Fortschritt-Berichte VDI

VDI

Reihe 3

Verfahrenstechnik

Dipl.-Ing. Manuel Dahmen,
Heinsberg

Nr. 954

Model-Based Design of Pure and Multicomponent Cellulosic Biofuels for Advanced Engine Concepts

Berichte aus der
Aachener Verfahrenstechnik - Prozesstechnik

RWTH Aachen University

Model-Based Design of Pure and Multicomponent Cellulosic Biofuels for Advanced Engine Concepts

Modellbasierter Entwurf von lignocellulosebasierten Biokraftstoffen für fortschrittliche Motorenkonzepte

Von der Fakultät für Maschinenwesen der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Manuel Dahmen

Berichter: Univ.-Prof. Dr.-Ing. Wolfgang Marquardt
Univ.-Prof. Dr.-Ing. (USA) Stefan Pischinger

Tag der mündlichen Prüfung: 27.09.2017

Fortschritt-Berichte VDI

Reihe 3

Verfahrenstechnik

Dipl.-Ing. Manuel Dahmen,
Heinsberg

Nr. 954

Model-Based Design of Pure and Multicomponent Cellulosic Biofuels for Advanced Engine Concepts

Berichte aus der
Aachener Verfahrenstechnik - Prozesstechnik

RWTH Aachen University

Dahmen, Manuel
Model-Based Design of Pure and Multicomponent Cellulosic Biofuels for Advanced Engine Concepts
Fortschr.-Ber. VDI Reihe 3 Nr. 954. Düsseldorf: VDI Verlag 2017.
292 Seiten, 71 Bilder, 48 Tabellen.
ISBN 978-3-18-395403-2, ISSN 0178-9503,
€ 100,00/VDI-Mitgliederpreis € 90,00.

Keywords: Fuel Design – Tailor-Made Biofuels – Cetane Number Prediction – Computer-Aided Molecular Design – Targeted Molecular Structure Generation – QSPR Modeling – Group Contribution Modeling – Optimization-Based Blend Design – Integrated Product and Pathway Design

This thesis describes model-based strategies for the identification of pure and multicomponent cellulosic biofuels that exhibit tailored properties for use in high-efficiency, low-emission internal combustion engines. Based on the principles of computer-aided molecular design, algorithmic exploration of the molecular search space by means of carbon- and energy-efficient refunctionalization of bio-derived platform chemicals is combined with quantitative structure-property relationship (QSPR) and group contribution modeling of key physicochemical fuel properties including the derived cetane number. The resulting virtual fuel screening approach is applied to the task of identifying oxygenated fuel candidates for both spark-ignition and compression-ignition engines. Optimization-based formulation of 100%-renewable biofuel blends is performed by means of integrated product and pathway design. Application of the novel design methodology yields biofuel mixtures that exhibit both the desired physicochemical properties and attractive process-related properties.

Bibliographische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter http://dnb.ddb.de abrufbar.

Bibliographic information published by the Deutsche Bibliothek
(German National Library)
The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at http://dnb.ddb.de.

D 82 (Diss. RWTH Aachen University, 2017)

Als Manuskript gedruckt. Printed in Germany.
ISSN 0178-9503
ISBN 978-3-18-395403-2

Vorwort

Die hier vorliegende Arbeit entstand während meiner Zeit als wissenschaftlicher Mitarbeiter an den Lehrstühlen für Prozesstechnik (02/2011−06/2014) und Systemverfahrenstechnik (07/2014−01/2017) der Aachener Verfahrenstechnik der RWTH Aachen. Mein ganz besonderer Dank gilt meinem Doktorvater, Herrn Univ.-Prof. Dr.-Ing. Wolfgang Marquardt, dessen Ideen, Offenheit und Vertrauen die freie und unabhängige Gestaltung meiner Forschungstätigkeit und damit auch der vorliegenden Arbeit überhaupt erst ermöglicht haben. Ich fühle mich ihm für seine Förderung und die stets angenehme und inspirierende Zusammenarbeit verbunden. Weiterhin danke ich Herrn Univ.-Prof. Dr.-Ing. (USA) Stefan Pischinger für die Übernahme des Koreferats und Herrn Univ.-Prof. Dr. rer. nat. Andreas Schuppert für die Übernahme des Prüfungsvorsitzes.

Ich bedanke mich bei Dr.-Ing. Florian Kremer, Dr.-Ing. Andreas Janssen, Dipl.-Wirt.-Ing. Benedikt Heuser, Dipl.-Ing. Fabian Hoppe, Dr. rer. nat. Stefanie Mersmann, Dr. rer. nat. Jakob Mottweiler, Univ.-Prof. Dr. rer. nat. Jürgen Klankermayer und Univ.-Prof. Dr.-Ing. (USA) Stefan Pischinger für den regelmäßigen fachlichen Austausch im Rahmen des Core Interaction Field (CIF) „Fuel Design“ des Exzellenzclusters „Tailor-Made Fuels from Biomass“ (TMFB). Ihre hilfreichen Anregungen hatten einen wesentlichen Anteil am Gelingen meiner Forschung und damit auch am Gelingen dieser Arbeit. Die Vielfalt der interdisziplinären Forschung im TMFB-Cluster habe ich als große Bereicherung empfunden.

Ein großer Dank geht an die Mitglieder der TMFB-Gruppe am Lehrstuhl und dabei insbesondere an Dr.-Ing. Manuel Hechinger, Dr.-Ing. Juan José Victoria Villeda, Dipl.-Ing. Kirsten Skiborowski und Dr.-Ing. Andreas Harwardt. Die enge Kooperation mit Euch und unsere vielen fachlichen Gespräche hatten einen großen Anteil am Gelingen dieser Arbeit. Die freundschaftliche und inspirierende Zusammenarbeit mit Euch wird eine bleibende schöne Erinnerung für mich sein. Darüber hinaus möchte ich mich bei allen Kollegen und den vielen Freunden am Lehrstuhl für die angenehme Atmosphäre und die gemeinsamen Aktivitäten bedanken. Es war eine tolle Zeit!

Der größte Dank gilt schließlich meiner Frau Tina und meinen Eltern für ihre Geduld und unermüdliche und liebevolle Unterstützung, auf die ich mich immer verlassen konnte.

Heinsberg, im September 2017 *Manuel Dahmen*

Contents

Notation VIII

Copyrights and Permissions XVI

Abstract XVII

Kurzfassung XVIII

1 Introduction 1

1.1 Structure of this thesis . . . 3

1.2 Previous publications of contents and results . . . 8

2 Basic concepts of fuel design 10

2.1 Physicochemical fuel properties and their impact on engine performance . . 11

2.1.1 Early-stage, rapid screening of fuel auto-ignition quality . . . 15

2.1.2 Cellulosic biofuels studied in research engines . . . 19

2.1.3 Perspectives on the definition of tailor-made fuels . . . 21

2.2 Exploitation of the structure of lignocellulosic biomass . . . 23

2.3 Carbon- and energy-efficient biofuel production . . . 29

2.4 Computer-aided design of chemical products and fuels . . . 33

2.4.1 Group contribution (GC) modeling . . . 34

2.4.2 Quantitative structure-property relationship (QSPR) modeling . . . 36

3 Targeted generation of candidate structures 40

3.1 Main principles in structure generation . . . 42

3.2 Relations to rule-based reaction network generation . . . 43

3.3 Pool-based scheme and implementation . . . 44

4 GC-based prediction of fuel auto-ignition quality 50

4.1 Materials and methods for the rapid screening . . . 51

4.2 Modeling strategy . . . 54

4.2.1 Selection of the general model structure . . . 55

4.2.2 Systematic model reduction . . . 62
4.2.3 Proposition of the final model . . . 66
4.2.4 Model validation . . . 71
4.3 Model application and uncertainty . . . 77
4.4 Conclusions and outlook . . . 81

5 QSPR-based prediction of physicochemical fuel properties 83
5.1 Modeling strategy . . . 83
5.1.1 Descriptor calculation, data pretreatment and outlier removal . . . 85
5.1.2 Model derivation, validation and application . . . 91
5.2 Comparison with GC-based prediction of thermophysical properties . . . 98
5.3 QSPR-based prediction of fuel auto-ignition quality . . . 102
5.4 Conclusions and outlook . . . 106

6 Model-based identification of biofuel candidates 107
6.1 Candidates for spark-ignition (SI) engines . . . 107
6.2 Candidates for compression-ignition (CI) engines . . . 120
6.3 Conclusions and outlook . . . 128

7 Blend formulation by simultaneous product and pathway design 130
7.1 Existing methodologies for computer-aided mixture design . . . 132
7.2 Blend design framework for tailor-made fuels . . . 132
7.3 Pathway model . . . 134
7.3.1 Extension of reaction network flux analysis . . . 134
7.3.2 Biomass supply . . . 139
7.3.3 Calculation of process performance measures . . . 140
7.4 Fuel property models . . . 142
7.4.1 Distillation curve model . . . 145
7.4.2 Reid vapor pressure model . . . 149
7.4.3 Phase stability criteria . . . 151
7.5 Problem formulation and solution strategy . . . 152
7.6 Case study: Blends for the spark-ignition (SI) engine . . . 159
7.6.1 Scenario description . . . 159
7.6.2 Results . . . 163
7.7 Conclusions and outlook . . . 172

8 Conclusions and outlook 175

Appendices 179

A Transformation rules for molecular structure generation **180**

B Experimental ignition delay data **188**

C GC model for the derived cetane number **196**

C.1 Group decomposition and descriptor data 196

C.2 Comparison of measurement data with results of GC model 204

C.3 Sensitivity equations and covariance matrix 209

D Derivation of QSPR models **212**

E Rational formulation of biofuel mixtures **214**

E.1 Application of Hoffmann-Florin equation 214

E.2 Application of COSTALD method 214

E.3 Trade-off analysis: Hydrogen demand vs. energy of fuel produced 215

E.4 Data and references for the case study 220

Bibliography **227**

Notation

Abbreviations

AD	applicability domain
ADC	Advanced Distillation Curve
AI	artificial intelligence
AIC	Akaike's information criterion
AICc	Akaike's information criterion corrected for small sample size
ARE	average relative error
ASG	ASG Analytik-Service GmbH
cMON	correlated motor octane number
cRON	correlated research octane number
C	carbon atom
CAMD	computer-aided molecular design
CFR	cooperative fuels research
CI	compression-ignition
CN	cetane number
CPD	conceptual process design
CPM	conversion pathway map
CRPM	chemical reaction pathway map
CV	cross-validation
DCN	derived cetane number
DIPPR	Design Institute for Physical Properties
DMTHF	dimethyltetrahydrofuran
ETBE	ethyl-tert-butyl ether
FIM	Fisher information matrix
FIT	Fuel Ignition Tester
GC	group contribution
GC^+	group contribution plus
GCM	group contribution method
H	hydrogen atom

HCCI	homogeneous charge compression ignition
HMF	hydroxymethylfurfural
IQT	Ignition Quality Tester
LHV	lower heating value
LLE	liquid-liquid equilibrium
LMO-CV	leave-many-out cross-validation
LOO-CV	leave-one-out cross-validation
MLR	multiple linear regression
MON	motor octane number
MTBE	methyl-tert-butyl ether
MTHF	methyltetrahydrofuran
NLP	nonlinear program
NMR	nuclear magnetic resonance
O	oxygen atom
ON	octane number
PC	principal component
PCA	principal component analysis
PCI	premixed compression-ignition
PLS	partial least squares
PNFA	process network flux analysis
QSAR	quantitative structure-activity relationship
QSPR	quantitative structure-property relationship
RCCI	reactivity controlled compression ignition
RMSE	root-mean-square error
RNFA	reaction network flux analysis
RON	research octane number
RVP	Reid vapor pressure
SCA	stochastic cluster analysis
SI	spark-ignition
TMFB	Tailor-Made Fuels from Biomass
TSI	threshold sooting index
VLE	vapor-liquid equilibrium

Latin

$\mathbf{A}$	matrix of stoichiometric coefficients
$AICc$	Akaike's information criterion corrected for small sample size

b	product flow	kmol/h
$\mathbf{b}$	vector of product flows	kmol/h
$\mathbf{c}$	vector of weights (PLS)	
C	pathway conversion	
$C_{ant,1} - C_{ant,7}$	parameters for the extended Antoine equation	
$C_{cost,a} - C_{cost,h}$	parameters for the COSTALD method	
$C_{dip,1} - C_{dip,4}$	parameters for the DIPPR 105 equation	
d	descriptor value	
$\mathbf{d}$	vector of descriptor values	
D	descriptor contribution (model parameter)	
$\mathbf{D}$	vector of descriptor contributions (model parameters)	
DCN	derived cetane number	
$\dot{E}_{fuel}$	energy flow of fuel produced	MJ/h
$\mathbf{E}$	matrix of $\mathbf{X}$ residuals (PCA / PLS)	
f	molar flow over pathway	kmol/h
$\mathbf{f}$	vector of molar flows over all pathways	kmol/h
$f_{COSTALD}$	COSTALD equation	kmol/m^3
f_{HF}	Hoffmann-Florin equation	kPa
f_{tdp}	tangent plane distance function	
F^{α}	F-distribution with confidence level α	
F	entry of the Fisher information matrix	
$\mathbf{F}$	Fisher information matrix	
g	group occurrence	
Δg	Gibb's energy change of mixing	J/mol
$\mathbf{g}$	vector of group occurrences	
G	group contribution (model parameter)	
$\mathbf{G}$	vector of group contributions (model parameters)	
H	enthalpy	kJ/kg, MJ/kg, kJ/mol, J/kmol
$\mathbf{H}$	Hessian matrix	
HHV	higher heating value	kJ/mol
$\mathbf{I}$	identity matrix	
K	sensitivity coefficient for octane index	
K	equilibrium ratio	
L	likelihood function	

LHV	lower heating value	MJ/kg, J/kmol, kJ/mol
m_{CO_2}	mass of carbon dioxide per MJ of fuel	g/MJ
$\dot{m}_{fuel}$	total mass flow of fuel produced	kg/h
$\dot{m}_{fuel,min}$	minimum total mass flow of fuel to be produced	kg/h
m_{H_2}	mass of hydrogen per mass of fuel	kg/kg
M	molar mass	g/mol, kg/kmol
n	amount of substance	mol
nC	number of carbon atoms (descriptor)	
n_C	number of palette compounds in a CPM	
$\dot{n}_{fuel}$	total mole flow of fuel produced	kmol/h
n_{H_2}	moles of hydrogen per mole of fuel	mol/mol
$n_{H_2,max}$	maximum hydrogen demand of fuel production (moles of hydrogen per mole of fuel)	mol/mol
$\dot{N}$	mole flow of vapor	kmol/s
N_c	number of compounds, i.e., length of property vector $\mathbf{y}$	
N_d	number of descriptors	
N_m	number of parameters in group contribution model	
N_{pc}	number of principal components in the model	
N_{pls}	number of PLS components in the model	
N_N	number of nodes in a CPM	
N_P	number of pathways in a CPM	
p	pressure	bar, kPa, Pa
$\mathbf{p}$	column of the loading matrix (PCA / PLS)	
P	parameter in group contribution or QSPR model	
$\mathbf{P}$	loading matrix (PCA / PLS)	
PC	set of palette compounds	
q	ignition delay prediction from perfect model	ln(ms)
$\mathbf{r}$	vector of model residuals (PLS)	various
R	universal gas constant	J/(mol·K)
$RMSE$	(weighted) root-mean-square error	
$\mathbf{s}$	vector of sensitivities	
std	standard deviation of a vector	various
S	pathway selectivity	

$\mathbf{S}$	diagonal matrix of singular values	
$SCBO$	sum of conventional bond orders (descriptor)	
$t^{1-(\alpha/2)}$	two-tails Student's t-distribution with confidence level α	
t	time	s
t	entry of a score vector	
$\mathbf{t}$	column of the score matrix (principal component)	
T	temperature	K, °C
$\mathbf{T}$	score matrix (PCA / PLS)	
T^2	Hotelling's T^2 diagnostic (PCA)	
$T10(m)/T50(m)$ /$T90(m)$	temperature on the distillation curve where 10 mol-% / 50 mol-% / 90 mol-% of the fuel have been evaporated	°C
$T10(v)/T50(v)$ /$T50(v)$	temperature on the distillation curve where 10 vol-% / 50 vol-% / 90 vol-% of the fuel have been evaporated	°C
u	model input	
$\mathbf{u}$	vector of model inputs	
$\mathbf{U}$	matrix of left singular values	
v	specific volume	m^3/kg
v	entry of eigenvector	
$\mathbf{v}$	eigenvector	
var	variation (PCA)	%
V	volume / molar volume	m^3, cm^3/mol, $m^3/kmol$
$\mathbf{V}$	matrix of right singular values	
V_f	volume fraction distilled	
V_F	entry of the covariance matrix	
$\mathbf{V_F}$	covariance matrix	
$V_R^{(0)}$	auxiliary variable in COSTALD method	
$V_R^{(\delta)}$	auxiliary variable in COSTALD method	
V_{shift}	normalized volume increment	
wt_{O_2}	oxygen content (weight fraction)	
$\mathbf{W}$	weight matrix (PLS)	
x'	mole fraction liquid phase	
x''	mole fraction vapor phase	
x	entry of a descriptor matrix/vector	

$\overline{x}$	mean of column-vector of descriptor data	
$\mathbf{x}$	column-vector of descriptor data	
$\mathbf{x}'$	vector of mole fractions liquid phase	
$\mathbf{X}$	matrix of descriptor data	
y	fuel property (various)	various
$\tilde{y}$	known fuel property data (various)	various
$\overline{\tilde{y}}$	mean of column-vector of known fuel property data (various)	various
$\mathbf{y}$	vector of fuel property data (various)	various
$\tilde{\mathbf{y}}$	vector of known fuel property data (various)	various
Δy	confidence interval for the prediction	ln(ms)
Y	pathway yield	
z	mole fraction of the blend	
$\mathbf{z}$	vector of mole fractions of the blend	

Greek

α	confidence level	
α	incremental step	
$\boldsymbol{\beta}$	vector of regression coefficients (PLS)	
γ	activity coefficient	
$\overline{\delta}$	similarity threshold	
δ	Euclidean distance in the PC space	
$\boldsymbol{\delta}$	vector of Euclidean distances in the PC space	
$\Delta\Theta$	confidence interval for model parameter	
$\tilde{\epsilon}$	normally distributed measurement error (zero mean)	ln(ms)
ζ	absolute parameter correlation	
η_{LHV}	LHV efficiency	MJ/MJ
Θ	parameter of group contribution model	
$\boldsymbol{\Theta}$	vector of model parameters	
χ	auxilliary variable	
λ	air-fuel equivalence ratio	
λ	eigenvalue	
μ	dynamic viscosity	Pa·s
ν	kinematic viscosity	mm^2/s
υ	stoichiometric coefficient	
ξ	mass fraction of the blend	
ρ_L	liquid density	kg/m^3

ρ_m	molar liquid density	kmol/m^3, mol/cm^3
σ	surface tension	N/m, mN/m
$\tilde{\sigma}$	measurement standard deviation	ln(ms)
$\tilde{\boldsymbol{\sigma}}$	vector of measurement standard deviations	ln(ms)
τ	IQT ignition delay	ms
ϕ	objective function in parameter estimation	
φ	length of vector	
ω	acentric factor	

Subscripts

$5NTN$	five nearest training neighbors
a	first-order structural group index
$a2$	second-order structural group index
$a3$	third-order structural group index
b	descriptor index
$boil$	normal boiling point
bp	bubble point pressure (approximated Reid vapor pressure)
c	combustion cycle index (measurement index)
$crit$	critical state
D	distillate
ext	external validation set
f	enthalpy of formation
h	pathway index
i	index (component / compound)
j	index (component / compound)
k	pathway index
l	index
L	liquid
max	maximum value
min	minimum value
$melt$	melting point
nAB	number of aromatic bonds (descriptor)
$nCCDB$	number of carbon-carbon double bonds (descriptor)
nTC	number of tertiary carbon atoms (descriptor)

nQC	number of quaternary carbon atoms (descriptor)
p	model parameter index
$P1$	biomass fractionation and depolymerization pathway
q	model parameter index
$Reid$	Reid vapor pressure
$S1$	biomass supply pathway
$S2$	hydrogen supply pathway
$thre$	threshold value
vap	enthalpy of vaporization
V	vapor
x	number of carbon atoms
y	number of hydrogen atoms
z	number of oxygen atoms

Superscripts

$*$	optimal value
0	reference state
cum	cumulative
j	eigenvector/eigenvalue index
lb	lower bound
$norm$	normalization factor
S	vapor pressure
t	model candidate index
$test$	test set (external validation set)
ub	upper bound

Copyrights and Permissions

Abstract

The present thesis describes model-based strategies for the identification of pure and multicomponent cellulosic biofuel candidates that exhibit tailored properties for use in high-efficiency, low-emission internal combustion engines. Following a description of the theoretical foundations of fuel design, an algorithm for the targeted generation of candidate structures is proposed that facilitates an exploration of the molecular search space by means of a rule-based approach resembling carbon- and energy-efficient chemo-catalytic refunctionalization of bio-derived platform chemicals. Model-based evaluation of the obtained structures is based on a group contribution method that is capable of predicting the derived cetane number (DCN) of oxygenated hydrocarbon species directly from molecular structure. Furthermore, the virtual fuel screening relies on tailored quantitative structure-property relationship (QSPR) models which can predict a range of important physicochemical fuel properties based on molecular descriptors computable from the two-dimensional molecular graph. The analysis of two comprehensive case studies reveals that compact ketones, furans and esters represent knock-resistant compounds which also exhibit favorable thermophysical properties deemed important for the in-cylinder mixture formation process in spark-ignition (SI) engines. In contrast, cyclic and acyclic ethers of moderate size readily auto-ignite and therefore represent first choice candidates for compression-ignition (CI) engines. Moreover, the high fuel oxygen contents, the low viscosities and the high volatilities of the ether compounds are expected to result in low levels of engine-out soot emissions. Finally, an optimization-based approach for the formulation of multicomponent biofuels by means of integrated product and pathway design is presented. Here, the objective is to maximize a process-related quantity, i.e., the energy of fuel produced (in terms of the lower heating value), and the constraints in the problem formulation allow to define target ranges for the blend's physicochemical properties. To account for non-ideal mixture behavior with respect to two important properties of SI engine fuels, i.e., the Reid vapor pressure and the distillation curve, the nonlinear program includes the UNIFAC group contribution model. Application of the new design methodology to a case study underlines the significance of performing combined product and pathway design, since only few investigated blends are found to exhibit both the desired fuel properties and attractive process-related properties.

Kurzfassung

Die vorliegende Arbeit beschreibt modellbasierte Strategien zur Identifikation lignocellulosebasierter Kraftstoffkandidaten mit vielversprechenden Eigenschaften für den Einsatz in hoch-effizienten und schadstoffarmen Verbrennungsmotoren. Auf Basis theoretischer Grundlagen zum molekularen Maßschneidern von Kraftstoffen wird dabei zunächst ein Algorithmus zur zielgerichteten Molekülstrukturgenerierung vorgestellt, der den Suchraum mittels eines regelbasierten Ansatzes systematisch aufspannt und dabei dem Konzept einer selektiven Refunktionalisierung biobasierter Plattformchemikalien folgt. Für die modellbasierte Bewertung der so erhaltenen Strukturen liefert die vorliegende Arbeit einen essentiellen Baustein in Form einer Gruppenbeitragsmethode zur Vorhersage der abgeleiteten Cetanzahl (engl. Abk. DCN). Daneben fußt die virtuelle Kraftstoffsuche vor allem auf maßgeschneiderten quantitativen Struktur-Eigenschafts-Beziehungen (engl. Abk. QSPR), die wesentliche Kraftstoffeigenschaften als Funktion molekularer Deskriptoren beschreiben. Die Analyse umfangreicher Fallstudien zeigt, dass kompakte Ketone, Furane und Ester sehr klopffeste Verbindungen darstellen, die zudem günstige Eigenschaften für die Gemischbildung im Ottomotor aufweisen. Für den Dieselmotor hingegen kommen vor allem cyclische und acyclische Ether mittlerer Größe in Frage, da diese Stoffe eine hohe Zündwilligkeit besitzen. Die hohen Sauerstoffgehalte, die vergleichsweise niedrigen Siedepunkte und die geringen Viskositäten der Etherkraftstoffe lassen zudem niedrige Partikelemissionen bei der Verbrennung im Dieselmotor erwarten. Schließlich wird ein optimierungsbasierter Ansatz vorgestellt, der ein integriertes Produkt- und Pfadentwurfsproblem zur Formulierung von Kraftstoffmischungen mit gewünschten Eigenschaften löst. Im Zielfunktional der Optimierung steht dabei mit der produzierten Energiemenge des Kraftstoffs (gemessen am Heizwert) eine prozessrelevante Größe. Die in den Nebenbedingungen des Problems auftretenden Stoffdatenmodelle erlauben die Beschränkung physikalisch-chemischer Kraftstoffeigenschaften und umfassen die UNIFAC-Gruppenbeitragsmethode, um die Einflüsse von nicht-idealem Mischungsverhalten auf Dampfdruck und Destillationskurve zu beschreiben. Die Anwendung der neuen Entwurfsmethode auf eine Fallstudie verdeutlicht die Wichtigkeit einer integrierten Betrachtung von Produkt- und Pfadentwurf, denn nur eine kleine Zahl der untersuchten Gemische weist neben den wünschenswerten Kraftstoffeigenschaften auch attraktive Prozesseigenschaften auf.

1 Introduction

Fuels derived from lignocellulosic biomass represent an important technological option for future sustainable mobile propulsion, as they are are thought to exert less negative impacts on biodiversity, land use and food production compared with the first-generation biofuels (Naik et al., 2010). These lignocellulosic fuels can be produced via three main routes: gasification, pyrolysis and hydrolysis (Lange, 2007). The two thermochemical approaches yield complex multicomponent mixtures of (oxygenated) hydrocarbon species. In contrast, hydrolysis and aqueous-phase processing of water-soluble sugars generally refer to a sequence of reactions that facilitates a selective synthesis of one or few oxygenated building blocks via chemo-catalytic and/or bio-catalytic routes (Serrano-Ruiz and Dumesic, 2012). As indicated in Figure 1.1, such oxygenated platforms can then be refunctionalized catalytically into oxygenated fuel components exhibiting tailored properties for high-efficiency and low-emission internal combustion engines (Janssen et al., 2011). Based on this idea, the Cluster of Excellence "Tailor-Made Fuels from Biomass" (TMFB) at RWTH Aachen University develops a comprehensive fuel design methodology that considers the fuel's molecular structure as the fundamental degree of freedom in optimizing both production and combustion of lignocellulosic biofuels (Janssen et al., 2011; Hoppe et al., 2016b).

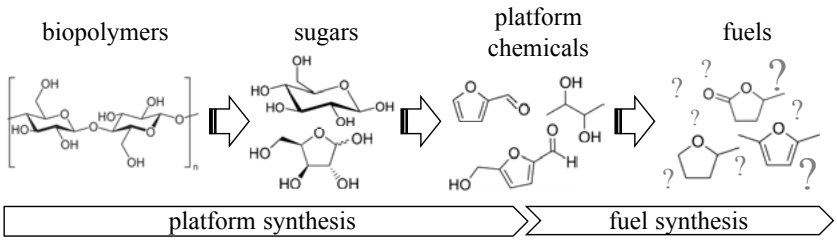


Figure 1.1: Two-step biomass-to-biofuel conversion strategy as pursued within TMFB (based on Janssen et al. (2011)): Aqueous-phase processing of cellulose and hemicellulose yielding platform chemicals followed by selective chemo-catalytic refunctionalization producing oxygenated fuel components.

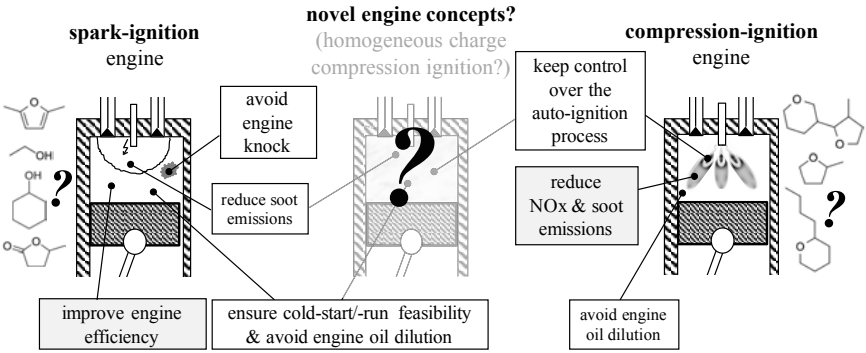


Figure 1.2: The primary objectives of fuel design as pursued within TMFB are (i) the improvement of engine efficiency (primarily spark-ignition engine) and (ii) the reduction of engine-out emissions of NOx and soot (primarily compression-ignition engine).

As illustrated in Figure 1.2, the primary objective of fuel design is to enable high-efficiency and clean combustion in both spark-ignition (SI) end compression-ignition (CI) engines. Burning pure ethanol instead of RON95 gasoline, for instance, allows to raise the effective compression ratio and therefore to increase the efficiency of the SI engine without encountering the issue of engine knock (Jeuland et al., 2004; Thewes et al., 2011b; Hoppe et al., 2016b). The efficiency increase is made possible by ethanol's high octane rating as well as ethanol's high enthalpy of vaporization which provides a high charge cooling effect. However, ethanol's high enthalpy of vaporization and boiling point also cause cold-start and cold-run problems, as an ethanol-rich fuel can condense on the cold cylinder wall thereby causing excessive emissions and engine oil dilution (Jeuland et al., 2004; Larsen et al., 2009; Thewes et al., 2011b; Hoppe et al., 2016b).

In case of the CI engine, the simultaneous reduction of engine-out soot and NOx emissions is the primary objective of TMFB's fuel design efforts (cf. Figure 1.2). To this end, an oxygenated fuel that has a low propensity to form soot is combined with exhaust gas recirculation to achieve low-temperature combustion (Janssen et al., 2011; Heuser et al., 2013a). Unlike diesel fuel, the oxygenate allows to shift the soot/NOx trade-off curve of CI engine combustion towards significantly lower values for both soot and NOx emissions (Boot et al., 2008; Janssen et al., 2011; Heuser et al., 2013a; Bhardwaj et al., 2013; García et al., 2016). In order to control the start of combustion in the CI engine, the fuel/air-mixture must auto-ignite briefly after the start of injection, whereas in the SI engine the fuel must withstand auto-ignition to avoid engine knock. The assessment of fuel

auto-ignition propensity therefore constitutes another important aspect of fuel design. As indicated in Figure 1.2, synergies between oxygenated species and more recent combustion concepts, i.e., homogeneous charge compression ignition (HCCI) and reactivity controlled compression ignition (RCCI), have also been explored in recent years (Yang et al., 2010; Contino et al., 2011b; Yang and Dec, 2013; Heuser et al., 2016).

Although the number of promising bio-derived platform molecules available for the production of tailor-made fuels is limited (Werpy et al., 2004; Bozell and Petersen, 2010; Sheldon, 2014), these building blocks can still be upgraded to a tremendous variety of molecular structures, in principle. Most of these structures have never been synthesized before and a solely experiment-based search for fuel components would be inappropriate. Instead, computer-aided molecular design (CAMD) (Joback, 1989; Gani, 2004a; Ng et al., 2015b) should be used to augment experimental and theoretical work because CAMD provides a way to systematically generate molecular structures that hold the promise of having the desired properties (Hechinger et al., 2010; Cholakov, 2011).

Originating from the research performed by the author during his time as a member of TMFB (from 2011 to early 2017), this thesis represents a continuation of previous research on model-based fuel design performed at the Institute for Process Systems Engineering at Aachener Verfahrenstechnik (Hechinger et al., 2010; Hechinger and Marquardt, 2010; Hechinger et al., 2012a,b; Dahmen et al., 2012; Victoria Villeda et al., 2012a; Voll and Marquardt, 2012a,b; Hechinger, 2014; Victoria Villeda et al., 2015; Victoria Villeda, 2017) and deals with computational methods supporting both the identification of tailored biofuel components and the rational formulation of mixtures of such components. To this end, three main concepts are outlined: (i) the targeted generation of candidate structures in an attempt to systematically explore the products from refunctionalization of bio-derived platform chemicals, (ii) a virtual fuel screening strategy based on predictive structure-property relationship modeling of key physicochemical fuel properties, and (iii) an optimization-driven approach to simultaneous product and pathway design of biofuel mixtures and their corresponding production routes. These methodologies are then applied to identify promising fuel candidates for both spark-ignition (SI) and compression-ignition (CI) engines.

