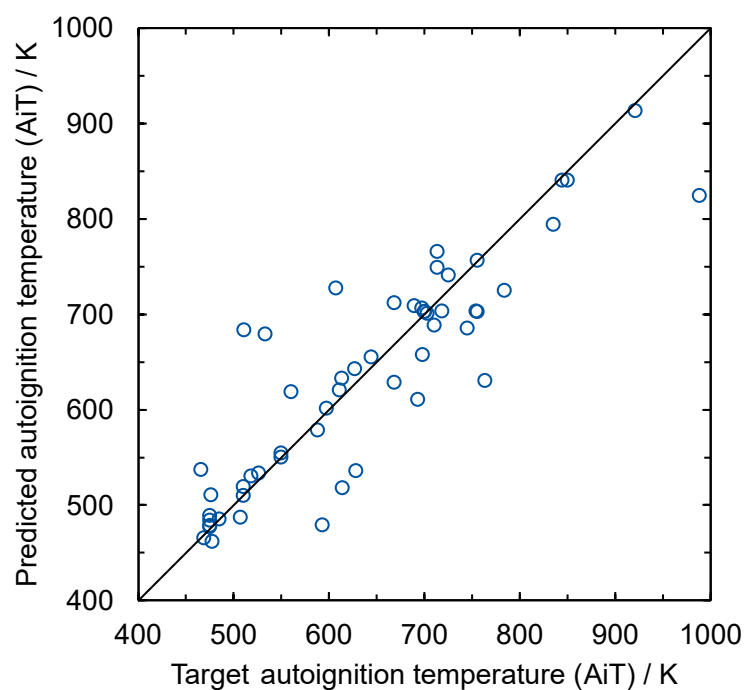


Figure D.2: Parity plot of predicted versus experimental autoignition temperature (AiT) of the test set.

- Bioconcentration factor (BCF, dimensionless; Arnot and Gobas, 2006) expressed as $\log(\text{BCF})$

Kernel configuration: SE + SE + SE + SE + WN

Number of data points for training: $N_{\text{Train}} = 366$

Number of data points for testing: $N_{\text{Test}} = 41$

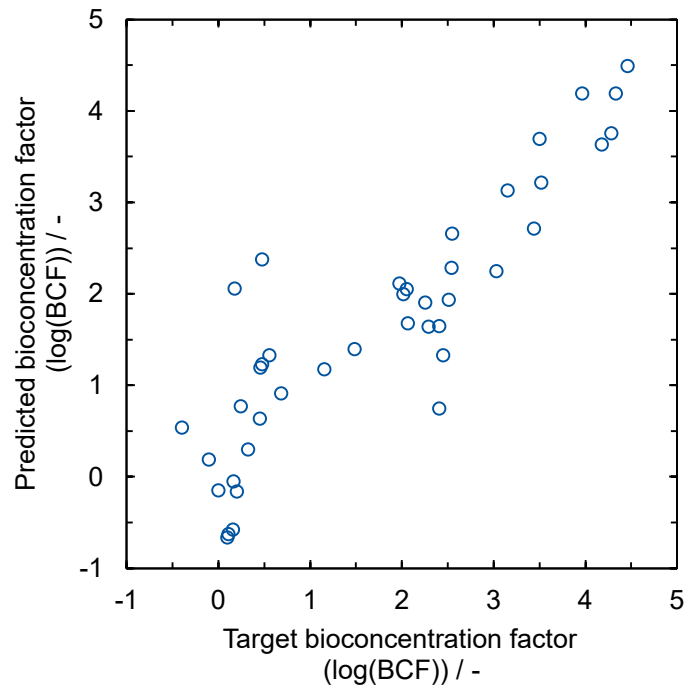


Figure D.3: Parity plot of predicted versus experimental bioconcentration factor (BCF) of the test set.

- Aqueous toxicity as lethal concentration for fathead minnow fish after 96 h exposure ($LC_{50}(FM)$; Ankley and Villeneuve, 2006) expressed as $-\log\left(\frac{LC_{50}(FM)}{\text{mol L}^{-1}}\right)$

Kernel configuration: Lin + SE · Lin + WN

Number of data points for training: $N_{\text{Train}} = 490$

Number of data points for testing: $N_{\text{Test}} = 54$

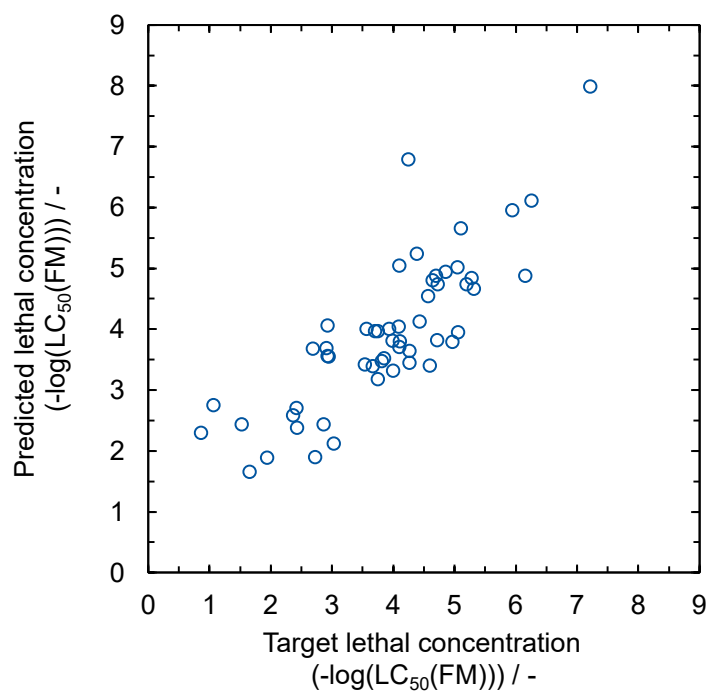


Figure D.4: Parity plot of predicted versus experimental aqueous toxicity ($LC_{50}(FM)$) of the test set.

- Oral toxicity as lethal dose for rats from oral intake (LD_{50} ; Walum, 1998) expressed as $-\log\left(\frac{LD_{50}}{\text{mol kg}^{-1}}\right)$

Kernel configuration: SE + SE + SE + SE + WN

Number of data points for training: $N_{\text{Train}} = 2157$

Number of data points for testing: $N_{\text{Test}} = 240$

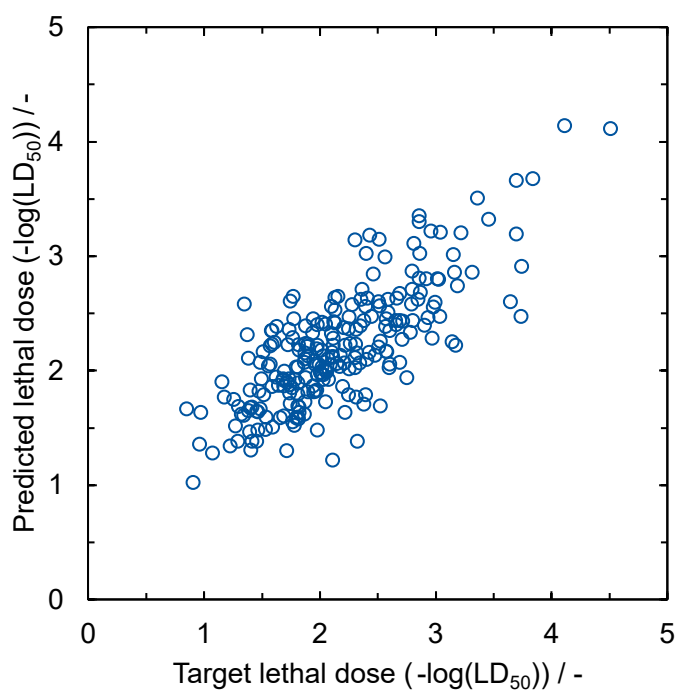


Figure D.5: Parity plot of predicted versus experimental oral toxicity (LD_{50}) of the test set.

- Permissible exposure limit (PEL) using the OSHA time-weighted average ($PEL_{OSHA-TWA}$; Spear and Selvin, 1989) expressed as $-\log\left(\frac{PEL_{OSHA-TWA}}{\text{mol m}^{-3}}\right)$

Kernel configuration: SE + SE + SE + SE + WN

Number of data points for training: $N_{\text{Train}} = 346$

Number of data points for testing: $N_{\text{Test}} = 38$

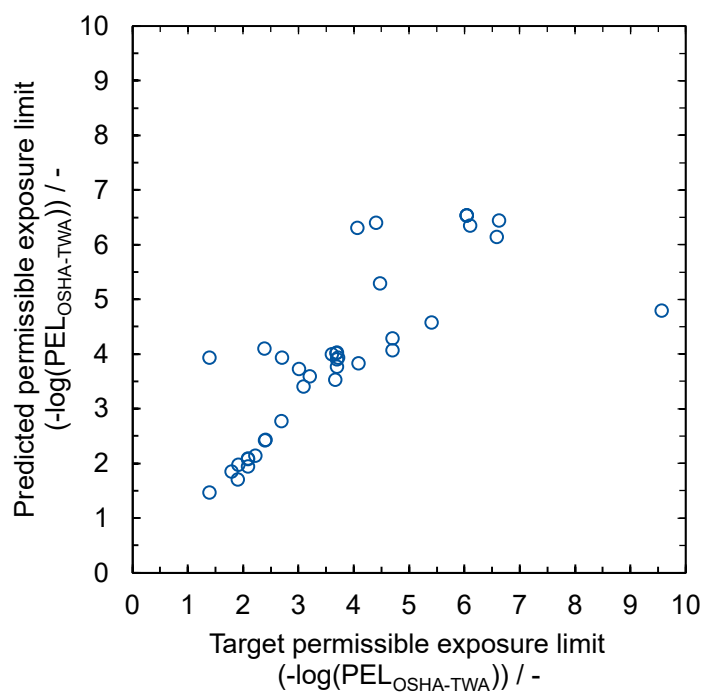


Figure D.6: Parity plot of predicted versus experimental permissible exposure limit ($PEL_{OSHA-TWA}$) of the test set.

- Unified Yield Sooting Index (uYSI, dimensionless; Das et al., 2018)

Kernel configuration: RQ · Lin + WN

Number of data points for training: $N_{\text{Train}} = 397$

Number of data points for testing: $N_{\text{Test}} = 44$

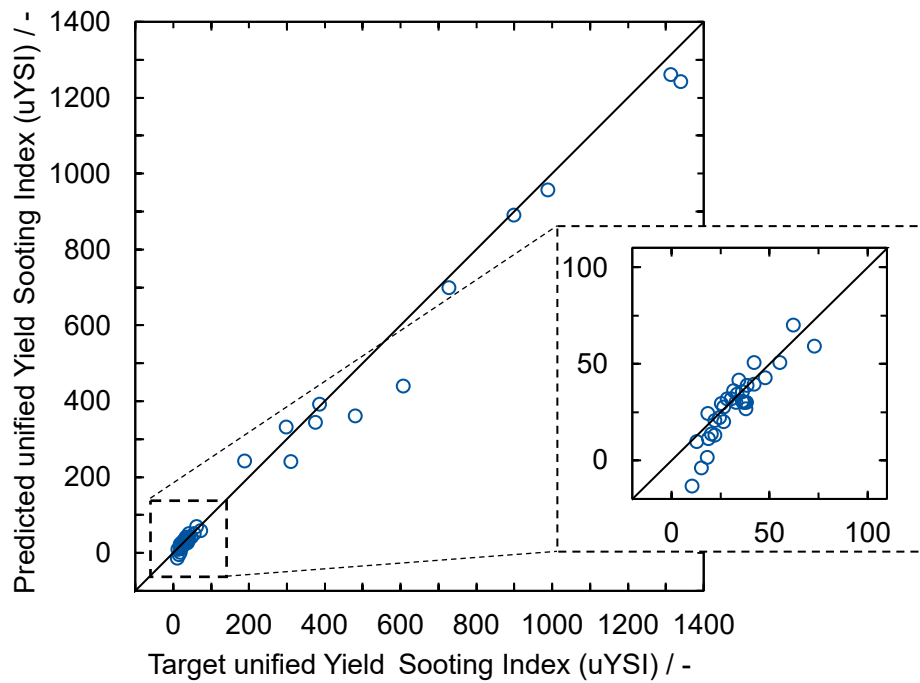


Figure D.7: Parity plot of predicted versus experimental unified yield sooting index (uYSI) of the test set.

- Gas-phase standard enthalpy of formation ($\Delta h_{f,\text{gas}}^0$) in kJ mol^{-1}

Kernel configuration: Lin + SE · Lin + WN

Number of data points for training: $N_{\text{Train}} = 697$

Number of data points for testing: $N_{\text{Test}} = 78$

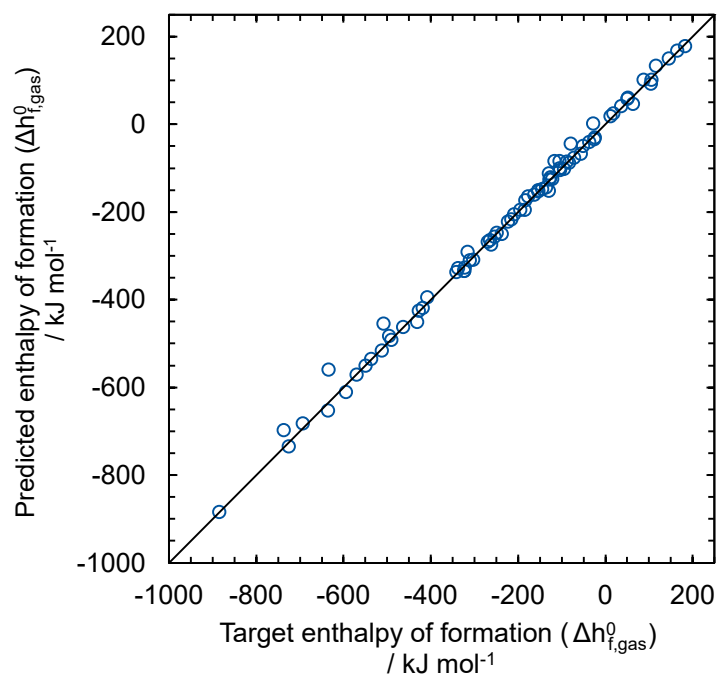


Figure D.8: Parity plot of predicted versus experimental gas-phase standard enthalpy of formation ($\Delta h_{f,\text{gas}}^0$) of the test set.