1.1 Structure of this thesis

In Chapter 2 the basic concepts of fuel design are introduced. To this end, the relationships between the physicochemical properties of (oxygenated) fuels and the performance of internal combustion engines are analyzed based on a short literature review. Furthermore, bio-derived intermediates, i.e., platform chemicals, frequently proposed in the context of

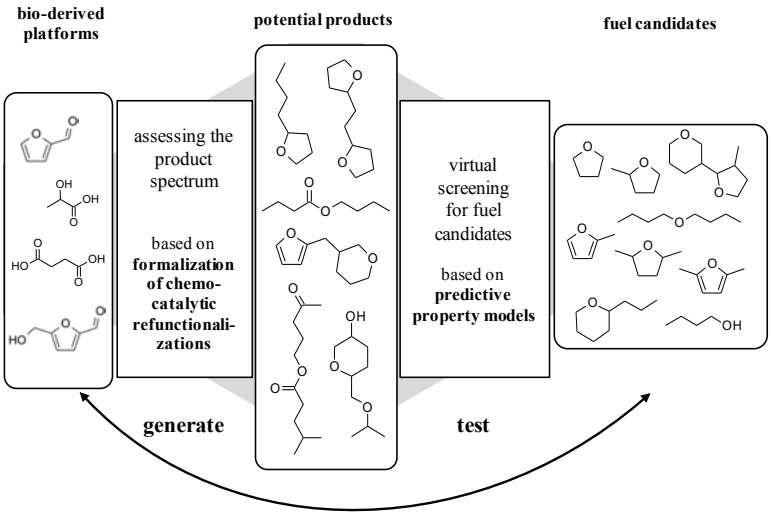


Figure 1.3: A generate & test realization of computer-aided molecular design (CAMD): Targeted generation of candidate structures followed by virtual fuel screening.

aqueous-phase bio-renewables processing are evaluated with regard to their potential in fuel production. The concept of carbon- and energy-efficient biofuel production is introduced here which acts as the guiding principle for the targeted generation of candidate structures proposed in the subsequent Chapter. Chapter 2 concludes with a brief literature review on methodologies for computer-aided design of chemical products and fuels.

In Chapter 3, the concepts of a novel molecular structure generator are described which links the molecular graphs of CAMD products to pre-defined platform chemicals by considering a few, simple transformation rules chosen to resemble carbon- and energy-efficient chemo-catalytic refunctionalization. As indicated in Figure 1.3, this algorithm allows to automatically generate a spectrum of potential products once sets of platforms and transformations have been specified by the modeler. The product spectrum can then be screened for biofuel candidates by employing computational property prediction.

Chapters 4 and 5 are entirely devoted to the mathematical modeling of the key physicochemical fuel properties that have been identified in Chapter 2. Due to their relatively simple nature resulting in high computational efficiency, group contribution (GC) models (Joback and Reid, 1987; Marrero and Gani, 2001) and quantitative structure-property relationship (QSPR) models (Katritzky et al., 1995, 2010) are readily applicable to the

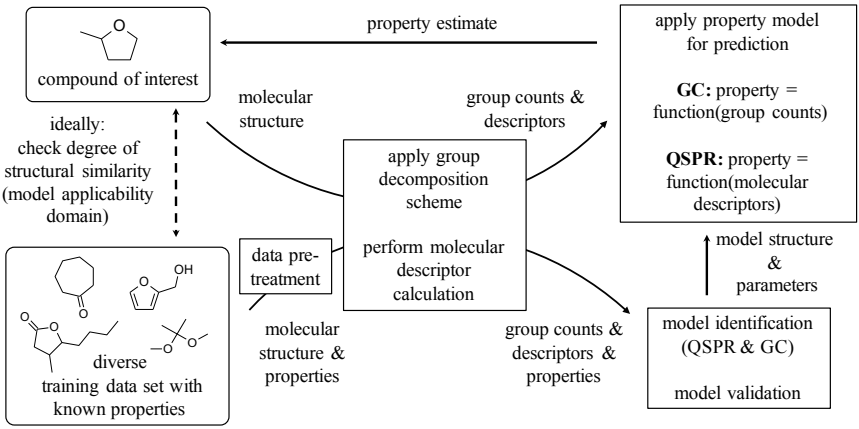


Figure 1.4: Computational property prediction by means of quantitative structure-property relationship (QSPR) and/or group contribution (GC) modeling.

property prediction task in the context of model-based fuel design (Hechinger et al., 2010; Cholakov, 2011; Dahmen et al., 2012; Saldana et al., 2013; Hechinger, 2014; Dahmen and Marquardt, 2016). GCs rely on the principle of group additivity, i.e., the assumption that the property of a molecule results from the summation of the contributions from each of the molecule's atoms or structural groups (Bader and Bayles, 2000). The group additivity principle is related to the similarity principle that constitutes the fundamental assumption behind QSPR modeling (Leonard and Roy, 2006; Tropsha and Golbraikh, 2010). The latter principle states that similar compounds, i.e., compounds exhibiting similar molecular descriptor data (Todeschini and Consonni, 2008) which can be computed based on the two- or three-dimensional molecular structure, will also exhibit similar properties (Leonard and Roy, 2006; Tropsha and Golbraikh, 2010). Figure 1.4 shows the main aspects of both GC and QSPR modeling covered in this thesis, i.e., data pretreatment and similarity analysis, group decomposition, descriptor calculation, model identification, model validation and property prediction.

As indicated in the lower half of Figure 1.4, QSPR and GC models are derived on the basis of sufficiently large and diverse collections of known molecular structures with known property data. Fortunately, such databases exist for many pure-component thermophysical properties relevant to fuel design. However, the assessment of fuel auto-ignition quality, which is fundamental in deciding on the target engine type, is often complicated in case of oxygenated species for two reasons. Whereas the combustion kinetics of first-

generation biofuels, i.e., ethanol and fatty acid methyl esters, have been studied extensively (Kohse-Höinghaus et al., 2010; Westbrook, 2013), mechanism development for long-chained alcohols, methyl esters and small cyclic and acyclic ethers is subject to current research (Herbinet et al., 2008; Tran et al., 2012; Sarathy et al., 2012; Westbrook, 2013; Sarathy et al., 2014; Cai et al., 2014, 2015; Dryer, 2015; Tripathi et al., 2017). With regard to a wider range of oxygenated fuels, such predictive kinetic models are expected to become available only in the long-term future. Furthermore, ignition delays of different fuels can only be compared in a meaningful way, if experimental boundary conditions are chosen to be identical or at least similar, because differences in mixture formation, air temperature and pressure in the combustion chamber also exert strong effects on the measured ignition delay and thus can blur the influence of the molecular structure. To address these challenges, Chapter 4 first describes the results from a rapid fuel screening campaign with an Ignition Quality Tester (IQT) (Allard et al., 1996, 1997; ASTM D6890, 2011). The IQT's standardized constant-volume combustion chamber is operated at a single, well-defined, engine-relevant condition that allows for a sound comparison of fuels ranging from high-octane gasoline-like fuels to high-cetane synthetic diesel fuels. Because of its ability to generate high-quality ignition delay data from 32 combustion cycles in less than 20 minutes given a sample of approximately 50 mL only, the IQT truly constitutes a rapid screening device. In contrast, determination of classical octane number (ON) and cetane number (CN) requires engine experiments involving a much larger sample ($\sim$ 1 L) and significantly more time (approximately a few hours) (Ghosh and Jaffe, 2006; Ghosh, 2008). Together with IQT data taken from the literature, a database has been established that covers 161 (oxygenated) hydrocarbons, i.e., acyclic and cyclic, branched and straight, saturated and unsaturated hydrocarbons as well as alcohols, ethers, esters, ketones, aldehydes, and aromatic and polyfunctional species. This database is subsequently used in Chapter 4 to derive a simple, yet predictive group contribution model for the so-called derived cetane number (DCN), i.e., a single scalar quantity that allows for a first characterization of a fuel's autoignition propensity in the context of (computational) fuel design.

Chapter 5 describes a generic strategy for QSPR modeling of different physicochemical fuel properties. For some of the properties considered in this thesis, the number of molecular descriptors available for modeling exceeds the number of training molecules with known property data. Moreover, in case of all properties, the degree of multicollinearity in the descriptor data is substantial. In order to derive predictive structure-property relationships on the basis of such data, principal component analysis (PCA) (Jolliffe, 2002) and partial least squares (PLS) regression (Höskuldsson, 1988) have been applied extensively in the literature (Cramer, 1993; Katritzky et al., 1995; Wold et al., 2001; Eriksson et al., 2006a; Katritzky et al., 2010). In Chapter 5, it is demonstrated that PCA and PLS can be used

to make predictions for all physicochemical fuel properties considered in this thesis. The predictive power of each QSPR model is confirmed by external validation, i.e., by testing the model on molecular structures that have not participated in model training (Tropsha et al., 2003). With regard to the virtual fuel screening (cf. Figure 1.3), an applicability domain (AD) concept has been implemented that allows to distinguish whether the QSPR model performs interpolation or extrapolation when it is used to predict the property of a specific target compound (Netzeva et al., 2005; Weaver and Gleeson, 2008). The Chapter concludes with a performance comparison between the QSPR models proposed here and two established group contribution methods taken from the publicly available literature.

Chapter 6 is dedicated to case studies where targeted structure generation and computational property prediction are applied to systematically explore and identify novel fuel candidates for spark-ignition and compression-ignition engines (cf. Figure 1.3). The most promising molecular motifs resulting from the virtual fuel screening are discussed in detail. While the discussion is primarily focussed on pure-component fuel candidates, some implications for blending applications are given.

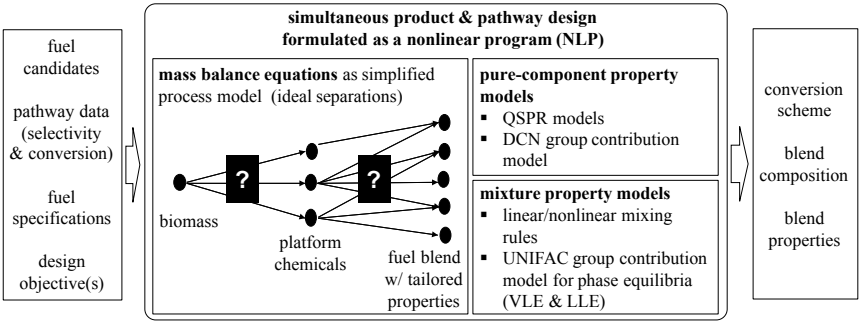


Figure 1.5: Optimization-driven approach for simultaneous product and pathway design of biofuel mixtures.

Chapter 7, finally, extends the scope beyond pure-component fuels and deals with the rational formulation of biofuel mixtures. Here, the blend design problem is stated as a simultaneous product and pathway design problem to take into account the fact that the additional degrees of freedom offered by a multicomponent fuel can be used to optimize properties of the production process, e.g., the mass or energy of fuel produced for a given quantity of biomass supplied. As can be seen from Figure 1.5, pathway selectivity and conversion data for all potential blend components are required to formulate a pathway model based on mass balance equations and the assumption of ideal separations, similar to

the methodology of reaction network flux analysis (RNFA) (Voll and Marquardt, 2012b). The pathway model is then augmented by GC and QSPR models for pure-component and mixture property prediction, including vapor-liquid equilibria (VLE) and liquid-liquid equilibria (LLE) models which allow to formulate constraints for the blend's distillation curve and to assess the blend's miscibility. The solution to the resulting nonlinear program (NLP) consists of a biofuel mixture with tailored properties and the corresponding optimal conversion scheme. The Chapter concludes with a case study of a blend formulation problem targeting a highly-boosted direct-injection spark-ignition engine.

Finally, Chapter 8 provides some conclusions on model-based design of pure and multicomponent cellulosic biofuels and is meant to give directions for further research in the field.

1.2 Previous publications of contents and results

This thesis has emerged from the research performed by the author during his time as a full-time researcher (Wissenschaftlicher Mitarbeiter) at the Institute for Process Systems Engineering at Aachener Verfahrenstechnik from February 2011 to January 2017. Throughout this entire period, the author has also been a member of TMFB. Most parts of this thesis have already been published:

- This Chapter (*Introduction*) is to some extent based on previous publications in *Energy & Fuels* (Dahmen and Marquardt, 2015, 2016, 2017).

- Parts of Chapter 2 (*Basic concepts of fuel design*), most parts of Chapter 3 (*Targeted generation of candidate structures*) and most parts of Chapter 6 (*Model-based identification of biofuel candidates*) as well as the contents of Appendix A (*Transformation rules for molecular structure generation*) have already been published in slightly modified forms in a contribution to *Energy & Fuels* (Dahmen and Marquardt, 2016).

- Parts of Chapter 2 (*Basic concepts of fuel design*), most parts of Chapter 4 (*GC-based prediction of fuel auto-ignition quality*), the contents of Appendix B (*Experimental ignition delay data*) and the contents of Appendix C (*GC model for the derived cetane number*) have already been published in slightly modified forms in a contribution to *Energy & Fuels* (Dahmen and Marquardt, 2015).

- Chapter 5 (*QSPR-based prediction of physicochemical fuel properties*) represents a continuation and a more detailed elaboration of the concepts published in *SAE International Journal of Fuels & Lubricants* (Dahmen et al., 2012). The basic principles

of the revised QSPR modeling approach as outlined in Chapter 5 have already been described briefly in a publication in *Energy & Fuels* (Dahmen and Marquardt, 2016). Statistical measures for the models derived in Chapter 5 have been published in two contributions to *Energy & Fuels* (Dahmen and Marquardt, 2016, 2017) where the models have been used to perform computational fuel design. Although being based on Dahmen et al. (2012) and Dahmen and Marquardt (2016, 2017), Chapter 5 has been almost entirely rewritten to provide a more rigorous theoretical description of the approach. Furthermore, the QSPR models are compared to existing property prediction methods.

- Most parts of Chapter 7 (*Blend formulation by simultaneous product and pathway design*) and the contents of Appendix E (*Rational formulation of biofuel mixtures*) have already been published in slightly modified forms in *Energy & Fuels* (Dahmen and Marquardt, 2017).

2 Basic concepts of fuel design

In this Chapter, the basic concepts behind fuel design, i.e., tailoring the molecular structure of a fuel to the specific needs of an internal combustion engine, shall be introduced. As depicted in Figure 2.1, fuel design does assume an understanding of how chemical structure determines physicochemical fuel properties, and how these properties impact engine performance. While these relationships are complex and not fully understood, some general trends can be derived from experience with fossil fuels of varying composition, but also from experience with a range of potential renewable fuels (Sorenson, 2001; Matijošius and Sokolovskij, 2009; Bradley, 2009; Starck et al., 2010; Kremer, 2011; Thewes et al., 2011b; Janssen et al., 2011; Christensen et al., 2011a; Reddemann et al., 2011; Rothamer and Donohue, 2013; Westbrook, 2013; Heuser et al., 2013a; Kalghatgi, 2014b; Baumgardner et al., 2015; McCormick et al., 2015; Hoppe et al., 2016a,b; Leitner et al., 2017). The first Section of this Chapter is dedicated to this discussion. Particular attention is focussed on the importance of fuel auto-ignition quality and its early assessment in the context of fuel design. The Section concludes with a brief review of cellulosic biofuels that have been studied in research engines and gives some perspectives on the definition of tailor-made fuels. Section 2.1 represents an aggregation of two contributions published separately in slightly modified forms in *Energy & Fuels* (Dahmen and Marquardt, 2015, 2016).

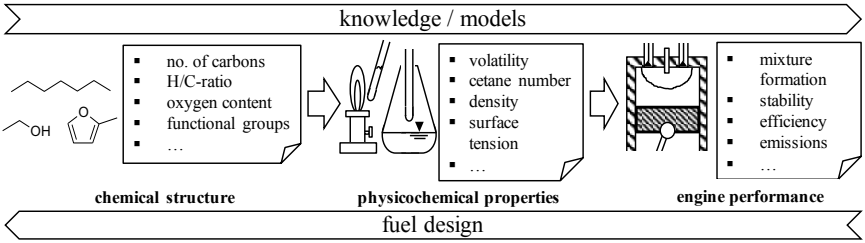


Figure 2.1: The inverse problem of fuel design: How do physicochemical fuel properties influence the performance of an internal combustion engine? Which molecular structures exhibit the desired properties?

Section 2.2 provides an overview on the production of bio-derived platform chemicals which constitute important intermediates in the synthesis of biochemicals and biofuels. To this end, building blocks frequently discussed in the literature are briefly reviewed. This forms the basis for Section 2.3 where a first evaluation of a platform's suitability for carbon- and energy-efficient biofuel production is presented. Sections 2.2 and 2.3 have already been published in a slightly modified form in *Energy & Fuels* (Dahmen and Marquardt, 2016).

Finally, Section 2.4 is meant to lay the foundations for the *model-based* fuel design concepts proposed in the later Chapters by briefly reviewing computational methods frequently employed in the design of chemical products and fuels, i.e., molecular structure generation, property modeling, and enumeration- and optimization-based design.

2.1 Physicochemical fuel properties and their impact on engine performance

Today's fuel standards, e.g., the European gasoline fuel standard EN 228 (2014) or the diesel fuel standard EN 590 (2014), guarantee that all fuel sold in the market is fully compatible with the existing car fleet. Furthermore, the standards are used to enforce legislative goals, e.g., the regulation of harmful emissions. However, if the full potential of oxygenated biofuels shall be exploited, the existing standards, most likely, are too restrictive. For instance, volatility, density and heating value differ significantly, if pure-component oxygenated fuels are compared to either gasoline (EN 228, 2014) or diesel (EN 590, 2014) fuel. There is no apparent reason why the historically established fossil fuel standards should define an "optimal" fuel, which is best-suited for running advanced internal combustion engines built to minimize emissions and maximize fuel efficiency.

Any formal definition of a tailored biofuel is complicated by the inherent complexities brought by the fuel/engine interaction. Additional complexity arises from the many degrees of freedom associated with different advanced combustion engine concepts. Still, some general relations between physicochemical fuel properties and engine performance can be derived and exploited for the purpose of fuel design as demonstrated by members of the TMFB cluster (Kremer, 2011; Klankermayer et al., 2011; Janssen et al., 2011; Thewes et al., 2011a,b; Victoria Villeda et al., 2012c; Dahmen et al., 2012; Hechinger et al., 2012b; Heuser et al., 2013a; Hechinger, 2014; Hoppe et al., 2016a,b; García et al., 2016; Dahmen and Marquardt, 2016; Leitner et al., 2017). Table 2.1 shows the physicochemical fuel properties that will be used throughout this thesis to identify fuel candidates by means of computational methods. This fuel definition has evolved from the regular exchange between the author of this thesis and members of the TMFB Core Interaction Field "Fuel Design", most notably Florian Kremer, Andreas Janssen, Benedikt Heuser, Fabian Hoppe

and Stefan Pischinger (Institute for Combustion Engines, RWTH Aachen University). The relevance of these properties is briefly discussed in the following.

Table 2.1: Important physicochemical properties of oxygenated fuel components and their impact on internal combustion engine performance.

fuel property	**SI engine**	**CI engine**
boiling point T_{boil} [°C]	oil dilution, cold start/run	soot, (oil dilution)
derived cetane number DCN [-]	compression ratio → efficiency	ignition control, (soot)
enthalpy of vaporization H_{vap} [kJ/kg]	cold start/run, charge cooling	—
oxygen content wt_{O_2} [%-wt]	(soot)	soot
lower heating value LHV [MJ/kg]	mileage per unit mass of fuel	
melting point T_{melt} [°C]	liquid fuel, winter operability	
liquid density ρ_L [kg/m^3]	mixture formation → emissions	
kinematic viscosity ν [mm^2/s]	mixture formation → emissions	
surface tension σ [mN/m]	mixture formation → emissions	

Liquid hydrocarbons have been the fuel of choice for transport applications over the past century due to their high volumetric energy density and ease in transportation, storage and handling (Kalghatgi, 2014b). And, although vehicles fueled by natural gas and liquefied petroleum gas have gained some market share, the focus of TMFB has always been on liquid renewable fuels. Hence, boiling and melting points are important properties in the design of biofuel components. Oxygenated fuels, however, exhibit a somewhat lower energy density than their fossil counterparts (Yanowitz et al., 2011). For instance, the volumetric energy density of ethanol is about a third lower than that of gasoline. This can be important for customer acceptance. However, from the technical point of view, this should not be a primary concern. Assuming a fixed volumetric intake of air, stoichiometric combustion in the SI engine requires more volume of fuel to be injected, if gasoline is substituted by ethanol. In fact, if this increase in volume is considered, the resulting in-cylinder energy, as measured by the mixture heating value, is slightly *higher* for a direct-injection spark-ignition engine running on ethanol instead of gasoline (Thewes et al., 2011b).

A fuel's volatility is especially critical for smooth operation of spark-ignition (SI) engines (Larsen et al., 2009; Yanowitz et al., 2011; Thewes et al., 2011b). Thewes et al. (2011b) report on high lube oil dilution for 1-butanol (boiling point of 118 °C) in their comparative assessment of alcohol-fuels. Besides the high boiling point, the high enthalpy of vaporization of the alcohol-fuels makes it difficult to perform engine cold-start, especially at low ambient temperatures (Jeuland et al., 2004; Thewes et al., 2011b; Hoppe et al., 2016a). However, a high latent heat of vaporization also has certain advantages: Once the warm-up driving phase has been completed, it provides a high charge cooling effect that enhances

filling efficiency (Jeuland et al., 2004; Milpied et al., 2009). Moreover, engine knock is less likely to occur in a cooler running engine (Larsen et al., 2009; Rothamer and Jennings, 2012).

Compared to SI engines, compression-ignition (CI) engines can burn less volatile fuel. Still, oil dilution can become a concern, if fuel components are too heavy. For instance, biodiesel, i.e., a mixture of long-chain alkyl esters, is less volatile than conventional diesel and hence wall impingement of liquid fuel becomes more likely, especially if advanced injection strategies are implemented to achieve low-temperature combustion aiming for reduced NOx and soot emissions (Dec, 2009; Fisher et al., 2010).

Fuel evaporation and mixing with air must occur in the short period between injection and ignition. In CI engines, the poor availability of oxygen in the fuel-rich regions of burning diesel jets triggers soot formation via the nucleation of intermediate hydrocarbon species which cannot be oxidized completely (Sorenson, 2001; Westbrook et al., 2006; Tree and Svensson, 2007; Dec, 2009; Kalghatgi, 2014b). As the liquid fuel is injected into the cylinder, it forms a cone-shaped spray that entrains hot gases from its surrounding. Soot formation is initiated when the fuel-rich mixture ignites and, eventually, the rich mixture burns out at the jet's periphery in a stoichiometric diffusion flame (Tree and Svensson, 2007; Dec, 2009). Since the diffusion flame limits the amount of oxygen available within the jet, non-sooting combustion requires that the air entrainment prior to the lift-off length, i.e., the distance between the nozzle and the beginning of the diffusion flame, is strong enough to form a mixture, which is too lean for excessive soot production (Pickett and Siebers, 2004; Tree and Svensson, 2007).

Several authors (Lee et al., 2002, 2005; Park et al., 2009, 2010; Wang et al., 2011; De Ojeda et al., 2011; Janssen et al., 2011; Heuser et al., 2013a,b, 2014; Wang et al., 2014; Liu et al., 2014) have concluded that a high fuel volatility, a low surface tension and/or a low viscosity can enhance fuel atomization and/or evaporation and thereby improve homogenization of the mixture. Others (Siebers, 1999; Tree and Svensson, 2007; Groendyk and Rothamer, 2015) have argued that, with today's high injection pressures, fuel atomization and interphase transport of mass and energy at droplet surfaces in many situations no longer constitute limiting factors, as diesel jets develop in a mixing-limited regime, i.e., a regime where fuel vaporization is predominantly controlled by the air entrainment and by the mixing of fuel with oxygen in the cross-sectional area of the spray. In the same line of thought, it has been hypothesized that a long ignition delay could improve mixture homogenization and thus could decrease soot production as its lower reactivity would yield a larger lift-off length (Boot et al., 2008; Donkerbroek et al., 2011; Janssen et al., 2011; García et al., 2016). However, further systematic investigations of fuel spray development at conditions representative of low-temperature diesel combustion must be performed on a

broader range of alternative (oxygenated) fuels to clarify and separate the impacts of the different physicochemical fuel properties on the mixture formation process in CI engines (Reddemann et al., 2010, 2011; García et al., 2016). If direct-injection SI engines are operated in the fuel-efficient stratified mode, mixture homogeneity is also less pronounced than in older SI engines using port fuel-injection systems. Similar to CI engines, it is believed that soot is formed under locally rich conditions here (Hemdal et al., 2011; Storch et al., 2015).

An interesting option for shifting the local carbon to oxygen ratio in the air/fuel-mixture towards the more favorable regime is to burn a fuel with a high oxygen content (Sorenson, 2001). Oxygenated fuels have indeed been found to effectively reduce emissions of particulate matter (Tree and Svensson, 2007; Janssen et al., 2011; Liu et al., 2012; García et al., 2016). While it seemed initially that the amount of soot reduction is controlled mainly by the fuel oxygen mass fraction (Sorenson, 2001), kinetic modeling suggests that different functional groups containing oxygen, e.g., alcohol, ether and ester, differ in their propensity to suppress the production of soot precursor species (Westbrook et al., 2006). On the basis of experimental observations, Boot et al. (2008) have hypothesized that oxygenates in which an oxygen atom is bonded to two carbon atoms (C−O−C) would have the highest potential to reduce soot emissions. The underlying idea is that the fuel oxygen sequesters the bonded carbon into partially oxidized C_1 or C_2 species and that therefore this carbon is no longer available for the formation of soot precursors like ethylene or acetylene (Boot et al., 2008). If the use of an oxygenated fuel is combined with exhaust gas recirculation, which can effectively reduce NOx emissions, the classical soot/NOx trade-off curve of CI engine combustion can be shifted towards significantly lower values for both soot and NOx emissions (Boot et al., 2008; Janssen et al., 2011; Heuser et al., 2013a; Bhardwaj et al., 2013; García et al., 2016). In SI engine combustion, ethanol-blended fuels have likewise revealed reduced emissions of particulate matter. However, for certain operating conditions, ethanol-containing fuel can increase soot emissions, because ethanol's physical properties, e.g., its high enthalpy of vaporization and its high viscosity, can have negative effects on the in-cylinder mixture formation process (Storch et al., 2015).

Ignition control is key for smooth and efficient engine operation. In the CI engine, the fuel needs to auto-ignite shortly after it has been injected into the compressed hot air. Fuels, that readily auto-ignite under temperature and pressure attained near the end of the compression stroke, exhibit a high cetane number (CN) (ASTM D613, 2015). In contrast, the air-fuel mixture is ignited by a spark from a spark plug in SI engines. Here, the fuel has to withstand auto-ignition during the compression stroke as well as in the so-called endgas, i.e., the region that has not yet been reached by the flame. Otherwise, knock, i.e., a steep rise in pressure not timed properly with the combustion event, can do severe

damage to the engine (Kalghatgi, 2005, 2015). Fuel antiknock quality is most commonly expressed by means of research octane number (RON) (ASTM D2699, 2013) and motor octane number (MON)(ASTM D2700, 2014). It is closely related to SI engine efficiency and power output. The trend towards downsizing and turbocharging, i.e., forcing as much air as possible into the cylinder, makes knock more likely to occur due to the resulting high temperature and pressure of the charge. In certain operational regimes, most modern high-performance SI engines are knock-limited for regular fuels (Kalghatgi, 2015). The kinetically controlled homogenous charge compression ignition (HCCI) engine is expected to combine the advantages of the SI engine (low emissions) and of the CI engine (high efficiency); it relies on the auto-ignition event taking place at a desired point in time for given temperature and pressure histories of premixed air and fuel in the cylinder for a certain varying load, intake air temperature and other operational boundary conditions (Yao et al., 2009; Kalghatgi, 2014b). Due to control problems and load restrictions the realization of a full HCCI engine is considered unlikely. However, (partially) premixed compression-ignition (PCI) combustion is a promising concept for the simultaneous reduction of soot and NOx emissions (Kalghatgi, 2014b): A longer ignition delay (corresponding to a lower CN) allows for earlier injection to provide more time for mixing fuel and oxygen before the combustion starts. This way, high levels of soot emissions can be avoided.

2.1.1 Early-stage, rapid screening of fuel auto-ignition quality

Single-cylinder cooperative fuels research (CFR) engines are used to experimentally determine RON, MON and CN (Kalghatgi, 2014b). The early screening of novel biofuel candidates, however, calls for a laboratory test method able to run on a small volume of fuel. The Ignition Quality Tester (IQT) described in ASTM D6890 (2011) determines the ignition quality of a fuel by measuring the ignition delay in a heated, temperature-controlled, constant-volume combustion chamber (Allard et al., 1997, 1996; Kalghatgi, 2014a). Advantages of this approach include excellent reproducibility, fast execution and low fuel demand (Ghosh and Jaffe, 2006; Ghosh, 2008). Moreover, since many pure-component (oxygenated) hydrocarbon fuels have been studied in this apparatus, there is already a collection of ignition delay data available in the literature. The IQT ignition delay is frequently re-expressed as *derived* cetane number (DCN). DCN is thought to represent an approximation of CN (Allard et al., 1997, 1996; ASTM D6890, 2011). Similarly, Perez and Boehman (2012) have suggested IQT correlations for RON and MON.

In the IQT apparatus, the injection of liquid fuel into compressed hot air is followed by simultaneously occurring evaporation, mixing and chemical reaction as illustrated in Figure 2.2. The buildup of radicals finally leads to the ignition event. Conditions, processes and ignition delays are similar to those observed in CI engines. Due to the precisely controlled

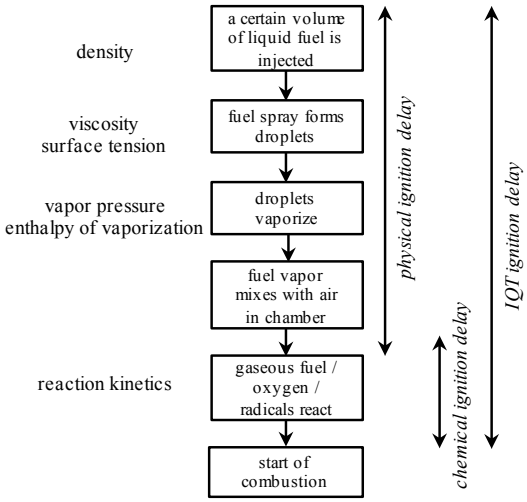


Figure 2.2: The processes occurring in the IQT and relevant fuel properties (based on Murphy et al. (2004)). The physical contribution to the ignition delay is usually completed within a few milliseconds (Ryan and Matheaus, 2003). The contribution of the chemical delay is similar in case of diesel-like fuels, but is much larger for gasoline-like fuels. Physical and chemical contribution to ignition delay overlap, since gas-phase reactions start as soon as the fuel vaporizes.

conditions in the IQT combustion chamber, the reproducibility errors on IQT derived CN are typically significantly lower than those of the ASTM D613 (2015) CFR engine (Allard et al., 1996, 1997; Ghosh and Jaffe, 2006; Ghosh, 2008). It should be noted that chemically different fuels such as tetrahydrofurans or ethers were not included in the derivation of the DCN equation (ASTM D6890, 2011). Still, high reactivity indicated by a diesel-like IQT ignition delay suggests further investigation of a fuel candidate towards use in CI engines. Such evidence can be generated by CN determination according to ASTM D613 (2015).

Perez and Boehman (2012) have based their IQT correlations for RON and MON on 21 surrogate fuels for gasoline. In contrast to their diesel-like counterparts, fuels with high ON show long delays (ranging from approximately 15 ms up to 100 ms in the IQT; cf. Appendix B). For some anti-knock agents such as toluene or methyl tert-butyl ether (MTBE) the ignition event is even absent, i.e., no ignition can be detected within 100 ms. The long delay is attributed to low reactivity because Ryan and Matheaus (2003) report evaporation and mixing in the IQT to last a few milliseconds only. Similarly, Bogin et

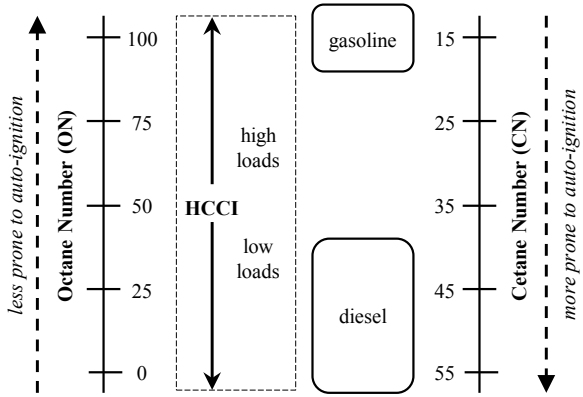


Figure 2.3: Qualitative sketch of the relationship between ON and CN (adapted from Kalghatgi (2005)). Typical ranges for ON / CN of gasoline and diesel fuels are given.

al. (Bogin Jr et al., 2011, 2013; Osecky, 2013), who have studied the mixture formation process in the IQT combustion chamber by means of computational fluid dynamics, have found that the fuel-air mixture in the IQT becomes pseudo-homogeneous for long ignition delays ($\sim$ 20 ms and longer). This conclusion has been shared by Perez and Boehman (2012). Auto-ignition in SI engines typically originates from the premixed endgas, i.e., the region where the air/fuel mixture has not yet been reached by the flame. Despite notable differences in mixture preparation and boundary conditions, ON and (D)CN are loosely negatively correlated as indicated in Figure 2.3 (Bowden et al., 1974; Ryan and Matheaus, 2003; Kalghatgi, 2005; Haas et al., 2011; Perez and Boehman, 2012). However, it is generally believed that ignition delay measured at a single temperature is a rather poor discriminator among different *high* ON fuels (Griffiths et al., 1997; Tanaka et al., 2003; Kalghatgi, 2005). Hence, if a long delay is measured in the IQT, the knock-resistance of the fuel candidate should be further quantified by determination of research octane number (RON) and motor octane number (MON) according to ASTM D2699 (2013) and ASTM D2700 (2014), respectively.

Ideal fuels for HCCI and PCI combustion comprise a medium reactivity, i.e., these fuels should have a moderate ON *and* a moderate CN (Kalghatgi, 2014b). On the IQT ignition delay scale, such a fuel is expected to lie somewhere in between the extremes defined by diesel-like and gasoline-like fuels. While CN determination is expected to offer further guidance for PCI combustion, auto-ignition in HCCI engines occurs over wide ranges of temperature and pressure. Most likely, detailed kinetic experimentation, e.g., in a shock

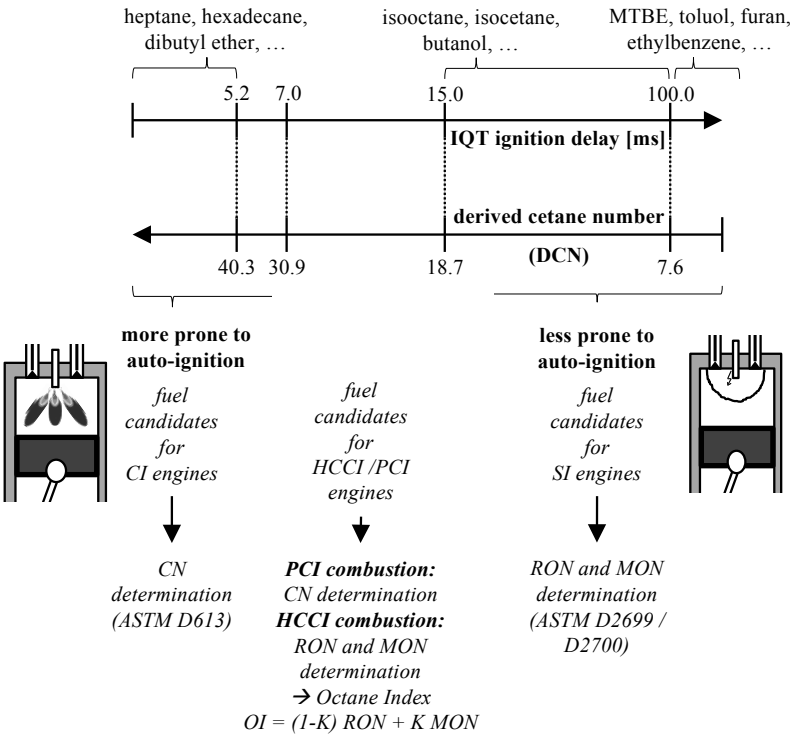


Figure 2.4: Qualitative sketch of the relationship between IQT ignition delay and derived cetane number (DCN). The axis do not scale linearly. Three segments on the IQT ignition delay scale are related to fuel candidates for different types of internal combustion engines (SI, PCI & HCCI, CI).

tube or in a rapid compression machine, is required to assess whether a novel compound can be used in this challenging type of combustion engine. A new control system for the IQT allows to study fuel auto-ignition for a wider range of pressure and temperature (Bogin Jr et al., 2011, 2013, 2014). This paves the way for the validation of chemical kinetic mechanisms by means of IQT measurements, if conditions and fuels are chosen to produce long ignition delays ($\sim$ 20 ms and larger) (Haas et al., 2011; Bogin Jr et al., 2011, 2013; Osecky, 2013; Bogin Jr et al., 2014). As an empirical alternative to detailed kinetic investigation, Kalghatgi (2005, 2014b) has proposed the so-called octane index, which can be obtained from rearranging RON, MON and a parameter K, which is not a fuel property, but depends on the particular type of engine.

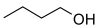
Figure 2.4 summarizes the interpretation of IQT data based on the aforementioned relationships between IQT ignition delay and auto-ignition in practical engines. To this end, three coarse segments on the IQT ignition delay scale are distinguished which refer to fuel candidates for CI, PCI & HCCI and SI engines, respectively. Figure 2.4 also suggests confirmatory analysis, e.g., CN determination in the CFR engine according to ASTM D613 (2015). It is expected that the classification is most reliable for extreme fuels, i.e., fuels comprising either a high CN (short IQT delay; cf. Figure 2.4, left) or a high ON (long or absent IQT delay; cf. Figure 2.4 right).

Although determination of CN, RON and MON can be seen as validation experiments, these cover a restricted range of possible engine operation only, thus highlighting the limitations of all conventional test procedures and empirical measures for the assessment of auto-ignition (Bradley, 2009; Kalghatgi, 2014b). Due to the complex interactions between fuel chemistry, thermophysical fuel properties and operational boundary conditions, the degrees of freedom of a particular internal combustion engine will most likely require optimization for a new fuel to unlock its full potential. This has been successfully demonstrated for 2-methylfuran (Thewes et al., 2011a; Hoppe et al., 2016b; Leitner et al., 2017), i.e., a novel pure-component biofuel (exhibiting a DCN of 9.1) for highly-boosted direct-injection SI engines.

2.1.2 Cellulosic biofuels studied in research engines

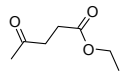
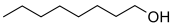
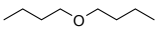
Based on the aforementioned relationships between physicochemical fuel properties and engine performance, a range of well known cellulosic biofuels has been identified and is briefly reviewed in Table 2.2. Besides the alcohols, the list includes furan derivatives, ethers and esters. Not all compounds exhibit ideal properties. γ-Valerolactone (Horváth et al., 2008) and ethyl levulinate (Christensen et al., 2011a), for instance, exhibit high RONs; however, poor volatilities strongly limit blending into gasoline (Yanowitz et al., 2011). The low DCN of ethyl levulinate does not allow it to replace a significant portion of fossil diesel fuel, either (Christensen et al., 2011a). The use of the alcohols is impaired by low volatilities and high enthalpies of vaporization (Thewes et al., 2011b). 2-Methyltetrahydrofuran (2-MTHF) can be synthesized efficiently from levulinic acid (Geilen et al., 2010). However, if 2-MTHF is blended into gasoline or ethanol, its poor RON (Christensen et al., 2011b) negatively affects the knock-resistance of the mixture. Likewise, the low DCN sets a limit for blending 2-MTHF into diesel fuel (Janssen et al., 2011).

Table 2.2: Cellulosic biofuels that have been studied in research engines.

compound	engine	main findings
 γ-valerolactone	SI	**pros:** high knock-resistance (RON 100), higher energy density than ethanol (Yanowitz et al., 2011); **cons:** very poor volatility (boiling point 218 °C) (Yanowitz et al., 2011); ⇒ blend component only
 2-methylfuran	SI	**pros:** high knock-resistance (RON 100.7), higher energy density than ethanol, improved cold start/run due to gasoline-like enthalpy of vaporization and low boiling point (64 °C) (Thewes et al., 2011a); **cons:** less knock-resistant than ethanol (Thewes et al., 2011a); **note:** the structurally similar 2,5-dimethylfuran has also been studied (Christensen et al., 2011b; Yanowitz et al., 2011; Rothamer and Jennings, 2012; Qian et al., 2015)
 ethanol	SI	**pros:** very high knock-resistance (RON 108.6) (Thewes et al., 2011a,b), charge cooling due to high enthalpy of vaporization (Jeuland et al., 2004); **cons:** low energy density, problematic cold start/run due to high enthalpy of vaporization and boiling point (78 °C) (Larsen et al., 2009; Thewes et al., 2011a,b)
 1-butanol	SI	**pros:** higher energy density than ethanol, less corrosive than ethanol (Jin et al., 2011; Yanowitz et al., 2011); **cons:** knock-resistance similar to RON95 but significantly worse than ethanol (Thewes et al., 2011b); problematic cold start/run and high oil dilution due to high boiling point (118 °C) (Thewes et al., 2011b; Yanowitz et al., 2011); **note:** butanol and pentanol isomers have also been studied (Yanowitz et al., 2011)
 ethyl-tert-butyl ether	SI	**pros:** extreme knock-resistance (RON 118), higher energy density than ethanol, gasoline-like enthalpy of vaporization, low boiling point (72 °C) (Lakó et al., 2008)
 2-methyltetrahydrofuran	SI/CI	**pros:** higher energy density than ethanol, high volatility (Yanowitz et al., 2011); **cons:** poor knock-resistance (RON 86) (Christensen et al., 2011b); poor derived cetane number (DCN 21.3; cf. Appendix B); ⇒ blend component only

Continued on next page

Table 2.2: Cellulosic biofuels that have been studied in research engines (continued).

compound	engine	main findings
 ethyl levulinate	SI/CI	**pros:** high knock-resistance (Christensen et al., 2011a); **cons:** very poor volatility (boiling point 209 °C), very low derived cetane number (DCN <5) (Yanowitz et al., 2011; Christensen et al., 2011a); ⇒ blend component only; **note:** the structurally similar butyl levulinate has also been studied (Yanowitz et al., 2011; Christensen et al., 2011a,b)
 1-octanol	CI	**pros:** high energy density, nearly soot-free combustion (Heuser et al., 2013a); **cons:** high hydrocarbon and high engine noise emissions due to long ignition delay (Heuser et al., 2013a) (DCN 33.9; cf. Appendix B); high viscosity (7.3 mPa·s at 25 °C) (Heuser et al., 2013a)
 dibutylether	CI	**pros:** high energy density, high volatility (boiling point 141 °C), nearly soot-free combustion (Heuser et al., 2013a); **note:** extremely prone to auto-ignition (Heuser et al., 2013a) (DCN ∼115; cf. Appendix B)

Three cellulosic biofuels from Table 2.2, namely ethyl-tert-butyl ether (ETBE), 2-methylfuran and di-butylether, can be considered pure-component biofuels with tailored properties. ETBE is derived from ethanol and isobutylene (Lakó et al., 2008). Large-scale fermentative production of isobutylene has been estimated to be feasible at reasonable cost (van Leeuwen et al., 2012) and the French company Global Bioenergies has announced the production of fully renewable ETBE in early 2017 (Global Bioenergies, 2017). 2-Methylfuran is more resistant to engine knock than RON95 gasoline. This allows it to increase engine efficiency by up to 18% (Hoppe et al., 2016b). Moreover, 2-methylfuran exhibits better properties in cold start/run than ethanol due to a more gasoline-like enthalpy of vaporization (Thewes et al., 2011a; Hoppe et al., 2016a,b). Di-butylether, finally, is an excellent fuel for the CI engine, since it burns nearly soot-free, even if exhaust gas recirculation is applied to reduce combustion temperatures below the threshold for NOx formation (Heuser et al., 2013a; Hoppe et al., 2016b).

2.1.3 Perspectives on the definition of tailor-made fuels

Due to the fact that the physicochemical properties mentioned in Table 2.1 can be predicted solely on the basis of the two-dimensional molecular graph, a vast amount of

molecular structures can be screened computationally, thus guiding experimental investigation towards the most promising fuel candidates. It shall be stressed, however, that the fuel/engine interaction is far too complex to be described entirely by the properties listed in Table 2.1. However, as will be shown in Chapter 6, these properties already allow to limit the range of potential fuel candidates to a few, distinct molecular motifs.

Within TMFB, efforts are taken to provide predictive models for two important additional fuel properties, i.e., the wear scar diameter as a measure for fuel lubricity (Masuch et al., 2011; Weinebeck and Murrenhoff, 2013) and the threshold sooting index (TSI) as an indicator for a fuel's intrinsic propensity to produce soot (Graziano et al., 2016). Yang et al. (2007) have reported that the TSI seems to correlate well with soot emissions from jet turbine combustion. Shortly thereafter, Pepiot-Desjardins et al. (2008) and then later Barrientos et al. (2013) have proposed first structure-property relations for TSI prediction based on the principle of group additivity. Dooley et al. (2012a) have successfully employed TSI in the rational formulation of surrogate fuels designed to emulate the combustion behavior of Jet-A fuel. However, Pitz and Mueller (2011) have noted that it is not yet clear whether fuel-induced changes in TSI quantitatively reflect fuel-induced changes in engine-out soot emissions encountered in practical automotive diesel engines. Recently, Graziano et al. (2016) have proposed to tackle modeling of raw soot emissions from diesel engines by a combination of three characteristic fuel numbers: the intrinsic tendency to produce soot, the ignition delay and few thermophysical properties that influence soot oxidation via their impact on the in-cylinder mixture formation process.

Perspectively, zero- or quasi-dimensional engine models (Verhelst and Sheppard, 2009; Kumar et al., 2013) could play a role in fuel design as they allow to assess the performance of alternative fuels in more detail, if the fuel-specific influences of gas temperature, pressure and composition on the ignition delay and/or flame speed can be represented by empirical or semi-empirical submodels. For instance, submodels for the laminar burning velocity of small alcohol-fuels have been proposed by Hechinger (Hechinger and Marquardt, 2010; Hechinger, 2014) and Verhelst and co-workers (Vancoillie et al., 2011). The latter model has been implemented in a quasi-dimensional engine simulation together with an algebraic knock model (Vancoillie et al., 2014). Similarly, Vandersickel et al. (2012) have developed a three-stage Arrhenius approach describing the ignition delay as a function of temperature, pressure, equivalence ratio and exhaust gas recirculation. However, due to the large amount of experimental data that is required for parametrization and validation of such submodels for the variety of molecular structures, engine simulation most likely will have to be reserved to the few, most promising fuel classes.

Beyond the properties mentioned above, there is a multitude of fuel properties like toxicity or storage stability which, although being important, are not considered in this

thesis because model-based evaluation of these properties is yet impossible for the variety of molecular structure considered here. Harnisch et al. (2013) have compiled a comprehensive list of laboratory-based test methods suitable for the investigation of fundamental properties, combustion engine related properties and health related properties of prospective liquid biofuel compounds. Within TMFB, research has been directed towards the toxicological and ecotoxicological potencies of selected biofuel componds, e.g., 2-methylfuran, 2-methyltetrahydrofuran and ethyl levulinate (Bluhm et al., 2012, 2016).

2.2 Exploitation of the structure of lignocellulosic biomass

Non-edible biomass consists of three major fractions: cellulose, hemicellulose and lignin (Huber et al., 2006; Lange, 2007). Lignin accounts for approximately 15-20 wt-% (Wyman et al., 2005) and is a highly branched large polyaromatic compound (Huber et al., 2006). Efficient selective catalytic extraction of lignin monomer units is difficult due to the variety of different interunit linkages (Azadi et al., 2013). Consequently, aqueous-phase catalytic processing of lignocellulosic biomass primarily targets the easier accessible carbohydrates found in cellulose and hemicellulose (Huber et al., 2006; Alonso et al., 2010). In this context, however, the lignin stream can be burned to provide process heat or electricity (Huber et al., 2006; Lange, 2007; Alonso et al., 2010; Serrano-Ruiz and Dumesic, 2011; Azadi et al., 2013). The production of renewable hydrogen for upgrading the carbohydrate streams constitutes another possible utilization of lignin (Azadi et al., 2013).

Cellulose is a linear polymer of glucose monomers and represents approximately 40-50 wt-% of dry biomass (Wyman et al., 2005; Lange, 2007; Serrano-Ruiz and Dumesic, 2012). In contrast, hemicellulose is an amorphous polymer of five different C_5 and C_6 sugars (25-35 wt-% of lignocellulose) (Wyman et al., 2005; Alonso et al., 2010; Serrano-Ruiz and Dumesic, 2012). Chemical and/or mechanical pretreatment is needed to break the lignin protection before the polysaccharides can be deconstructed into monosaccharides such as glucose ($C_6H_{12}O_6$) and xylose ($C_5H_{10}O_5$) by enzymatic or acid hydrolysis (Serrano-Ruiz and Dumesic, 2011, 2012). After depolymerization, the sugars have to be deoxygenated and/or joined in order to obtain a liquid biofuel (Serrano-Ruiz and Dumesic, 2011, 2012). The production of cellulosic biofuel as a two-step process, i.e., chemo- or bio-catalytic conversion of (hemi-)cellulose into high-volume intermediates and subsequent chemo-catalytic upgrading to fuels, is sketched in Figure 2.5. Full deoxygenation of biomass is possible, however neither necessary nor meaningful, since a small amount of oxygen can be utilized to tailor fuel properties (Lange, 2007).

Chemical conversion of sugars into valuable products mainly proceeds via the formation of furfural and hydroxymethylfurfural (HMF) (Alonso et al., 2010; Serrano-Ruiz and

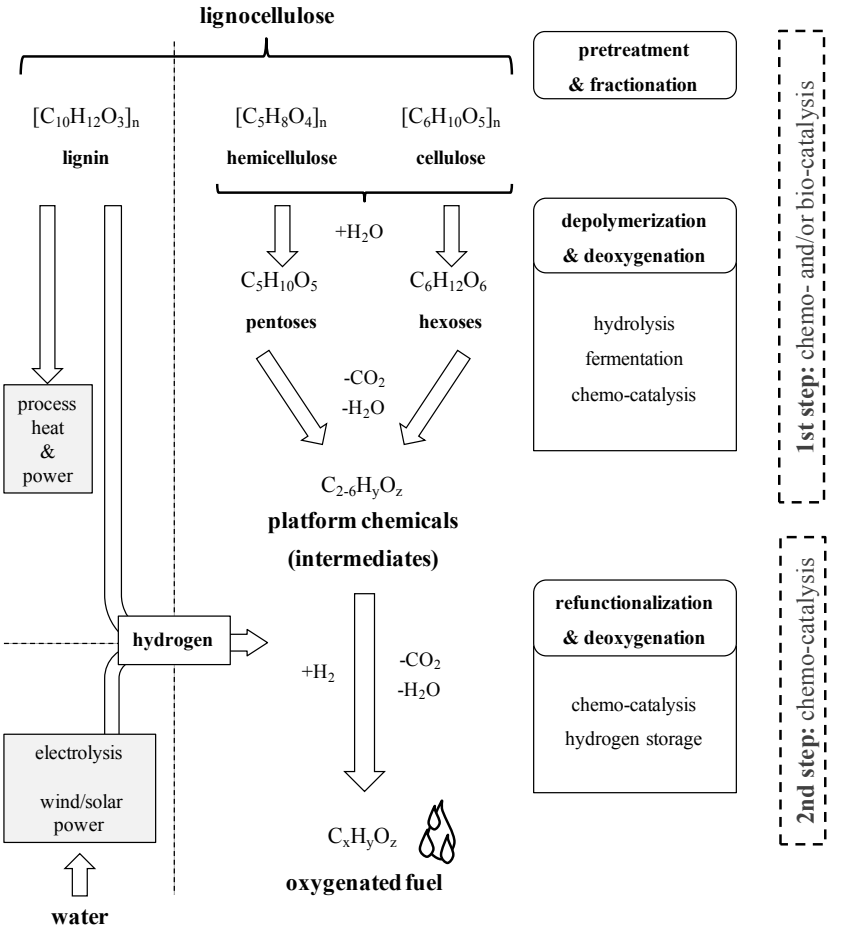


Figure 2.5: Production of cellulosic biofuel as a two-step process, i.e., chemo- or bio-catalytic conversion of (hemi-)cellulose into high-volume intermediates and subsequent chemo-catalytic upgrading to fuels.

Table 2.3: Potential platform molecules derived from chemo- (C) or bio- (B) catalytic processing of carbohydrates.

compound		**theor. yield** [g/g_{sugar}]	**real yield** [g/g_{sugar}] **(% of max.)**	**remark(s)**
			hexoses	
ethanol	B	0.51	0.50 (98%) (Straathof, 2014)	high productivity and titer (Jang et al., 2012; Klein-Marcuschamer and Blanch, 2013; Straathof, 2014)
lactic acid	B	1.00	1.00 (100%) (Straathof, 2014)	high productivity and titer (Jang et al., 2012; Straathof, 2014)
acetone	B	0.48	0.11 (23%) (Straathof, 2014)	co-products: ethanol and 1-butanol (Green, 2011; Jang et al., 2012, 2013)
1-butanol	B	0.41	0.36 (88%) (Straathof, 2014)	*E. coli* → high yield, moderate titer (Shen et al., 2011; Straathof, 2014) *Clostridium beijerinckii* → high productivity (Ezeji et al., 2007; Straathof, 2014)
itaconic acid	B	0.72	0.62 (86%) (Straathof, 2014)	low productivity (Jang et al., 2012; Straathof, 2014), aerobic (Straathof, 2014), O_2 as additional substrate
citric acid	B	1.07	0.88 (83%) (Straathof, 2014)	low productivity, aerobic (Straathof, 2014), O_2 as additional substrate
succinic acid	B	1.12	1.1 (98%) (Straathof, 2014)	CO_2 as additional substrate, high titer, moderate/high productivity (Straathof, 2014)
isobutanol	B	0.41	0.35 (85%) (Straathof, 2014)	low productivity (Jang et al., 2012; Straathof, 2014)
2,3-butanediol	B	0.50	0.48 (96%) (Straathof, 2014)	high titer (Jang et al., 2012; Straathof, 2014), high productivity (Jang et al., 2012; Straathof, 2014), aerobic, O_2 as additional substrate
acetic acid	B	1.00	0.8 (80%) (Straathof, 2014)	high titer (Jang et al., 2012; Straathof, 2014), low productivity (Jang et al., 2012; Straathof, 2014)
butyric acid	B	0.59	0.46 (78%) (Straathof, 2014)	low/moderate productivity (Jang et al., 2012; Straathof, 2014), high titer (Jang et al., 2012; Straathof, 2014)

Continued on next page

Table 2.3: Potential platform molecules derived from chemo- (C) or bio- (B) catalytic processing of carbohydrates (continued).

compound		**theor. yield** [g/g$_{sugar}$]	**real yield** [g/g$_{sugar}$] **(% of max.)**	**remark(s)**
hexoses (continued)				
acetic acid	B	1.00	0.8 (80%) (Straathof, 2014)	high titer, low productivity (Jang et al., 2012; Straathof, 2014)
butyric acid	B	0.59	0.46 (78%) (Straathof, 2014)	low/moderate productivity, high titer (Jang et al., 2012; Straathof, 2014)
1,3-propanediol	B	0.63	0.51 (81%) (Straathof, 2014)	high titer, high productivity(Jang et al., 2012; Straathof, 2014)
3-hydroxypropionic acid	B	1.00	0.53 (53%) (Lynch et al., 2013; Straathof, 2014)	high titer, low productivity (Lynch et al., 2013; Straathof, 2014)
fumaric acid	B	1.29	0.85 (66%) (Cao et al., 1996; Engel et al., 2008)	CO_2 as additional substrate, high titer, high productivity (Cao et al., 1996; Engel et al., 2008)
malic acid	B	1.49	0.94 (63%) (Zelle et al., 2008; Straathof, 2014)	CO_2 as additional substrate, high titer, low productivity (Jang et al., 2012; Straathof, 2014)
hydroxymethylfurfural	C	0.70	∼0.63 (∼90%) (Yang et al., 2011; van Putten et al., 2013)	biphasic solvent, heterogeneous catalyst (Rackemann and Doherty, 2011; van Putten et al., 2013) (fructose (Shimizu et al., 2009; Yang et al., 2011) / glucose (Fan et al., 2011))
levulinic acid	C	0.64	∼0.51 (∼79%) (BioMetics, Inc., 2002; Chang et al., 2006; Rackemann and Doherty, 2011; Climent et al., 2014)	homogeneous catalysts, heterogeneous catalysts less successful (Rackemann and Doherty, 2011)

Continued on next page

Table 2.3: Potential platform molecules derived from chemo- (C) or bio- (B) catalytic processing of carbohydrates (continued).

compound		**theor. yield** [g/g$_{sugar}$]	**real yield** [g/g$_{sugar}$] **(% of max.)**	**remark(s)**
hexoses (continued)				
γ-valerolactone	C	0.56	~0.40 (~72%) (Alonso et al., 2013)	levulinic acid as intermediate, non-precious metal catalysis (Wright and Palkovits, 2012; Alonso et al., 2013)
pentoses				
furfural	C	0.64	~0.51 (~80%) (Gürbüz et al., 2013)	γ-valerolactone as solvent, heterogeneous catalyst (Gürbüz et al., 2013)
levulinic acid[a]	C	0.77	~0.56 (~72%) (Rackemann and Doherty, 2011)	furfural and furfuryl alcohol as intermediates (Rackemann and Doherty, 2011; Climent et al., 2014)
γ-valerolactone[a]	C	0.67	~0.43 (~64%) (Alonso et al., 2013)	levulinic acid as intermediate, non-precious metal catalysis (Wright and Palkovits, 2012; Alonso et al., 2013)

[a] requires hydrogen as additional substrate

Dumesic, 2011; Gallezot, 2012). Production of furfural based on C_5 sugars has been industrialized (Mamman et al., 2008). In contrast, HMF synthesis is more complicated. It is performed best with fructose, requiring an extra isomerization step for the more abundant glucose (Chheda et al., 2007; Serrano-Ruiz and Dumesic, 2012). Unwanted side reactions constitute another problem encountered in chemical synthesis of furanic platform chemicals. Biphasic reactor concepts have shown promising results. Here, HMF is continuously extracted by an organic solvent in order to avoid consecutive reactions in the aqueous phase (Román-Leshkov et al., 2006; Alonso et al., 2010; Serrano-Ruiz and Dumesic, 2012). Both furfural and HMF can be upgraded to yield levulinic acid or γ-valerolactone, two chemicals which are sometimes considered platform chemicals on their own (Wright and Palkovits, 2012; Alonso et al., 2013).

Bio-catalytic conversion of carbohydrates into platform chemicals represents an interesting alternative due to the high selectivity, mild conditions and low exergy losses (Jäger and Büchs, 2012). Synthetic biology and metabolic engineering are employed in the optimization of bacteria and yeast with respect to high yield production of advanced biofuels (Wackett, 2011; McEwen and Atsumi, 2012; Rabinovitch-Deere et al., 2013). Unfortunately, microbial cells tend to be less productive, if mixtures of hydrolyzed monosaccharides are used instead of pure glucose (Straathof, 2014). Furthermore, since glucose is the preferred sugar, the pentoses often accumulate in the fermentation media until the preferred sugar is completely consumed (Kim et al., 2010). However, some (genetically modified) strains are capable of co-fermenting both hexoses and pentoses such that sequential, separate fermentation steps are not required (Jäger and Büchs, 2012).

A variety of platform chemicals derived from lignocellulosic biomass has been proposed in recent years (Werpy et al., 2004; Bozell and Petersen, 2010; Straathof, 2014; Sheldon, 2014). Table 2.3 includes oxygenated products from fermentation of glucose as well as derivatives of furfural and HMF. Data provided in this table have been sourced mainly from two review papers published by Jang et al. (2012) and Straathof (2014). Glycerol is a large-volume by-product of biodiesel production and its valorisation to commodity chemicals is therefore attracting interest (Serrano-Ruiz and Dumesic, 2012; Sheldon, 2014). However, the triglyceride feedstock belongs to the edible, non-abundant biomass. Hence, glycerol is not included in Table 2.3. For the sake of simplicity, all compounds in Table 2.3 are grouped by two main types of substrates, namely hexoses and pentoses. Note that fermentation products constitute the vast majority of intermediates, as organisms are well-skilled to produce a variety of acids, diols and polyfunctional compounds. These need to be upgraded in order to obtain fuel candidates, thus calling for the two-step process outlined in Figure 2.5.

2.3 Carbon- and energy-efficient biofuel production

The theoretical maximum yield for any intermediate or product $C_xH_yO_z$ of glucose can be calculated from the general reaction stoichiometry (Cherubini and Stromman, 2010; Straathof, 2014)

$$C_6H_{12}O_6 + \upsilon_{O_2} \cdot O_2 \rightarrow \upsilon_{product} \cdot C_xH_yO_z + \upsilon_{CO2} \cdot CO_2 + \upsilon_{H_2O} \cdot H_2O \tag{2.1}$$

involving stoichiometric coefficients υ_{O_2}, $\upsilon_{product}$, υ_{CO_2} and υ_{H_2O}.

This yield is pathway-independent and solely determined by the elemental composition of reactants and products. It can be considered a benchmark for any real technology, since it allows to estimate how close the performance of a practical process approaches the ideal case (Cherubini and Stromman, 2010). In economically feasible biofuel production, raw material cost is considered to dominate all other costs (Vickers et al., 2012; Klein-Marcuschamer and Blanch, 2013). An ideal process would achieve the theoretical maximum yield and would not exhibit any cost related to the transformation of biomass into liquid fuel such as energy input, catalyst, labor, waste treatment etc. Since the material costs are inescapable, the minimum production cost can be estimated based on this assumption (Klein-Marcuschamer and Blanch, 2013). Note that υ_{CO_2} for the computation of the theoretical maximum yield in Eqn. (2.1) can be negative, if the microorganisms can utilize CO_2 as additional substrate (Straathof, 2014).

Next to the theoretical and practical yields, the lower heating value efficiency is an important criterion for the selection of intermediates in biofuel production. One kilogram (kg) of glucose has an energy content of 14.1 MJ, based on the lower heating value (LHV) which is computed as the heat of combustion minus the enthalpy of vaporization of water formed during combustion at standard state (Lide, 2003). Given the stoichiometry of ethanol production (1 mol glucose is converted into 2 mol ethanol and 2 mol CO_2), 1 kg of glucose can be converted into 0.51 kg of ethanol, if it is assumed that the maximum theoretical yield is achieved (Straathof, 2014). The energy content of these 0.51 kg of ethanol is about 13.7 MJ (Lide, 2003). Thus, if the LHV energy efficiency is defined as

$$\text{LHV efficiency } [\%] = 100\ [\%] \cdot \frac{\text{mass of fuel produced [kg]} \cdot \text{LHV of fuel [MJ/kg]}}{\text{mass of glucose spent [kg]} \cdot \text{LHV of glucose [MJ/kg]}}\ , \tag{2.2}$$

it becomes clear that in ethanol production the energy provided by the glucose is almost entirely conserved in the resulting liquid fuel. Petrus and Noordermeer (2006) have argued that conversion of sugars to biofuels generally proceeds at highest LHV efficiencies, if

the necessary deoxygenation is carried out by removal of CO_2 and/or water. The latter compounds are combustion products and therefore, by definition, have a heating value of zero. The rationale behind the LHV efficiency is that biomass is cumbersomely grown, harvested and transported. A low sugar-to-fuel LHV efficiency can only be offset, if more biomass is produced or if external hydrogen is supplied. Moreover, Lange (2007) observed that, irrespectively of the technology chosen, LHV energy efficiency seems to inversely correlate with the capital cost of a biomass conversion plant.

The LHV efficiency not only depends on the mass yield, but also on the changes to the atomic composition. The significance of the choice of the "right" platform chemical for biofuel production can be illustrated with the help of the simple LHV model developed by Hechinger et al. (2010):

$$LHV\ [\mathrm{MJ/kg}] = \frac{73.147 + 795.727 \cdot nC - 187.697 \cdot SCBO}{M\ [\mathrm{g/mol}]} \tag{2.3}$$

Model equation (2.3) is applicable to $C_xH_yO_z$ compounds with a wide range of carbon atoms and has been derived and cross-validated using the DIPPR database of organic compounds (AIChE, 2012). M denotes the molar mass and nC is the number of carbon atoms. $SCBO$ is the abbreviation for the sum of conventional bond orders of all bonds not involving hydrogen, i.e., 1, 2, 1.5 for single, double and aromatic bonds respectively (Todeschini and Consonni, 2008). For instance, nC=6 and $SCBO$=12 result in a LHV of 14.4 MJ/kg for glucose (literature value (Lide, 2003) 14.1 MJ/kg, 2.1% prediction error). The prediction for ethanol is 28 MJ/kg (literature value (Lide, 2003) 26.8 MJ/kg, 4.5% prediction error). Hence, the predicted theoretical limit for LHV efficiency of ethanol ($C_2H_6O_1$) production becomes 99% based on the stoichiometry υ_{O_2}=0, $\upsilon_{product}$=2, υ_{CO_2}=2 and υ_{H_2O}=0 in Eqn. (2.1).