D.4 Contribution of fuel properties to expected engine efficiency increase

The predicted engine efficiency increase is attributed to the four considered fuel properties to varying degrees (Figures D.9 and D.10).

In the pure-component design, RON and Δh_{vap} contribute positively for all top 10 candidates, while octane sensitivity ($OS = RON - MON$) always contributes negatively (Figure D.9). For most candidates, the RON is the strongest contributor. The first five candidates are esters, for which both low octane sensitivity and low LBV impact efficiency negatively. For the seventh candidate fuel 1-butanol, Δh_{vap} has the biggest impact on efficiency.

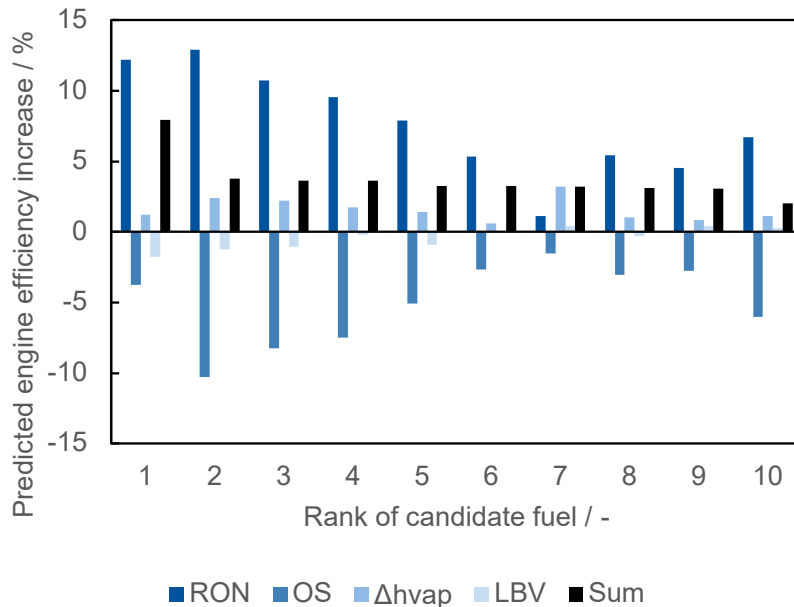


Figure D.9: Contributions of fuel properties to expected engine efficiency increase for the top 10 pure-component fuels.

RON: Research Octane Number, OS: Octane Sensitivity, Δh_{vap} : enthalpy of vaporisation, LBV: Laminar Burning Velocity, Sum: Sum over all four individual properties

For the binary ethanol blends, all four properties (RON, OS, Δh_{vap} , and LBV) positively impact engine efficiency (Figure D.10). The majority of the blends are composed of ethanol, and ethanol is well known for its high RON, OS, Δh_{vap} , and LBV. As the relevance of OS for engine efficiency strongly depends on the K-value, the engine

efficiency increase is plotted for two choices of the K-value: $K = -1.5$ and $K = -0.5$. Due to the linear contribution of the K-value, the predicted efficiency increase due to OS is simply divided by three, lowering the total relative efficiency increase of the top blend from 19.5% ($K = -1.5$) to 14.9% ($K = -0.5$). Noteworthy, a different choice for the K-value also changes the ranking. Considering the uncertainty of the K-factor, the ranking provides recommendations and is not a finalised assessment. To further prioritise among the vast number of promising blend candidates, expert knowledge could be utilised, e.g. on the potential for large-scale production.

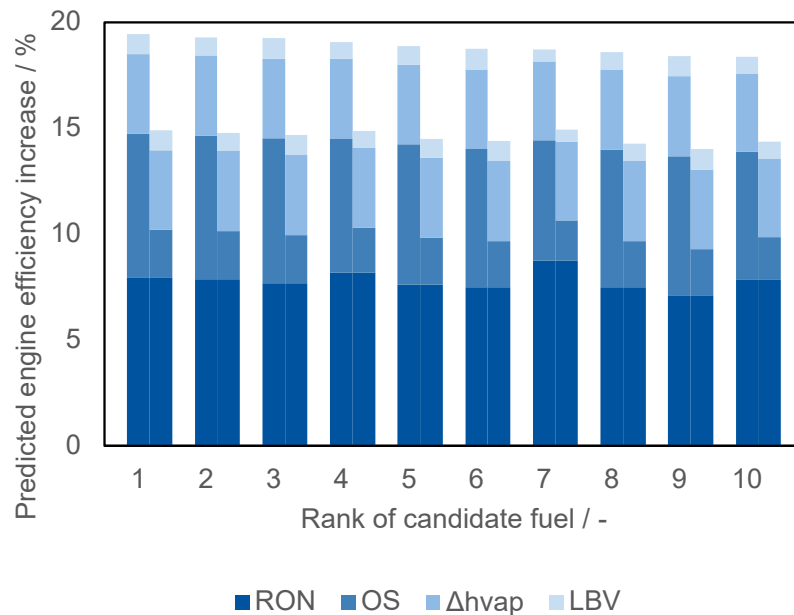


Figure D.10: Contributions of fuel properties to expected engine efficiency increase for the blend design for the top 10 binary blends with ethanol. The contributions are plotted for two K-values, -1.5 (left bar) and -0.5 (right bar), highlighting the influence of the K-value on the predicted engine efficiency increase.

RON: Research Octane Number, OS: Octane Sensitivity, Δh_{vap} : enthalpy of vaporisation, LBV: Laminar Burning Velocity.

D.5 Influence of property constraints on the number of feasible candidate blends

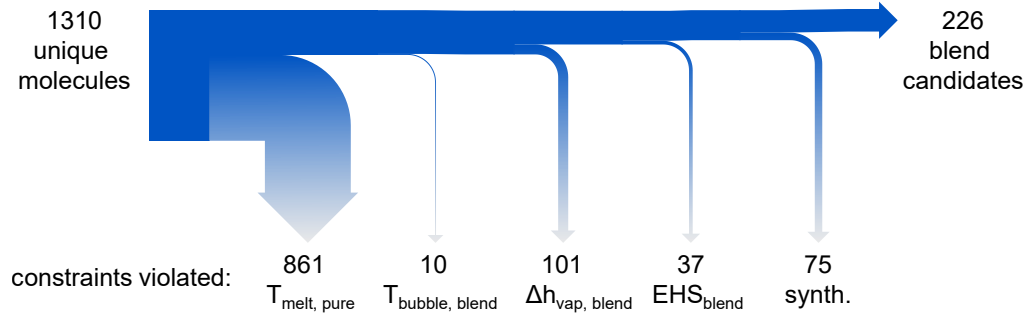


Figure D.11: Influence of property constraints on the number of feasible candidate blends.

D.6 Calculation and property data for the estimation of direct CO₂ emissions

Starting from Equation 6.1 in Section 6.2.4, the following formulation is used to calculate the fuel consumption of each fuel blend:

$$\frac{\text{merit}}{100\%} = \frac{\eta_{\text{blend}}}{\eta_{\text{RON95}}} - 1 = \frac{\frac{\dot{P}_{\text{engine}}^1}{\dot{m}_{\text{blend}} \cdot \text{LHV}_{\text{blend}}}}{\frac{\dot{P}_{\text{engine}}^1}{\dot{m}_{\text{RON95}} \cdot \text{LHV}_{\text{RON95}}}} - 1 \quad (\text{D.1})$$

$$\Leftrightarrow \dot{m}_{\text{blend}} = \frac{\dot{m}_{\text{RON95}} \cdot \text{LHV}_{\text{RON95}}}{\frac{\text{merit}}{100\%} + 1} \cdot \frac{1}{\text{LHV}_{\text{blend}}} \quad (\text{D.2})$$

For the fuel consumption of RON95 gasoline \dot{m}_{RON95} , $\dot{V}_{\text{RON95}} = 7 \text{ L}/100 \text{ km}$ is assumed. Assuming constant engine operation at the optimal operating point achieving maximum engine efficiency, as given by the merit value for each candidate blend, the CO₂ emissions for driving 100 km with a candidate blend equal:

$$\dot{m}_{\text{CO}_2} = M_{\text{CO}_2} \cdot \frac{\dot{m}_{\text{blend}}}{M_{\text{blend}}} \cdot N_{\text{C,blend}} \quad (\text{D.3})$$

The average number of carbon atoms per molecule in the blend $N_{C,\text{blend}}$ and the LHV of the blend LHV_{blend} are obtained from:

$$N_{C,\text{blend}} = \sum_{i=1}^2 x_i \cdot N_i \quad \text{and} \quad (\text{D.4})$$

$$LHV_{\text{blend}} = \sum_{i=1}^2 x_i \cdot LHV_i, \quad (\text{D.5})$$

where x_i is the molar fraction of molecule i in the blend and N_i is the number of carbon atoms of molecule i .

For the calculation, the following property data is used:

Property	Value	Reference
lower heating value (LHV)	42.13 kJ kg ⁻¹	Leitner et al. (2017)
density (ρ)	737 kg m ⁻³	Leitner et al. (2017)
molar mass (M)	100 kg kmol ⁻¹	vom Lehn et al. (2021b)
composition in mass fractions		Leitner et al. (2017)
w_C	0.8447	
w_H	0.1327	
w_O	0.0226	

Bibliography

- Ackermann, P., Braun, K. E., Burkardt, P., Heger, S., König, A., Morsch, P., Lehrheuer, B., Surger, M., Völker, S., Blank, L. M., Du, M., Heufer, K. A., Roß-Nickoll, M., Viell, J., von der Aßen, N., Mitsos, A., Pischinger, S., and Dahmen, M. (2021). Designed to Be Green, Economic, and Efficient: A Ketone-Ester-Alcohol-Alkane Blend for Future Spark-Ignition Engines. *ChemSusChem*, 14(23):5254–5264.
- Adjiman, C. S., Sahinidis, N. V., Vlachos, D. G., Bakshi, B., Maravelias, C. T., and Georgakis, C. (2021). Process Systems Engineering Perspective on the Design of Materials and Molecules. *Industrial & Engineering Chemistry Research*, 60(14):5194–5206.
- Adu, I. K., Sugiyama, H., Fischer, U., and Hungerbühler, K. (2008). Comparison of methods for assessing environmental, health and safety (EHS) hazards in early phases of chemical process design. *Process Safety and Environmental Protection*, 86(2):77–93.
- Aikawa, K., Sakurai, T., and Jetter, J. J. (2010). Development of a Predictive Model for Gasoline Vehicle Particulate Matter Emissions. *SAE International Journal of Fuels and Lubricants*, 3(2):610–622.
- Alexander, D. L. J., Tropsha, A., and Winkler, D. A. (2015). Beware of R^2 : simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *Journal of Chemical Information and Modeling*, 55(7):1316–1322.
- Alshehri, A. S., Gani, R., and You, F. (2020). Deep learning and knowledge-based methods for computer-aided molecular design—toward a unified approach: State-of-the-art and future directions. *Computers & Chemical Engineering*, 141:107005.
- Alshehri, A. S., Tula, A. K., You, F., and Gani, R. (2021). Next generation pure component property estimation models: With and without machine learning techniques. *AIChE Journal*, e17469.
- Altenburger, R., Nendza, M., and Schüürmann, G. (2003). Mixture toxicity and its modeling by quantitative structure-activity relationships. *Environmental Toxicology and Chemistry*, 22(8):1900–1915.

- Alwosheel, A., van Cranenburgh, S., and Chorus, C. G. (2018). Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*, 28:167–182.
- American Society for Testing and Materials (2000). *ASTM Standard Test Method E659-78*. The American Society for Testing and Materials, West Conshohocken, PA, United States.
- American Society for Testing and Materials (2021). *Standard Specification for Automotive Spark-Ignition Engine Fuel 75.160.20*. The American Society for Testing and Materials, West Conshohocken, PA, United States.
- Anastas, P. T. and Zimmerman, J. B. (2003). Peer Reviewed: Design Through the 12 Principles of Green Engineering. *Environmental Science & Technology*, 37(5):94A–101A.
- Ankley, G. T. and Villeneuve, D. L. (2006). The fathead minnow in aquatic toxicology: Past, present and future. *Aquatic Toxicology*, 78(1):91–102.
- Arnot, J. A. and Gobas, F. A. (2006). A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. *Environmental Reviews*, 14(4):257–297.
- Asprion, N., Böttcher, R., Mairhofer, J., Yliruka, M., Höller, J., Schwientek, J., Vanaret, C., and Bortz, M. (2019). Implementation and Application of Model-Based Design of Experiments in a Flowsheet Simulator. *Journal of Chemical & Engineering Data*, 65(3):1135–1145.
- Atkins, P. W. and Friedman, R. (2011). *Molecular quantum mechanics*. Oxford University Press, Oxford, United Kingdom, 5. edition.
- Atkinson, A. C. and Bogacka, B. (2002). Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics. *Chemometrics and Intelligent Laboratory Systems*, 61(1-2):17–33.
- Atkinson, A. C., Donev, A. N., and Tobias, R. (2006). *Optimum experimental designs, with SAS*, volume 34 of *Oxford statistical science series*. Oxford University Press, Oxford, United Kingdom.
- Austin, N. D., Sahinidis, N. V., and Trahan, D. W. (2017). A COSMO-based approach to computer-aided mixture design. *Chemical Engineering Science*, 159:93–105.