In contrast, 3-methyltetrahydrofuran has been proposed as a biofuel candidate derived from itaconic acid (Geilen et al., 2010). However, the predicted LHV energy efficiency for the step glucose to itaconic acid is only 76%. In the aerobic fermentation of itaconic acid, oxygen is not only removed as CO_2 and water, but also introduced to the molecular structure. Looking at the stoichiometry, this can be interpreted as a partial oxidation. In consequence, itaconic acid does not carry all the energy contained in the substrate glucose anymore. Subsequent steps to eliminate water and/or CO_2 with or without hydrogenation cannot return this energy, as can be seen from Figure 2.6. If hydrogen is used in the upgrading of carbohydrate streams, its LHV (120 MJ/kg) should be included in the calculation, since the hydrogen has to be provided either externally or via production from residual biomass (cf. Figure 2.5).

Predicted LHV efficiencies for other platform chemicals are shown in Table 2.4. The calculations are based on both maximum theoretical yield and practical yield. Note that

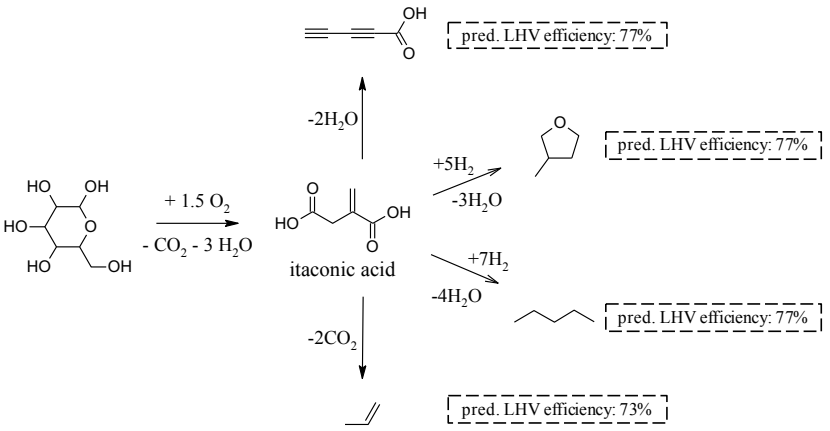


Figure 2.6: Itaconic acid as intermediate in biofuel production and predicted LHV efficiencies (glucose/hydrogen to end product).

values slightly larger than one are obtained for some compounds due to model inaccuracy. Similar to itaconic acid, citric acid and 2,3-butanediol are associated with inescapable losses due to the unfavorable changes to the atomic composition.

As can be seen from Tables 2.3 and 2.4, there are significant gaps between theoretical mass yields and reported practical mass yields for the majority of intermediates. Thus, real LHV efficiencies fall short of 100%. Likewise, practical chemical conversion into furfural and HMF does not proceed at the theoretical maximum yield. For economically feasible production of advanced biofuels, a multitude of problems has to be solved. In fermentation processes, yields, but also productivity and titer have to be improved (Klein-Marcuschamer and Blanch, 2013; Van Dien, 2013). It has been estimated that microbial fermentation processes with productivities below 2 g/(L·h) cannot be commercialized due to high capital cost for the fermenters (Van Dien, 2013). With respect to downstream separation cost, a minimum titer of 50 g/L is desired for the production of basic and intermediate chemicals (Van Dien, 2013). If engineering of enzymes and strains has yielded sufficiently fast production pathways, strategies like cell immobilization, retention or recycling and in-situ product recovery can be employed to optimize a bio-catalytic production system (Straathof, 2014). In chemo-catalytic conversion, heterogeneous non-precious metal catalysis is expected to significantly reduce processing costs (Carlini et al., 2004; Fan et al., 2011; Yang et al., 2011; Wright and Palkovits, 2012).

The bio-derived platform chemicals still comprise a high oxygen to carbon ratio. The elimination of CO_2 is one possibility to yield liquid fuel. However, higher mass-specific

Table 2.4: Potential platform molecules and predicted sugar to chemical LHV efficiencies.

compound	sum formula	pred. LHV [MJ /kg]	pred. LHV efficiency at theor. yield [×100%]	pred. LHV efficiency at real yield [×100%]
hexoses				
ethanol	$C_2H_6O_1$	28.0	0.99	0.97
lactic acid	$C_3H_6O_3$	14.8	1.03 (1.00)	1.00
acetone	$C_3H_6O_1$	29.4	0.99	0.23[a]
1-butanol	$C_4H_{10}O_1$	33.8	0.97	0.85
itaconic acid	$C_5H_6O_4$	15.3	0.77[b]	0.66
citric acid	$C_6H_8O_7$	10.6	0.78[b]	0.64
succinic acid	$C_4H_6O_4$	13.3	1.04 (1.00)	0.98
isobutanol	$C_4H_{10}O_1$	33.8	0.97	0.83
2,3-butanediol	$C_4H_{10}O_2$	25.7	0.89[b]	0.85
acetic acid	$C_2H_4O_2$	15.2	1.06 (1.00)	0.80
butyric acid	$C_4H_8O_2$	24.2	0.99	0.78
1,3-propanediol	$C_3H_8O_2$	22.5	0.99	0.80
3-hydroxypropionic acid	$C_3H_6O_3$	14.8	1.03 (1.00)	0.53
fumaric acid	$C_4H_4O_4$	11.9	1.06 (1.00)	0.66
malic acid	$C_4H_6O_5$	10.3	1.06 (1.00)	0.63
hydroxymethylfurfural	$C_6H_6O_3$	19.8	0.96	0.86
levulinic acid	$C_5H_8O_3$	20.3	0.91[c]	0.72[c]
γ-valerolactone	$C_5H_8O_2$	25.5	0.98	0.71
pentoses				
furfural	$C_5H_4O_2$	21.7	0.96	0.77
levulinic acid	$C_5H_8O_3$	20.3	0.98[d]	0.71[d]
γ-valerolactone	$C_5H_8O_2$	25.5	0.96[d]	0.62[d]

[a] LHV of co-products ethanol and 1-butanol not considered

[b] aerobic fermentation

[c] LHV of sideproduct formic acid not considered

[d] LHV of hydrogen (120 MJ/kg) as additional substrate considered

energy densities are obtained, if hydrogen is used for deoxygenation and for saturation (cf. Eqn. (2.3)). Depending on the process and its environment, different levels of hydrogenation can be attractive (cf. Figure 2.5). The supply of hydrogen by steam reforming of fossil energy carriers, although being the cheapest choice nowadays (Abbasi and Abbasi, 2011), obviously contradicts the vision of carbon-neutral biofuel production and utilization. Supply via gasification of residual biomass is linked to high cost (Holladay et al., 2009). Pyrolysis (Baumlin et al., 2006), aqueous phase reforming (Cortright et al., 2002) and biological hydrogen (Kapdan and Kargi, 2006) present alternative options (Holladay et al., 2009; Abbasi and Abbasi, 2011). Finally, hydrogen for biofuel synthesis can be produced from water electrolysis using (intermittent) renewable electricity (Shinnar and Citro, 2006; Agrawal et al., 2007; Muradov and Veziroğlu, 2008).

2.4 Computer-aided design of chemical products and fuels

Computational methods and tools supporting the rational design of single-molecule chemical products are commonly summarized under the term computer-aided molecular design (CAMD) (Achenie et al., 2003; Gani and Ng, 2015). CAMD aims at solving the inverse problem of property prediction, thus the purpose of CAMD is to identify molecular structures that meet certain pre-defined target properties. Typical examples of CAMD products include industrial solvents, refrigerants, ionic liquids and pharmaceutical agents (Ng et al., 2015b). In recent years, CAMD techniques have been applied to the design of pure-component biofuels as well (Hechinger et al., 2010; Victoria Villeda et al., 2012a; Hechinger et al., 2012b; Dahmen et al., 2012; Hechinger, 2014; Victoria Villeda et al., 2015; Hoppe et al., 2016b; Dahmen and Marquardt, 2016).

Various CAMD approaches have been developed over the years and it is common practice to classify CAMD techniques into different categories, e.g., generate & test and mathematical programming (Achenie et al., 2003; Ng et al., 2015b). The generate & test approach (Gani et al., 1991; Constantinou et al., 1996; Achenie et al., 2003) describes a two-step process. Fist, candidate structures are generated algorithmically, e.g., by joining pre-defined functional groups into all chemically feasible molecules. Typically, an enumeration-based strategy is pursued here. Then, in the second step, the generated structures are tested, i.e., their physicochemical properties are predicted and compared against the target values.

In contrast, the mathematical programming approach (Duvedi and Achenie, 1996; Churi and Achenie, 1996; Samudra and Sahinidis, 2013; Zhang et al., 2015a) aims at solving the inverse problem directly by combining structure generation and property evaluation in one optimization problem. Again, the basic principle here is that the molecular structure of the product can be assembled from a given set of structural groups. However, this time

numerical optimization is used to form feasible molecules on the basis of formalized bonding requirements. Along with property constraints based on the group additivity principle, the entire design problem can thus be written as a single mathematical optimization problem, i.e., a mixed-integer nonlinear program (Duvedi and Achenie, 1996; Achenie et al., 2003). The numerical solution to this problem corresponds to a feasible molecular structure that is predicted to exhibit the desired physicochemical properties.

There are additional CAMD categories, e.g., evolutionary strategies, and often algorithms from different categories are combined in a hybrid approach where a hierarchy of sub-problems is solved due to the fact that the original problem can be decomposed (Gani, 2004a,b). Excellent overviews of CAMD approaches and applications have been published by Achenie et al. (2003) and Ng et al. (2015b). The applicability of CAMD depends on the availability of mathematical models describing the target properties as functions of molecular structure. Besides structure generation, property modeling is therefore an essential part of every model-based molecular design. In principle, a variety of property models ranging from simple empirical models to physically motivated models, e.g., a combination of quantum-mechanics with equations of state, can be employed as part of CAMD. Molecular detail, properties and corresponding models can be organized in a hierarchical fashion to improve computational efficiency (Harper et al., 1999; Harper and Gani, 2000; Gani, 2004a). A comprehensive review of the different modeling approaches available in the context of computer-aided fuel design has been compiled by Hechinger (2014). In practice, however, group contribution methods (GCMs) and quantitative structure-property relationships (QSPRs) are most commonly used due to their relatively simple algebraic nature allowing for the rapid evaluation of large numbers of molecular structures. The fundamentals of GCM and QSPR are briefly introduced in the following two subsections, since all the property modeling efforts described in the present thesis are based on GCM and QSPR.

2.4.1 Group contribution (GC) modeling

Group contribution methods (GCMs) explicitly relate information about the molecular structure of a compound to the macroscopic property of interest by means of one (or few) algebraic equation(s). The fundamental assumption in group contribution modeling is that a property y_i of a molecule i can be explained by additive contributions from different molecular groups (Constantinou and Gani, 1994; Bader and Bayles, 2000).

Most often, the model structure

$$y_i = \sum_a g_{a,i} G_a + P, \tag{2.4}$$

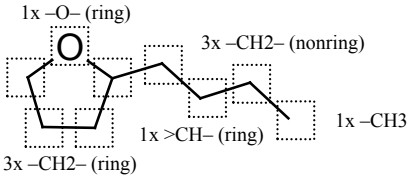


Figure 2.7: Group decomposition scheme of Joback's method (Joback and Reid, 1987) applied to butyl-tetrahydrofuran.

which is linear in the parameters G_a and P, is pragmatically selected as a starting point because of its simplicity. The parameter G_a represents the contribution of the structural group a with integer occurrence $g_{a,i}$. Parameter P represents a contribution that is common to all molecules. G_a and P need to be determined by means of regression analysis performed on $\tilde{\mathbf{y}}$, i.e., the vector of known (experimental) property data.

Instead of Eqn. (2.4), nonlinear functions can be applied. For instance, Joback and Reid (1987) have suggested to calculate the critical temperature T_{crit} of molecule i based on

$$T_{crit,i} = T_{boil,i}\left(P_{T_{crit},1} + P_{T_{crit},2}\left(\sum_a g_{a,i}G_{a,T_{crit}}\right) - \left(\sum_a g_{a,i}G_{a,T_{crit}}\right)^2\right)^{-1}. \tag{2.5}$$

Note that an estimate of the normal boiling point $T_{boil,i}$ is used alongside the group contributions $g_{a,i}$ in this model. Similar to Eqn. (2.4), $P_{T_{crit},1}$ and $P_{T_{crit},2}$ are constants. $G_{a,T_{crit}}$ are the group contributions.

Due to a modest number of groups, the application of Joback's GCM is straightforward, as can be seen from the example of butyl-tetrahydrofuran shown in Figure 2.7. In Joback's scheme, however, certain isomers may yield an identical group decomposition, and hence cannot be distinguished. With the aim of improving the accuracy of GCM, in particular with regard to the description of isomers and polyfunctional compounds, Gani and coworkers have pioneered multi-level GCMs which rely on a larger number of (more complex) structural groups that are organized in a hierarchical manner (Constantinou and Gani, 1994; Marrero and Gani, 2001; Hukkerikar et al., 2012). To this end, the linear model from Eqn. (2.4) has been slightly modified to distinguish first-, second- and third-order groups:

$$y_i = \underbrace{\sum_a g_{a,i}G_a}_{\text{first-order groups}} + \chi_{a2}\cdot\underbrace{\sum_{a2} g_{a2,i}G_{a2}}_{\text{second-order groups}} + \chi_{a3}\cdot\underbrace{\sum_{a3} g_{a3,i}G_{a3}}_{\text{third-order groups}} + P. \tag{2.6}$$

Based on a comprehensive database of thermophysical properties, Marrero and Gani (2001) have performed a three-step regression procedure on Eqn. (2.6): In the first level

of estimation, the constants χ_{a2} and χ_{a3} were set to zero. Similar to Joback's GCM, the 182 first-order groups have been chosen to describe the entire molecule, i.e., each fragment of a given molecule shall be covered by exactly one first-order group. In contrast, the 122 second-order groups have been allowed to overlap each other. Furthermore, simple compounds do not require any second-order groups at all. The second level of regression has been carried out by setting χ_{a2} to unity and χ_{a3} to zero and by fixating the contributions G_a obtained in the first run. This way, the second-order groups are thought to act as corrections for the first-level property estimation. Finally, the same principle has been applied to 66 third-order groups, which had been chosen to describe large molecular fragments in polycyclic compounds.

A significant fraction of the parameters in the Marrero and Gani (2001) model, including certain parameters for first-order group contributions, could not be determined by regression due to missing property data. To overcome this limitation, Gani et al. (2005) have proposed the so-called group contribution plus (GC^+) methodology, i.e., a combination of group contribution and atom connectivity index modeling. Here, the connectivity index model is used to predict the missing group contributions from the Marrero and Gani model (Villalba, 2009; Hukkerikar et al., 2012; Hukkerikar, 2013). Gonzalez et al. (2007) have transferred this concept to the prediction of missing group interaction parameters in the UNIFAC method (Fredenslund et al., 1975), i.e., a GCM frequently used for the prediction of liquid-phase activity coefficients. However, the general applicability of the approach for prediction of missing group interaction parameters in UNIFAC is heavily debated (Mohs et al., 2009; Gani and González, 2009; Villalba, 2009; Mustaffa et al., 2011).

2.4.2 Quantitative structure-property relationship (QSPR) modeling

The aim of quantitative structure-property relationship (QSPR) or quantitative structure-*activity* relationship (QSAR) modeling is to identify a mathematical relationship between the property (or the activity) of interest and a set of so-called molecular descriptors, i.e., scalar quantities that can be derived from the two- or three-dimensional structure of a given molecule (Katritzky et al., 1995; Todeschini and Consonni, 2008). Commercially available software packages like CODESSA PRO (Katritzky et al., 2005) and DragonX (Todeschini et al., 2009) support the calculation of hundreds or even thousands of such descriptors belonging to various classes, e.g., the classes of constitutional, topological or quantum-chemical descriptors (Katritzky et al., 1995; Todeschini and Consonni, 2008). Examples for constitutional descriptors include the molecular weight of a compound, the number of atoms of a certain type and the number of bonds of a certain type. The branching of a molecule is described by topological descriptors, i.e., descriptors related to molecular connectivity. The determination of quantum-chemical descriptors is more computationally

demanding, because the full spatial molecular arrangement needs to be known. Although the required geometry optimization can be performed by means of simplified molecular mechanics or semi-empirical methods, precise three-dimensional coordinates are obtained only from *ab initio* quantum-chemical methods (Hechinger et al., 2012c). Note that functional group counts can constitute molecular descriptors as well.

Most QSPR models rely on the assumption that the property of interest y depends linearly on few, selected molecular descriptors (Katritzky et al., 1995, 2010). Hence, the general model equation for a compound i reads

$$y_i = P + d_{1,i}D_1 + d_{2,i}D_2 + ... + d_{N_d,i}D_{N_d} \ , \tag{2.7}$$

where $P, D_1, D_2, ...D_{N_d}$ are the model parameters and $d_{1,i}, d_{2,i}, ..., d_{N_d,i}$ are the numerical values obtained for the N_d descriptors used in the particular model. Based on experimental property data $\tilde{\mathbf{y}}$, the N_d 'significant' descriptors are chosen by some form of (heuristic) stepwise multi-parameter regression analysis (Katritzky et al., 1995, 2010; Yousefinejad and Hemmateenejad, 2015). Many algorithms have been proposed for this variable selection task, however, there is no general consensus on a 'best' way of choosing a multivariate model (Olden and Jackson, 2000; Whittingham et al., 2006; Sauerbrei et al., 2007; Murtaugh, 2009). Instead of multiple linear regression (MLR), projection methods like principal component analysis (PCA) (Wold et al., 1987; Jolliffe, 2002) or partial least squares (PLS) (Höskuldsson, 1988; Wold et al., 2001) are often preferred, if the descriptor data suffer from a high degree of collinearity, or if the ratio of the number of training molecules to the number of variables, i.e., the descriptors, is small (Katritzky et al., 1995, 2010). Methods based on artificial intelligence (AI), e.g., genetic algorithms (Rogers and Hopfinger, 1994), artificial neural networks (Wikel and Dow, 1993) or simulated annealing (Sutter et al., 1995), have been applied to the variable selection problem as well. Some of the AI methods, e.g., artificial neural networks (Oinuma et al., 1990), can also be utilized for model construction and are, in principle, able to reveal nonlinear relationships between the molecular descriptors and the property of interest. Recently, Yousefinejad and Hemmateenejad (2015) have provided an excellent review of the different strategies for variable selection and model construction in the context of QSPR/QSAR modeling.

Overfitting and chance correlation are two major pitfalls in QSPR/QSAR modeling (Dearden et al., 2009; Cherkasov et al., 2014). Overfitted models include more terms, i.e., descriptors, than are necessary to describe the property y. The irrelevant descriptors can have a negative impact on the predictive capability of the model, because they add noise to the predictions (Hawkins, 2004). If variable selection strategies are employed, the descriptors are typically chosen such that the resulting model equation is highly significant by some standard statistical criteria (Katritzky et al., 2010). Since standard criteria

do not take into account the number of descriptors actually screened for inclusion, they can be highly misleading (Topliss and Costello, 1972; Topliss and Edwards, 1979). As a consequence, there is a substantial risk that the identified correlation has occurred purely by chance. The possibility of obtaining chance correlations is particularly high, if the number of compounds is small compared to the number of potential descriptors (Topliss and Costello, 1972). For MLR studies, five compounds per available descriptor constitutes a frequently recommended minimum ratio (Dearden et al., 2009; Cherkasov et al., 2014; Yousefinejad and Hemmateenejad, 2015).

The problems of overfitting and chance correlation can be mitigated to some extent by choosing appropriate algorithms for model construction, e.g., by considering PLS if the descriptors are numerous and collinear (Clark and Cramer, 1993; Eriksson et al., 2003; Katritzky et al., 2010). Validation strategies are key in revealing both overfitting and chance correlation phenomena. Careful model validation is absolutely essential as it is a much more reliable way of estimating the predictive power of a QSPR/QSAR model than is the examination of the goodness-of-fit with respect to the training data (Tropsha et al., 2003; Eriksson et al., 2003; Hawkins, 2004; Gramatica, 2007; Roy, 2007). In recent years, it has become common to distinguish between internal and external validation (Gramatica, 2007). In internal validation, the entire data set is used both to fit the model and to estimate the prediction error. Cross-validation (CV) is most frequently applied for this task (Hawkins et al., 2003). In contrast, external validation refers to the use of an entirely independent data set to test the predictive power of the derived model (Eriksson et al., 2003; Tropsha et al., 2003). To facilitate an external validation, the modeling data set is split into a training set and a test set before modeling is initiated. Alternatively, external validation data can be acquired after modeling has been completed, e.g., by performing a new round of experiments. The successive application of internal and external validation is now widely recognized as the gold standard in QSPR/QSAR modeling (Tropsha et al., 2003; Roy, 2007; Dearden et al., 2009; Tropsha, 2010; Gramatica, 2014).

Most training data sets are created by retrieving property data from databases of known physicochemical properties. Since these data sets do not result from optimal experimental design, they are more or less structurally limited, i.e., the compounds cover only certain parts of the chemical space (Jaworska et al., 2005). The model applicability domain (AD) concept relies on the assumption that, in general, interpolation is more reliable than extrapolation. In the context of QSPR/QSAR modeling this implies that predictions are considered to be most accurate, if the target compounds are 'similar' to the training compounds (Netzeva et al., 2005). In order to quantify similarity, the descriptor space is examined. The strengths and limitations of different similarity measures for use in QSPR/QSAR studies have been compared multiple times (Netzeva et al., 2005; Jaworska et al., 2005;

Weaver and Gleeson, 2008), however, there is no AD definition that is considered 'best' in all cases. It is therefore up to the modeler to select an AD concept that is suitable for the training data and the prediction task at hand (Sahigara et al., 2012).

It shall be noted that several authors have formulated best practices for QSPR/QSAR modeling (Golbraikh and Tropsha, 2002a; Tropsha et al., 2003; Eriksson et al., 2003; Walker et al., 2003; Gramatica, 2007; Tropsha, 2010; Cherkasov et al., 2014) and that lists of typical errors and pitfalls in QSPR/QSAR studies have been compiled by Cronin and Schultz (2003), Dearden et al. (2009) and Scior et al. (2009).

3 Targeted generation of candidate structures

In earlier work on biofuel CAMD (Hechinger et al., 2012a,b; Dahmen et al., 2012; Victoria Villeda et al., 2012b,c; Hechinger, 2014; Weinebeck et al., 2014; Hoppe et al., 2016b), the molecular structure generator Molgen (Gugisch et al., 2015) had been used to generate candidate structures in a systematic fashion. Molgen enables the deterministic generation of all mathematically feasible $C_xH_yO_z$ structures solely based on the valence rules. A bad-list of unwanted (sub-)structures can be supplied to remove certain structures from the output, e.g., peroxides ($-O-O-$) or structures with patterns considered chemically infeasible (Hechinger, 2014; Gugisch et al., 2015). The strength of the Molgen generator, i.e., its ability to explore the "full" molecular search space, however, comes at the price of high combinatorial explosion even for molecules of relatively small sizes. Little structural variations, e.g., an additional methyl group or a slightly different position of a carbon-carbon double bond, result in a large number of structurally similar compounds exhibiting similar properties. In the context of fuel design, the number of molecules predicted to satisfy all imposed property constraints easily becomes huge and a further ranking of fuel candidates solely on the basis of property values gets difficult (Dahmen et al., 2012). Most importantly, there is no "connection" between the promising molecules and any potential substrates. Consequently, the experienced chemist will have to work through a huge list of molecules to find potential targets for synthesis.

Similarly, in the case of optimization-based CAMD, it is difficult to formulate a comprehensive set of mathematical constraints that link CAMD products to multiple pre-defined substrates such that this link can be interpreted as a potential synthetic pathway from substrates to products. As pointed out by Ng et al. (2015b), molecular design that considers the reactions involved to actually synthesize a product remains a relatively unexplored research area. Addressing this challenge, Chemmangattuvalappil and Eden (2013) have recently proposed a new algorithm that solves molecular design problems in reactive systems by tracking changes in molecular signature descriptors, which can be related to different types of chemical reactions. This way, modifications in the chemical structure can be correlated to changes in the property, because the molecular signatures also serve as a

basis for property estimation by means of inverse quantitative structure-property/activity relationship modeling (Visco et al., 2002).

With the aim of focussing CAMD efforts on products resulting from carbon- and energy-efficient upgrading of bio-derived platforms, a rule-based generator of molecular structures is introduced in this Chapter. In contrast to Molgen, the targeted algorithm starts on a user-defined list of bio-derived intermediates, i.e., the platform molecules, and iteratively expands the product spectrum by applying a few, simple transformation rules reflecting changes that could occur, in principle, due to chemo-catalytic refunctionalization.

The structure generator has been inspired by rule-based generation of reaction mechanisms in chemical (Corey and Wipke, 1969; Broadbelt et al., 1994; Buxton et al., 1997; Song, 2004; Rangarajan et al., 2010; Victoria Villeda, 2017) and metabolic engineering (Hatzimanikatis et al., 2005; Yim et al., 2011) and had originally been developed as a complement to ReNeGen (Victoria Villeda, 2017), i.e., a reaction network generator for use in biofuel pathway exploration and analysis developed at the Institute for Process Systems Engineering at Aachener Verfahrenstechnik. The refunctionalization rules for the structure generator (cf. Appendix A), which were chosen to preserve the full carbon skeleton of the platform chemicals in order to focus entirely on functional group transformations (Buxton et al., 1997), have been compiled in collaboration by the author of this thesis and Juan José Victoria Villeda, the author of ReNeGen, during their time at the Institute for Process Systems Engineering. Chemical feasibilities of specific refunctionalizations have also been discussed with Stefanie Mersmann (Institute of Organic Chemistry, RWTH Aachen University) and Jürgen Klankermayer (Institut für Technische und Makromolekulare Chemie, RWTH Aachen University). The structure generator has been implemented in Matlab (The MathWorks Inc., 2016). Juan José Victoria Villeda has proposed to preserve the carbon skeleton and has contributed implementations of Morgan's algorithm (Morgan, 1965; Figueras, 1993) as well as routines for structure import/export from/to SMILES code (Weininger, 1988; Weininger et al., 1989) based on an interface to the OpenBabel software package (O'Boyle et al., 2011).

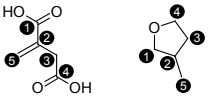
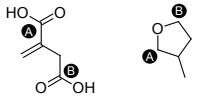
Early versions of the structure generator have been presented at conferences (Dahmen et al., 2013a,b; Klankermayer et al., 2013; Marquardt et al., 2013) and large parts of this Chapter have already been published in Energy & Fuels in a slightly modified form (Dahmen and Marquardt, 2016). CAMD approaches relying on similar ways of targeted structure generation, starting from bio-derived platforms and following pre-specified transformations, have been reported by Bergez-Lacoste et al. (2014), Moity et al. (2014, 2016) and Gerbaud et al. (2017) with respect to the computational identification of bio-solvents.

3.1 Main principles in structure generation

In order to resemble carbon- and energy-efficient processing, functional transformations should adhere to the three main principles depicted in Table 3.1, thus following the line of thought presented in Section 2.3, i.e., the energy density of a fuel is tied to the number of carbon atoms and the amount of bonded hydrogen.

The first principle ensures that each feasible combination of precursor molecule(s) and product molecule satisfies the reaction stoichiometry, where hydrogen and water are the only additional reactants. Moreover, waste is restricted to CO_2 and water, because their enthalpy of combustion is zero. The first principle therefore resembles both the concept of atom-efficiency, i.e., one of the fundamental cornerstones of green chemistry (Trost, 1991; Constable et al., 2002; Li and Trost, 2008; Anastas and Eghbali, 2010), and the concept of energy-efficiency (Petrus and Noordermeer, 2006).

Table 3.1: Three main principles for the generation of molecular graphs from pre-defined intermediates as part of model-based fuel design.

	main principle	**rationale**	**example: itaconic acid → 3-MTHF (Geilen et al., 2010)**
1.	stoichiometry should only cover substrate, product, H_2, CO_2 and H_2O	general feasibility, LHV & atom efficiency	$C_5H_6O_4 + 5\ H_2 \rightarrow C_5H_{10}O_1 + 3\ H_2O$
2.	preservation of carbon-carbon bonds (no rearrangement)	selectivity ⇒ LHV & atom efficiency	
3.	refunctionalization instead of arbitrary oxygen functionality	selectivity ⇒ LHV & atom efficiency	

The second principle rules out isomerization. Similar to Buxton et al. (1997), the carbon skeleton of the intermediates shall be preserved in the course of molecule generation, as the focus lies on functional group transformations. It is assumed here that, in the majority of cases, the preservation of the existing carbon-carbon bonds will improve the chances that a refunctionalization can be carried out with high selectivity, i.e., a fundamental prerequisite for atom- and carbon-efficient fuel production.

The third principle refers to the location of oxygen in the fuel molecule. It ensures that oxygen functionality, e.g., an alcohol group, cannot be attached at an arbitrary position of the carbon skeleton, but must result from a feasible functional transformation applied to a precursor molecule, e.g., a carboxylic acid. In this manner, oxygen atoms can only

occur at distinct positions, which are ultimately given by the starting materials, i.e., the bio-derived platform chemicals.

3.2 Relations to rule-based reaction network generation

Compared to rule-based generators of reaction networks like ReNeGen (Victoria Villeda, 2017), the algorithm of the structure generator presented here is quite simple. This is due to the fact that there is no network structure to be processed or stored. As indicated in Figure 3.1, only the involved platforms are memorized along with each product to be generated by the algorithm. The needed information can be retrieved easily from the immediate predecessor molecule(s). This simple strategy allows for the computationally efficient generation of large numbers of molecular structures over many iterations by avoiding the complexities associated with the consideration of reaction pathways.

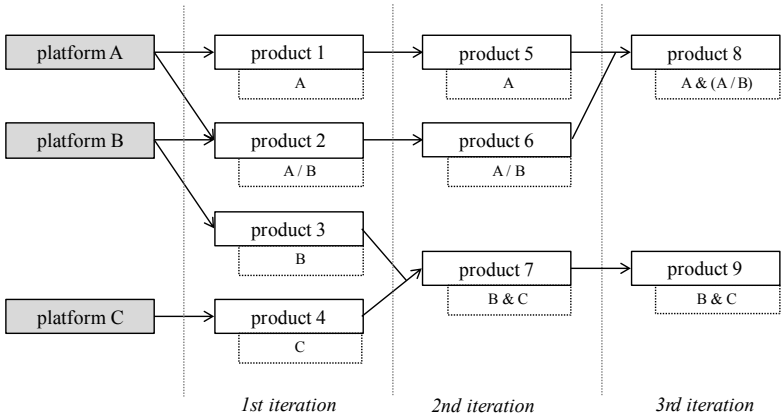


Figure 3.1: Basic function of the molecular structure generator: Instead of processing a network structure (corresponding to the arrows in the graph), only the information on involved platforms (i.e., the information in the dashed rectangles) is stored for each product to be generated by the algorithm.

As a part of a systematic methodology for computer-aided product design in biorefineries, Chemmangattuvalappil and Ng (2013) have proposed the concept of a chemical reaction pathway map (CRPM) by (i) categorizing organic chemicals based on ten functional groups, and (ii) by connecting the categories via 27 potential reaction pathways. In terms of the idea of exploring feasible transformations in bioprocessing by focussing attention on specific functional groups, the CRPM approach has some similarities to the

structure generator described here. In fact, structure generation in CAMD and reaction mechanism generation can be regarded as complementary tools in model-based fuel design. A mechanism generator, e.g., the software-package RING developed by Daoutidis and coworkers (Rangarajan et al., 2010, 2012; Daoutidis et al., 2013; Rangarajan et al., 2014a,c), or ReNeGen (Victoria Villeda, 2017) can be employed to explore the synthetic feasibility of substrate/product combinations more rigorously, i.e., by considering specific reaction pathways and conditions. Many schemes for heterogeneous catalytic reactions relevant to bio-renewable processing have been formalized for RING, making it particularly useful in this context. It allows for on-the-fly estimation of thermochemistries, activation barriers and even kinetic rates (Marvin et al., 2013; Rangarajan et al., 2014b). Recently, variants of reaction network flux analysis (RNFA) (Voll and Marquardt, 2012b) have been combined with both RING and ReNeGen to automatically generate and evaluate synthesis routes for the production of biofuels (Marvin et al., 2013; Victoria Villeda, 2017).

3.3 Pool-based scheme and implementation

The implementation of the structure generator is based on a simple representation of molecular graphs allowing for an efficient manipulation, thus strongly reducing algorithmic complexity and computational effort. The three principles from Table 3.1, in particular the fact that no rearrangement of existing carbon-carbon bonds shall occur, allow to encode molecular graphs by the intuitive scheme shown in Table 3.2. Here, the hydrocarbon skeleton is decomposed into straight carbon chains which superimpose at branching positions. Note that most bio-derived platform chemicals are straight-chain molecules, because neither glucose nor xylose are branched. The carbon atoms in each chain are numbered consecutively and functionality is attached at distinct positions. If one (multiple) carbon chain(s) is (are) connected via an oxygen atom with itself (each other), this carbon-oxygen-carbon motif is referred to as *C-O-C coupling*. *C/C identities* are used to describe the intersection of straight carbon chains. Exemplary encodings of molecular graphs can be found in Table 3.3. Delocalization of electrons in aromatic rings, e.g., in a furan, is not specified separately, but is recognized internally. An ester is composed of an =O group next to a C-O-C coupling.