- Baehr, C., Smith, G. J., Sleeman, D., Zevaco, T. A., Raffelt, K., and Dahmen, N. (2022). Aldehydes and ketones in pyrolysis oil: analytical determination and their role in the aging process. *RSC Advances*, 12(12):7374–7382.
- Bakhiya, N. and Appel, K. E. (2010). Toxicity and carcinogenicity of furan in human diet. *Archives of Toxicology*, 84(7):563–578.
- Bakshi, B. R. (2019). Toward Sustainable Chemical Engineering: The Role of Process Systems Engineering. *Annual Review of Chemical and Biomolecular Engineering*, 10:265–288.
- Bard, Y. (1974). *Nonlinear parameter estimation*. Academic Press Inc., New York, NY, United States.
- Bardi, M., Di Lella, A., and Bruneaux, G. (2019). A novel approach for quantitative measurements of preferential evaporation of fuel by means of two-tracer laser induced fluorescence. *Fuel*, 239:521–533.
- Bausa, J. and Marquardt, W. (2000). Quick and reliable phase stability test in VLLE flash calculations by homotopy continuation. *Computers & Chemical Engineering*, 24(11):2447–2456.
- Bausa, J., Watzdorf, R. v., and Marquardt, W. (1998). Shortcut methods for nonideal multicomponent distillation: I. Simple columns. *AIChE Journal*, 44(10):2181–2198.
- Baxevanidis, P., Papadokonstantakis, S., Kokossis, A., and Marcoulaki, E. (2021). Group contribution–based LCA models to enable screening for environmentally benign novel chemicals in CAMD applications. *AIChE Journal*, e17544.
- Bazyleva, A., Abildskov, J., Anderko, A., Baudouin, O., Chernyak, Y., de Hemptinne, J.-C., Diky, V., Dohrn, R., Richard, J. E., Jacquemin, J., Jaubert, J.-N., Joback, K. G., Kattner, U. R., Kontogeorgis, G., Loria, H., Mathias, P. M., O’Connell, J. P., Schröer, W., Smith, G. J., Soto, A., Wang, S., and Weir, R. D. (2021). Good Reporting Practice for Thermophysical and Thermochemical Property Measurements (IUPAC Technical Report). *Pure and Applied Chemistry*, 93(2):253–272.
- Becke, A. D. (1993). Density–functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652.
- Behr, A., Ebbinghaus, P., and Naendrup, F. (2004). Process Concepts for the Transition Metal Catalyzed Syntheses of Formic Acid and Dimethylformamide Based on Carbon Dioxide. *Chemical Engineering & Technology*, 27(5):495–501.

- Behrens, R., von Harbou, E., Thiel, W. R., Böttinger, W., Ingram, T., Sieder, G., and Hasse, H. (2017). Monoalkylcarbonate Formation in Methyldiethanolamine–H₂O–CO₂. *Industrial & Engineering Chemistry Research*, 56(31):9006–9015.
- Biegler, L. T., Grossmann, I. E., and Westerberg, A. W. (1997). *Systematic Methods of Chemical Process Design*. Prentice Hall international series in the physical and chemical engineering sciences. Prentice Hall PTR, Upper Saddle River, NJ, United States.
- BP Europa (2021). BP Benzin Bleifrei 95: Safety Data Sheet No. SCH2106. BP Europa SE Hamburg, Zweigniederlassung BP (Switzerland) Zug, https://www.bp.com/content/dam/bp/country-sites/de_ch/switzerland/home/produkt-und-services/sicherheitsdatenblaetter/bp_benzin_bleifrei_95_eng.pdf, (accessed on December 2, 2021).
- Brown, B. J., Hanson, M. E., Liverman, D. M., and Merideth, R. W. (1987). Global sustainability: Toward definition. *Environ. Manage.*, 11(6):713–719.
- Bulicka, J. and Prochazka, J. (1976). Diffusion coefficients in some ternary systems. *Journal of Chemical & Engineering Data*, 21(4):452–456.
- Burger, B., Maffettone, P. M., Gusev, V. V., Aitchison, C. M., Bai, Y., Wang, X., Li, X., Alston, B. M., Li, B., Clowes, R., Rankin, N., Harris, B., Sprick, R. S., and Cooper, A. I. (2020). A mobile robotic chemist. *Nature*, 583(7815):237–241.
- Burger, J., Papaioannou, V., Gopinath, S., Jackson, G., Galindo, A., and Adjiman, C. S. (2015). A hierarchical method to integrated solvent and process design of physical CO₂ absorption using the SAFT- γ Mie approach. *AIChE Journal*, 61(10):3249–3269.
- Burkardt, P., Ottenwälder, T., König, A., Viell, J., Mitsos, A., Wouters, C., Marquardt, W., Pischinger, S., and Dahmen, M. (2021). Toward co-optimization of renewable fuel blend production and combustion in ultra-high efficiency SI engines. *International Journal of Engine Research*, 146808742110409.
- Calvo-Serrano, R., González-Miquel, M., Papadokostantakis, S., and Guillén-Gosálbez, G. (2018). Predicting the cradle-to-gate environmental impact of chemicals from molecular descriptors and thermodynamic properties via mixed-integer programming. *Computers & Chemical Engineering*, 108:179–193.
- Calvo-Serrano, R. and Guillén-Gosálbez, G. (2018). Streamlined Life Cycle Assessment under uncertainty integrating a network of the petrochemical industry and

- optimization techniques: Ecoinvent vs mathematical modeling. *ACS Sustainable Chemistry & Engineering*, 6(5):7109–7118.
- Canals, L. M. i., Azapagic, A., Doka, G., Jefferies, D., King, H., Mutel, C., Nemecek, T., Roches, A., Sim, S., and Stichnothe, H. (2011). Approaches for Addressing Life Cycle Assessment Data Gaps for Bio-based Products. *Journal of Industrial Ecology*, 15(5):707–725.
- Carney, J. G., Cunningham, P., and Bhagwan, U. (1999). Confidence and prediction intervals for neural network ensembles. *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339)*, vol. 2:1215–1218, IEEE.
- Chai, S., Song, Z., Zhou, T., Zhang, L., and Qi, Z. (2022). Computer-aided molecular design of solvents for chemical separation processes. *Current Opinion in Chemical Engineering*, 35:100732.
- Chai, S., Zhang, L., Du, J., Tula, A. K., Gani, R., and Eden, M. R. (2021). A Versatile Modeling Framework for Integrated Chemical Product Design. *Industrial & Engineering Chemistry Research*, 60(1):436–456.
- Chemangattuvalappil, N. G. (2020). Development of solvent design methodologies using computer-aided molecular design tools. *Current Opinion in Chemical Engineering*, 27:51–59.
- Chen, G., Song, Z., and Qi, Z. (2021). Transformer-convolutional neural network for surface charge density profile prediction: Enabling high-throughput solvent screening with COSMO-SAC. *Chemical Engineering Science*, 246:117002.
- Chen, L. and Stone, R. (2011). Measurement of Enthalpies of Vaporization of Isooctane and Ethanol Blends and Their Effects on PM Emissions from a GDI Engine. *Energy & Fuels*, 25(3):1254–1259.
- Chen, Q. and Grossmann, I. E. (2017). Recent developments and challenges in optimization-based process synthesis. *Annual Review of Chemical and Biomolecular Engineering*, 8:249–283.
- Chickos, J. S. and Acree, W. E. (2003). Enthalpies of Vaporization of Organic and Organometallic Compounds, 1880–2002. *Journal of Physical and Chemical Reference Data*, 32(2):519–878.
- Chilton, T. H. and Colburn, A. P. (1935). Distillation and Absorption in Packed Columns: A Convenient Design and Correlation Method. *Industrial & Engineering Chemistry*, 27(3):255–260.

- Clark, J. H. (2017). From waste to wealth using green chemistry: The way to long term stability. *Current Opinion in Green and Sustainable Chemistry*, 8:10–13.
- Clarke, C. J., Tu, W.-C., Levers, O., Bröhl, A., and Hallett, J. P. (2018). Green and Sustainable Solvents in Chemical Processes. *Chemical reviews*, 118(2):747–800.
- Clayton, A. D., Schweidtmann, A. M., Clemens, G., Manson, J. A., Taylor, C. J., Niño, C. G., Chamberlain, T. W., Kapur, N., Blacker, A. J., Lapkin, A. A., and Bourne, R. A. (2020). Automated self-optimisation of multi-step reaction and separation processes using machine learning. *Chemical Engineering Journal*, 384:123340.
- Coley, C. W., Thomas, D. A., Lummiss, J. A. M., Jaworski, J. N., Breen, C. P., Schultz, V., Hart, T., Fishman, J. S., Rogers, L., Gao, H., Hicklin, R. W., Plehiers, P. P., Byington, J., Piotti, J. S., Green, W. H., Hart, A. J., Jamison, T. F., and Jensen, K. F. (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science*, 365(6453):eaax1566.
- COSMOlogic GmbH & Co. KG (2017a). COSMOconf16 (4.1), Leverkusen, Germany.
- COSMOlogic GmbH & Co. KG (2017b). COSMOtherm17 (C30-1701), Leverkusen, Germany.
- COSMOlogic GmbH & Co. KG (2017c). TURBOMOLE 7.2, Leverkusen, Germany.
- Craig, D. K., Baskett, R. L., Davis, J. S., Dukes, L., Hansen, D. J., Petrocchi, A. J., Powell, T. J., Sutherland, P. J., and Tuccinardi, T. E. (1999). Recommended default methodology for analysis of airborne exposures to mixtures of chemicals in emergencies. *Applied Occupational and Environmental Hygiene*, 14(9):609–617.
- Dahmen, M. and Marquardt, W. (2015). A novel group contribution method for the prediction of the derived cetane number of oxygenated hydrocarbons. *Energy & Fuels*, 29(9):5781–5801.
- Dahmen, M. and Marquardt, W. (2016). Model-Based Design of Tailor-Made Biofuels. *Energy & Fuels*, 30(2):1109–1134.
- Dahmen, M. and Marquardt, W. (2017). Model-based formulation of biofuel blends by simultaneous product and pathway design. *Energy & Fuels*, 31(4):4096–4121.
- Das, D. D., St. John, P. C., McEnally, C. S., Kim, S., and Pfefferle, L. D. (2018). Measuring and predicting sooting tendencies of oxygenates, alkanes, alkenes, cycloalkanes, and aromatics on a unified scale. *Combustion and Flame*, 190:349–364.

- Dechambre, D., Pauls, C., Greiner, L., Leonhard, K., and Bardow, A. (2014a). Towards automated characterisation of liquid–liquid equilibria. *Fluid Phase Equilibria*, 362:328–334.
- Dechambre, D., Wolff, L., Pauls, C., and Bardow, A. (2014b). Optimal Experimental Design for the Characterization of Liquid–Liquid Equilibria. *Industrial & Engineering Chemistry Research*, 53(50):19620–19627.
- Deutz, S., Bongartz, D., Heuser, B., Kätelhön, A., Schulze Langenhorst, L., Omari, A., Walters, M., Klankermayer, J., Leitner, W., Mitsos, A., Pischinger, S., and Bardow, A. (2018). Cleaner production of cleaner fuels: wind-to-wheel – environmental assessment of CO₂-based oxymethylene ether as a drop-in fuel. *Energy & Environmental Science*, 11(2):331–343.
- DIN German Institute for Standardization (2017). *Automotive fuels - Unleaded petrol - Requirements and test methods; German version DIN EN 228:2012+A1:2017*. Beuth Verlag GmbH, Berlin.
- Dirrenberger, P., Glaude, P. A., Bounaceur, R., Le Gall, H., Da Cruz, A. P., Konnov, A. A., and Battin-Leclerc, F. (2014). Laminar burning velocity of gasolines with addition of ethanol. *Fuel*, 115:162–169.
- Dong, Q., Chirico, R. D., Yan, X., Hong, X., and Frenkel, M. (2005). Uncertainty Reporting for Experimental Thermodynamic Properties. *Journal of Chemical & Engineering Data*, 50(2):546–550.
- Dong, X., Liao, Z., Sun, J., Huang, Z., Jiang, B., Wang, J., and Yang, Y. (2020). Simultaneous Optimization for Organic Rankine Cycle Design and Heat Integration. *Industrial & Engineering Chemistry Research*, 59(46):20455–20471.
- Dong, Y., Zhu, R., Guo, Y., and Lei, Z. (2018). A United Chemical Thermodynamic Model: COSMO-UNIFAC. *Industrial & Engineering Chemistry Research*, 57(46):15954–15958.
- Döntgen, M., Przybylski-Freund, M.-D., Kröger, L. C., Kopp, W. A., Ismail, A. E., and Leonhard, K. (2015). Automated discovery of reaction pathways, rate constants, and transition states using reactive molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 11(6):2517–2524.
- Döntgen, M., Schmalz, F., Kopp, W. A., Kröger, L. C., and Leonhard, K. (2018). Automated Chemical Kinetic Modeling via Hybrid Reactive Molecular Dynamics and

- Quantum Chemistry Simulations. *Journal of Chemical Information and Modeling*, 58(7):1343–1355.
- Douglas, J. M. (1985). A hierarchical decision procedure for process synthesis. *AIChE Journal*, 31(3):353–362.
- Douguet, D., Munier-Lehmann, H., Labesse, G., and Pochet, S. (2005). LEA3D: a computer-aided ligand design for structure-based drug design. *Journal of Medicinal Chemistry*, 48(7):2457–2468.
- Draper, N. R. and Smith, H. (1998). *Applied regression analysis*. John Wiley & Sons, Inc., Hoboken, NJ, United States.
- Duarte, B. P., Atkinson, A. C., Granjo, J. F., and Oliveira, N. M. (2021). A model-based framework assisting the design of vapor-liquid equilibrium experimental plans. *Computers & Chemical Engineering*, 145:107168.
- Duvenaud, D. (2014). *Automatic model construction with Gaussian processes*. PhD thesis, Apollo - University of Cambridge Repository.
- Duvenaud, D., Lloyd, J., Grosse, R., Tenenbaum, J., and Zoubin, G. (2013). Structure Discovery in Nonparametric Regression through Compositional Kernel Search. In Dasgupta, S. and McAllester, D., editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1166–1174, Atlanta, GA, United States.
- ecoinvent (2019a). Market for cooling energy: Global, ecoinvent database version 3.6.
- ecoinvent (2019b). Market group for electricity, medium voltage: Europe without Switzerland, ecoinvent database version 3.6.
- Elsido, C., Martelli, E., and Grossmann, I. E. (2019). A bilevel decomposition method for the simultaneous heat integration and synthesis of steam/organic Rankine cycles. *Computers & Chemical Engineering*, 128:228–245.
- Elsido, C., Mian, A., and Martelli, E. (2017). A systematic methodology for the techno-economic optimization of Organic Rankine Cycles. *Energy Procedia*, 129:26–33.
- Emmenegger, M. F. (2000). Drying, natural gas: Rest-of-World, ecoinvent database version 3.6.

- Enders, S., Kahl, H., and Winkelmann, J. (2007). Surface Tension of the Ternary System Water + Acetone + Toluene. *Journal of Chemical & Engineering Data*, 52(3):1072–1079.
- European Commission-Joint Research Centre (2011). *International reference life cycle data system (ILCD) handbook: Recommendations for Life Cycle Impact Assessment in the European context*, volume 24571 of *EUR, Scientific and technical research series*. Publications Office, Luxembourg, 1st edition.
- Faber, F. A., Hutchison, L., Huang, B., Gilmer, J., Schoenholz, S. S., Dahl, G. E., Vinyals, O., Kearnes, S., Riley, P. F., and von Lilienfeld, O. A. (2017). Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *Journal of Chemical Theory and Computation*, 13(11):5255–5264.
- Farrell, J. T., Zigler, B. T., Ratcliff, M. A., Miles, P., Kolodziej, C., Sjoberg, M., Sluder, S., Szybist, J., Wagner, S., Splitter, D., Pihl, J., Toops, T., Debusk, M., Storey, J., and Vuilleumier, D. (2018). Co-Optimization of Fuels & Engines: Efficiency Merit Function for Spark-Ignition Engines; Revisions and Improvements Based on FY16-17 Research: Technical Report No. DOE/GO-102018-5041, National Renewable Energy Laboratory (NREL), Golden, CO, United States.
- Fedorov, V. V. and Leonov, S. L. (2014). *Optimal design for nonlinear response models*. Chapman & Hall / CRC Biostatistics Series. Taylor & Francis Group, Boca Raton, FL, United States, 1st edition.
- Fioroni, G., Fouts, L., Luecke, J., Vardon, D., Huq, N., Christensen, E., Huo, X., Alleman, T., McCormick, R., Kass, M., Polikarpov, E., Kukkadapu, G., and Whitesides, R. A. (2019). Screening of Potential Biomass-Derived Streams as Fuel Blendstocks for Mixing Controlled Compression Ignition Combustion. *SAE International Journal of Advances and Current Practices in Mobility*, 1(3):1117–1138.
- Fleitmann, L., Ackermann, P., Schilling, J., Kleinekorte, J., Rittig, J. G., vom Lehn, F., Schweidtmann, A. M., Pitsch, H., Leonhard, K., Mitsos, A., Bardow, A., and Dahmen, M. (2023a). Molecular Design of Fuels for Maximum Spark-Ignition Engine Efficiency by Combining Predictive Thermodynamics and Machine Learning. *Energy & Fuels*, 37(3):2213–2229.
- Fleitmann, L., Gertig, C., Scheffczyk, J., Schilling, J., Leonhard, K., and Bardow, A. (2023b). From Molecules to Heat-Integrated Processes: Computer-Aided Design of Solvents and Processes Using Quantum Chemistry. *Chemie Ingenieur Technik*, 95(3):368–380.

- Fleitmann, L., Kleinekorte, J., Leonhard, K., and Bardow, A. (2021a). COSMO-susCAMPD: Sustainable solvents from combining computer-aided molecular and process design with predictive life cycle assessment. *Chemical Engineering Science*, 245:116863.
- Fleitmann, L., Pyschik, J., Wolff, L., and Bardow, A. (2021b). Optimal physical property data for process simulations by optimal experimental design. In *31st European Symposium on Computer Aided Process Engineering*, volume 50 of *Computer Aided Chemical Engineering*, pages 851–857. Elsevier, Amsterdam, The Netherlands.
- Fleitmann, L., Pyschik, J., Wolff, L., Schilling, J., and Bardow, A. (2022). Optimal experimental design of physical property measurements for optimal chemical process simulations. *Fluid Phase Equilibria*, 113420.
- Forte, E., Kulkarni, A., Burger, J., Bortz, M., Küfer, K.-H., and Hasse, H. (2020). Multi-criteria optimization for parametrizing excess Gibbs energy models. *Fluid Phase Equilibria*, 522:112676.
- Forte, E., von Harbou, E., Burger, J., Asprion, N., and Bortz, M. (2017). Optimal Design of Laboratory and Pilot-Plant Experiments Using Multiobjective Optimization. *Chemie Ingenieur Technik*, 89(5):645–654.
- Franceschini, G. and Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, 63(19):4846–4872.
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009). Gaussian09 Rev. D.01, Wallingford, CT, United States.

- Gani, R. (2004). Chemical product design: challenges and opportunities. *Computers & Chemical Engineering*, 28(12):2441–2457.
- Gani, R. (2019). Group contribution-based property estimation methods: advances and perspectives. *Current Opinion in Chemical Engineering*, 23:184–196.
- Gani, R. and Ng, K. M. (2015). Product design – Molecules, devices, functional products, and formulated products. *Computers & Chemical Engineering*, 81:70–79.
- Gani, R. and Zhang, L. (2020). Editorial overview: Frontiers of Chemical Engineering: Chemical Product Design. *Current Opinion in Chemical Engineering*, 27:A1–A3.
- Gao, W. and Coley, C. W. (2020). The Synthesizability of Molecules Proposed by Generative Models. *Journal of Chemical Information and Modeling*, 60(12):5714–5723.
- García-Ródenas, R., García-García, J. C., López-Fidalgo, J., Martín-Baos, J. Á., and Wong, W. K. (2020). A comparison of general-purpose optimization algorithms for finding optimal approximate experimental designs. *Computational Statistics & Data Analysis*, 144:106844.
- Garrido-Merchán, E. C. and Hernández-Lobato, D. (2020). Dealing with categorical and integer-valued variables in Bayesian Optimization with Gaussian processes. *Neurocomputing*, 380:20–35.
- Gertig, C., Fleitmann, L., Hemprich, C., Hense, J., Bardow, A., and Leonhard, K. (2021). CAT-COSMO-CAMPD: Integrated in silico design of catalysts and processes based on quantum chemistry. *Computers & Chemical Engineering*, 153:107438.
- Gertig, C., Fleitmann, L., Schilling, J., Leonhard, K., and Bardow, A. (2020a). Rx-COSMO-CAMPD: Enhancing Reactions by Integrated Computer-Aided Design of Solvents and Processes based on Quantum Chemistry. *Chemie Ingenieur Technik*, 92(10):1489–1500.
- Gertig, C., Leonhard, K., and Bardow, A. (2020b). Computer-aided molecular and processes design based on quantum chemistry: current status and future prospects. *Current Opinion in Chemical Engineering*, 27:89–97.
- Gevers, M. and Ljung, L. (1986). Optimal experiment designs with respect to the intended model application. *Automatica*, 22(5):543–554.

- Ghysels, Verstraelen, T., Hemelsoet, K., Waroquier, M., and van Speybroeck, V. (2010). TAMkin: a versatile package for vibrational analysis and chemical kinetics. *Journal of Chemical Information and Modeling*, 50(9):1736–1750.
- Glasco, D. L. and Bell, J. G. (2021). Electrochemical oscillations during the oxidation of 2-(dimethylamino)ethanethiol. *Journal of Electroanalytical Chemistry*, 902:115822.
- Gmehling, J. and Kleiber, M. (2014). Vapor–Liquid Equilibrium and Physical Properties for Distillation. In *Distillation*, pages 45–95. Academic Press Inc., New York, NY, United States.
- Goedkoop, M., Heijungs, R., Huijbregts, M., Schryver, Struijs, J., and van Zelm, R. (2009). ReCiPe 2008. *A life cycle impact assessment method which comprises harmonised category indicators at the midpoint and the endpoint level*, 1:1–126.
- Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., and Aspuru-Guzik, A. (2018). Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science*, 4(2):268–276.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. The MIT press, Cambridge, MA, United States.
- Gopinath, S., Jackson, G., Galindo, A., and Adjiman, C. S. (2016). Outer approximation algorithm with physical domain reduction for computer-aided molecular and separation process design. *AIChE Journal*, 62(9):3484–3504.
- Gottschalk, H. C., Poblitzki, A., Suhm, M. A., Al-Mogren, M. M., Antony, J., Auer, A. A., Baptista, L., Benoit, D. M., Bistoni, G., Bohle, F., Dahmani, R., Firaha, D., Grimme, S., Hansen, A., Harding, M. E., Hochlaf, M., Holzer, C., Jansen, G., Klopper, W., Kopp, W. A., Kröger, L. C., Leonhard, K., Mouhib, H., Neese, F., Pereira, M. N., Ulusoy, I. S., Wuttke, A., and Mata, R. A. (2018). The furan microsolvation blind challenge for quantum chemical methods: First steps. *The Journal of Chemical Physics*, 148(1):014301.
- Griffiths, R.-R. and Hernández-Lobato, J. M. (2020). Constrained Bayesian optimization for automatic chemical design using variational autoencoders. *Chemical Science*, 11(2):577–586.
- Grossmann, I. E. and Harjunkoski, I. (2019). Process Systems Engineering: Academic and industrial perspectives. *Computers & Chemical Engineering*, 126:474–484.

- Gschwend, D., Müller, S., Wokaun, A., and Vogel, F. (2018). Optimum Fuel for Spark Ignition Engines from Lignin Pyrolysis Oil. *Energy & Fuels*, 32(9):9388–9398.
- Gschwend, D., Soltic, P., Edinger, P., Wokaun, A., and Vogel, F. (2017). Performance evaluation of gasoline alternatives using a thermodynamic spark-ignition engine model. *Sustainable Energy Fuels*, 1(9):1991–2005.
- Gschwend, D., Soltic, P., Wokaun, A., and Vogel, F. (2019). Review and Performance Evaluation of Fifty Alternative Liquid Fuels for Spark-Ignition Engines. *Energy & Fuels*, 33(3):2186–2196.
- Gugisch, R., Kerber, A., Kohnert, A., Laue, R., Meringer, M., Rücker, C., and Wassermann, A. (2015). MOLGEN 5.0, a molecular structure generator. In C. Basak, S., Restrepo, G., and L. Villaveces, J., editors, *Advances in Mathematical Chemistry and Applications*, pages 113–138. Bentham Science Publishers, Amsterdam, The Netherlands.
- Guillén-Gosálbez, G., You, F., Galán-Martín, Á., Pozo, C., and Grossmann, I. E. (2019). Process systems engineering thinking and tools applied to sustainability problems: current landscape and future opportunities. *Current Opinion in Chemical Engineering*, 26:170–179.
- Han, C. and Chaloner, K. (2003). D- and c-optimal designs for exponential regression models used in viral dynamics and other applications. *Journal of Statistical Planning and Inference*, 115(2):585–601.
- Häse, F., Aldeghi, M., Hickman, R. J., Roch, L. M., and Aspuru-Guzik, A. (2021). Gryffin : An algorithm for Bayesian optimization of categorical variables informed by expert knowledge. *Applied Physics Reviews*, 8(3):031406.
- Häse, F., Roch, L. M., and Aspuru-Guzik, A. (2019). Next-Generation Experimentation with Self-Driving Laboratories. *Trends in Chemistry*, 1(3):282–291.
- Hashim, H., Narayanasamy, M., Yunus, N. A., Shiun, L. J., Muis, Z. A., and Ho, W. S. (2017). A cleaner and greener fuel: Biofuel blend formulation and emission assessment. *Journal of Cleaner Production*, 146:208–217.
- Hauschild, M. Z. and Huijbregts, M. A., editors (2015). *Life cycle impact assessment. LCA Compendium*. Springer, Dordrecht, The Netherlands.
- Hechinger, M. (2014). *Model-based identification of promising biofuel candidates for spark-ignited engines: Dissertation, RWTH Aachen University*, volume 940

- of *Berichte aus der Aachener Verfahrenstechnik - Prozesstechnik*. VDI-Verlage, Düsseldorf, Germany.
- Hechinger, M., Dahmen, M., Villeda, J. J. V., and Marquardt, W. (2012). Rigorous Generation and Model-Based Selection of Future Biofuel Candidates. In Iftekhar A. Karimi and Rajagopalan Srinivasan, editors, *11th International Symposium on Process Systems Engineering*, volume 31 of *Computer Aided Chemical Engineering*, pages 1341–1345. Elsevier.
- Hellweg, S., Fischer, U., Scheringer, M., and Hungerbühler, K. (2004). Environmental assessment of chemicals: methods and application to a case study of organic solvents. *Green Chemistry*, 6(8):418–427.
- Hellweg, S. and Milà i Canals, L. (2014). Emerging approaches, challenges and opportunities in life cycle assessment. *Science*, 344(6188):1109–1113.
- Hischier, R. (2019). Toluene production, liquid: Europe without Switzerland, ecoinvent database version 3.6.
- Holland-Letz, T. (2017). On the combination of c- and D-optimal designs: General approaches and applications in dose-response studies. *Biometrics*, 73(1):206–213.
- Holland-Letz, T., Gunkel, N., Amtmann, E., and Kopp-Schneider, A. (2018). Parametric modeling and optimal experimental designs for estimating isobolograms for drug interactions in toxicology. *Journal of biopharmaceutical statistics*, 28(4):763–777.
- Holland-Letz, T. and Kopp-Schneider, A. (2018). Optimal experimental designs for estimating the drug combination index in toxicology. *Computational Statistics & Data Analysis*, 117:182–193.
- Hopp, M. and Gross, J. (2017). Thermal Conductivity of Real Substances from Excess Entropy Scaling Using PCP-SAFT. *Industrial & Engineering Chemistry Research*, 56(15):4527–4538.
- Hopp, M. and Gross, J. (2019). Thermal Conductivity from Entropy Scaling: A Group-Contribution Method. *Industrial & Engineering Chemistry Research*, 58(44):20441–20449.
- Hoppe, F., Burke, U., Thewes, M., Heufer, A., Kremer, F., and Pischinger, S. (2016a). Tailor-Made Fuels from Biomass: Potentials of 2-butanone and 2-methylfuran in direct injection spark ignition engines. *Fuel*, 167:106–117.