The general flowsheet of the molecular structure generator is depicted in Figure 3.2. It reflects the idea of a pool of molecules that undergo refunctionalization. Starting with a user-defined set of intermediates, transformation operators address either *refunctionalization*, i.e., the change of one or multiple functional group(s) including ring closure or cleavage, or *aggregation*, i.e., two molecules are joined into a larger one, without rearranging existing carbon-carbon bonds. The sequence of transformations is aligned with the

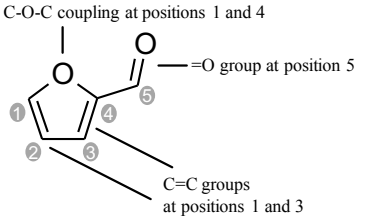
Table 3.2: Encoding scheme for a molecular graph.

Any molecular structure is encoded by three elements:		
	1. N_I straight carbon chains	
	2. N_J C−O−C couplings used to connect carbon chains	
	3. N_K C/C identities used to connect carbon chains	
chains		
1	nC	number of carbon atoms
	$\mathbf{p}_{-OH}$	position vector of alcohol groups
	$\mathbf{p}_{=O}$	position vector of ketone/aldehyde groups
	$\mathbf{p}_{COOH}$	position vector of acid groups
	$\mathbf{p}_{C=C}$	position vector of carbon-carbon double-bonds
2	...	
...	...	
N_I	...	
C-O-C couplings		
1	$(\chi_1,\chi_1^a)-(\chi_2,\chi_2^b)$	carbons χ_1^a and χ_2^b of chains χ_1 and χ_2 are linked via oxygen
...	...	
N_J	...	
C/C identities		
1	$(\chi_1,\chi_1^a)-(\chi_2,\chi_2^b)$	carbon χ_1^a of chain χ_1 is identical to carbon χ_2^b of chain χ_2
...	...	
N_K	...	

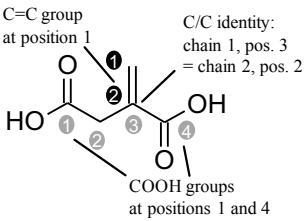
aim of reducing the generation of duplicate molecules while ensuring that each possible combination of transformations and intermediates is covered. Removal of duplicates is still needed at distinct steps in the process and is performed with the help of Morgan's algorithm (Morgan, 1965; Figueras, 1993). The hydrogen demand as well as the number of water and carbon dioxide molecules produced are tracked throughout the generation process for each substrate/product combination individually.

An exemplary refunctionalization rule is given in Table 3.4. This rule can be applied to any molecular graph comprising two or more alcohol groups. The purpose is to form a ring involving oxygen, i.e., a cyclic ether. A set of auxiliary routines has been implemented for tasks like counting the number of attached hydrogens, returning ring sizes, performing

Table 3.3: Furfural and itaconic acid encoded by the scheme depicted in Table 3.2.



furfural
(one straight carbon chain ❶ to ❺)



itaconic acid
(two straight carbon chains ❶ to ❹ and ❶ to ❷)

chains			**chains**		
1	nC	5	1	nC	4
	$\mathbf{p}_{-OH}$	[ ]		$\mathbf{p}_{-OH}$	[ ]
	$\mathbf{p}_{=O}$	[5]		$\mathbf{p}_{=O}$	[ ]
	$\mathbf{p}_{COOH}$	[ ]		$\mathbf{p}_{COOH}$	[1, 4]
	$\mathbf{p}_{C=C}$	[1, 3]		$\mathbf{p}_{C=C}$	[ ]
			2	nC	2
				$\mathbf{p}_{-OH}$	[ ]
				$\mathbf{p}_{=O}$	[ ]
				$\mathbf{p}_{COOH}$	[ ]
				$\mathbf{p}_{C=C}$	[1]
C-O-C couplings			**C-O-C couplings**		
1	(1,1)−(1,4)		∅		
C/C identities			**C/C identities**		
∅			1	(1,3)−(2,2)	

subpath detection etc. Note that the rule is formulated to directly act on the encoding scheme shown in Table 3.2.

Corma et al. (2007), Alonso et al. (2010), Serrano-Ruiz et al. (2011), Serrano-Ruiz and Dumesic (2011), Lange et al. (2010, 2012) and Climent et al. (2014) have reviewed catalytic strategies to produce biofuels from carbohydrates. On the basis of these collections, rules for structural modification that comply with the three main principles from Table 3.1 have been formalized. These rules can be found in Appendix A. In an attempt to broaden the product spectrum beyond C_6 derivatives, adjustments to the molecular weight can be made by aldol condensation (Julis and Leitner, 2012; Luska et al., 2014) (C−C coupling) or by etherification/esterification (C−O−C coupling). Structurally, these reactions are aggregations of two carbon skeletons without isomerization or carbon loss.

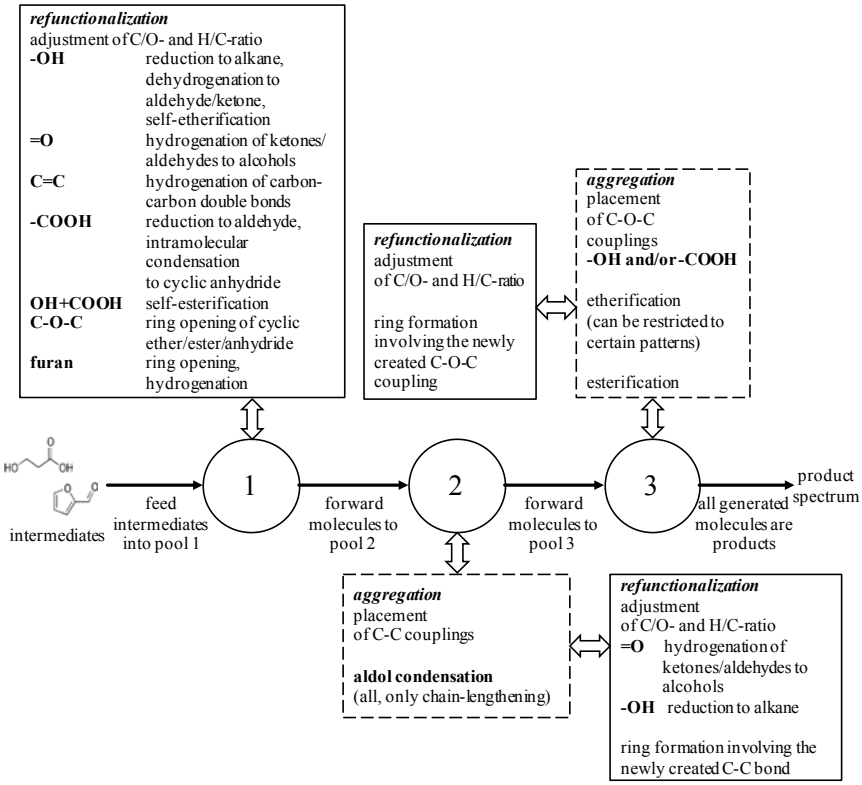


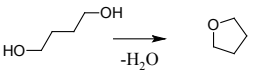
Figure 3.2: Pool-based approach followed in molecular structure generation.

Refunctionalization and aggregation rules, however, give rise to a combinatorial growth. An automatic pruning of the product spectrum is achieved through *a priori* estimation of boiling and melting points by means of Joback's group contribution method (Joback and Reid, 1987). For instance, if two alcohols A and B can undergo etherification, the *a priori* estimate of the normal boiling point T_{boil} (in °C) of the product C is

$$T_{boil,C} = T_{boil,A} + T_{boil,B} - 2 \cdot 92.88 + 22.42 \ . \tag{3.1}$$

Since both normal boiling point and the melting point are linear combinations of group weights, the aggregated molecule will be composed of all the groups contained in both A and B, except for the two alcohol groups (−OH group weight of 92.88), which are

Table 3.4: Exemplary illustration of a refunctionalization rule. The complete list of rules can be found in Appendix A.

 example: 1,4-butanediol is refunctionalized yielding tetrahydrofuran	**cyclodehydration of diols to cyclic ether (self-etherification)** 1. C^1 and C^2 are carbons in the same molecule, that a. both carry an $-$OH group b. are not identical ($C^1 \neq C^2$) c. have at least k carbons/oxygens inbetween (no rings smaller than k+3) d. have a maximum of l carbons/oxygens inbetween (no rings larger then l+3) e. have only one existing path between each other (not part of an existing ring) OR have two existing paths between each other, but one path is a subset of the other path (allows for the formation of two rings attached to each other) 2. remove the $-$OH groups from C^1 and C^2 3. make a C-O-C coupling between C^1 and C^2, i.e., C^1-O-C^2 **[-H_2O]**

replaced by a single non-ring ether group ($-$O$-$ group weight of 22.42) (Joback and Reid, 1987). Similarly, the effect of any refunctionalization on boiling and melting point can be assessed. Additional pruning is achieved by considering structural constraints, e.g., a maximum number of carbon atoms or a maximum ring size. A complete scenario definition for the molecular structure generator contains a list of bio-derived intermediates together with a list of active transformation rules and constraints.

Figure 3.3 shows exemplary products generated by the rule-based approach starting from only two intermediates, namely furfural and acetone. According to the workflow depicted in Figure 3.2, the products are generated in three consecutive steps: (i) refunctionalization, (ii) aggregation by carbon-carbon coupling and refunctionalization, and (iii) aggregation by carbon-oxygen-carbon coupling and refunctionalization. Adhering to the main principles from Table 3.1, the algorithm generates acyclic and cyclic ethers, ketones, aldehydes, alcohols, furans and polyfunctional compounds.

It shall be stressed that, in general, it cannot be expected that the number of transformation steps proposed by the structure generator is similar to the number of reaction steps. This shall be exemplified by the reaction pathway for 2-methyltetrahydrofuran (2-MTHF) production starting from levulinic acid as proposed by Geilen et al. (2010). This pathway

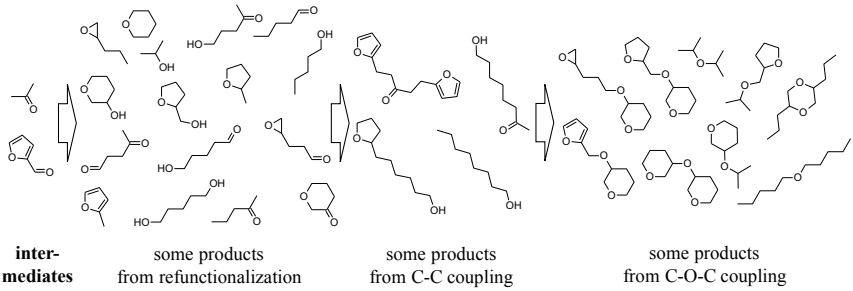


Figure 3.3: Exemplary products automatically generated based on two bio-derived intermediates, namely furfural and acetone.

describes a sequence of seven elementary reactions involving six intermediates, one of which is a protonated species. While the five nonprotonated intermediates are also considered potential products of levulinic acid by the structure generator, only three steps are required in the scheme presented here to transform levulinic acid into 2-MTHF. Furthermore, the catalytic system designed by Geilen et al. (2010) facilitates a one-pot synthesis (Hayashi, 2016), i.e., with a proper combination of ligand, additive, solvent and reaction temperature the entire sequence of reactions can be performed in one reaction device. Because such catalytic systems eliminate the need for purification and recycling of intermediates, they can drastically simplify the process flowsheet, ultimately leading to lower operating cost, waste production, as well as a lower consumption of auxiliary chemicals and energy (Bruggink et al., 2003). Since sophisticated reaction engineering is required to develop these catalytic systems, experienced chemists will still have to select the most attractive targets from the list of CAMD fuel candidates manually. However, compared to the valence-rule based structure generation, the targeted approach provides a much smaller list of alternatives to choose from. Furthermore, the transformation rules can be re-configured or extended easily. Hence, the human expert can utilize the computational tool to systematically explore and refine scenarios for the production of tailored biofuel components. Subsequent use of complementary tools like RING or ReNeGen to generate, explore and optimize the reaction pathways associated with certain substrate-product-combinations is conceivable, however, beyond the scope of this thesis.

4 GC-based prediction of fuel auto-ignition quality

The early assessment of fuel auto-ignition quality is of paramount importance for any fuel design methodology (cf. discussions in Chapter 1 and in Subsection 2.1.1). Although multiple models for the prediction of pure-component octane number (ON) and cetane number (CN) from molecular structure have been proposed over the course of the last 30 years (DeFries et al., 1987; Meusinger and Moros, 1999; Yang et al., 2001; Albahri, 2003; Santana et al., 2006; Smolenskii et al., 2008; Lapidus et al., 2008; Katritzky et al., 2010; Creton et al., 2010; Abdul Jameel et al., 2016), their applicability ranges are largely limited to non-oxygenated hydrocarbon species, i.e., the constituents of fossil fuels. Specifically, the validity of the earlier CN models is restricted to n-alkanes, iso-alkanes and singly substituted alkylbenzenes (DeFries et al., 1987), to alkanes and cycloalkanes (Lapidus et al., 2008; Smolenskii et al., 2008), or even to iso-alkanes only (Yang et al., 2001). The model of Smolenskii et al. (2008) also returns some strange CN values, e.g., -146.12 for 3-methyl-3-ethylhexane. More recently, Creton et al. (2010) have proposed separate QSPR models for four classes of molecules, i.e., alkanes, cycloalkanes, alkenes and aromatics, and Abdul Jameel et al. (2016) have trained a multiple linear regression model on CN and derived cetane number (DCN) data of 71 hydrocarbons and 54 hydrocarbon blends. The few attempts to extend the range of validity of ON and CN models to oxygenates have generally suffered from a lack of ignition data of adequate quality for the variety of molecular structures. Taylor et al. (2004), Saldana et al. (2011), Dahmen et al. (2012) and Sennott et al. (2013a,b) have built QSPR models by gathering data mainly from the experimental CN compendium released by Murphy et al. (2004), which contains some data on oxygenates. However, only a small set of CN from this compendium had been measured in an ASTM D613 (2015) cooperative fuels research (CFR) engine, whereas a much larger set of correlated CN had been derived from *cetene* numbers, from ignition delay data acquired in combustion bomb experiments and from mixture data (Murphy et al., 2004).

The limitations of existing models for fuel auto-ignition quality, i.e., ON, CN or DCN, thus can largely be attributed to the lack of a high-quality training database holding ignition delay data from a single well-defined experimental set-up for the variety of (oxy-

genated) hydrocarbon species relevant to biofuel design. With the aim of assembling such a database, DCN data from the literature have been augmented with data from a rapid fuel screening campaign performed with an Ignition Quality Tester (IQT). The specific advantages of IQT-derived ignition delay and DCN in the context of fuel design have already been discussed in Chapter 1 and Subsection 2.1.1. The results from the screening campaign are briefly summarized at the beginning of this Chapter. The established database is then used to develop a model for IQT ignition delay and DCN prediction based on the principle of group additivity. To ease the use of the model, all inputs can be calculated (manually) solely on the basis of the two-dimensional molecular structure. Likewise computation of the DCN is facilitated by means of a few simple equations. As such, the model can be easily incorporated into almost any manual or computational procedure evaluating the potential of a molecular entity as a novel biofuel component.

This Chapter is structured as follows: Materials and methods for the experimental screening are briefly summarized in Section 4.1. Thereafter, key assumptions and decisions in the course of modeling are described in detail in Section 4.2. The initial model formulation suffers from high parametric uncertainty and correlation, thus necessitating a stepwise model reduction. The final model is validated by means of cross-validation and external validation. Application of the model is demonstrated in Section 4.3 by means of three examples. Finally, some conclusions are provided. This Chapter has already been published in a slightly modified form in *Energy & Fuels* (Dahmen and Marquardt, 2015).

4.1 Materials and methods for the rapid screening

The fuel screening campaign has been executed with an IQT according to ASTM D6890 (2011) by ASG Analytik-Service GmbH (abbreviated below as ASG) in Neusäss, Germany. ASTM D6890 (2011) requires that the charge air pressure in the IQT combustion chamber must equal 2.1 MPa. Moreover, the charge air temperature as recorded by a thermocouple shall be 818 ± 30 K. Then, liquid fuel is injected into the chamber and the ignition delay is defined as the time between the start of fuel injection, i.e., the rise of the needle, and the start of significant heat release indicated by an increase in pressure beyond the pre-injection level. The IQT device is calibrated by adjusting the set-point of the chamber outer surface temperature controller until two reference substances, namely n-heptane and methylcyclohexane, yield pre-defined ignition delays (ASTM D6890, 2011; Alnajjar et al., 2010). According to ASTM D6890 (2011), the average over 32 combustion cycles is taken as the final ignition delay τ of a certain fuel. The so-called derived cetane number (DCN) can then be calculated as

Table 4.1: IQT ignition delay data and DCN data collected for the purpose of modeling.

	no. of compounds found in literature	no. of compounds screened by ASG
n-alkanes	6	2
iso-alkanes	14	2
cyclo-alkanes	8	2
alkenes	9	2
cyclo-alkenes	1	6
alcohols	9	8
aldehydes	0	6
ketones	2	9
esters	12	6
acyclic ethers	3	10
furans	3	2
tetrahydrofurans	2	3
other cyclic ethers	3	3
lactones	0	4
benzene compounds	2	6
more than one oxygen functionality	6	10
total	**80**	**81**

$$\mathrm{DCN} = 4.460 + 186.6/\tau \ , \tag{4.1}$$

if 3.1 ms $\leq \tau \leq$ 6.5 ms. Ignition delays outside this range are transformed into DCN according to

$$\mathrm{DCN} = 83.99(\tau - 1.512)^{-0.658} + 3.547 \tag{4.2}$$

instead (ASTM D6890, 2011). Combustion gases are released after each cycle.

It shall be noted that ASTM D7170 (2011) describes the so-called Fuel Ignition Tester (FIT), which is another constant-volume combustion chamber set-up for the determination of DCN. However, as a result of different boundary conditions, FIT-DCN values can differ significantly from IQT-DCN values (Hui et al., 2012). Therefore, whenever it is referred to DCN throughout this thesis, the intended reference is to IQT-DCN (ASTM D6890, 2011).

Fuel samples that had been bought and sent to ASG had a minimum purity of 96%. Some samples contained small concentrations of oxidation inhibitors for stabilization. Ta-

Table 4.2: Variation in ignition delay and DCN in case of 3-pentanone and γ-undecanolactone with respect to peroxide number measured according to ISO 3960 (2010). All measurements have been conducted by ASG.

	mmeq O/kg*	delay τ [ms]	DCN
3-pentanone[a]	22	8.51	26.9
3-pentanone[b]	<1	14.04	19.5
γ-undecanolactone[a]	23	3.57	56.7
γ-undecanolactone[b]	<1	3.88	52.6

* milliequivalent oxygen per kg fuel (mmeq O/kg)

[a] as supplied by the manufacturer (cf. Appendix B)

[b] after column chromatography

ble 4.1 summarizes ignition delay data provided by ASG and DCN data collected from the literature, in particular, from the two National Renewable Energy Laboratory CN compendia (Murphy et al., 2004; Yanowitz et al., 2014). In total, data on 161 fuels have been collected for the purpose of modeling. The whole data set, including supplier, purity and stabilizer information, is provided in Appendix B. Due to strong effects of mixture formation, gas temperature and pressure on the ignition delay, CN data deduced from ignition delays obtained in other than IQT combustion experiments have been deliberately neglected.

The main criteria for selection of compounds for IQT testing were commercial availability, a reasonable price, sufficient quantity and purity. Furthermore, diversity with respect to chemical structure was a main goal. Here, diversity has been assessed by grouping compounds into functional classes, which were defined to represent typical (oxygenated) products from lignocellulosic biomass (cf. Table 4.1). The author of this thesis has also received suggestions for the selection of test compounds from Jakob Mottweiler (Institute of Organic Chemistry, RWTH Aachen University).

Although the impact of peroxides forming in many fuel samples is rarely addressed in the literature on combustion kinetic experiments, their ignition-enhancing effect is well-known, for instance, from studies on biodiesel combustion (Graboski and McCormick, 1998). Yanowitz et al. (2014) also hypothesize that peroxides might have compromised some of the data collected in their CN compendium. It is impossible to reliably *estimate* the amount of peroxides in a fuel sample or the influence on the measured ignition delay just from similarity of molecular structures. While certain molecular structures might favor peroxide formation, the fuel history, i.e., its preparation, its purification and its storage, is typically unknown. A dedicated exemplary study was performed for 3-pentanone and γ-undecanolactone. As can be seen from Table 4.2, this investigation suggests a notable

distortion of IQT ignition delay and DCN, if the peroxide level is very high. As a consequence, the amount of peroxides has been measured by ASG for all fuel samples according to ISO 3960 (2010), once the author of this thesis had become aware of this issue during the experimental campaign (cf. Appendix B). ISO 3960 (2010) describes the iodometric determination of the peroxide concentration in vegetable and animal fats and oils as follows: The fuel sample is dissolved in isooctane and acetic acid first. Subsequently, potassium iodine is added. Oxygen, chemically bonded as peroxide, liberates iodine from potassium iodine; this iodine can be detected visually by titration with sodium thiosulfate. Mueller et al. (2012) and Yanowitz et al. (2014) have reported on the use of silica gel columns to remove ignition-accelerating species from hydrocarbons and/or non-polar oxygenates. The silica gel adsorbs polar contaminants like water or oxygenates. Aluminium oxide (alumina) constitutes another adsorbent frequently used in column chromatography. Prior to the IQT testing, alumina column chromatography had been performed by ASG to obtain accurate results, if the peroxide number according to ISO 3960 (2010) was found to be moderate or high (cf. Appendix B). The findings from Table 4.2 strongly suggest that peroxide testing should be performed routinely prior to combustion experiments, at least, if unsaturated and/or oxygenated species are investigated.

4.2 Modeling strategy

The modeling procedure is divided into four consecutive stages which are depicted in Table 4.3. In stage 1, the general model structure is introduced based on the theory of group contribution modeling. Since parameter estimates of the initial model suffer from high uncertainty and correlation, focus is then directed towards the rational reduction of the number of model inputs and parameters in stage 2. In stage 3, three hypotheses are introduced allowing for the formulation of the final IQT ignition delay model. This model is validated in stage 4. The methods denoted in Table 4.3 are briefly introduced in the following subsections to illustrate decision making in the course of modeling. The interested reader is referred to the cited literature for background information.

As can be seen from Figure 4.1 (top), experimental ignition delays τ are neither uniformly nor normally distributed in the range zero to 100 ms. It is generally recommended to transform such data prior to modeling (Eriksson et al., 2006b). A frequently employed transformation is the so-called log-transformation (Eriksson et al., 2006b): The logarithmic delays $\ln \tau$ shown in Figure 4.1 (center) represent a better approximation of the normal distribution. Note that it was decided not to model the DCN directly. The rationale is that τ represents the raw property measured in the IQT experiment, while DCN is merely a correlated quantity proposed to represent an estimate of CN in the CFR engine.

Table 4.3: Modeling procedure for the IQT ignition delay.

stage	task(s)	method(s)
1	generate ideas on the **general model structure**	group contribution modeling, residual analysis
2	generate ideas on **how to reduce parameter uncertainty and correlation**	confidence interval analysis, eigenvalue analysis, correlation analysis
3	introduce **three hypotheses** to enhance model performance	residual analysis, correlation analysis, information theory
4	**validate** final model	cross-validation, external validation

Typically, DCN is published instead of τ. Consequently, DCNs found in literature have been converted into logarithmic ignition delays using Eqns. (4.1) and (4.2) (cf. Appendix B). The relationship between $\ln \tau$ and DCN as defined in ASTM D6890 (2011) is depicted in Figure 4.1 (bottom): Eqns. (4.1) and (4.2) yield similar DCNs for $\ln \tau > 1.1$, i.e., DCN < 70. However, if $\ln \tau < 1.1$, i.e., DCN > 70, Eqn. (4.2) will give a higher DCN than Eqn. (4.1).

4.2.1 Selection of the general model structure

The first application of the group additivity principle to CN prediction of n-alkanes, isoalkanes and singly substituted alkylbenzenes was reported by DeFries et al. (1987). Since then, different spectroscopic methods such as nuclear magnetic resonance (NMR) spectroscopy have been utilized in the correlation of fuel composition, e.g., the fraction of carbon atoms in distinct functional group categories, to CN of petroleum-derived products (DeFries et al., 1987; Gulder and Glavincevski, 1986; Cookson and Smith, 1990; Cookson et al., 1993; Yang et al., 2002; Ghosh and Jaffe, 2006; Ghosh, 2008; Dryer, 2015). More recently, Mueller et al. (2012) have applied NMR analysis to compare and to match a surrogate fuel's compositional characteristics to those of reference diesel fuels produced from real-world refinery streams by considering eleven characteristic carbon-types, e.g., primary carbon (−CH3), secondary carbon (−CH2−) or quaternary carbon (aliphatic carbon). The same carbon-type classification has been used in the formulation of gasoline surrogate fuels by Ahmed et al. (2015). Dooley et al. (2010, 2012a,b), Won et al. (2014) and Dryer et al. (2014) have observed nearly identical distributions of methylene (−CH2−), methyl (−CH3) and benzyl functional groups in jet fuels and corresponding surrogate fuels. The surrogates had been formulated to match four combustion property targets for jet fuel, i.e., the DCN, the H/C molar ratio, the threshold sooting index (TSI), and the average molec-

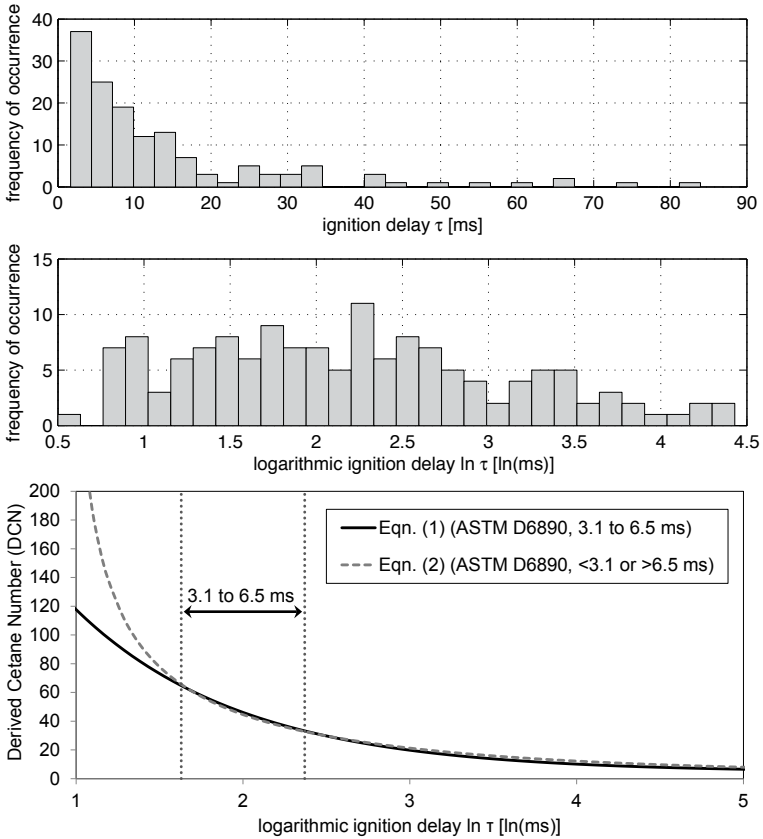


Figure 4.1: Histograms for non-scaled experimental ignition delay data τ (top) and for logarithmic ignition delay data $\ln \tau$ (center). Relationship between logarithmic ignition delay $\ln \tau$ and DCN based on the two correlations described in ASTM D6890 (2011) (cf. Eqns. (4.1) and (4.2)) (bottom).

Table 4.4: Nineteen structural groups from Joback's GCM (Joback and Reid, 1987) used for modeling of the IQT ignition delay.

hydrocarbon (non-ring)
−CH3, −CH2−, >CH−, >C<, =CH2, =CH−, =C<
hydrocarbon (ring)
−CH2−, >CH−, >C<, =CH−, =C<
oxygen (non-ring)
−OH, −O−, >C=O, O=CH−, −COO−
oxygen (ring)
−O−, >C=O

ular weight (Dooley et al., 2012b). Based on these encouraging findings, it is believed that surrogate fuels of relatively low complexity (in terms of actual molecular composition) can be formulated by emulating the real fuel's distribution of key distinct functional groups (Dryer et al., 2014; Dryer, 2015).

Still, significant uncertainty remains over the most suitable choice of functional groups or molecular fragments to be considered for the purpose of modeling of a certain target property, especially if a wide variety of (oxygenated) molecular structures shall be covered. Not surprisingly then, many different sets of groups have been used to model pure-component thermophysical and combustion-related phenomena (e.g., Benson and Buss (1958); Fredenslund et al. (1975); Klincewicz and Reid (1984); Joback and Reid (1987); DeFries et al. (1987); Constantinou and Gani (1994); Meusinger and Moros (1999); Marrero and Gani (2001); Yang et al. (2001); Albahri (2003); Pepiot-Desjardins et al. (2008)). In the present contribution, the experimental ignition delay data have been collected to cover all structural groups defined in Joback's GCM, except for the acid group (COOH), the alkyne groups (#CH and #C−) and the consecutive double bonds group (=C=). If exclusively $C_xH_yO_z$ compounds are considered, only nineteen of the original forty-one Joback groups remain relevant. These are listed in Table 4.4. In contrast to Joback's GCM, no distinction between alcohols and phenols is made.

Except for a few small molecules like hydrogen or methane, the molecular graph of any feasible $C_xH_yO_z$ compound can be decomposed into Joback's groups, where each atom is covered by exactly one group. It shall be noted that the eleven carbon-types considered by Mueller et al. (2012) and Ahmed et al. (2015) in the characterization of diesel and gasoline fuels can be linked to the Joback groups from Table 4.4.

For certain molecular graphs, different software implementations of Joback's method arrive at different group decompositions. For instance, the software package ICAS-ProPred (CAPEC, 2012) decomposes lactones into cyclic ether and cyclic ketone groups, while the

Dortmund Database web-interface reports ester groups instead (Dortmund Data Bank Software & Separation Technology GmbH, 2014b). Differences can also be noted in the treatment of carbonates and anhydrides (Dortmund Data Bank Software & Separation Technology GmbH, 2014a,c; CAPEC, 2012). In this thesis, the decomposition is consistent with that of ICAS-ProPred (CAPEC, 2012). Consequently, lactones are decomposed into cyclic ether and cyclic ketone groups, carbonates into ester and ether groups and anhydrides into ether and ketone groups (cf. Appendix C.1).

The basic formulation of a group contribution model (cf. Eqn. (2.4)) can be extended to allow for model inputs other than the group occurrences $g_{a,i}$. For instance, a major drawback of Joback's GCM is its inability to recognize aromatic bonds. To overcome this limitation, the descriptor $d_{nAB,i}$, i.e., the number of aromatic bonds in molecule i (Todeschini et al., 2009), may be introduced into Eqn. (2.4) to result in

$$y_i = \sum_a g_{a,i} G_a + d_{nAB,i} D_{nAB} + P \ , \tag{4.3}$$

where the parameter D_{nAB} represents the contribution of the descriptor to the explanation of y_i. A generalized formulation of the linear model allowing for multiple descriptors $d_{b,i}$ reads as

$$y_i = \sum_a g_{a,i} G_a + \sum_b d_{b,i} D_b + P \ . \tag{4.4}$$

In the following, different mathematical models $y_i^t\left(\mathbf{G}^t, \mathbf{D}^t, P^t, \mathbf{g}_i, \mathbf{d}_i\right)$ describing the logarithmic ignition delay $y_i^t = \ln \tau_i$ are distinguished by means of the model candidate index t. The vector $\mathbf{G}^t$ holds the parameters related to the structural groups. $\mathbf{D}^t$ denotes the vector of parameters related to the descriptors and P^t refers to the additional model parameter. All parameters of model t can be concatenated in $\mathbf{\Theta}^t = \left[\mathbf{G}^{t\mathrm{T}}, \mathbf{D}^{t\mathrm{T}}, P^t\right]^{\mathrm{T}} \in \mathbb{R}^{N_m^t}$, where N_m^t denotes the total number of model parameters. The input vectors $\mathbf{g}_i$ and $\mathbf{d}_i$ are summarized as $\mathbf{u}_i = \left[\mathbf{g_i}^{\mathrm{T}}, \mathbf{d_i}^{\mathrm{T}}\right]^{\mathrm{T}}$, where $\mathbf{g}_i$ describes the integer group occurrences in compound i and vector $\mathbf{d}_i$ holds the values of the descriptors for compound i, respectively.