- Hoppe, F., Heuser, B., Thewes, M., Kremer, F., Pischinger, S., Dahmen, M., Hechinger, M., and Marquardt, W. (2016b). Tailor-made fuels for future engine concepts. *International Journal of Engine Research*, 17(1):16–27.
- Houska, B., Telen, D., Logist, F., Diehl, M., and van Impe, J. F. (2015). An economic objective for the optimal experiment design of nonlinear dynamic processes. *Automatica*, 51:98–103.
- Hukkerikar, A. S., Kalakul, S., Sarup, B., Young, D. M., Sin, G., and Gani, R. (2012a). Estimation of Environment-Related Properties of Chemicals for Design of Sustainable Processes: Development of Group-Contribution+ (GC+) Property Models and Uncertainty Analysis. *Journal of Chemical Information and Modeling*, 52(11):2823–2839.
- Hukkerikar, A. S., Sarup, B., ten Kate, A., Abildskov, J., Sin, G., and Gani, R. (2012b). Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilibria*, 321:25–43.
- Huo, X., Huq, N. A., Stunkel, J., Cleveland, N. S., Starace, A. K., Settle, A. E., York, A. M., Nelson, R. S., Brandner, D. G., Fouts, L., St. John, P. C., Christensen, E. D., Luecke, J., Mack, J. H., McEnally, C. S., Cherry, P. A., Pfefferle, L. D., Strathmann, T. J., Salvachúa, D., Kim, S., McCormick, R. L., Beckham, G. T., and Vardon, D. R. (2019). Tailoring diesel bioblendstock from integrated catalytic upgrading of carboxylic acids: a “fuel property first” approach. *Green Chemistry*, 21(21):5813–5827.
- Huq, N. A., Huo, X., Hafenstine, G. R., Tift, S. M., Stunkel, J., Christensen, E. D., Fioroni, G. M., Fouts, L., McCormick, R. L., Cherry, P. A., McEnally, C. S., Pfefferle, L. D., Wiatrowski, M. R., Benavides, P. T., Bidy, M. J., Connatser, R. M., Kass, M. D., Alleman, T. L., St John, P. C., Kim, S., and Vardon, D. R. (2019). Performance-advantaged ether diesel bioblendstock production by a priori design. *Proceedings of the National Academy of Sciences of the United States of America*, 116(52):26421–26430.
- International Energy Agency (2018). *The Future of Petrochemicals: Towards more sustainable plastics and fertilisers*. IEA Publications, Paris, France.
- International Energy Agency (2021). *World Energy Outlook*. IEA Publications, Paris, France.

- ISO 14040 (2006). Life Cycle Assessment: principles and framework. *Environmental management*.
- Itani, L. M., Bruneaux, G., Di Lella, A., and Schulz, C. (2015). Two-tracer LIF imaging of preferential evaporation of multi-component gasoline fuel sprays under engine conditions. *Proceedings of the Combustion Institute*, 35(3):2915–2922.
- Jens, C. M., Müller, L., Leonhard, K., and Bardow, A. (2019). To Integrate or Not to Integrate—Techno-Economic and Life Cycle Assessment of CO₂ Capture and Conversion to Methyl Formate Using Methanol. *ACS Sustainable Chemistry & Engineering*, 7(14):12270–12280.
- Jens, C. M., Nowakowski, K., Scheffczyk, J., Leonhard, K., and Bardow, A. (2016). CO from CO₂ and fluctuating renewable energy via formic-acid derivatives. *Green Chemistry*, 18(20):5621–5629.
- Jhamb, S., Liang, X., Gani, R., and Kontogeorgis, G. M. (2019). Systematic Model-Based Methodology for Substitution of Hazardous Chemicals. *ACS Sustainable Chemistry & Engineering*, 7(8):7652–7666.
- Jimenez-Gonzalez, C. (2019). Life cycle considerations of solvents. *Current Opinion in Green and Sustainable Chemistry*, 18:66–71.
- Jiménez-González, C., Kim, S., and Overcash, M. R. (2000). Methodology for developing gate-to-gate Life cycle inventory information. *The International Journal of Life Cycle Assessment*, 5(3):153–159.
- Jirasek, F., Alves, R. A. S., Damay, J., Vandermeulen, R. A., Bamler, R., Bortz, M., Mandt, S., Kloft, M., and Hasse, H. (2020). Machine Learning in Thermodynamics: Prediction of Activity Coefficients by Matrix Completion. *The Journal of Physical Chemistry Letters*, 11(3):981–985.
- Jirasek, F. and Hasse, H. (2021). Perspective: Machine Learning of Thermophysical Properties. *Fluid Phase Equilibria*, 549:113206.
- Joback, K. G. and Reid, R. C. (1987). Estimation of pure-component properties from group-contributions. *Chemical Engineering Communications*, 57(1-6):233–243.
- Jonuzaj, S. and Adjiman, C. S. (2017). Designing optimal mixtures using generalized disjunctive programming: Hull relaxations. *Chemical Engineering Science*, 159:106–130.

- Jonuzaj, S., Cui, J., and Adjiman, C. S. (2019). Computer-aided design of optimal environmentally benign solvent-based adhesive products. *Computers & Chemical Engineering*, 130:106518.
- Jonuzaj, S., Gupta, A., and Adjiman, C. S. (2018). The design of optimal mixtures from atom groups using Generalized Disjunctive Programming. *Computers & Chemical Engineering*, 116:401–421.
- Jungbluth, N. (2019). Heavy fuel oil, burned in refinery furnace: Europe without Switzerland, ecoinvent database version 3.6.
- Kaiser, S. and Engell, S. (2020). Integrating Superstructure Optimization under Uncertainty and Optimal Experimental Design in early Stage Process Development. In *30th European Symposium on Computer Aided Process Engineering*, volume 48 of *Computer Aided Chemical Engineering*, pages 799–804. Elsevier, Amsterdam, The Netherlands.
- Kaiser, S., Menzel, T., and Engell, S. (2021). Focusing experiments in the early phase process design by process optimization and global sensitivity analysis. In *31st European Symposium on Computer Aided Process Engineering*, volume 50 of *Computer Aided Chemical Engineering*, pages 899–904. Elsevier, Amsterdam, The Netherlands.
- Kalakul, S., Zhang, L., Fang, Z., Choudhury, H. A., Intikhab, S., Elbashir, N., Eden, M. R., and Gani, R. (2018). Computer aided chemical product design – ProCAPD and tailor-made blended products. *Computers & Chemical Engineering*, 116:37–55.
- Kalghatgi, G. T. (2001). Fuel Anti-Knock Quality - Part I. Engine Studies: SAE Technical Paper 2001-01-3584. SAE Technical Paper Series. SAE International, 400 Commonwealth Drive, Warrendale, PA, United States.
- Kaminski, S., Kirgios, E., Bardow, A., and Leonhard, K. (2017). Improved Property Predictions by Combination of Predictive Models. *Industrial & Engineering Chemistry Research*, 56(11):3098–3106.
- Kaminski, S. and Leonhard, K. (2020). SEPP: Segment-Based Equation of State Parameter Prediction. *Journal of Chemical & Engineering Data*, 65(12):5830–5843.
- Karka, P., Papadokostantakis, S., and Kokossis, A. (2022). Digitizing sustainable process development: From ex-post to ex-ante LCA using machine-learning to evaluate bio-based process technologies ahead of detailed design. *Chemical Engineering Science*, 250:117339.

- Kashinath, S. A. A., Manan, Z. A., Hashim, H., and Alwi, S. R. W. (2012). Design of green diesel from biofuels using computer aided technique. *Computers & Chemical Engineering*, 41:88–92.
- Kassai, M., Aksu, C., Shiraishi, T., Cracknell, R., and Shibuya, M. (2019). Mechanism Analysis on the Effect of Fuel Properties on Knocking Performance at Boosted Conditions: SAE Technical Paper 2019-01-0035. SAE Technical Paper Series. SAE International, 400 Commonwealth Drive, Warrendale, PA, United States.
- Keßler, T., Kunde, C., Linke, S., Sundmacher, K., and Kienle, A. (2021). Integrated Computer-Aided Molecular and Process Design: Green Solvents for the Hydroformylation of Long-Chain Olefines. *Chemical Engineering Science*, 117243.
- Kirkpatrick, J., McMorro, B., Turban, D. H. P., Gaunt, A. L., Spencer, J. S., Matthews, A. G. D. G., Obika, A., Thiry, L., Fortunato, M., Pfau, D., Castellanos, L. R., Petersen, S., Nelson, A. W. R., Kohli, P., Mori-Sánchez, P., Hassabis, D., and Cohen, A. J. (2021). Pushing the frontiers of density functionals by solving the fractional electron problem. *Science*, 374(6573):1385–1389.
- Klamt, A. (2003). Prediction of the mutual solubilities of hydrocarbons and water with COSMO-RS. *Fluid Phase Equilibria*, 206(1-2):223–235.
- Klamt, A., Eckert, F., and Arlt, W. (2010). COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures. *Annual Review of Chemical and Biomolecular Engineering*, 1:101–122.
- Kleinekorte, J., Fleitmann, L., Bachmann, M., Kätelhön, A., Barbosa-Póvoa, A., von der Assen, N., and Bardow, A. (2020). Life Cycle Assessment for the Design of Chemical Processes, Products, and Supply Chains. *Annual Review of Chemical and Biomolecular Engineering*, 11:203–233.
- Kleinekorte, J., Kröger, L., Leonhard, K., and Bardow, A. (2019). A neural network-based framework to predict process-specific environmental impacts. In Kiss, A. A., Zondervan, E., Lakerveld, R., and Özkan, L., editors, *29th European Symposium on Computer Aided Process Engineering*, volume 47 of *Computer-Aided Chemical Engineering*, pages 1447–1452. Elsevier, Amsterdam, The Netherlands.
- Kohn, W. and Sham, L. J. (1965). Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140(4A):A1133–A1138.

- Kong, L., Sen, S. M., Henao, C. A., Dumesic, J. A., and Maravelias, C. T. (2016). A superstructure-based framework for simultaneous process synthesis, heat integration, and utility plant design. *Computers & Chemical Engineering*, 91:68–84.
- König, A., Marquardt, W., Mitsos, A., Viell, J., and Dahmen, M. (2020a). Integrated design of renewable fuels and their production processes: recent advances and challenges. *Current Opinion in Chemical Engineering*, 27:45–50.
- König, A., Neidhardt, L., Viell, J., Mitsos, A., and Dahmen, M. (2020b). Integrated design of processes and products: Optimal renewable fuels. *Computers & Chemical Engineering*, 134:106712.
- König, A., Siska, M., Schweidtmann, A. M., Rittig, J. G., Viell, J., Mitsos, A., and Dahmen, M. (2021). Designing production-optimal alternative fuels for conventional, flexible-fuel, and ultra-high efficiency engines. *Chemical Engineering Science*, 237:116562.
- Kontogeorgis, G. M., Dohrn, R., Economou, I. G., de Hemptinne, J.-C., ten Kate, A., Kuitunen, S., Mooijer, M., Žilnik, L. F., and Vesovic, V. (2021). Industrial Requirements for Thermodynamic and Transport Properties: 2020. *Industrial & Engineering Chemistry Research*, 60(13):4987–5013.
- Kontogeorgis, G. M., Mattei, M., Ng, K. M., and Gani, R. (2019). An Integrated Approach for the Design of Emulsified Products. *AIChE Journal*, 65(1):75–86.
- Kranz, P. and Kaiser, S. A. (2019). LIF-based imaging of preferential evaporation of a multi-component gasoline surrogate in a direct-injection engine. *Proceedings of the Combustion Institute*, 37(2):1365–1372.
- Krenn, M., Häse, F., Nigam, A., Friederich, P., and Aspuru-Guzik, A. (2020). Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation. *Machine Learning: Science and Technology*, 1(4):045024.
- Krep, L., Roy, I. S., Kopp, W., Schmalz, F., Huang, C., and Leonhard, K. (2022). Efficient Reaction Space Exploration with ChemTraYzer-TAD. *Journal of Chemical Information and Modeling*, 62(4):890–902.
- Kröger, L. C., Müller, S., Smirnova, I., and Leonhard, K. (2020). Prediction of Solvation Free Energies of Ionic Solutes in Neutral Solvents. *The Journal of Physical Chemistry. A*, 124(20):4171–4181.

- Kruber, K. F., Grüters, T., and Skiborowski, M. (2021). Advanced hybrid optimization methods for the design of complex separation processes. *Computers & Chemical Engineering*, 107257.
- Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *The annals of mathematical statistics*, 22(1):79–86.
- Kuzhagaliyeva, N., Horváth, S., Williams, J., Nicolle, A., and Sarathy, S. M. (2022). Artificial intelligence-driven design of fuel mixtures. *Communications Chemistry*, 5(1):1–10.
- Kuzmanović, B., van Delden, M. L., Kuipers, N. J. M., and de Haan, A. B. (2003). Fully Automated Workstation for Liquid–Liquid Equilibrium Measurements. *Journal of Chemical & Engineering Data*, 48(5):1237–1244.
- Larsen, U., Johansen, T., and Schramm, J. (2009). Ethanol as a Future Fuel for Road Transportation: Main report. International Energy Agency, <https://www.osti.gov/etdeweb/biblio/21580860>.
- Leach, F., Chapman, E., Jetter, J. J., Rubino, L., Christensen, E. D., St. John, P. C., Fioroni, G. M., and McCormick, R. L. (2021). A Review and Perspective on Particulate Matter Indices Linking Fuel Composition to Particulate Emissions from Gasoline Engines. *SAE International Journal of Fuels and Lubricants*, 15(1):3–28.
- Lee, U., Burre, J., Caspari, A., Kleinekorte, J., Schweidtmann, A. M., and Mitsos, A. (2016). Techno-economic Optimization of a Green-Field Post-Combustion CO₂ Capture Process Using Superstructure and Rate-Based Models. *Industrial & Engineering Chemistry Research*, 55(46):12014–12026.
- Leitner, W., Klankermayer, J., Pischinger, S., Pitsch, H., and Kohse-Höinghaus, K. (2017). Advanced Biofuels and Beyond: Chemistry Solutions for Propulsion and Production. *Angewandte Chemie (International ed.)*, 56(20):5412–5452.
- Li, R., Herreros, J. M., Tsolakis, A., and Yang, W. (2021). Machine learning and deep learning enabled fuel sooting tendency prediction from molecular structure. *Journal of Molecular Graphics & Modelling*, 111:108083.
- Li, R., Herreros, J. M., Tsolakis, A., and Yang, W. (2022). Integrated machine learning-quantitative structure property relationship (ML-QSPR) and chemical kinetics for high throughput fuel screening toward internal combustion engine. *Fuel*, 307:121908.

- Liesche, G., Schack, D., and Sundmacher, K. (2019). The FluxMax approach for simultaneous process synthesis and heat integration: Production of hydrogen cyanide. *AIChE Journal*, 65(7):e16554.
- Lin, S.-T., Chang, J., Wang, S., Goddard, W. A., and Sandler, S. I. (2004). Prediction of Vapor Pressures and Enthalpies of Vaporization Using a COSMO Solvation Model. *The Journal of Physical Chemistry A*, 108(36):7429–7439.
- Lindsey, C. and Sheather, S. (2010). Variable selection in linear regression. *The Stata Journal*, 10(4):650–669.
- Linke, S., McBride, K., and Sundmacher, K. (2020). Systematic green solvent selection for the hydroformylation of long chained alkenes. *ACS Sustainable Chemistry & Engineering*, 8(29):10795–10811.
- Liu, Q., Zhang, L., Liu, L., Du, J., Meng, Q., and Gani, R. (2019a). Computer-aided reaction solvent design based on transition state theory and COSMO-SAC. *Chemical Engineering Science*, 202:300–317.
- Liu, Q., Zhang, L., Liu, L., Du, J., Tula, A. K., Eden, M., and Gani, R. (2019b). OptCAMD: An optimization-based framework and tool for molecular and mixture product design. *Computers & Chemical Engineering*, 124:285–301.
- Liu, Q., Zhang, L., Tang, K., Feng, Y., Zhang, J., Zhuang, Y., Liu, L., and Du, J. (2019c). Computer-aided reaction solvent design considering inertness using group contribution-based reaction thermodynamic model. *Chemical Engineering Research and Design*, 152:123–133.
- Liu, Q., Zhang, L., Tang, K., Liu, L., Du, J., Meng, Q., and Gani, R. (2021). Machine learning-based atom contribution method for the prediction of surface charge density profiles and solvent design. *AIChE Journal*, 67(2):e17110.
- Lloyd, J., Duvenaud, D., Grosse, R., Tenenbaum, J., and Ghahramani, Z. (2014). Automatic Construction and Natural-Language Description of Nonparametric Regression Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 28(1).
- Loschen, C. and Klamt, A. (2016). New Developments in Prediction of Solid-State Solubility and Cocrystallization Using COSMO-RS Theory. In Abramov, Y. A., editor, *Computational pharmaceutical solid state chemistry*, pages 211–233. John Wiley & Sons, Inc., Hoboken, NJ, United States.

- Lucia, S. and Paulen, R. (2014). Robust Nonlinear Model Predictive Control with Reduction of Uncertainty via Robust Optimal Experiment Design. *IFAC Proceedings Volumes*, 47(3):1904–1909.
- Majer, V. and Svoboda, V. (1985). *Enthalpies of vaporization of organic compounds: A critical review and data compilation*, volume 32 of *Chemical data series*. Blackwell Scientific Publications, Oxford, United Kingdom.
- Marrero, J. and Gani, R. (2001). Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria*, 183-184:183–208.
- Martín, M. and Adams II, T. A. (2019). Challenges and future directions for process and product synthesis and design. *Computers & Chemical Engineering*, 128:421–436.
- Marvin, W. A., Rangarajan, S., and Daoutidis, P. (2013). Automated Generation and Optimal Selection of Biofuel-Gasoline Blends and Their Synthesis Routes. *Energy & Fuels*, 27(6):3585–3594.
- Mathias, P. M. (2016). Effect of VLE uncertainties on the design of separation sequences by distillation – Study of the benzene–chloroform–acetone system. *Fluid Phase Equilibria*, 408:265–272.
- McBride, K., Linke, S., Xu, S., and Sundmacher, K. (2018). Computer Aided Design of Green Thermomorphic Solvent Systems for Homogeneous Catalyst Recovery. In *13th International Symposium on Process Systems Engineering (PSE 2018)*, volume 44 of *Computer Aided Chemical Engineering*, pages 1783–1788. Elsevier, Amsterdam, The Netherlands.
- McCormick, R. L., Fioroni, G., Fouts, L., Christensen, E., Yanowitz, J., Polikarpov, E., Albrecht, K., Gaspar, D. J., Gladden, J., and George, A. (2017). Selection Criteria and Screening of Potential Biomass-Derived Streams as Fuel Blendstocks for Advanced Spark-Ignition Engines. *SAE International Journal of Fuels and Lubricants*, 10(2):442–460.
- McEnally, C. S., Das, D. D., and Pfefferle, L. D. (2017). Detailed YSI Database Volume 2.tab. In *Yield Sooting Index Database Volume 2: Sooting Tendencies of a Wide Range of Fuel Compounds on a Unified Scale*. Harvard Dataverse, V1.
- Mitsos, A., Asprion, N., Floudas, C. A., Bortz, M., Baldea, M., Bonvin, D., Caspari, A., and Schäfer, P. (2018). Challenges in process optimization for new feedstocks and energy sources. *Computers & Chemical Engineering*, 113:209–221.

- Mitsos, A., Bollas, G. M., and Barton, P. I. (2009). Bilevel optimization formulation for parameter estimation in liquid–liquid phase equilibrium problems. *Chemical Engineering Science*, 64(3):548–559.
- Moro, S., Chipman, J. K., Wegener, J.-W., Hamberger, C., Dekant, W., and Mally, A. (2012). Furan in heat-treated foods: formation, exposure, toxicity, and aspects of risk assessment. *Molecular Nutrition & Food Research*, 56(8):1197–1211.
- Mukkula, A. R. G., Mateáš, M., Fikar, M., and Paulen, R. (2021). Robust multi-stage model-based design of optimal experiments for nonlinear estimation. *Computers & Chemical Engineering*, 107499.
- Müller, J. (2016). MISO: mixed-integer surrogate optimization framework. *Optimization and Engineering*, 17(1):177–203.
- Müller, S. (2019). Flexible heuristic algorithm for automatic molecule fragmentation: application to the UNIFAC group contribution model. *Journal of Cheminformatics*, 11(1):57.
- Murat Sen, S., Henao, C. A., Braden, D. J., Dumesic, J. A., and Maravelias, C. T. (2012). Catalytic conversion of lignocellulosic biomass to fuels: Process development and techno-economic evaluation. *Chemical Engineering Science*, 67(1):57–67.
- Nagata, I. (1984). Liquid-liquid equilibria for four ternary systems containing methanol and cyclohexane. *Fluid Phase Equilibria*, 18(1):83–92.
- Nagata, I. (1987). Liquid-liquid equilibria for ternary acetonitrile-ethanol-saturated hydrocarbon and acetonitrile-1-propanol-saturated hydrocarbon mixtures. *Thermochimica Acta*, 119(2):357–368.
- Ng, L. Y., Andiappan, V., Chemmangattuvalappil, N. G., and Ng, D. K. (2015a). A systematic methodology for optimal mixture design in an integrated biorefinery. *Computers & Chemical Engineering*, 81:288–309.
- Ng, L. Y., Chong, F. K., and Chemmangattuvalappil, N. G. (2015b). Challenges and opportunities in computer-aided molecular design. *Computers & Chemical Engineering*, 81:115–129.
- Nguyen, N.-K. and Miller, A. J. (1992). A review of some exchange algorithms for constructing discrete D-optimal designs. *Computational Statistics & Data Analysis*, 14(4):489–498.

- Nigam, A. K., Friederich, P., Krenn, M., and Aspuru-Guzik, A. (2019). Augmenting Genetic Algorithms with Deep Neural Networks for Exploring the Chemical Space. *arXiv*, preprint:arXiv:1909.11655.
- Nigam, A. K., Pollice, R., and Aspuru-Guzik, A. (2021). JANUS: Parallel Tempered Genetic Algorithm Guided by Deep Neural Networks for Inverse Molecular Design. *arXiv*, preprint:arXiv:2106.04011.
- NIST Chemistry WebBook (2021). Eds. P.J. Linstrom and W.G. Mallard: NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg MD, 20899, United States.
- Ooi, J., Ng, D. K., and Chemmangattuvalappil, N. G. (2018). Optimal molecular design towards an environmental friendly solvent recovery process. *Computers & Chemical Engineering*, 117:391–409.
- Ooi, J., Ng, D. K. S., and Chemmangattuvalappil, N. G. (2019). A Systematic Molecular Design Framework with the Consideration of Competing Solvent Recovery Processes. *Industrial & Engineering Chemistry Research*, 58(29):13210–13226.
- Ooi, Y. J., Aung, K. N. G., Chong, J. W., Tan, R. R., Aviso, K. B., and Chemmangattuvalappil, N. G. (2022). Design of fragrance molecules using computer-aided molecular design with machine learning. *Computers & Chemical Engineering*, 157:107585.
- Otto, S. A., Kadin, M., Casini, M., Torres, M. A., and Blenckner, T. (2018). A quantitative framework for selecting and validating food web indicators. *Ecological Indicators*, 84:619–631.
- Papadopoulos, A. I., Perdomo, F. A., Tzirakis, F., Shavaliyeva, G., Tsvintzelis, I., Kazepidis, P., Nessi, E., Papadokonstantakis, S., Seferlis, P., Galindo, A., Jackson, G., and Adjiman, C. S. (2020). Molecular engineering of sustainable phase-change solvents: From digital design to scaling-up for CO₂ capture. *Chemical Engineering Journal*, 127624.
- Papadopoulos, A. I., Stijepovic, M., and Linke, P. (2010). On the systematic design and selection of optimal working fluids for Organic Rankine Cycles. *Applied Thermal Engineering*, 30(6-7):760–769.
- Papadopoulos, A. I., Tsvintzelis, I., Linke, P., and Seferlis, P. (2018). Computer-Aided Molecular Design: Fundamentals, Methods, and Applications. In Reedijk, J.,

- editor, *Elsevier Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*. Elsevier, Waltham, MA, United States.
- Papoulias, S. A. and Grossmann, I. E. (1983). A structural optimization approach in process synthesis—II: Heat recovery networks. *Computers & Chemical Engineering*, 7(6):707–721.
- Parvatker, A. G. and Eckelman, M. J. (2020). Simulation-Based Estimates of Life Cycle Inventory Gate-to-Gate Process Energy Use for 151 Organic Chemical Syntheses. *ACS Sustainable Chemistry & Engineering*, 8(23):8519–8536.
- Pereira, F. E., Keskes, E., Galindo, A., Jackson, G., and Adjiman, C. S. (2011). Integrated solvent and process design using a SAFT-VR thermodynamic description: High-pressure separation of carbon dioxide and methane. *Computers & Chemical Engineering*, 35(3):474–491.
- Pereira, R. and Pasa, V. (2006). Effect of mono-olefins and diolefins on the stability of automotive gasoline. *Fuel*, 85(12-13):1860–1865.
- Pistikopoulos, E. N., Barbosa-Povoa, A., Lee, J. H., Misener, R., Mitsos, A., Reklaitis, G. V., Venkatasubramanian, V., You, F., and Gani, R. (2021). Process systems engineering – The generation next? *Computers & Chemical Engineering*, 147:107252.
- Ploskas, N. and Sahinidis, N. V. (2021). Review and comparison of algorithms and software for mixed-integer derivative-free optimization. *Journal of Global Optimization*, 82:433–462.
- Pronzato, L. and Walter, E. (1990). Experiment design for bounded-error models. *Mathematics and Computers in Simulation*, 32(5-6):571–584.
- Recker, S., Kerimoglu, N., Harwardt, A., Tkacheva, O., and Marquardt, W. (2013). On the integration of model identification and process optimization. In *23rd European Symposium on Computer Aided Process Engineering*, volume 32 of *Computer Aided Chemical Engineering*, pages 1021–1026. Elsevier, Amsterdam, The Netherlands.
- Redepenning, C. and Marquardt, W. (2017). Pinch-based shortcut method for the conceptual design of adiabatic absorption columns. *AIChE Journal*, 63(4):1213–1225.
- Redepenning, C., Recker, S., and Marquardt, W. (2017). Pinch-based shortcut method for the conceptual design of isothermal extraction columns. *AIChE Journal*, 63(4):1236–1245.