In addition to the **linear model structure**, i.e.,

$$y_i^t\left(\mathbf{\Theta}^t, \mathbf{u}_i\right) = \sum_a g_{a,i} G_a^t + \sum_b d_{b,i} D_b^t + P^t \ , \tag{4.5}$$

the **nonlinear model structure** is considered

$$y_i^t\left(\mathbf{\Theta}^t, \mathbf{u}_i\right) = \exp\left[\sum_a g_{a,i} G_a^t + \sum_b d_{b,i} D_b^t + P^t\right] \ , \tag{4.6}$$

which can be rearranged to yield

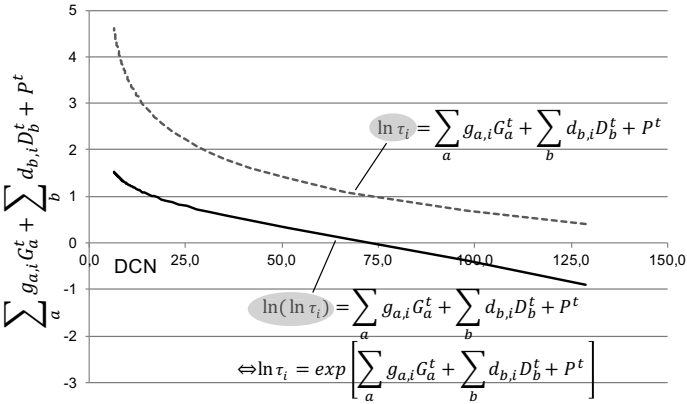


Figure 4.2: DCN vs. $\ln \tau_i$ and $\ln(\ln \tau_i)$ for τ_i varied between 1.5 ms and 100 ms. DCN is computed via Eqn. (4.1) over the entire range $1.5 \leq \tau_i \leq 100$.

$$\ln y_i^t \left(\boldsymbol{\Theta}^t, \mathbf{u}_i\right) = \sum_a g_{a,i} G_a^t + \sum_b d_{b,i} D_b^t + P^t \tag{4.7}$$

to clarify its motivation: If τ_i is varied between zero and 100 ms, both $\ln \tau = y_i^t \left(\boldsymbol{\Theta}^t, \mathbf{u}_i\right) = ...$ (Eqn. (4.5)) and $\ln(\ln \tau) = \ln y_i^t \left(\boldsymbol{\Theta}^t, \mathbf{u}_i\right) = ...$ (Eqn. (4.7)) can be plotted against DCN as shown in Figure 4.2. It turns out, there is a nearly linear relationship between $\ln(\ln \tau)$ and DCN for DCN larger than 20. The nonlinear structure proposed in Eqn. (4.6) to model $\ln \tau$ is motivated by the assumption of an approximately linear relationship between DCN and $\mathbf{u_i}$, i.e., the structural group occurrences $g_{a,i}$ and the descriptor values $d_{b,i}$.

Numerical optimization is commonly employed to determine the values of the parameters $\boldsymbol{\Theta}^t$ (Joback and Reid, 1987; Constantinou and Gani, 1994; Marrero and Gani, 2001). Here, the parameter estimation problem is formulated as the minimization

$$\min_{\boldsymbol{\Theta}^t} \phi^t = \frac{1}{2} \sum_{i=1}^{\varphi(\tilde{\mathbf{y}})} \frac{1}{\tilde{\sigma}_i^2} \left[y_i^t \left(\boldsymbol{\Theta}^t, \mathbf{u}_i\right) - \tilde{y}_i\right]^2 \ , \tag{4.8}$$

where $\varphi(\tilde{\mathbf{y}})$ measures the number of elements in the vector of experimental logarithmic ignition delay data $\tilde{\mathbf{y}}$. Optimal parameter estimates are denoted $\boldsymbol{\Theta}^{t*}$ in the following.

The weights $1/{\tilde{\sigma}_i}^2$ in the objective function ϕ^t (cf. Eqn. (4.8)) can be used to emphasize certain experiments in contrast to others using the standard deviation $\tilde{\sigma}_i$ as a measure of the reliability of the experimental data. Since the IQT performs 32 combustion cycles for each compound i, the standard deviation

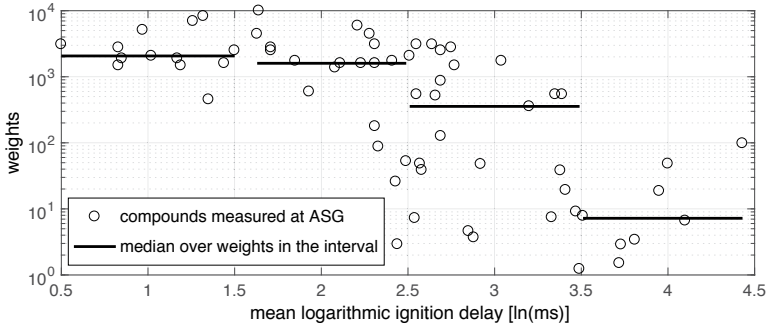


Figure 4.3: Weights ${1/\tilde{\sigma}_i}^2$ (inverse variances in the minimization objective) for the compounds measured at ASG computed from Eqn. (4.9). The black lines indicate the median values for the four intervals [0.0,..,1.5], (1.5,..,2.5], (2.5,..,3.5] and (3.5,..,5.0].

$$\tilde{\sigma}_i = \sqrt{\frac{\sum_{c=1}^{32} \left(\tilde{y}_{i,c} - \tilde{y}_i\right)^2}{31}} \tag{4.9}$$

can be computed based on $\tilde{y}_{i,c}$, i.e., the logarithmic delay for the individual cycle c. On average, measurement uncertainty is correlated to the ignition delay, i.e., the standard deviation $\tilde{\sigma}_i$ is small for short delays and then increases as the delay becomes larger. This can be seen from Figure 4.3, where the circles represent the weights $1/\tilde{\sigma}_i^2$ computed from Eqn. (4.9). For the literature data, however, according information on measurement uncertainty is missing. To compensate for this deficiency, the logarithmic ignition delay scale has been divided into four intervals and the median of the weights $1/\tilde{\sigma}_i^2$ obtained from the experiments performed at ASG has been calculated for each interval. These median weights are indicated by the black lines in Figure 4.3 and have been pragmatically assigned to the compounds retrieved from literature.

Based on the measurement data $\tilde{\mathbf{y}}$ and $\tilde{\boldsymbol{\sigma}}$, a weighted root-mean-square error (RMSE) can be computed via

$$RMSE^t = \sqrt{\frac{\sum_{i=1}^{\varphi(\tilde{\mathbf{y}})} \frac{1}{\tilde{\sigma}_i^2} \left[y_i^t\left(\boldsymbol{\Theta}^{t*}, \mathbf{u}_i\right) - \tilde{y}_i\right]^2}{\varphi(\tilde{\mathbf{y}})}} \tag{4.10}$$

to evaluate the quality of fit offered by model candidate y_i^t.

In a first step, residual analysis is applied to discriminate between the linear and non-linear model structures (cf. Eqns. (4.5) and (4.6)). To this end, identical input variables are supplied, i.e., the occurrences $g_{a,i}$ of all groups listed in Table 4.4 and the number

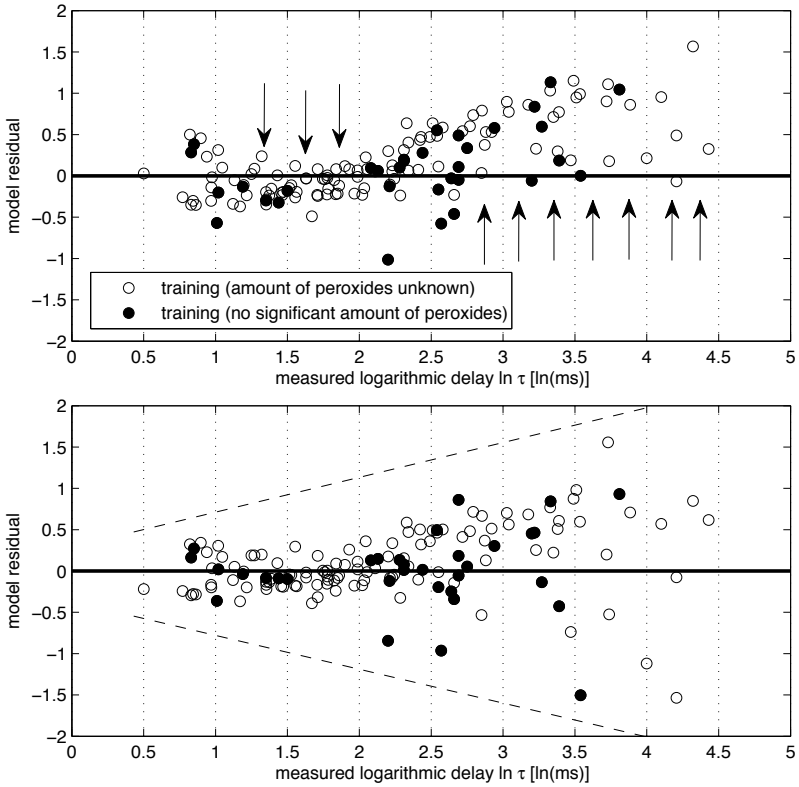


Figure 4.4: Residual plots ($\tilde{y}_i - y_i^t$) for the linear model (Eqn. (4.5)) (top) and the nonlinear model (Eqn. (4.6)) (bottom). Identical model inputs (occurrences $g_{a,i}$ of all groups listed in Table 4.4 and the descriptor $d_{nAB,i}$, i.e., the number of aromatic bonds).

of aromatic bonds $d_{nAB,i}$. Both models comprise an identical number of parameters, but the nonlinear structure exhibits a RMSE which is 17.3% lower than that for the linear structure. The residuals $(\tilde{y}_i - y_i^t)$ are compared graphically for both models in Figure 4.4. For the nonlinear model, residuals become larger as ignition delay and measurement uncertainty increase (cf. Figure 4.4, bottom). This is a desired consequence of the weighted least squares formulation depicted in Eqn. (4.8). In contrast, longer ignition delays are consistently underpredicted by the linear model, while for small-to-medium delays a trend towards overprediction can be noted (cf. Figure 4.4, top). Apparently, the exponential term in Eqn. (4.6) is able to resolve this issue, without introducing additional model parameters. As can be seen from Figure 4.4, the peroxide status (low peroxide level vs. unknown peroxide level) cannot be used to discriminate between the two model structures. Possibly, the current share of data with a known peroxide status is simply too small ($\sim$ 23%) and/or model accuracy is insufficient to graphically reveal a noticeable correlation between peroxide status and model residuals.

4.2.2 Systematic model reduction

If the number of measured properties, i.e., $\varphi(\tilde{\mathbf{y}})$, is small, the estimated model parameters $\boldsymbol{\Theta}^{t*}$ are often uncertain and/or correlated due to a lack of appropriate experimental data. Confidence intervals for parameter estimates can be derived from the Fisher information matrix (FIM)

$$F_{p,q}^t(\boldsymbol{\Theta}^{t*}, \mathbf{u}) = \sum_{i=1}^{\varphi(\tilde{\mathbf{y}})} \frac{\partial y_i^t(\boldsymbol{\Theta}^{t*}, \mathbf{u}_i)}{\partial \Theta_p^{t*}} \frac{1}{\tilde{\sigma}_i^2} \frac{\partial y_i^t(\boldsymbol{\Theta}^{t*}, \mathbf{u}_i)}{\partial \Theta_q^{t*}} \ , \tag{4.11}$$

which approximates the Hessian of the objective function ϕ^t by excluding the second-order derivatives of y_i^t (Walter and Pronzato, 1997). The FIM $\mathbf{F}^t$ links the sensitivities of the model with respect to its parameters with information on measurement uncertainty (Petersen et al., 2001; McLean and McAuley, 2012). It can be used to construct a linear approximation of the covariance matrix (Donaldson and Schnabel, 1987; Marsili-Libelli et al., 2003)

$$\mathbf{V}_{\mathbf{F}^t}(\boldsymbol{\Theta}^{t*}, \mathbf{u}) = \left(\frac{\sum_{i=1}^{\varphi(\tilde{\mathbf{y}})} \frac{1}{\tilde{\sigma}_i^2} \left(y_i^t(\boldsymbol{\Theta}^{t*}, \mathbf{u}_i) - \tilde{y}_i \right)^2}{\varphi(\tilde{\mathbf{y}}) - N_m^t} \right) \cdot \mathbf{F}^t(\boldsymbol{\Theta}^{t*}, \mathbf{u})^{-1} \tag{4.12}$$

allowing for the definition of a confidence interval $\Delta\Theta_p^{t*}$ for the estimate of the individual parameter p (Donaldson and Schnabel, 1987; Marsili-Libelli et al., 2003):

$$\Delta\Theta_p^{t*}(\boldsymbol{\Theta}^{t*}, \mathbf{u}) = t_{\varphi(\tilde{\mathbf{y}})-N_m^t}^{1-(\alpha/2)} \sqrt{V_{F_{p,p}^t}(\boldsymbol{\Theta}^{t*}, \mathbf{u})} \tag{4.13}$$

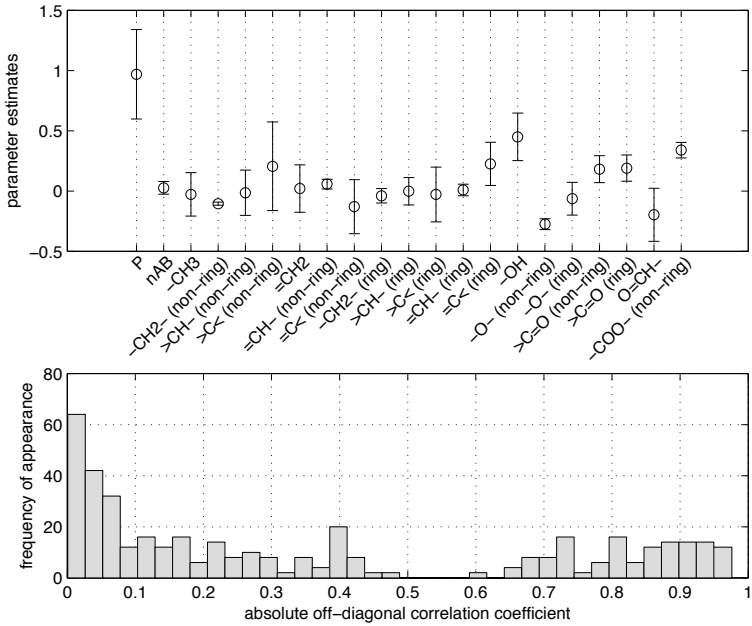


Figure 4.5: Nonlinear model (cf. Eqn. (4.6)) including all structural groups from Table 4.4 and the descriptor $d_{nAB,i}$, i.e., the number of aromatic bonds. Parameter estimates and 95% confidence intervals (top) and distribution of absolute off-diagonal parameter correlation coefficients (bottom).

Here, $t^{1-(\alpha/2)}_{\varphi(\tilde{\mathbf{y}})-N^t_m}$ denotes the two-tails Student's t-distribution for a given confidence level α and $\varphi(\tilde{\mathbf{y}}) - N^t_m$ degrees of freedom. If experimental screening and parameter estimation are repeated many times, the 95% confidence intervals will enclose the true parameter values with 95% probability (Bard, 1974). The size of the confidence interval is therefore an indicator for parametric uncertainty in the estimated model.

In case of the nonlinear model proposed in stage 1, large confidence intervals indicate high parametric uncertainty, as can be seen from Figure 4.5 (top). Moreover, parameter estimates are correlated, i.e., changes in one parameter estimate can be counteracted by changes in others. The absolute correlation $\zeta^t_{p,q}$ between two parameters $p \neq q$ has been computed based on the covariance matrix via (Franceschini and Macchietto, 2008)

$$\zeta^t_{p,q}(\boldsymbol{\Theta}^{t*}, \mathbf{u}) = \left| \frac{V_{F^t_{p,q}}(\boldsymbol{\Theta}^{t*}, \mathbf{u})}{\sqrt{V_{F^t_{p,p}}(\boldsymbol{\Theta}^{t*}, \mathbf{u})}\sqrt{V_{F^t_{q,q}}(\boldsymbol{\Theta}^{t*}, \mathbf{u})}} \right| . \tag{4.14}$$

Table 4.5: Rational model simplification aiming at reduced parameter uncertainty and correlation.

t	model inputs	N_m^t	$RMSE^t$	$\lambda_{min}(\mathbf{F}^t)$	mean/ max. corr. coeff.
2	all groups, $d_{nAB,i}$ (**basic nonlinear model)**	21	8.75	692	0.35/ 0.98
3	added $d_{nCCDB,i}$, removed all groups containing C=C double bonds (*reduction 1*)	17	9.07	6407	0.20/ 0.86
4	>CH− (ring), >CH− (non-ring), >C< (ring) and >C< (non-ring) removed (*reduction 2*)	13	9.57	14571	0.18/ 0.81
5	−CH2− (ring) removed (*reduction 3*) (**reduced model**)	12	10.15	22053	0.15/ 0.73

The histogram of $\zeta_{p,q}^t$ is depicted in Figure 4.5 (bottom). Some parameter correlations approach the theoretical maximum value of one indicating severe problems with the model.

Model simplification is applied to reduce parametric uncertainty and correlation. As a simple scalar measure for parametric uncertainty, the smallest eigenvalue of the FIM, i.e., $\lambda_{min}(\mathbf{F}^t)$, is considered here. This choice is motivated by the eigenvalue-approach described by Quaiser and Mönnigmann (2009): $\lambda^{t,j}$ and $\mathbf{v}^{t,j}$ denote the j-th eigenvalue and eigenvector of the FIM, respectively. All eigenvalues are sorted such that $\lambda_{min}(\mathbf{F}^t) = \lambda^{t,1} \leq \lambda^{t,2} \leq ... \leq \lambda^{t,N_m^t}$ and all eigenvectors are normalized, i.e., ${\mathbf{v}^{t,j}}^{\mathrm{T}}\mathbf{v}^{t,j} = 1$. Then, a movement in the parameter space along direction $\mathbf{v}^{t,j}$ originating at $\boldsymbol{\Theta}^{t*}$ given some real $\alpha > 0$ can be expressed as (Quaiser and Mönnigmann, 2009):

$$\phi^t(\boldsymbol{\Theta}^{t*} + \alpha\mathbf{v}^{t,j}) = \phi^t(\boldsymbol{\Theta}^{t*}) + \frac{1}{2}\alpha^2{\mathbf{v}^{t,j}}^{\mathrm{T}}\mathbf{F}^t\mathbf{v}^{t,j} = \phi^t(\boldsymbol{\Theta}^{t*}) + \frac{1}{2}\alpha^2\lambda^{t,j} \tag{4.15}$$

Thus, the smaller $\lambda_{min}(\mathbf{F}^t)$, the smaller is the change in the objective function ϕ^t obtained for a given α. A larger $\lambda_{min}(\mathbf{F}^t)$ therefore indicates less parametric uncertainty.

The first model simplification applies to the five structural groups describing C=C double bonds (cf. Table 4.4). These five groups are substituted by the the total number of C=C double bonds in a molecule ($d_{nCCDB,i}$). Thus, the constitution of a C=C double-bond (e.g., chain end or branching) is no longer accounted for in the model (**reduction 1**). This results in a modest 3.7% increase in the RMSE, however, $\lambda_{min}(\mathbf{F}^t)$ is improved by one order of magnitude, as can be seen from Table 4.5 (model t=3).

From Eqn. (4.15) it becomes obvious, that the smallest change in ϕ^t is obtained, if the parameter values are modified in the direction of $\mathbf{v}^{t,1}$, i.e., the eigenvector belonging to the smallest eigenvalue $\lambda_{min}(\mathbf{F}^t) = \lambda^{t,1}$. After substitution of the C=C double bond groups, $\mathbf{v}^{t,1}$ does not point explicitly into the direction of a single parameter, e.g., $\mathbf{v}^{t,1} = [0, .., 0, 1, 0, .., 0]^{\mathrm{T}}$, but contains non-zero elements of different magnitude at all positions. Quaiser and Mönnigmann (2009) suggest to use

$$|v_l^{t,1}| = \max_l \left(|v_1^{t,1}|, |v_2^{t,1}|, .., |v_{N_m^t}^{t,1}| \right) \tag{4.16}$$

in order to select the parameter l associated with the highest uncertainty. Here, this parameter l refers to the >C< (non-ring) group. The three highest correlations to other parameters reported for this group are: 0.83 to −CH3, 0.74 to >CH− (non-ring) and 0.68 to >CH− (ring). For obvious reasons, there must be a strong relationship between the number of −CH3 groups and the number of branchings in a hydrocarbon skeleton. Hence, the four branching groups, i.e., >C< (non-ring), >C< (ring), >CH− (non-ring), >CH− (ring), are dropped in an attempt to further simplify the model, where only the −CH3 group is retained to indirectly account for branching (**reduction 2**). This yields another modest 5.5% increase in the RMSE, however, $\lambda_{min}(\mathbf{F}^t)$ more than doubles, as can be seen from Table 4.5 (model t=4).

The methyl (−CH3) and methylene (−CH2−) groups play a central role in low temperature auto-ignition chemistry (Dooley et al., 2012a; Won et al., 2014; Dryer, 2015). The number of −CH2− groups influences the rate of alkylperoxy radical isomerization, which controls radical production in this oxidation regime (Westbrook, 2000; Oehlschlaeger et al., 2009; Dooley et al., 2012a). The number of −CH3 groups is directly related to the degree of branching (Dooley et al., 2012a) and it is well known that a larger number of methyl groups will decrease CN (O'Connor et al., 1992; Heck et al., 1998; Santana et al., 2006). For highly branched fuel molecules, rates and number of possible alkylperoxy radical isomerization reactions decrease as fewer low-energy transition state rings can be formed (Westbrook, 2000). Thus, the relative proportion of −CH3 and −CH2− groups is thought to determine the composition of the radical pool produced and has been proposed as an indicator for the gas phase kinetic reactivity at low temperatures (Dooley et al., 2012a; Won et al., 2014; Dryer, 2015).

The final model reduction step is achieved by exclusion of the −CH2− (ring) group due to high correlations with the −CH2− (non-ring) (0.69) and the −CH3 (0.54) groups (**reduction 3**). After this third model simplification, the only parameter correlations exceeding a value of 0.5 are those involving parameter P^t. In total, the three simplifications have proven successful for significantly decreasing parametric uncertainty as indicated by the steep rise in the smallest eigenvalue $\lambda_{min}(\mathbf{F}^t)$ given in Table 4.5. Along with the model

Table 4.6: Estimator for the vapor pressure p^S in bar at 298 K. Estimates for the normal boiling point T_{boil} in K, the critical temperature T_{crit} in K and the critical pressure p_{crit} in bar are retrieved from Joback's GCM (Joback and Reid, 1987).

$$f_{T_{boil}} = \frac{1}{T_{boil}} - 7.9151 \cdot 10^{-3} + 2.6726 \cdot 10^{-3} \cdot log_{10}(T_{boil}) - 0.8625 \cdot 10^{-6} \cdot T_{boil}$$
$$f_{T_{crit}} = \frac{1}{T_{crit}} - 7.9151 \cdot 10^{-3} + 2.6726 \cdot 10^{-3} \cdot log_{10}(T_{crit}) - 0.8625 \cdot 10^{-6} \cdot T_{crit}$$
$$\chi_1 = 11.52608845 - \ln\left(\frac{1.01325}{p_{crit}}\right) \cdot \frac{f_{T_{boil}}}{f_{T_{boil}} - f_{T_{crit}}}$$
$$\chi_2 = \frac{\ln(1.01325/p_{crit})}{f_{T_{boil}} - f_{T_{crit}}}$$
$$p^S = \exp\left[\chi_1 + \chi_2 \cdot 1.7962 \cdot 10^{-3} - 13.81551056\right] \cdot 10$$

inputs, the number of parameters has been reduced from 21 to just 12, thereby increasing the RMSE by 16% in total. Based on this reduced, less correlated model, three additional hypotheses are introduced in the following stage to yield the final IQT ignition delay model.

4.2.3 Proposition of the final model

Assuming the existence of a correlation between the vapor pressure p^S and the *physical* contribution to the IQT ignition delay (as a part of τ, cf. Figure 2.2), the descriptor p^S (in bar at 298 K) is added to the reduced model (**hypothesis 1**). Experimental data on p^S, however, may not always be available. To overcome this limitation, an estimator for p^S based on the Hoffmann-Florin equation (Hoffmann and Florin, 1943; Gmehling et al., 2012) is given in Table 4.6. The same estimator has been employed for the determination of p^S for *all* training compounds (cf. Appendix C.1) in order to provide consistency with the future use of the IQT ignition delay model. Since p^S may span multiple orders of magnitude, the natural logarithm of the vapor pressure, i.e., $\ln\left(d_{i,p^S}\right)$, is added to the reduced model.

If the CNs of paraffins and olefins are plotted over the number of carbon atoms, an asymptotic trend can be noted (Fehér et al., 2014): The CN monotonically increases, however, the increase per additional carbon becomes smaller as the molecule gets larger. To reflect this behavior, a logarithmic term for the −CH2− (non-ring) group input is proposed (**hypothesis 2**), i.e., $\ln\left(g_{-CH2-(non-ring),i} + 1\right)$ is used instead of $g_{-CH2-(non-ring),i}$ in Eqn. (4.6). Here, the nearly linear relationship between $\ln(\ln\tau)$ and DCN for DCN larger than 20 is exploited (cf. Figure 4.2).

Isocetane (2,2,4,4,6,8,8-heptamethylnonane), a reference compound for the determination of CN, is a large, saturated hydrocarbon ($C_{16}H_{34}$). Due to its branched structure, isocetane has a high fraction of difficult-to-abstract primary C-H-bonds. Moreover, isocetane has few secondary C-H sites fostering alkylperoxy radical isomerization reactions leading to degenerate chain branching (Heck et al., 1998; Westbrook, 2000; Oehlschlaeger

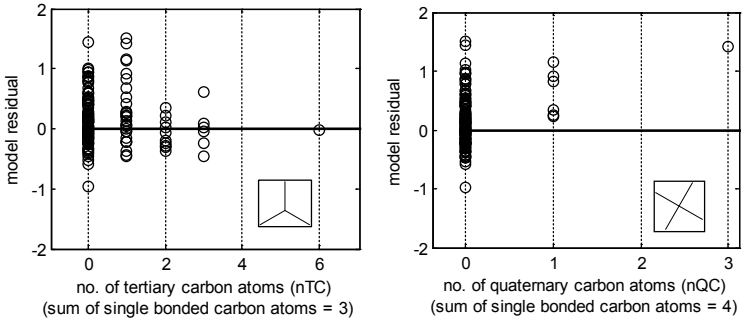


Figure 4.6: Residuals of the reduced model ($\tilde{y}_i - y_i^5$) with respect to the number of tertiary or quaternary carbon atoms.

et al., 2009; Westbrook et al., 2011). As a consequence, isocetane's DCN is as low as 14.2 (cf. Appendix B). For a similar reason, isooctane (2,2,4-trimethylpentane, C_7H_{16}) has a DCN of 17.0 (cf. Appendix B). In the reduced model, both molecules are described entirely by the number of −CH2− (non-ring) groups (3 vs. 1) and the number of −CH3 groups (9 vs. 5). However, ignition delays for such highly branched molecules are underestimated by the reduced model. Dooley et al. (2012a), who have highlighted the importance of the methyl (−CH2−) and methylene (−CH3) groups in their group additivity analysis of surrogate fuels, have not considered tertiary and quaternary carbon atoms. Indeed, as can be seen from Figure 4.6 (left), residuals of the reduced model scatter around zero, mostly independent of the number of tertiary carbon atoms. However, there seems to be a correlation between model residuals and the number of quaternary carbon atoms (cf. Figure 4.6, right). These carbon atoms are attached to four other carbon atoms via single bonds.

In a recent investigation performed by Mueller et al. (2012), aliphatic quaternary carbon atoms have not been detected by NMR analysis in two practical diesel fuels. However, surrogate fuels formulated based on the methodology proposed by Mueller et al. (2012) contained up to 29 mol-% of isocetane, a molecule with three quaternary carbon atoms. A priori, isocetane had been chosen together with eight other hydrocarbon palette compounds to emulate the behavior of the two fossil diesel fuels. On a carbon-type basis, there are considerable differences between surrogate and reference fuels considered by Mueller et al. (2012), not only with regard to aliphatic quaternary carbon. Moreover, it is unclear whether a mismatch in aliphatic quaternary carbon can be compensated by discrepancies in other carbon-types. Thus, the work of Mueller et al. (2012) does not constitute clear evidence against using the number of quaternary carbon atoms to improve the performance

Table 4.7: Reduced model (t=5) together with model candidates based on hypotheses 1 to 3.

t	description	N_m^t	$RMSE^t$	$AICc^t$	$\lambda_{min}(\mathbf{F}^t)$	mean/ max. corr. coeff.
5	cf. Table 4.5 (**reduced model**)	12	10.15	14218	22053	0.15/ 0.73
6	added $d_{p^S,i}$ *(hypothesis 1)*	13	9.45	12242	15109	0.22/ 0.63
7	logarithmic −CH2− (non-ring) *(hypotheses 1 & 2)*	13	8.93	10877	16622	0.20/ 0.62
8	added $d_{nQC,i}$ *(hypotheses 1 to 3)* (**final model**)	14	8.51	9811	16530	0.18/ 0.60

of the group contribution model. In fact, non-zero contributions from quaternary carbon atoms have been included in earlier CN models (DeFries et al., 1987; Yang et al., 2001) and ON models (Meusinger and Moros, 1999; Albahri, 2003) derived for iso-alkanes. Based on the IQT data (cf. Figure 4.6), the number of quaternary carbon atoms $d_{nQC,i}$ is added to the model in order to obtain a more aggressive DCN reduction due to heavy branching (**hypothesis 3**). Both non-ring and ring quaternary carbons are counted for in $d_{nQC,i}$.

Since the inclusion of an additional model parameter typically improves the RMSE, Akaike's information criterion (AIC) (Walter and Pronzato, 1997; Burnham and Anderson, 2002; Michalik et al., 2009) is used to estimate whether this improvement does also reduce the information loss against the "true" relationship between molecular structure and ignition delay. The AIC represents a trade-off between the likelihood function L^t, i.e., a measure for the quality of fit, and the number of model parameters N_m^t. If the sample size is small (generally in case of $\varphi(\tilde{\mathbf{y}})/N_m^t < 40$), a correction term is introduced to the original AIC formulation (Burnham and Anderson, 2002): Assuming independent, normally distributed residuals with zero mean and variances $\tilde{\sigma}_i$, the Akaike information criterion corrected for a small sample size (AICc) is calculated as (Hurvich and Tsai, 1989; Walter and Pronzato, 1997; Burnham and Anderson, 2002; Michalik et al., 2009)

$$AICc^t = -2 \cdot \ln\left[L^t(\boldsymbol{\Theta}^{t*}, \tilde{\boldsymbol{\sigma}}, \tilde{\mathbf{y}}, \mathbf{u})\right] + 2N_m^t \left(\frac{\varphi(\tilde{\mathbf{y}})}{\varphi(\tilde{\mathbf{y}}) - N_m^t - 1}\right) \tag{4.17}$$

$$= 2\left[\ln\left(\sqrt{2\pi}\right)\varphi(\tilde{\mathbf{y}}) + \sum_{i=1}^{\varphi(\tilde{\mathbf{y}})} \ln\left(\tilde{\sigma}_i\right)\right] + 2\sum_{i=1}^{\varphi(\tilde{\mathbf{y}})}\left(\frac{\left(y_i^t(\boldsymbol{\Theta}^{t*}, \mathbf{u}_i) - \tilde{y}_i\right)^2}{2\tilde{\sigma}_i^2}\right) + 2N_m^t\left(\frac{\varphi(\tilde{\mathbf{y}})}{\varphi(\tilde{\mathbf{y}}) - N_m^t - 1}\right) . \tag{4.18}$$

As a rule of thumb, a difference of more than 10 indicates that the model comprising the

higher AICc has essentially no empirical support (Burnham and Anderson, 2002). It can be seen from Table 4.7, that all three hypotheses result in large improvements in the AICc. Moreover, the benign correlation structure obtained from the previous model reduction is not negatively affected.