- Remmert, S., Campbell, S., Cracknell, R., Schuetze, A., Lewis, A., Giles, K., Akehurst, S., Turner, J., Popplewell, A., and Patel, R. (2014). Octane Appetite: The Relevance of a Lower Limit to the MON Specification in a Downsized, Highly Boosted DISI Engine. *SAE International Journal of Fuels and Lubricants*, 7(3):743–755.
- Renon, H. and Prausnitz, J. M. (1968). Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE Journal*, 14(1):135–144.
- Righi, S., Baioli, F., Dal Pozzo, A., and Tugnoli, A. (2018). Integrating life cycle inventory and process design techniques for the early estimate of energy and material consumption data. *Energies*, 11(4):970.
- Rios, L. M. and Sahinidis, N. V. (2013). Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247–1293.
- Rittig, J. G., Ritzert, M., Schweidtmann, A. M., Winkler, S., Weber, J. M., Morsch, P., Heufer, K. A., Grohe, M., Mitsos, A., and Dahmen, M. (2022). Graph machine learning for design of high-octane fuels. *AIChE Journal*, e17971.
- Rojas, C. R., Welsh, J. S., Goodwin, G. C., and Feuer, A. (2007). Robust optimal experiment design for system identification. *Automatica*, 43(6):993–1008.
- Ruiz, E. M. (2019). Treatment of spent solvent mixture, hazardous waste incineration: Europe without Switzerland, Ecoinvent database version 3.6.
- Ryu, J., Kong, L., Pastore de Lima, A. E., and Maravelias, C. T. (2020). A generalized superstructure-based framework for process synthesis. *Computers & Chemical Engineering*, 133:106653.
- Samaras, Z. C., Andersson, J., Bergmann, A., Hausberger, S., Toumasatos, Z., Keskinen, J., Haisch, C., Kontses, A., Ntziachristos, L. D., Landl, L., Mamakos, A., and Bainschab, M. (2020). Measuring Automotive Exhaust Particles Down to 10 nm. *SAE International Journal of Advances and Current Practices in Mobility*, 3(1):539–550.
- Samudra, A. P. and Sahinidis, N. V. (2013). Optimization-based framework for computer-aided molecular design. *AIChE Journal*, 59(10):3686–3701.
- Sanchez Medina, E. I., Linke, S., Stoll, M., and Sundmacher, K. (2022). Graph neural networks for the prediction of infinite dilution activity coefficients. *Digital Discovery*, 1:216–225.

- Sattler, K. and Feindt, H. J., editors (1995). *Thermal Separation Processes: Chapter 01 - Basic Concepts*. Wiley-VCH Verlag GmbH, Weinheim, Germany.
- Schack, D., Liesche, G., and Sundmacher, K. (2020). The FluxMax approach: Simultaneous flux optimization and heat integration by discretization of thermodynamic state space illustrated on methanol synthesis process. *Chemical Engineering Science*, 215:115382.
- Scheffczyk, J., Fleitmann, L., Schwarz, A., Lampe, M., Bardow, A., and Leonhard, K. (2017a). COSMO-CAMD: A framework for optimization-based computer-aided molecular design using COSMO-RS. *Chemical Engineering Science*, 159:84–92.
- Scheffczyk, J., Schäfer, P., Fleitmann, L., Thien, J., Redepenning, C., Leonhard, K., Marquardt, W., and Bardow, A. (2018). COSMO-CAMPD: a framework for integrated design of molecules and processes based on COSMO-RS. *Molecular Systems Design & Engineering*, 3(4):645–657.
- Scheffczyk, J., Schäfer, P., Jens, C. M., Leonhard, K., and Bardow, A. (2017b). Integrated process and solvent design using COSMO-RS for the production of CO from CO₂ and H₂. In Espuña, A., Graells, M., and Puigjaner, L., editors, *Computer Aided Chemical Engineering vol. 40*, Computer Aided Chemical Engineering, pages 1765–1770. Elsevier, Amsterdam, The Netherlands.
- Schilling, J., Horend, C., and Bardow, A. (2020). Integrating superstructure-based design of molecules, processes, and flowsheets. *AIChE Journal*, 66(5):e16903.
- Schilling, J., Tillmanns, D., Lampe, M., Hopp, M., Gross, J., and Bardow, A. (2017). From molecules to dollars: integrating molecular design into thermo-economic process design using consistent thermodynamic modeling. *Molecular Systems Design & Engineering*, 2(3):301–320.
- Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Hunter, C. A., Bekas, C., and Lee, A. A. (2019). Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Central Science*, 5(9):1572–1583.
- Schwaller, P., Petraglia, R., Zullo, V., Nair, V. H., Haeuselmann, R. A., Pisoni, R., Bekas, C., Iuliano, A., and Laino, T. (2020). Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chemical Science*, 11(12):3316–3325.
- Schweidtmann, A. M., Clayton, A. D., Holmes, N., Bradford, E., Bourne, R. A., and Lapkin, A. A. (2018). Machine learning meets continuous flow chemistry: Automated

- optimization towards the Pareto front of multiple objectives. *Chemical Engineering Journal*, 352:277–282.
- Schweidtmann, A. M., Rittig, J. G., König, A., Grohe, M., Mitsos, A., and Dahmen, M. (2020). Graph Neural Networks for Prediction of Fuel Ignition Quality. *Energy & Fuels*, 34(9):11395–11407.
- Sharma, N. and Liu, Y. A. (2022). A hybrid science-guided machine learning approach for modeling chemical processes: A review. *AIChE Journal*, e17609.
- Shirazi, S. A., Abdollahipoor, B., Windom, B., Reardon, K. F., and Foust, T. D. (2020). Effects of blending C3-C4 alcohols on motor gasoline properties and performance of spark ignition engines: A review. *Fuel Processing Technology*, 197:106194.
- Singh, E., Mohammed, A., Gorbatenko, I., and Sarathy, M. (2021). On the Relevance of Octane Sensitivity in Heavily Downsized Spark-Ignited Engines. In *15th International Conference on Engines & Vehicles*. SAE International.
- Skiborowski, M., Bausa, J., and Marquardt, W. (2016). A Unifying Approach for the Calculation of Azeotropes and Pinch Points in Homogeneous and Heterogeneous Mixtures. *Industrial & Engineering Chemistry Research*, 55(24):6815–6834.
- Smith, K. E. C., Schmidt, S. N., Dom, N., Blust, R., Holmstrup, M., and Mayer, P. (2013). Baseline toxic mixtures of non-toxic chemicals: "solubility addition" increases exposure for solid hydrophobic chemicals. *Environmental Science & Technology*, 47(4):2026–2033.
- Smith, R. (2005). *Chemical process design and integration*. John Wiley & Sons Ltd, Chichester, West Sussex, United Kingdom.
- Smith, R. L., Ruiz-Mercado, G. J., Meyer, D. E., Gonzalez, M. A., Abraham, J. P., Barrett, W. M., and Randall, P. M. (2017). Coupling Computer-Aided Process Simulation and Estimations of Emissions and Land Use for Rapid Life Cycle Inventory Modeling. *ACS Sustainable Chemistry & Engineering*, 5(5):3786–3794.
- Soh, L. and Eckelman, M. J. (2016). Green Solvents in Biomass Processing. *ACS Sustainable Chemistry & Engineering*, 4(11):5821–5837.
- Song, R., Keller, A. A., and Suh, S. (2017). Rapid life-cycle impact screening using artificial neural networks. *Environmental Science & Technology*, 51(18):10777–10785.

- Song, Z., Hu, X., Wu, H., Mei, M., Linke, S., Zhou, T., Qi, Z., and Sundmacher, K. (2020). Systematic Screening of Deep Eutectic Solvents as Sustainable Separation Media Exemplified by the CO₂ Capture Process. *ACS Sustainable Chemistry & Engineering*, 8(23):8741–8751.
- Sousa, T., Correia, J., Pereira, V., and Rocha, M. (2021). Generative Deep Learning for Targeted Compound Design. *Journal of Chemical Information and Modeling*, 61(11):5343–5361.
- Spear, R. C. and Selvin, S. (1989). OSHA’s permissible exposure limits: regulatory compliance versus health risk. *Risk Analysis*, 9(4):579–586.
- Squire, W. and Trapp, G. (1998). Using Complex Variables to Estimate Derivatives of Real Functions. *SIAM Review*, 40(1):110–112.
- St. John, P. C., Bartlett, M., and Kim, S. (2022). Group-contribution predictions of Yield Sooting Index (YSI) - YSI Estimator, National Renewable Energy Laboratory (NREL). <https://ysi.ml.nrel.gov/>, (accessed on March 06, 2022).
- Steiner, S., Wolf, J., Glatzel, S., Andreou, A., Granda, J. M., Keenan, G., Hinkley, T., Aragon-Camarasa, G., Kitson, P. J., Angelone, D., and Cronin, L. (2019). Organic synthesis in a modular robotic system driven by a chemical programming language. *Science*, 363(6423):eaav2211.
- Stephens, P. J., Devlin, F. J., Chabalowski, C. F., and Frisch, M. J. (1994). Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry*, 98(45):11623–11627.
- Strigle, R. F. (1994). *Packed Tower Design and Applications: Random and Structured Packings*. Gulf Publishing Company, Houston, TX, United States, 2nd edition.
- Sun, Y. and Sahinidis, N. V. (2022). Computer-aided retrosynthetic design: fundamentals, tools, and outlook. *Current Opinion in Chemical Engineering*, 35:100721.
- Sun, Y., Sahinidis, N. V., Sundaram, A., and Cheon, M.-S. (2020). Derivative-free optimization for chemical product design. *Current Opinion in Chemical Engineering*, 27:98–106.
- Supronowicz, W., Ignatyev, I. A., Lolli, G., Wolf, A., Zhao, L., and Mleczko, L. (2015). Formic acid: a future bridge between the power and chemical industries. *Green Chemistry*, 17(5):2904–2911.

- Szybist, J. P., Busch, S., McCormick, R. L., Pihl, J. A., Splitter, D. A., Ratcliff, M. A., Kolodziej, C. P., Storey, J. M., Moses-DeBusk, M., Vuilleumier, D., Sjöberg, M., Sluder, C. S., Rockstroh, T., and Miles, P. (2021). What fuel properties enable higher thermal efficiency in spark-ignited engines? *Progress in Energy and Combustion Science*, 82:100876.
- Taylor, R. and Krishna, R. (1993). *Multicomponent Mass Transfer*. John Wiley & Sons, Inc., Hoboken, NJ, United States.
- Telen, D., Houska, B., Logist, F., and van Impe, J. (2016). Multi-purpose economic optimal experiment design applied to model based optimal control. *Computers & Chemical Engineering*, 94:212–220.
- Telen, D., Houska, B., Vallerio, M., Logist, F., and van Impe, J. (2017). A study of integrated experiment design for NMPC applied to the Droop model. *Chemical Engineering Science*, 160:370–383.
- Ten, J. Y., Hassim, M. H., and Chemmangattuvalappil, N. G. (2020). Integration of safety and health aspects in a simultaneous process and molecular design framework. *Chemical Engineering Research and Design*, 153:849–864.
- Ten, J. Y., Hassim, M. H., Ng, D. K. S., and Chemmangattuvalappil, N. G. (2017). A molecular design methodology by the simultaneous optimisation of performance, safety and health aspects. *Chemical Engineering Science*, 159:140–153.
- Ten, J. Y., Liew, Z. H., Oh, X. Y., Hassim, M. H., and Chemmangattuvalappil, N. (2021). Computer-Aided Molecular Design of Optimal Sustainable Solvent for Liquid-Liquid Extraction. *Process Integration and Optimization for Sustainability*.
- The MathWorks Inc. (2018a). MATLAB (Release R2018b): Global Optimization Toolbox, Natick, MA, United States.
- The MathWorks Inc. (2018b). MATLAB (Release R2018b): Optimization Toolbox, Natick, MA, United States.
- The MathWorks Inc. (2018c). MATLAB (Release R2018b): Statistics and Machine Learning Toolbox, Natick, MA, United States.
- The MathWorks Inc. (2019). MATLAB (Release R2019a): Optimization Toolbox, Natick, MA, United States.

- Thewes, M., Muether, M., Pischinger, S., Budde, M., Brunn, A., Sehr, A., Adomeit, P., and Klankermayer, J. (2011). Analysis of the Impact of 2-Methylfuran on Mixture Formation and Combustion in a Direct-Injection Spark-Ignition Engine. *Energy & Fuels*, 25(12):5549–5561.
- Thewes, M., Müther, M., Brassat, A., Pischinger, S., and Sehr, A. (2012). Analysis of the Effect of Bio-Fuels on the Combustion in a Downsized DI SI Engine. *SAE International Journal of Fuels and Lubricants*, 5(1):274–288.
- Thien, J., Peters, C., Brands, T., Koß, H.-J., and Bardow, A. (2017). Efficient Determination of Liquid–Liquid Equilibria Using Microfluidics and Raman Microspectroscopy. *Industrial & Engineering Chemistry Research*, 56(46):13905–13910.
- Thien, J., Reinpold, L., Brands, T., Koß, H.-J., and Bardow, A. (2020). Automated Physical Property Measurements from Calibration to Data Analysis: Microfluidic Platform for Liquid–Liquid Equilibrium Using Raman Microspectroscopy. *Journal of Chemical & Engineering Data*, 65(2):319–327.
- Thinkstep AG (2017). GaBi: LCA software and LCI database, Stuttgart, Germany.
- Tobita, H. (2000). Polymerization Processes, 1. Fundamentals. In *Ullmann's Encyclopedia of Industrial Chemistry*, pages 1–50. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
- Tyn, M. T. and Calus, W. F. (1975). Temperature and Concentration Dependence of Mutual Diffusion Coefficients of Some Binary Liquid Systems. *Journal of Chemical & Engineering Data*, 20:310–316.
- Uhlemann, J., Costa, R., and Charpentier, J.-C. (2019). Product Design and Engineering in Chemical Engineering: Past, Present State, and Future. *Chemical Engineering & Technology*, 42(11):2258–2274.
- United Nations Environment Programme (2019). *Global Chemicals Outlook II: From Legacies to Innovative Solutions: Implementing the 2030 Agenda for Sustainable Development*.
- Vaidya, P. D. and Kenig, E. Y. (2007). CO₂-Alkanolamine Reaction Kinetics: A Review of Recent Studies. *Chemical Engineering & Technology*, 30(11):1467–1474.
- van Kleef, L. M., Oyewunmi, O. A., and Markides, C. N. (2019). Multi-objective thermo-economic optimization of organic Rankine cycle (ORC) power systems in waste-heat recovery applications using computer-aided molecular design techniques. *Applied Energy*, 251:112513.

- van Ness, H. C. (1995). Thermodynamics in the treatment of vapor/liquid equilibrium (VLE) data. *Pure and Applied Chemistry*, 67(6):859–872.
- vom Lehn, F., Cai, L., Copa Cáceres, B., and Pitsch, H. (2021a). Exploring the fuel structure dependence of laminar burning velocity: A machine learning based group contribution approach. *Combustion and Flame*, 232:111525.
- vom Lehn, F., Cai, L., Tripathi, R., Broda, R., and Pitsch, H. (2021b). A property database of fuel compounds with emphasis on spark-ignition engine applications. *Applications in Energy and Combustion Science*, 5:100018.
- Walters, W. P. and Barzilay, R. (2021). Applications of Deep Learning in Molecule Generation and Molecular Property Prediction. *Accounts of Chemical Research*, 54(2):263–270.
- Walum, E. (1998). Acute oral toxicity. *Environmental Health Perspectives*, 106(2):497–503.
- Walz, O., Djelassi, H., Caspari, A., and Mitsos, A. (2018). Bounded-error optimal experimental design via global solution of constrained min–max program. *Computers & Chemical Engineering*, 111:92–101.
- Walz, O., Djelassi, H., and Mitsos, A. (2019). Optimal experimental design for optimal process design: A trilevel optimization formulation. *AIChE Journal*, 66(1):e16788.
- Watson, O. L., Jonuzaj, S., McGinty, J., Sefcik, J., Galindo, A., Jackson, G., and Adjiman, C. S. (2021). Computer Aided Design of Solvent Blends for Hybrid Cooling and Antisolvent Crystallization of Active Pharmaceutical Ingredients. *Organic Process Research & Development*, 25(5):1123–1142.
- Wernet, G., Hellweg, S., Fischer, U., Papadokostantakis, S., and Hungerbühler, K. (2008). Molecular-structure-based models of chemical inventories using neural networks. *Environmental Science & Technology*, 42(17):6717–6722.
- Wernet, G., Papadokostantakis, S., Hellweg, S., and Hungerbühler, K. (2009). Bridging data gaps in environmental assessments: Modeling impacts of fine and basic chemical production. *Green Chemistry*, 11(11):1826–1831.
- Westbrook, C. K., Pitz, W. J., and Curran, H. J. (2006). Chemical Kinetic Modeling Study of the Effects of Oxygenated Hydrocarbons on Soot Emissions from Diesel Engines. *The Journal of Physical Chemistry A*, 110(21):6912–6922.

- White, M. T., Oyewunmi, O. A., Chatzopoulou, M. A., Pantaleo, A. M., Haslam, A. J., and Markides, C. N. (2018). Computer-aided working-fluid design, thermodynamic optimisation and thermoeconomic assessment of ORC systems for waste-heat recovery. *Energy*, 161:1181–1198.
- White, M. T., Oyewunmi, O. A., Haslam, A. J., and Markides, C. N. (2017). Industrial waste-heat recovery through integrated computer-aided working-fluid and ORC system optimisation using SAFT- γ Mie. *Energy Conversion and Management*, 150:851–869.
- Winter, B., Winter, C., Esper, T., Schilling, J., and Bardow, A. (2023). SPT-NRTL: A physics-guided machine learning model to predict thermodynamically consistent activity coefficients. *Fluid Phase Equilibria*, 568:113731.
- Winter, B., Winter, C., Schilling, J., and Bardow, A. (2022). A smile is all you need: predicting limiting activity coefficients from SMILES with natural language processing. *Digital Discovery*, 1:859–869.
- Wolff, L., Koß, H.-J., and Bardow, A. (2016). The optimal diffusion experiment. *Chemical Engineering Science*, 152:392–402.
- Wolff, L. W. M. (2021). *From model-based experimental design and analysis of diffusion and liquid-liquid equilibria to process applications: Dissertation, RWTH Aachen University*, volume 30 of *Aachener Beiträge zur technischen Thermodynamik*. Wissenschaftsverlag Mainz GmbH, Aachen, Germany.
- Wynn, H. P. (1972). Results in the Theory and Construction of D -Optimum Experimental Designs. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):133–147.
- Yalkowsky, S. and He, Y. (2003). *Handbook of Aqueous Solubility Data: 1st Edition*. CRC Press, Boca Raton, FL, United States.
- Yu, Y. (2010). Monotonic convergence of a general algorithm for computing optimal designs. *The Annals of Statistics*, 38(3):1593–1606.
- Yunus, N. A., Gernaey, K. V., Woodley, J. M., and Gani, R. (2014). A systematic methodology for design of tailor-made blended products. *Computers & Chemical Engineering*, 66:201–213.
- Zhang, L., Babi, D. K., and Gani, R. (2016). New Vistas in Chemical Product and Process Design. *Annual Review of Chemical and Biomolecular Engineering*, 7:557–582.

- Zhang, L., Kalakul, S., Liu, L., Elbashir, N. O., Du, J., and Gani, R. (2018a). A Computer-Aided Methodology for Mixture-Blend Design. Applications to Tailor-Made Design of Surrogate Fuels. *Industrial & Engineering Chemistry Research*, 57(20):7008–7020.
- Zhang, L., Mao, H., Liu, L., Du, J., and Gani, R. (2018b). A machine learning based computer-aided molecular design/screening methodology for fragrance molecules. *Computers & Chemical Engineering*, 115:295–308.
- Zhang, L., Mao, H., Liu, Q., and Gani, R. (2020a). Chemical product design – recent advances and perspectives. *Current Opinion in Chemical Engineering*, 27:22–34.
- Zhang, X., Ding, X., Song, Z., Zhou, T., and Sundmacher, K. (2021a). Integrated ionic liquid and rate-based absorption process design for gas separation: Global optimization using hybrid models. *AIChE Journal*, 67(10):e17340.
- Zhang, X., Zhang, L., Fung, K. Y., Bakshi, B. R., and Ng, K. M. (2020b). Sustainable product design: A life-cycle approach. *Chemical Engineering Science*, 217:115508.
- Zhang, X., Zhou, T., and Ng, K. M. (2021b). Optimization-based cosmetic formulation: Integration of mechanistic model, surrogate model, and heuristics. *AIChE Journal*, 67(1):e17064.
- Zhang, X., Zhou, T., Zhang, L., Fung, K. Y., and Ng, K. M. (2019). Food Product Design: A Hybrid Machine Learning and Mechanistic Modeling Approach. *Industrial & Engineering Chemistry Research*, 58(36):16743–16752.
- Zhang, Z. (2016). Synthesis of γ -Valerolactone from Carbohydrates and its Applications. *ChemSusChem*, 9(2):156–171.
- Zheng, J., Zhao, Y., and Truhlar, D. G. (2009). The DBH24/08 Database and Its Use to Assess Electronic Structure Model Chemistries for Chemical Reaction Barrier Heights. *Journal of Chemical Theory and Computation*, 5(4):808–821.
- Zhou, T., Gani, R., and Sundmacher, K. (2021). Hybrid Data-Driven and Mechanistic Modeling Approaches for Multiscale Material and Process Design. *Engineering*, 7(9):1231–1238.
- Zhou, T., McBride, K., Linke, S., Song, Z., and Sundmacher, K. (2020). Computer-aided solvent selection and design for efficient chemical processes. *Current Opinion in Chemical Engineering*, 27:35–44.

- Zhou, T., Qi, Z., and Sundmacher, K. (2014). Model-based method for the screening of solvents for chemical reactions. *Chemical Engineering Science*, 115:177–185.
- Zhou, T., Song, Z., and Sundmacher, K. (2019a). Big Data Creates New Opportunities for Materials Research: A Review on Methods and Applications of Machine Learning for Materials Design. *Engineering*, 5(6):1017–1026.
- Zhou, T., Song, Z., Zhang, X., Gani, R., and Sundmacher, K. (2019b). Optimal Solvent Design for Extractive Distillation Processes: A Multiobjective Optimization-Based Hierarchical Framework. *Industrial & Engineering Chemistry Research*, 58(15):5777–5786.
- Zhou, T., Zhou, Y., and Sundmacher, K. (2017). A hybrid stochastic–deterministic optimization approach for integrated solvent and process design. *Chemical Engineering Science*, 159:207–216.
- Zhu, R., Taheri, M., Zhang, J., and Lei, Z. (2020). Extension of the COSMO-UNIFAC Thermodynamic Model. *Industrial & Engineering Chemistry Research*, 59(4):1693–1701.
- Zimmerman, J. B., Anastas, P. T., Erythropel, H. C., and Leitner, W. (2020). Designing for a green chemistry future. *Science*, 367(6476):397–400.

Publications

Publications directly related to this thesis

The content of the following publications has been integrated in this thesis:

Articles in peer-reviewed journals

Fleitmann, L., Ackermann, P., Schilling, J., Kleinekorte, J., Rittig, J. G., vom Lehn, F., Schweidtmann, A. M., Pitsch, H., Leonhard, K., Mitsos, A., Bardow, A., and Dahmen, M. (2023a). Molecular Design of Fuels for Maximum Spark-Ignition Engine Efficiency by Combining Predictive Thermodynamics and Machine Learning. *Energy & Fuels*, 37(3):2213–2229.

Fleitmann, L., Gertig, C., Scheffczyk, J., Schilling, J., Leonhard, K., and Bardow, A. (2023b). From Molecules to Heat-Integrated Processes: Computer-Aided Design of Solvents and Processes Using Quantum Chemistry. *Chemie Ingenieur Technik*, 95(3):368–380.

Fleitmann, L., Kleinekorte, J., Leonhard, K., and Bardow, A. (2021). COSMO-susCAMPD: Sustainable solvents from combining computer-aided molecular and process design with predictive life cycle assessment. *Chemical Engineering Science*, 245:116863.

Fleitmann, L., Pyschik, J., Wolff, L., Schilling, J., and Bardow, A. (2022). Optimal experimental design of physical property measurements for optimal chemical process simulations. *Fluid Phase Equilibria*, 113420.

Peer-reviewed conference contributions

Fleitmann, L., Pyschik, J., Wolff, L., and Bardow, A. (2021). Optimal physical property data for process simulations by optimal experimental design. In Turkey, M. and Gani, R., editors, *31th European Symposium on Computer Aided Process Engineering*, volume 50 of *Computer-aided chemical engineering*, pages 851–857. Elsevier, Amsterdam, The Netherlands.

Fleitmann, L., Scheffczyk, J., Schäfer, P., Jens, C., Leonhard, K., and Bardow, A. (2018). Integrated design of solvents in hybrid reaction-separation processes using COSMO-RS. *Chemical Engineering Transactions*, 69:559–564.

Publications indirectly related to this thesis

Articles in peer-reviewed journals

Frumkin, J. A., Fleitmann, L., and Doherty, M. F. (2019). Ultimate Reaction Selectivity Limits for Intensified Reactor–Separators. *Industrial & Engineering Chemistry Research*, 58(15):6042–6048.

Gertig, C., Fleitmann, L., Hemprich, C., Hense, J., Bardow, A., and Leonhard, K. (2021). CAT-COSMO-CAMPD: Integrated in silico design of catalysts and processes based on quantum chemistry. *Computers & Chemical Engineering*, 153:107438.

Gertig, C., Fleitmann, L., Schilling, J., Leonhard, K., and Bardow, A. (2020). Rx–COSMO–CAMPD: Enhancing Reactions by Integrated Computer–Aided Design of Solvents and Processes based on Quantum Chemistry. *Chemie Ingenieur Technik*, 92(10):1489–1500.

Gertig, C., Kröger, L., Fleitmann, L., Scheffczyk, J., Bardow, A., and Leonhard, K. (2019). Rx-COSMO-CAMD: Computer-Aided Molecular Design of Reaction Solvents Based on Predictive Kinetics from Quantum Chemistry. *Industrial & Engineering Chemistry Research*, 58(51):22835–22846.

Haus, M. O., Winter, B., Fleitmann, L., Palkovits, R., and Bardow, A. (2022). Making more from bio-based platforms: life cycle assessment and techno-economic analysis of N -vinyl-2-pyrrolidone from succinic acid. *Green Chemistry*, 24(17):6671–6684.

Kleinekorte, J., Fleitmann, L., Bachmann, M., Kätelhön, A., Barbosa-Póvoa, A., von der Assen, N., and Bardow, A. (2020). Life Cycle Assessment for the Design of Chemical Processes, Products, and Supply Chains. *Annual Review of Chemical and Biomolecular Engineering*, 11:203–233.

Scheffczyk, J., Fleitmann, L., Schwarz, A., Lampe, M., Bardow, A., and Leonhard, K. (2017). COSMO-CAMD: A framework for optimization-based computer-aided molecular design using COSMO-RS. *Chemical Engineering Science*, 159:84–92.

Scheffczyk, J., Schäfer, P., Fleitmann, L., Thien, J., Redepenning, C., Leonhard, K., Marquardt, W., and Bardow, A. (2018). COSMO-CAMPD: a framework for integrated design of molecules and processes based on COSMO-RS. *Molecular Systems Design & Engineering*, 3(4):645–657.

Peer-reviewed conference contributions

Gertig, C., Fleitmann, L., Hemprich, C., Hense, J., Bardow, A., and Leonhard, K. (2020). Integrated In Silico Design of Catalysts and Processes based on Quantum Chemistry. In Pierucci, S., Manenti, F., Bozzano, G. L., and Manca, D., editors, *30th European Symposium on Computer Aided Process Engineering*, volume 48 of *Computer-aided chemical engineering*, pages 889–894. Elsevier, Amsterdam, The Netherlands.

Scheffczyk, J., Fleitmann, L., Schwarz, A., Bardow, A., and Leonhard, K. (2016). Computer-Aided Molecular Design by Combining Genetic Algorithms and COSMO-RS. In Kravanja, Z. and Bogataj, M., editors, *26th European Symposium on Computer Aided Process Engineering*, volume 38 of *Computer-aided chemical engineering*, pages 115–120. Elsevier, Amsterdam, The Netherlands.