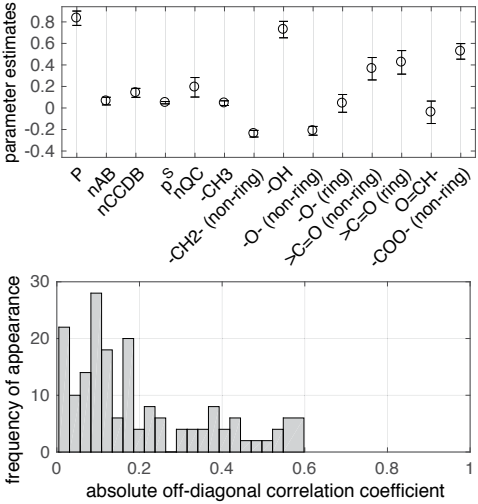


Figure 4.7: IQT ignition delay model: Parameter estimates and 95% confidence intervals (top). Distribution of parameter correlation coefficients (bottom).

Model number 8 constitutes the final IQT ignition delay model. In contrast to the original nonlinear model, small confidence intervals indicate low parametric uncertainty, as can be seen from Figure 4.7 (top). The histogram of parameter correlations is depicted in Figure 4.7 (bottom). Here, the maximum correlation coefficient is smaller than 0.6. This suggests that different physicochemical effects are well-separated in the regressors. Parity plots for predictions vs. measurements are depicted in Figure 4.8. As a general trend, model residuals increase as ignition delay and measurement uncertainty become larger. Due to the scaling induced by Eqn. (4.1), scattering in the DCN space is more uniform. Eqn. (4.1) has been used for error calculations and parity plot illustrations even beyond the range provided in ASTM D6890 (2011). Since the disagreement between the two DCN equations is substantial only for DCN larger than 70 (cf. Figure 4.1) and few fuels ignite that fast, the general plausibility of both Eqns. (4.1) and (4.2) in this regime is questionable. The strong nonlinearity of Eqn. (4.2), however, gives rise to huge deviations in the DCN space resulting from relatively small deviations in the logarithmic ignition delay.

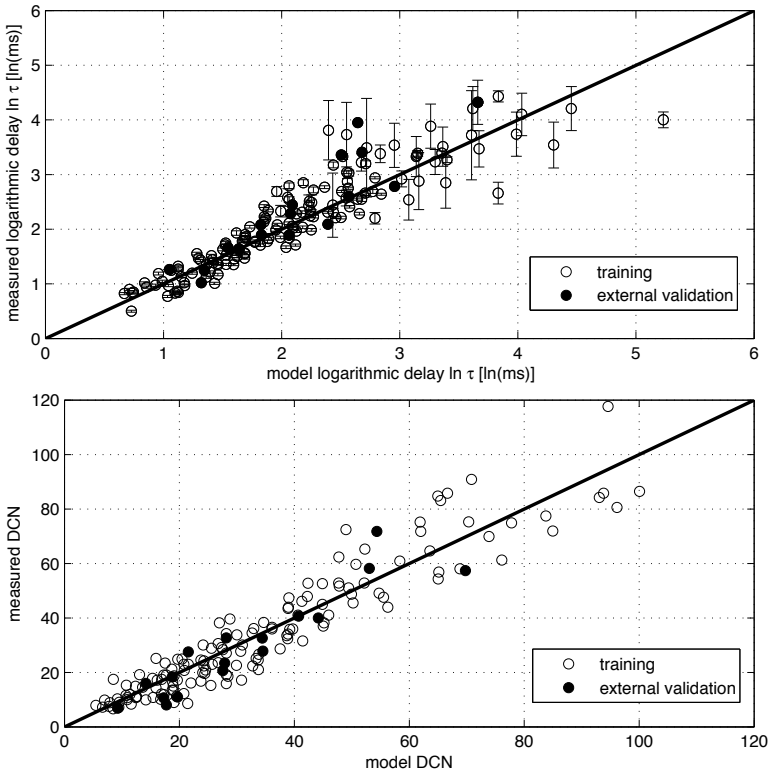


Figure 4.8: IQT ignition delay model: Parity plot for ignition delays (top). Error bars indicate the measurement standard deviation ($\pm\tilde{\sigma}$). Parity plot for DCNs (bottom). Note: $\ln\tau$ has been converted to DCN exclusively based on Eqn. (4.1).

Table 4.8: IQT ignition delay model: Main assumptions.

1	The IQT ignition delay τ of medium-sized liquid (oxygenated) hydrocarbons can be estimated from additive structural group and molecular descriptor effects according to the general model structure of $\ln \tau = \exp\left[\sum_a (\text{group effects}) + \sum_b (\text{descriptor effects}) + P\right]$, where P is a constant parameter.
2	The non-oxygenated hydrocarbon part of the molecule is described by the no. of −CH3 and −CH2− (non-ring) group(s), the no. of C=C double bonds, the no. of quaternary carbon atoms, the no. of aromatic bonds and the estimated vapor pressure.
3	The oxygenated part is described by −OH, −O− (non-ring), >C=O (non-ring), O=CH−, −COO− (non-ring), −O− (ring) and >C=O (ring) groups and the estimated vapor pressure.
4	All group occurrences and descriptors enter $\sum_a(...)$ or $\sum_b(...)$ linearly, except for the −CH2− (non-ring) group and the vapor pressure.

The main assumptions behind the proposed model are briefly summarized in Table 4.8. A detailed description of the model is given in Table 4.9. The equations needed to facilitate the computation of the IQT ignition delay τ for a given $C_xH_yO_z$ structure can be found here. These equations can be rearranged to yield the general structure introduced in Eqn. (4.6). Furthermore, Table 4.9 summarizes the model inputs, i.e., the group occurrences and descriptors. An example is given for each model input.

4.2.4 Model validation

Model validation is performed by a combination of cross-validation and external validation. Leave-one-out cross-validation (LOO-CV) is an estimator of the expected prediction error which has a low bias. However, this estimator can have a high variance (Hastie et al., 2009), i.e., the actual prediction error for an independent test sample can be quite different. Leave-multiple-out cross-validation (LMO-CV), in contrast, has a low variance, but can be biased upwards, i.e., LMO-CV can overestimate the true prediction error (Hastie et al., 2009). In case of the IQT ignition delay model, a variant of LMO-CV, i.e., the *repeated* leave-multiple-out cross-validation, is applied. Here, the compounds in the training data set are randomly assigned to sets 1 and 2 in each iteration such that set 1 contains 80% of the training compounds, while the remaining 20% are located in set 2. Data set 1 is used for parameter estimation yielding $\boldsymbol{\Theta}_1^*$ and data set 2 is used for prediction, i.e., model predictions $y_i(\boldsymbol{\Theta}_1^*, \mathbf{u}_{2,i})$ are compared to experimental data $\tilde{y}_{2,i}$. In total, ten thousand iterations were performed to obtain an estimate on the expected prediction error. Table

Table 4.9: IQT ignition delay model: Model equations and inputs for the computation of the IQT ignition delay τ_i.

model equations:		
$\ln(\tau_i) = y(\mathbf{g}_i, \mathbf{d}_i) = \exp\left[\text{Carbon}_i + \text{Oxygen}_i + \text{Oxygen(ring)}_i + \text{Descriptors}_i + 0.8341\right]$		
$\text{Carbon}_i = 0.0449 \cdot g_{-CH3,i} - 0.2389 \cdot \ln\left(g_{-CH2-(non-ring),i} + 1\right)$		
$\text{Oxygen}_i = 0.7286 \cdot g_{-OH,i} - 0.0401 \cdot g_{O=CH-,i} - 0.2123 \cdot g_{-O-(non-ring),i}$ $+0.3649 \cdot g_{>C=O(non-ring),i} + 0.5260 \cdot g_{-COO-(non-ring),i}$		
$\text{Oxygen(ring)}_i = 0.0426 \cdot g_{-O-(ring),i} + 0.4241 \cdot g_{>C=O(ring),i}$		
$\text{Descriptors}_i = 0.0639 \cdot d_{nAB,i} + 0.1404 \cdot d_{nCCDB,i} + 0.1923 \cdot d_{nQC,i} + 0.0492 \cdot \ln\left(d_{p^S,i}\right)$		
model inputs $\mathbf{g}_i$, $\mathbf{d}_i$ for a molecule i as follows:		examples:
$g_{-CH3,i}$	no. of −CH3 groups	g_{-CH3}=2 for n-heptane
$g_{-CH2-(non-ring),i}$	no. of −CH2− (non-ring) groups	$g_{-CH2-(non-ring)}$=5 for n-heptane
$g_{-OH,i}$	no. of −OH groups	g_{-OH}=1 for ethanol
$g_{O=CH-,i}$	no. of O=CH− groups	$g_{O=CH-}$=1 for hexanal
$g_{-O-(non-ring),i}$	no. of −O− (non-ring) groups	$g_{-O-(non-ring)}$=1 for dibutylether
$g_{>C=O(non-ring),i}$	no. of >C=O (non-ring) groups	$g_{>C=O(non-ring)}$=1 for 2-butanone
$g_{-COO-(non-ring),i}$	no. of −COO− (non-ring) groups	$g_{-COO-(non-ring)}$=1 for methyl hexanoate
$g_{-O-(ring),i}$	no. of −O− (ring) groups	$g_{-O-(ring)}$=1 for tetrahydrofuran
$g_{>C=O(ring),i}$	no. of >C=O (ring) groups	$g_{>C=O(ring)}$=1 for cyclopentanone
$d_{nAB,i}$	no. of aromatic bonds	d_{nAB}=6 for benzene
$d_{nCCDB,i}$	no. of carbon-carbon double bonds	d_{nCCDB}=2 for furan
$d_{nQC,i}$	no. of quaternary carbon atoms*	d_{nQC}=1 for 2,2-dimethylbutane
$d_{p^S,i}$	vapor pressure (cf. Table 4.6)	d_{p^S}=0.0024 for hexyl acetate
τ_i	IQT ignition delay in ms	
$d_{p^S,i}$	vapor pressure in bar at 298 K	

*quaternary carbon atoms are carbon atoms that are attached to four other carbon atoms via single bonds; both non-ring and ring quaternary carbons are counted for in $d_{nQC,i}$

Table 4.10: IQT ignition delay model: Statistical performance measures.

	training data set	LMO cross-validation data set	external validation data set
root-mean-square error (RMSE)	8.51	10.54	10.40
mean absolute error in DCN space	5.8	6.6	5.8
maximum absolute error in DCN space	23.4	33.4	17.4

Note: DCN error calculations are based exclusively on Eqn. (4.1).

4.10 summarizes statistical performance measures for the training and the cross-validation data sets.

It is generally recommended to have measurement data $\tilde{\mathbf{y}}_{ext}$ for an additional set of compounds, the so-called external validation set, which has not been involved in the derivation of the model. Such an external validation set had been separated prior to modeling (cf. Appendix B). From each class of compounds depicted in Table 4.1, one randomly chosen molecule had been placed in the external validation set. Two molecules were chosen randomly from the heterogeneous class of compounds comprising more than one type of oxygen functionality. Selection from functional classes has been motivated by the concepts of molecular diversity and similarity between training and test set compounds (Golbraikh et al., 2003; Leonard and Roy, 2006).

As shown in Table 4.10, a mean absolute DCN prediction error of 5.8 is obtained for the external validation set. With the exception of hexanal and diethoxymethane, the error does not exceed the value of 10, as can be seen from Figure 4.9. Compared to the training data set, the RMSE is 22% higher for the external validation set and 24% higher in case of LMO-CV (cf. Table 4.10). These numbers agree with the expectation of a too optimistic RMSE for the training data set. Similar values for LMO-CV and external validation suggest that the model has good predictive power (Golbraikh and Tropsha, 2002a).

A second independent data set can be used to test the model: No ignition within 100 ms was detected for few compounds, e.g., for gasoline anti-knock agents methyl tert-butyl ether (MTBE) and toluene. This absence of auto-ignition is expressed as DCN<7. Generally, the model performs well in the prediction of these extreme compounds as indicated in Figure 4.10. However, a rather weak estimate is obtained for MTBE. A comparison between MTBE and the structurally similar 2,2-dimethylbutane is presented in Figure 4.11. MTBE's DCN is much lower suggesting that the −O− group *inhibits* auto-ignition. However, this finding contradicts the role of the ether group that has been found in other compounds, which is to *promote* auto-ignition (e.g., in dibutylether; further examples can be found in Appendix B). Attempts to improve the prediction of MTBE by adding a

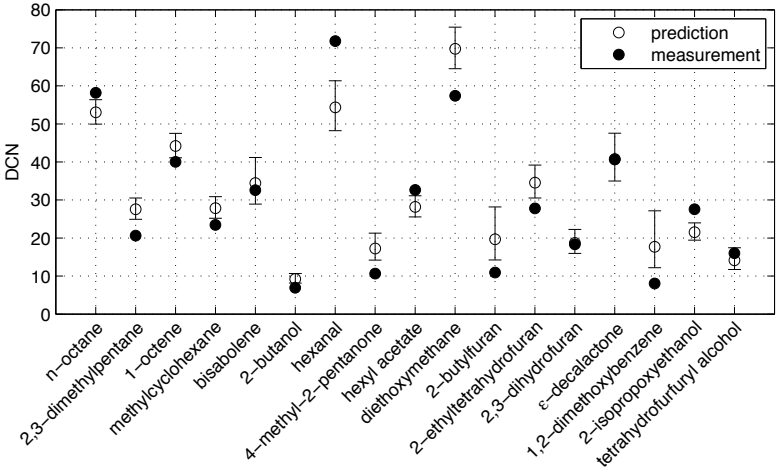


Figure 4.9: IQT ignition delay model: Predicted and measured DCNs for the compounds in the external validation set together with the 95% confidence intervals for the predictions. Note: $\ln \tau$ has been converted to DCN exclusively based on Eqn. (4.1).

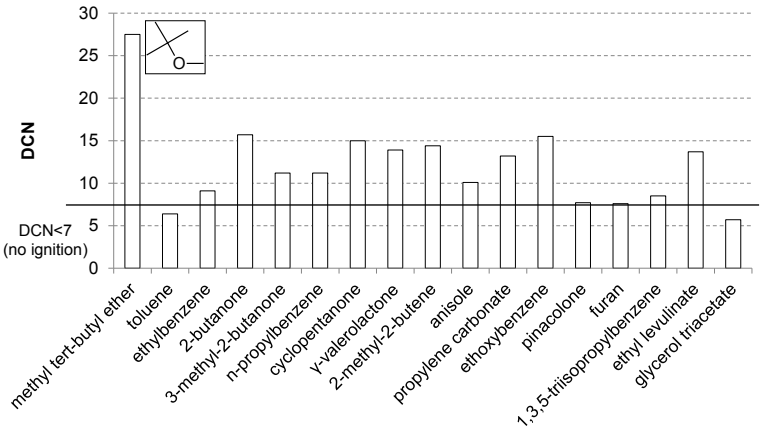


Figure 4.10: IQT ignition delay model: Predicted DCN for the compounds that did not ignite in the IQT (DCN<7). Data for 1,3,5-triisopropylbenzene, ethyl levulinate and glycerol triacetate have been taken from the compendium of Yanowitz et al. (2014). The other compounds have been screened by ASG. Note that DCN is computed from Eqn. (4.1).

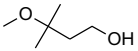
<7	23.5	8.6	**measurement**
			
methyl tert-butyl ether	2,2-dimethylbutane	3-methoxy-3-methyl-1-butanol	
27.5	17.5	21.4	**model**
18.5	17.5	14.1	**model (–O– group ignored)**

Figure 4.11: Measured DCN (top) vs. predicted DCN (bottom) for ethers comprising an −O− group next to a tertiary carbon atom (according to Eqn. (4.1)). 2,2-Dimethylbutane is a hydrocarbon structurally similar to methyl tert-butyl ether.

descriptor $d_{nTC,i}$, i.e., the number of tertiary carbon atoms, were not successful. This suggests an interaction effect between neighboring groups, however, more experimental data is needed to support or reject this hypothesis. For the time being, simply neglecting the contribution of the −O− group next to a tertiary carbon atom improves the prediction for MTBE. The same applies for 3-methoxy-3-methyl-1-butanol, i.e., the only other molecule in the database comprising an −O− group next to a tertiary carbon atom (cf. Figure 4.11).

The residual plots in Figure 4.12 reveal no apparent systematic trend in residuals ($\tilde{y}_i - y_i$) with respect to different characteristics of hydrocarbons, e.g., molecular weight, number of tertiary carbon atoms or number of −CH3 groups. Measured and predicted data for the individual compounds can be found in Appendix C.2.

It shall be noted, that model validation techniques assume, that data sets have been drawn in a representative manner from the overall set of molecular structures of potential interest. In other words, the error estimate cannot be transferred to highly dissimilar compounds. To the author's knowledge, no reliable applicability domain concept or similarity threshold for use in GCM has been proposed so far. Still, Table 4.11 provides some guidance on the applicability range of the proposed model. It summarizes few important characteristics of those molecules that were used to derive the model. Properties of a novel molecule can be compared against those reported in Table 4.11, before the model is applied for prediction.

The user should also inspect a novel molecule for functional groups more complex than those defined in Table 4.9. For instance, a DCN of 25.7 is predicted for diethyl carbonate, a fuel that did not ignite in the IQT (DCN<7). However, carbonate esters were not included in the training set. As a result, the model does not recognize the carbonate ester (−OC(=O)O−) group as a functional group of its own. Instead, it decomposes this

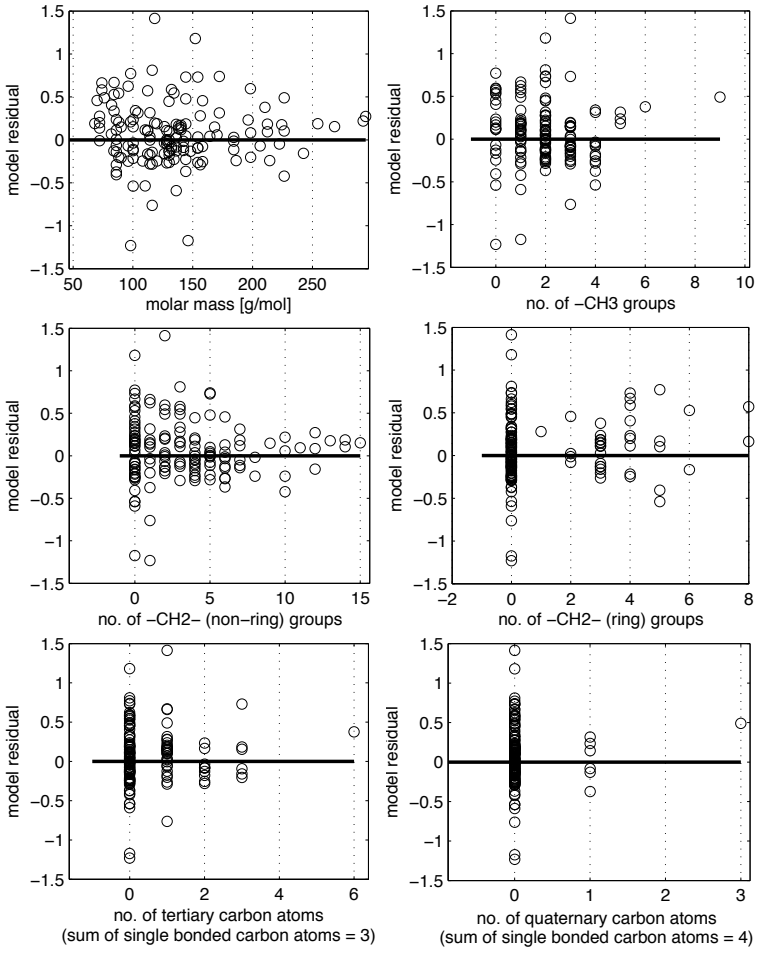


Figure 4.12: IQT ignition delay model: Residuals ($\tilde{y}_i - y_i$).

Table 4.11: IQT ignition delay model: Properties of training molecules.

no. of −CH3 groups	0 to 9
no. of −CH2− (non-ring) groups	0 to 15
no. of −CH2− (ring) groups	0 to 8
no. of −OH, −O− (ring), >C=O (non-ring), >C=O (ring), O=CH−, −COO− (non-ring) groups	0 to 1
no. of −O− (non-ring) groups	0 to 3
no. of aromatic bonds	0 to 12
no. of C=C double bonds	0 to 6
vapor pressure [ln(bar)] (Hoffmann-Florin equation, cf. Table 4.6)	-17.8 to -0.3
molecular weight [g/mol]	68.1 to 294.5
ring size (heavy atoms)	4 to 8
no. of benzene motifs	0 to 2
normal boiling point [K] (Joback and Reid, 1987)	307 to 705

complex group into two smaller groups, i.e., one ester (−COO−) group and one ether (−O−) group. Hence, if the model shall be applied to a carbonate ester, one must assume that the superposition of the effects from −O− and −COO− groups adequately describes the effect of the more complex carbonate ester (−OC(=O)O−) functional group.

4.3 Model application and uncertainty

Three examples shall illustrate the prediction of the IQT ignition delay τ and the DCN. Model equations and inputs are taken from Table 4.9. The group decomposition for hexyl acetate (HA) is derived from the molecular structure depicted in Figure 4.13 (left): 5 × −CH2−, 2 × −CH3, 1 × −COO−. Following the explanations given in Table 4.6, the vapor pressure p^S is computed based on critical properties (T_{crit} = 613 K, p_{crit} = 25.5 bar) and the normal boiling point (T_{boil} = 441 K): $\ln(p^S)$ = -6.0518 ln(bar).

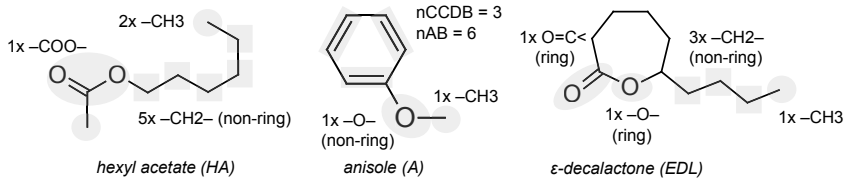


Figure 4.13: Group decomposition illustrated for hexyl acetate, anisole and ϵ-decalactone.

The contributions from the hydrocarbon groups, from the oxygen groups and from the descriptors can be assembled with Table 4.9 as follows:

$$\begin{aligned}\text{Carbon}_{HA} =& 0.0449 \cdot g_{-CH3,HA} - 0.2389 \cdot \ln\left(g_{-CH2-(non-ring),HA} + 1\right) \\ =& 0.0449 \cdot 2 + (-0.2389) \cdot \ln(5+1) \end{aligned} \tag{4.19}$$

$$\text{Oxygen}_{HA} = 0.5260 \cdot g_{-COO-(non-ring),HA} = 0.5260 \cdot 1 \tag{4.20}$$

$$\text{Descriptors}_{HA} = 0.0492 \cdot \ln(d_{p^S,HA}) = 0.0492 \cdot (-6.0518) \tag{4.21}$$

These contributions allow for the computation of the ignition delay τ_{HA} (cf. Table 4.9):

$$\begin{aligned} y_{HA} =& \exp[\text{Carbon}_{HA} + \text{Oxygen}_{HA} + \text{Descriptors}_{HA} + 0.8341] \\ =& \exp[2 \cdot 0.0449 + \ln(5+1) \cdot (-0.2389) \\ &+ 1 \cdot 0.5260 - 6.0518 \cdot 0.0492 + 0.8341] = 2.06 \end{aligned} \tag{4.22}$$

$$\Rightarrow \ln \tau_{HA} = y_{HA} \Rightarrow \tau_{HA} = 7.87 \text{ ms} \tag{4.23}$$

Applying Eqn. (4.2) defined in ASTM D6890 (2011) yields a DCN of 28.4. The measured DCN for hexyl acetate is 32.2.

It is possible to compute a confidence interval for the model prediction $\ln \tau_i = y_i$ based on the FIM. This procedure has been proposed and successfully applied in the context of group contribution modeling just recently (Hukkerikar et al., 2012). While $q_i(\mathbf{u}_i)$ describes the "true" IQT ignition delay for given model inputs $\mathbf{u}_i$, i.e., the set of group occurrences and descriptor values for compound i, the measured delay $\tilde{y}_i$ is assumed to be compromised by normally distributed errors with zero mean $\tilde{\epsilon}_i(\mathbf{u}_i)$, i.e.,

$$\tilde{y}_i(\mathbf{u}_i) = q_i(\mathbf{u}_i) + \tilde{\epsilon}_i(\mathbf{u}_i) \ . \tag{4.24}$$

The model $y_i(\mathbf{u}_i)$ from Table 4.9 represents an approximation of $q_i(\mathbf{u}_i)$, such that $|q_i(\mathbf{u}_i) - y_i(\mathbf{u}_i)| > 0$. The accuracy of this approximation can be quantified with a confidence interval Δy_i computed as (Seber and Wild, 2003; Rasmussen, 2003)

$$\Delta y_i = t^{1-(\alpha/2)}_{\varphi(\tilde{\mathbf{y}})-14} \left(\mathbf{s}_i^{\mathrm{T}}(\mathbf{u}_i, \boldsymbol{\Theta}^*) \cdot \mathbf{V}_{\mathbf{F}}(\boldsymbol{\Theta}^*, \mathbf{u}) \cdot \mathbf{s}_i(\mathbf{u}_i, \boldsymbol{\Theta}^*)\right)^{1/2} \ , \tag{4.25}$$

where $\mathbf{V}_{\mathbf{F}}(\boldsymbol{\Theta}^*, \mathbf{u})$ denotes the covariance matrix constructed from the training data $\tilde{\mathbf{y}}$, $\tilde{\boldsymbol{\sigma}}$ and $\mathbf{u}$ (cf. Eqn. (4.14)). In addition, the sensitivity

$$\mathbf{s}_i^{\mathrm{T}}(\mathbf{u}_i, \boldsymbol{\Theta}^*) = \left[\frac{\partial y_i(\mathbf{u}_i, \boldsymbol{\Theta}^*)}{\partial \Theta_1^*}, .., \frac{\partial y_i(\mathbf{u}_i, \boldsymbol{\Theta}^*)}{\partial \Theta_{14}^*}\right] \tag{4.26}$$

is needed. The equations for $\mathbf{s}_i^{\mathrm{T}}(\mathbf{u}_i, \boldsymbol{\Theta}^*)$ and the numerical values for the covariance matrix can be found in Appendix C.3.

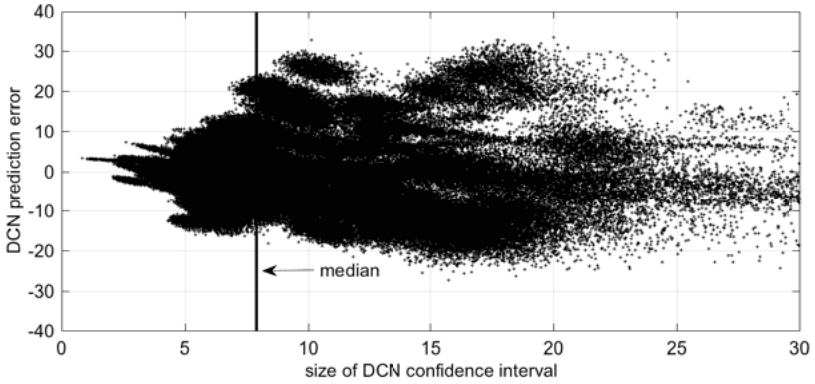


Figure 4.14: IQT ignition delay model: Size of 95% confidence interval vs. DCN prediction error for ten thousand runs of leave-multiple-out cross-validation (LMO-CV). The median size of the confidence intervals is 7.9. The maximum DCN prediction error is 33.4. Note that the x-axis has been cut.

The role of the confidence interval can be illustrated by subtracting the model $y_i(\mathbf{u}_i)$ from both sides of Eqn. (4.24):

$$\underbrace{\tilde{y}_i(\mathbf{u}_i) - y_i(\mathbf{u}_i)}_{\text{residual}} = \underbrace{[q_i(\mathbf{u}_i) - y_i(\mathbf{u}_i)]}_{\text{quantified by the confidence interval}} + \tilde{\epsilon}_i(\mathbf{u}_i) \tag{4.27}$$

The confidence interval does not account for the error term $\tilde{\epsilon}_i(\mathbf{u}_i)$ acting on a measurement $\tilde{y}_i(\mathbf{u}_i)$ (Carney et al., 1999; Rasmussen, 2003). Furthermore, the zero mean assumption for the errors $\tilde{\epsilon}$ might be violated due to impurities, especially peroxides. Ten thousand runs of LMO-CV were performed to check, whether the size of the confidence interval is related to the DCN prediction error, despite these limitations. As can be seen from Figure 4.14, the median size of the confidence intervals is 7.9, i.e., 50% of the intervals are smaller than 7.9. The left part of Figure 4.14 suggests, that for these intervals there is a relationship between interval size and DCN prediction error. For intervals larger than the median, however, no apparent relationship can be found in Figure 4.14. The actual probability of a measurement $\tilde{y}_i(\mathbf{u}_i)$ falling into the computed 95% confidence interval $y_i(\mathbf{u}_i) \pm \Delta y_i(\mathbf{u}_i)$ is approximately 40% in LMO-CV. If these intervals are extended by ± 5 in the DCN space, the hit ratio becomes about three out of four. If the intervals are extended by ± 10, less than one out of ten measured DCNs is found outside the provided range.

Returning to the example of hexyl acetate (HA), the 95% confidence interval can be computed based on the sensitivity equations $\mathbf{s}_i^{\mathrm{T}}$ and the covariance matrix $\mathbf{V_F}\left(\boldsymbol{\Theta}^*, \mathbf{u}\right)$:

$$\begin{aligned}
\mathbf{s}_{HA}^{\mathrm{T}} =& [y_{HA}, 0, 0, \ln(d_{p^S,HA}) \cdot y_{HA}, 0, g_{-CH3,HA} \cdot y_{HA}, \\
& \ln(g_{-CH2-(non-ring),HA} + 1) \cdot y_{HA}, 0, 0, 0, 0, 0, 0, \\
& g_{-COO-(non-ring),HA} \cdot y_{HA}] \\
=& [2.06, 0, 0, -6.0518 \cdot 2.06, 0, 2 \cdot 2.06, \\
& \ln(5+1) \cdot 2.06, 0, 0, 0, 0, 0, 0, 1 \cdot 2.06] \\
=& [2.06, 0, 0, -12.48, 0, 4.13, 3.70, 0, 0, 0, 0, 0, 0, 0, 2.06] \qquad (4.28)
\end{aligned}$$

$$\Delta y_{HA} = t_{144-14}^{1-(\alpha/2)} \left(\mathbf{s}_{HA}^{\mathrm{T}} \mathbf{V_F}\left(\boldsymbol{\Theta}^*, \mathbf{u}\right) \mathbf{s}_{HA}\right)^{1/2} \qquad (4.29)$$

for α=0.05 one obtains $t_{144-14}^{1-(\alpha/2)} = 1.9784$

$$\Leftrightarrow \Delta y_{HA} = 1.9784 \cdot \left(\mathbf{s}_{HA}^{\mathrm{T}} \mathbf{V_F}\left(\boldsymbol{\Theta}^*, \mathbf{u}\right) \mathbf{s}_{HA}\right)^{1/2} = 0.12 \qquad (4.30)$$

$$\Rightarrow y_{HA} \pm \Delta y_{HA} = 2.06 \pm 0.12 \qquad (4.31)$$

$$\Rightarrow 7.00 \leq \tau_{HA}\ [\mathrm{ms}] \leq 8.84 \qquad (4.32)$$

$$\Rightarrow 26.2 \leq DCN \leq 30.9 \qquad (4.33)$$

As a second example, ϵ-decalactone (EDL) shall illustrate the computation for a cyclic species. The following information is derived from the molecular structure depicted in Figure 4.13 (center): 3 × −CH2− (non-ring), 1 × −CH3, 1 × >C=O (ring), 1 × −O− (ring), T_{crit} = 767 K, p_{crit} = 27.1 bar, T_{boil} = 547 K, $\ln(p^S)$ = -10.5823 ln(bar). Note that the four −CH2− (ring) groups in ϵ-decalactone (EDL) are of no relevance since this group is not a model input. The computation of τ based on the equations given in Table 4.9 yields

$$\begin{aligned}
y_{EDL} =& \exp[1 \cdot 0.0449 + \ln(3+1) \cdot (-0.2389) \qquad (4.34) \\
& + 1 \cdot 0.4241 + 1 \cdot 0.0426 - 10.5823 \cdot 0.0492 + 0.8341] \\
\Rightarrow y_{EDL} =& 1.64 = \ln \tau_{EDL} \Rightarrow \tau_{EDL} = 5.15\ \mathrm{ms}. \qquad (4.35)
\end{aligned}$$

Applying Eqn. (4.1) defined in ASTM D6890 (2011) returns a DCN of 40.7. Utilizing the sensitivity equations and the covariance matrix from Appendix C.3, the 95% confidence interval for the DCN prediction is computed as

$$y_{EDL} \pm \Delta y_{EDL} = 1.64 \pm 0.17 \qquad (4.36)$$

$$\Rightarrow 4.33 \leq \tau_{EDL}\ [\mathrm{ms}] \leq 6.12 \qquad (4.37)$$

$$\Rightarrow 35.0 \leq DCN \leq 47.5\ . \qquad (4.38)$$

The measured DCN for ϵ-decalactone (EDL) is 40.7.

Finally, anisole (A) shall serve as an example for an aromatic species. The following information on anisole (A) can be retrieved from Figure 4.13 (right): 1 × −CH3, 1 × −O− (non-ring), nAB = 6, nCCDB = 3, T_{crit} = 620 K, p_{crit} = 40.3 bar, T_{boil} = 409 K, ln(p^S) = -4.4172 ln(bar). The computation of the IQT ignition delay yields

$$\begin{aligned} y_A =& \exp[1 \cdot 0.0449 + 1 \cdot (-0.2123) + 3 \cdot 0.1404 \\ & - 4.4172 \cdot 0.0492 + 6 \cdot 0.0639 + 0.8341] \end{aligned} \tag{4.39}$$

$$\Rightarrow y_A = 3.50 = \ln \tau_A \Rightarrow \tau_A = 33.26 \text{ ms}. \tag{4.40}$$

Applying Eqn. (4.2) defined in ASTM D6890 (2011) returns a DCN of 12.2. The 95% confidence intervals are

$$y_A \pm \Delta y_A = 3.50 \pm 0.69 \tag{4.41}$$

$$\Rightarrow 16.71 \leq \tau_A \text{ [ms]} \leq 66.18 \tag{4.42}$$

$$\Rightarrow 9.0 \leq DCN \leq 17.6 \text{ .} \tag{4.43}$$

The measured DCN for anisole is <7 (no ignition).

4.4 Conclusions and outlook

To overcome the significant limitations of existing ON, CN and DCN databases and models, a high-quality collection of ignition delay data from a single, well-defined experimental set-up has been put together. Half of this collection consists of pure-component DCN data taken from the literature. The other half has resulted from a rapid fuel screening campaign that has been organized by the author of this thesis and that has been executed by ASG. The compounds were chosen to resemble the structural features of potential products from lignocellulosic biomass, specifically focussing on oxygenated species.

On the basis of this data set, a simple, yet predictive group contribution model has been derived to estimate the IQT ignition delay and DCN. This is the first auto-ignition model in the publicly available literature that can be applied to a wide range of bio-derived compounds such as acyclic and cyclic, branched and straight, saturated and unsaturated hydrocarbons as well as alcohols, ethers, esters, ketones, aldehydes, and aromatic and polyfunctional compounds. The calculation procedure does neither require involved computational tools nor any information that could not be derived from the two-dimensional molecular graph. The model prediction yields a first estimate of the auto-ignition propensity of a fuel candidate, if a sufficiently large and pure sample is not yet available.

In the view of the complexity of low-temperature combustion chemistry and the overlapping mixture formation process, the accuracy of prediction provided by the simple model is quite remarkable. Still, the study suggests, that the amount of available measurement data constrains model complexity, since the inclusion of additional structural groups results in significant parametric uncertainty. Large functional groups, stereochemistry or group interaction effects might alter the ignition delay, but had to be neglected in the present model. Thus, it will be important to increase the size of the database in the future. Ideally, DCN data with unknown fuel sample peroxide status should be replaced with remeasured data.

5 QSPR-based prediction of physicochemical fuel properties

This Chapter deals with the development of predictive quantitative structure-property relationship (QSPR) models for the physicochemical fuel properties listed in Table 2.1. It represents a continuation and a more detailed elaboration of a QSPR modeling strategy that has been published in *SAE International Journal of Fuels & Lubricants* (Dahmen et al., 2012). The original concept as outlined by Dahmen et al. (2012) has also been explained by Hechinger (2014). Later, modifications and improvements to the QSPR modeling concept have been made by the author of this thesis with respect to data preparation, outlier removal, data set split, validation and the applicability domain concept. The revised approach is outlined in this Chapter. Its basic principles have already been briefly described in a publication in *Energy & Fuels* (Dahmen and Marquardt, 2016). Statistical measures for the critical data QSPR models have also been published previously in *Energy & Fuels* (Dahmen and Marquardt, 2017). Although being based on the papers of Dahmen et al. (2012) and Dahmen and Marquardt (2016, 2017), the present Chapter has been almost entirely rewritten to provide a more rigorous theoretical description of the QSPR modeling approach. Furthermore, the QSPR models are compared to well-established group contribution models taken from the publicly available literature (Joback and Reid, 1987; Hukkerikar et al., 2012; Hukkerikar, 2013). The processes of extracting property data from the DIPPR database (AIChE, 2012) and of computing molecular descriptors by means of DragonX (Todeschini et al., 2009) have evolved from the ideas of Manuel Hechinger (Hechinger et al., 2010; Hechinger, 2014).

5.1 Modeling strategy

The general workflow followed throughout this Section is depicted in Figure 5.1. Two projection techniques, i.e., principal component analysis (PCA) (Jolliffe, 2002) and partial least squares (PLS) regression (Höskuldsson, 1988), constitute the cornerstones of the approach. The modeling strategy presented here has been inspired by the works of Eriksson and co-workers on the role of PCA/PLS in the derivation of statistically sound quantitative

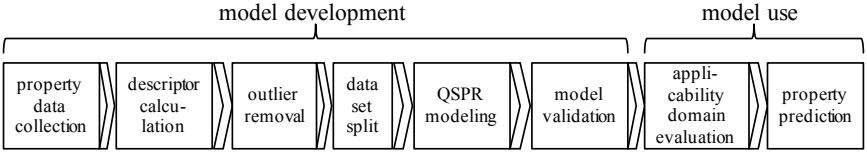


Figure 5.1: Workflow for the development and use of QSPR models followed in this Chapter.

structure-activity relationship (QSAR) models (Wikström et al., 1998; Wold et al., 2001; Eriksson et al., 2003, 2006a,b) and by the works of Golbraikh, Tropsha and co-workers on best practices for the derivation of truly predictive QSARs (Tropsha et al., 2003; Tropsha and Golbraikh, 2007, 2010; Tropsha, 2010; Cherkasov et al., 2014).

Thermophysical property data for model development have been collected from the Design Institute for Physical Properties (DIPPR) 801 database (AIChE, 2012), i.e., a widely used collection of pure-component property data for a large number of chemical structures. Queried data have been restricted to compounds containing carbon, oxygen and hydrogen atoms only. Since molecular graphs with less than three carbon atoms are not expected to represent novel fuel candidates, such small compounds have been excluded from the data sets. The DIPPR database is an *evaluated* database, making it particularly useful for the development of GCMs and QSPRs (Constantinou and Gani, 1994; Knotts et al., 2001; Ericksen et al., 2002; Hechinger et al., 2010; Hechinger, 2014). As a result of the evaluation process, DIPPR assigns the attribute "accepted" to property data which is considered reliable. Moreover, an accuracy level, i.e., an upper bound on the expected relative error, is provided for each accepted value.

Table 5.1 shows the number of structures retrieved from the DIPPR database in case of each thermophysical fuel property mentioned in Table 2.1. Critical properties T_{crit}, p_{crit} and V_{crit} have also been included. Note that structure and corresponding property data have been extracted from the DIPPR database and imported into Matlab (The MathWorks Inc., 2016) by means of a software tool developed by Manuel Hechinger and Luisa Schulze Langenhorst (Hechinger and Schulze Langenhorst, 2012). For the purpose of modeling, only "accepted" data with an uncertainty of $\leq$10% have been considered. In the following, $\tilde{\mathbf{y}} \in \mathbb{R}^{N_c}$ denotes the vector of known property data for a given property, where N_c refers to the number of compounds which is identical to the number of property values.

The distribution of the property data $\tilde{\mathbf{y}}$ has been examined manually in case of each property. For instance, the histogram of the viscosity data $\tilde{\mathbf{y}}(\mu)$ in Figure 5.2 reveals four compounds with exceptionally high viscosities. Note that the log-transformation (Eriksson et al., 2006b) has been applied to obtain a distribution, which is closer to normality, as

Table 5.1: Number of $C_{x>2}H_yO_z$ structures retrieved from the DIPPR 801 database (AIChE, 2012) with $\leq$10% relative measurement error in the property data; "accepted" data only.

thermophysical property	unit	no. of compounds	reference state
boiling point T_{boil}	K	977	1 atm
melting point T_{melt}	K	954	1 atm
lower heating value LHV	J/kmol	971	298 K
liquid molar density ρ_m	kmol/m^3	707	298 K
enthalpy of vaporization H_{vap}	J/kmol	487	298 K
surface tension σ	N/m	574	298 K
dynamic viscosity μ	Pa·s	473	298 K
critical temperature T_{crit}	K	1007	–
critical pressure p_{crit}	Pa	963	–
critical volume V_{crit}	m^3/kmol	387	–

the unscaled viscosity data spans multiple orders of magnitude. The four high viscous compounds have been removed from the μ data set, as they can hardly comply with the basic assumption behind QSPR modeling, i.e., the existence of similar compounds having similar properties (Leonard and Roy, 2006; Tropsha and Golbraikh, 2010). Manual inspection of the other $\tilde{\mathbf{y}}$ has prompted the exclusion of four compounds with isolated property values from the LHV data set, one compound from the T_{melt} data set, three compounds from the H_{vap} data set and seven compounds from the p_{crit} data set. The list of excluded species can be found in Appendix D.

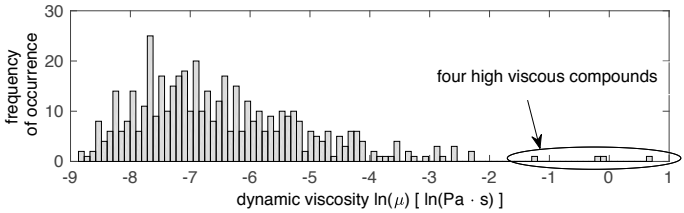


Figure 5.2: The histogram of the natural logarithm of the dynamic viscosity property data $\tilde{\mathbf{y}}(\mu)$ reveals four property outliers.

5.1.1 Descriptor calculation, data pretreatment and outlier removal

Nine-hundred zero- to two-dimensional molecular descriptors have been computed for each molecular structure with the commercial descriptor calculation software DragonX (blocks 1 to 10, 17 and 18 in DragonX version 1.4.4; cf. Todeschini et al. (2009)) and pretreatment

Table 5.2: Number of compounds N_c, number of descriptors N_d, ratio of compounds to descriptors N_c/N_d and numerical rank of pretreated descriptor matrix $\mathbf{X} \in \mathbb{R}^{N_c \times N_d}$. The numerical rank has been computed with Matlab's (The MathWorks Inc., 2016) *rank* function which is based on the singular value decomposition.

	N_c	N_d	N_c/N_d	rank of $\mathbf{X}$
T_{boil}	977	657	1.49	568
T_{melt}	953	657	1.45	567
LHV	967	658	1.47	570
ρ_m	707	655	1.08	555
H_{vap}	484	649	0.75	472
σ	573	654	0.88	550
μ	469	647	0.72	458
T_{crit}	1007	658	1.53	571
p_{crit}	963	658	1.46	568
V_{crit}	387	647	0.60	381

of the resulting descriptor matrix $\mathbf{X} \in \mathbb{R}^{N_c \times 900}$ has been performed as follows: First, $\mathbf{X}$ has been pruned by removing columns with non-varying descriptor data. Subsequently, the reduced matrix has been inspected for missing values, as certain molecular descriptors cannot be evaluated for arbitrary compounds. If the calculation of descriptor l had failed in one or multiple instances, the entire column $\mathbf{x}_l$ has been removed from $\mathbf{X}$. Since the results of PCA and PLS analysis depend on the scaling of the descriptor data (Wold et al., 2001), $\mathbf{X}$ has been autoscaled column-wise to unit-variance and zero mean according to (van den Berg et al., 2006; Tropsha and Golbraikh, 2010)

$$\overline{x}_l = \frac{1}{N_c} \sum_{i=1}^{N_c} x_{l,i} \ , \tag{5.1}$$

$$std(\mathbf{x}_l) = \left(\frac{1}{N_c - 1} \sum_{i=1}^{N_c} \left(x_{l,i} - \overline{x}_l \right)^2 \right)^{\frac{1}{2}} \ , \tag{5.2}$$

$$\mathbf{x}_l \leftarrow \frac{\mathbf{x}_l - \overline{x}_l}{std(\mathbf{x}_l)} \ , \tag{5.3}$$

where the "$\leftarrow$" denotes that the scaling is performed on $\mathbf{x}_l \in \mathbb{R}^{N_c}$, i.e., the column-vector of descriptor data for descriptor l. To this end, the mean $\overline{x}_l$ and the standard deviation $std(\mathbf{x}_l)$ of the unscaled vector $\mathbf{x}_l$ are calculated first. Autoscaling renders all descriptors equally important since *a priori* there is a lack of knowledge on the relative importance of the different descriptors (Wold et al., 2001).

Table 5.2 shows the characteristics of $\mathbf{X} \in \mathbb{R}^{N_c \times N_d}$, i.e., the pretreated descriptor matrix,

where N_c is the number of compounds and N_d is the number of descriptors. The ratio N_c/N_d is highly unfavorable for variable selection strategies (cf. discussion in Section 2.4.2). Moreover, the low numerical rank is an indicator for the high degree of collinearity in $\mathbf{X}$. To deal with both the unfavorable N_c/N_d ratio and the interrelated nature of the descriptor data, PCA and PLS regression have been employed for further analysis and model construction. PCA can be viewed as a dimension reduction technique as it projects $\mathbf{X}$ down onto few orthogonal latent variables, i.e., the so-called principal components, which are chosen to retain most of the variation present in the original data (Jolliffe, 2002). The central concept behind PCA is the decomposition of $\mathbf{X}$ into a product of two matrices, i.e.,

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} \,, \tag{5.4}$$

where $\mathbf{T}$ is the so-called $N_c \times N_d$ *score* matrix and $\mathbf{P}^{\mathrm{T}}$ denotes the so-called $N_d \times N_d$ *loading* matrix (Wold et al., 1987; Eriksson et al., 2006a). The columns of the score matrix $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_{N_d}]$ are the principal components (PCs), which are orthogonal to each other. $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_{N_d}]$ is an orthogonal matrix, i.e., $\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{P}\mathbf{P}^{\mathrm{T}} = \mathbf{I}$ and $\mathbf{P}^{\mathrm{T}} = \mathbf{P}^{-1}$. In this thesis, the PCs have been computed with Matlab's (The MathWorks Inc., 2016) *princomp* function which is based on the singular value decomposition of $\mathbf{X}$, i.e.,

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} \,, \tag{5.5}$$

where $\mathbf{U}$ and $\mathbf{V}$ denote the matrices of the left and right singular vectors and $\mathbf{S}$ is a diagonal matrix holding the singular values (Wold et al., 1987; Abdi, 2010; Bro and Smilde, 2014).

To achieve the dimension reduction, one typically considers only N_{pc} principal components, i.e., the matrices $\mathbf{T}$ and $\mathbf{P}$ are truncated such that $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_{N_{pc}}] \in \mathbb{R}^{N_c \times N_{pc}}$ and $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_{N_{pc}}] \in \mathbb{R}^{N_d \times N_{pc}}$ approximate the descriptor matrix $\mathbf{X}$. This can be written as

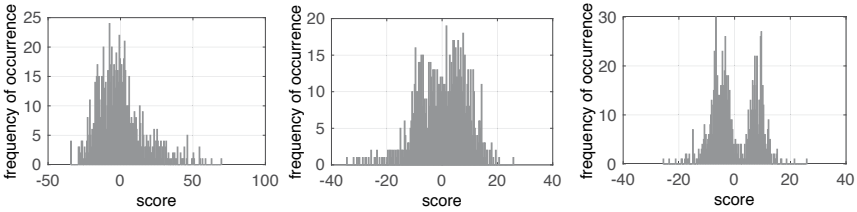
$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \,, \tag{5.6}$$

where $\mathbf{E}$ denotes the residual matrix (Wold et al., 1987; Eriksson et al., 2006a). In the context of QSPR/QSAR modeling, the first few principal components of the descriptor data set $\mathbf{X}$ are sometimes called *principal properties*, because it is assumed that they summarize the major structural and chemical properties of the compounds under consideration (Eriksson et al., 2000, 2006a). A graphical analysis of the first two or three PCs therefore reveals relationships among the compounds with regard to the presence of clusters or strong outliers, i.e., compounds with a special character (Wold et al., 1987; Eriksson et al., 2006a; Bro and Smilde, 2014).

It is common practice to report the cumulative variation $var^{cum}(N_{pc})$ in the original data $\mathbf{X}$, which can be explained by the first N_{pc} principal components (Jolliffe, 2002):

$$var^{cum}(N_{pc}) = 100 \left(\sum_{l=1}^{N_{pc}} (std(\mathbf{t}_l))^2 \right) / \left(\sum_{l=1}^{N_d} (std(\mathbf{t}_l))^2 \right) \quad [\%] \tag{5.7}$$

In Eqn. (5.7), $std(\mathbf{t}_l)^2$ denotes the variance, i.e., the square of the standard deviation (cf. Eqn. (5.2)), of the l-th principal component. For similarity analysis by means of principal properties, in the present thesis, N_{pc} has been chosen as the minimum number of PCs required to explain at least 60% of the variation in $\mathbf{X}$ as defined in Eqn. (5.7). This follows earlier QSAR studies (Eriksson et al., 2000; Wold et al., 2001; Eriksson et al., 2006a,b), where threshold values between 50% and 70% had been employed to detect clusters and outliers, typically resulting in a two- or three-dimensional analysis.



(a) 1st PC (39%) **(b)** 2nd PC (12%) **(c)** 3rd PC (9%)

Figure 5.3: Histograms of the first three principal components (PCs) computed for the autoscaled heating value descriptor matrix $\mathbf{X}(LHV)$. The variation in $\mathbf{X}(LHV)$ that is explained by each PC is given in parentheses ($var^{cum}(3) = 60.8\%$).

As can be seen from Figure 5.3, three PCs are indeed sufficient to explain 60.8% of the variation in the autoscaled heating value descriptor data set $\mathbf{X}(LHV)$. Moreover, the PCs are approximately normally distributed, except for the third PC which discriminates the oxygenated and the non-oxygenated hydrocarbons. With approximately normally distributed data, Hotelling's T^2 diagnostic can be applied to detect strong outliers based on their distance to the center of the PC data (Eriksson et al., 2006a,b; Bro and Smilde, 2014). To this end, T_i^2 for compound i is calculated based on the scores $[\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_{N_{pc}}]$ as (Eriksson et al., 2006b)

$$T_i^2 = \sum_{l=1}^{N_{pc}} \frac{t_{l,i}^2}{std(\mathbf{t}_l)^2} \ . \tag{5.8}$$

Table 5.3: Modeling data sets $\mathbf{X}$ obtained **after pretreatment and outlier removal**. Data for LHV, ρ_m, T_{boil}, T_{melt}, H_{vap}, σ and μ have already been published in *Energy & Fuels* (Dahmen and Marquardt, 2016).

	LHV	ρ_m	T_{boil}	T_{melt}	H_{vap}	σ	μ	T_{crit}	p_{crit}	V_{crit}
no. of compounds										
... retrieved from DIPPR database	971	707	977	954	487	573	473	1007	963	387
... w/o property outliers	967	707	977	953	484	573	469	1007	956	387
... w/o PCA outliers	932	676	942	914	480	558	459	967	928	380
no. of PCs needed to describe 60% of variation in $\mathbf{X}$	3	3	3	3	3	3	3	3	3	3
share of training compounds above 95% Hotelling's T^2 threshold in PCA model [%]	6.4	4.6	6.2	6.0	5.2	4.7	6.1	6.3	5.9	5.5

Compound i is considered to lie outside the confidence region of the PCA model, if T_i^2 exceeds a certain threshold value

$$T^2{}_{thre} = \frac{N_{pc}(N_c^2 - 1)}{N_c(N_c - N_{pc})} \cdot F^{\alpha}(N_{pc}, N_c - N_{pc}) \ , \tag{5.9}$$

where $F^{\alpha}(N_{pc}, N_c - N_{pc})$ denotes the F-distribution with N_{pc} and $N_c - N_{pc}$ degrees of freedom and confidence level α (Wikström et al., 1998; Chiang et al., 2000). In a two-dimensional PC space, the confidence region is an ellipse with the origin in the center of the data (Eriksson et al., 2006b).

In this thesis, the T^2 diagnostic constitutes the central part of an iterative algorithm for automatic outlier detection and removal depicted in Figure 5.4. Following data pretreatment, a first PCA model is created and analyzed by means of Hotelling's T^2 statistic. If one or multiple compounds are located outside the 99.9% confidence region, the compound i exhibiting the highest T_i^2 value is removed and a new PCA model is built. This procedure is applied repeatedly until all outliers by this definition are excluded, thus yielding the final modeling data sets $\mathbf{X}$ / $\tilde{\mathbf{y}}$. As can be seen from Table 5.3, only reasonable shares of compounds are located outside the 95% confidence regions after outlier exclusion has been

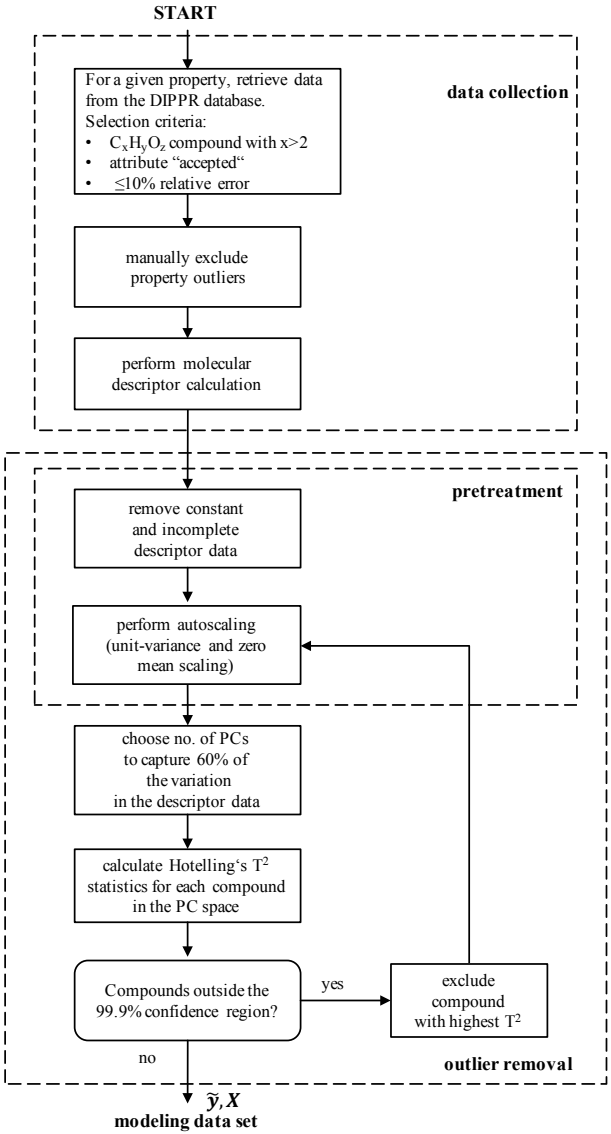
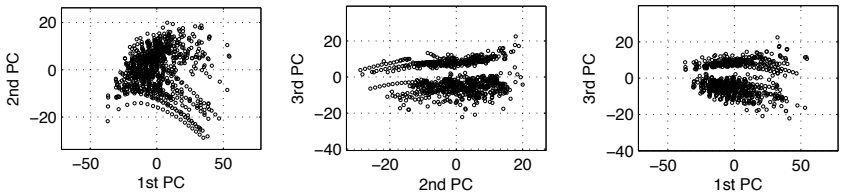


Figure 5.4: Workflow for data collection, data pretreatment and outlier removal yielding thermophysical property data $\tilde{\mathbf{y}}$ and corresponding descriptor matrices $\mathbf{X}$ for subsequent QSPR modeling.

performed. As a consequence of outlier removal, carbon, hydrogen and oxygen ranges have been significantly reduced, e.g., from $C_{3-36}H_{2-74}O_{0-8}$ down to $C_{3-26}H_{2-42}O_{0-7}$ in case of the normal boiling point data set. The list of excluded compounds contains exclusively very large species (cf. Appendix D). Since fuel candidates are expected to exhibit much lower molecular weights, it is considered unlikely that outlier exclusion reduces the fuel-relevant applicability range of a QSPR model to be derived from data sets $\mathbf{X}$ / $\tilde{\mathbf{y}}$.



(a) 1st PC vs. 2nd PC **(b)** 2nd PC vs. 3rd PC **(c)** 1st PC vs. 3rd PC

Figure 5.5: Score plots for the first three principal components (PCs) of $\mathbf{X}(LHV)$ ($var^{cum}(3) = 60.7\%$) after outlier exclusion.

As can be seen from the score plots in Figure 5.5, the chemical compounds in the LHV data set are well distributed in the PC space after outlier exclusion has been performed, and there are no isolated compounds far away from the center of the data. Similar conclusions can be drawn for the other thermophysical properties. This suggests that $\mathbf{X}$ and $\tilde{\mathbf{y}}$ resulting from the workflow depicted in Figure 5.4 constitute a sound foundation for the derivation of QSPR models by means of PLS.

5.1.2 Model derivation, validation and application

Unlike (bio-)chemical activities, which are thought to be primarily determined by the presence of a specific fragment in the molecular structure, physical properties of molecules are believed to be global properties, i.e., the idea is that every atom in a molecule somehow contributes to the measured property (Netzeva et al., 2005). Following this line of thought, the principal properties, i.e., the first few principal components, have been utilized for data partitioning into training and test sets. To this end, about 15% of the rows from $\mathbf{X}$ / $\tilde{\mathbf{y}}$ have been moved into new test data sets $\mathbf{X}^{test}$ / $\tilde{\mathbf{y}}^{test}$. The test sets have been set aside in order to perform external validation (Roy, 2007; Tropsha, 2010) at a later stage.

Golbraikh and Tropsha (2002b), Tropsha et al. (2003), Golbraikh et al. (2003) and Leonard and Roy (2006) have suggested that a rational selection of training and test sets should be based on the proximity of the test set compounds to the training set compounds (and vice versa) in the space of original descriptors or principal components. Furthermore,

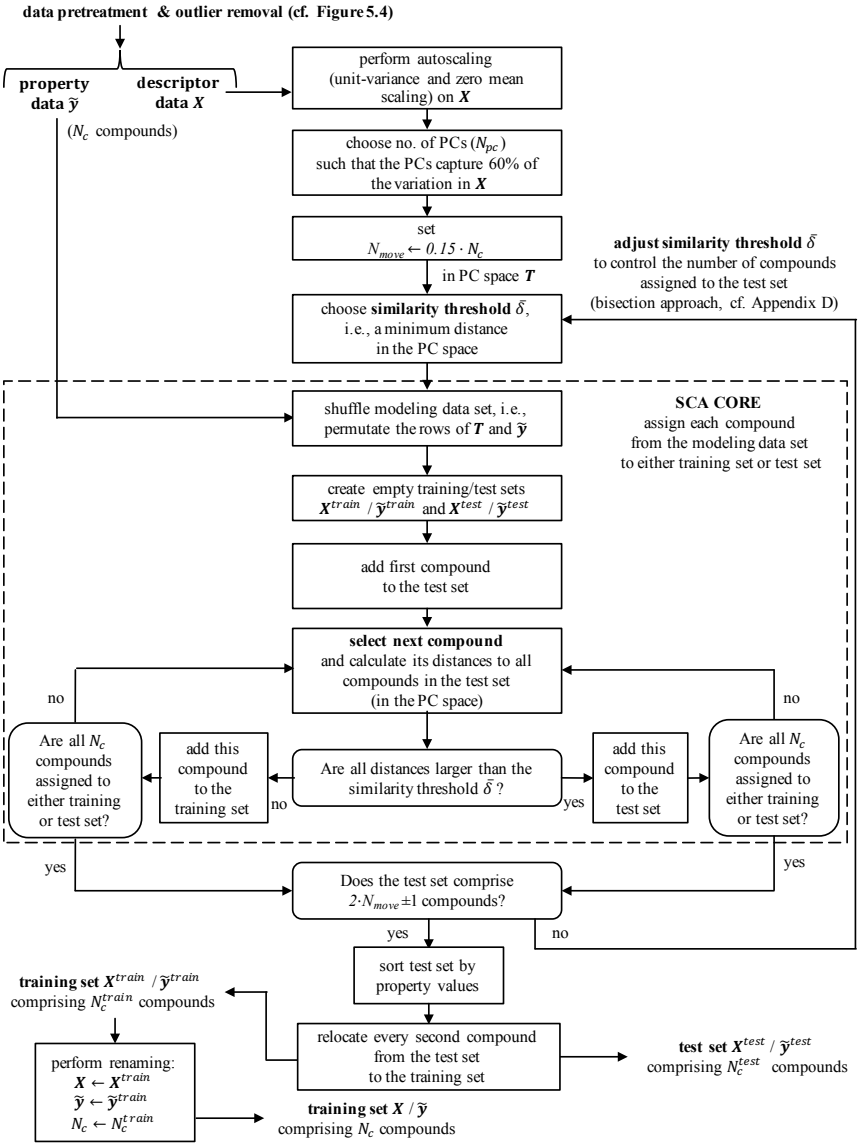


Figure 5.6: Workflow of the modified SCA algorithm used to relocate 15% of the rows from $\mathbf{X}$ / $\tilde{\mathbf{y}}$ into test sets $\mathbf{X}^{test}$ / $\tilde{\mathbf{y}}^{test}$ (N_c^{test} compounds).

the training data should be distributed within the entire PC space to ensure that the validity range of the QSPR model is maximized (Golbraikh and Tropsha, 2002b).

In this thesis, a variant of stochastic cluster analysis (SCA) (Reynolds et al., 1998; Tropsha and Golbraikh, 2010) has been applied to the selection of test set compounds i and j such that the Euclidean distance $\delta_{i,j}$ between i and j in the PC space is larger than a certain similarity threshold $\overline{\delta}$, i.e.,

$$\delta_{i,j} = \sqrt{\sum_{l=1}^{N_{pc}} \left(t_{l,i} - t_{l,j}\right)^2} \stackrel{!}{>} \overline{\delta} \; . \tag{5.10}$$

The workflow of the modified SCA algorithm is depicted in Figure 5.6. The inner core has been taken from Tropsha and Golbraikh (2010), whereas the similarity threshold $\overline{\delta}$ is controlled slightly different with the help of a bisection approach (cf. Appendix D). This way, $\overline{\delta}$ is iteratively adjusted such that 15% of the modeling data become a diverse test set. In the final step of the algorithm, the preliminary test set (comprising about 30% of the modeling data) is sorted by the property values and every second compound is relocated to the training set. This step resembles the concept of activity rank-based selection in QSAR (Golbraikh and Tropsha, 2002b; Leonard and Roy, 2006) and aims at avoiding that the property range of the test set extends considerably beyond the property range of the training set. As indicated in Figure 5.7, the modified SCA algorithm returns descriptor data sets $\mathbf{X}$ (subsequently used for QSPR model building) and $\mathbf{X}^{test}$ (subsequently used to assess the predictive power of the QSPR model by means of external validation) which occupy roughly the same molecular domain, i.e., the same three-dimensional volume in the PC space.

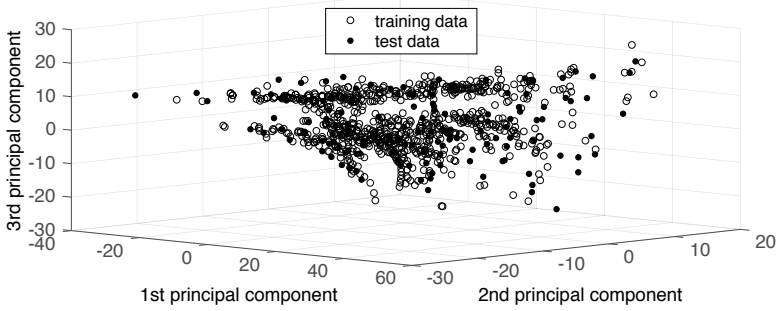


Figure 5.7: Exemplary distribution of training data and test data in the PC space of $\mathbf{X}$ as generated by the modified SCA algorithm (cf. Figure 5.6).

The structure-property relationships, i.e., the actual QSPR models, have been determined by PLS regression following the concepts presented by Eriksson et al. (2006a,b). In contrast to multiple linear regression (MLR), but similar to PCA, the projection technique PLS yields robust results, even if the ratio of the number of training compounds to the number of descriptors, i.e., N_c/N_d, is small and/or descriptor data $\mathbf{X}$ suffer from a high degree of collinearity (Cramer, 1993; Clark and Cramer, 1993; Wold et al., 2001). Like PCA, PLS regression is sensitive to scaling (Wold et al., 2001; Eriksson et al., 2006b). Thus, auto-scaling (cf. Eqns. (5.1)-(5.3)) is applied to all column-vectors $\mathbf{x}_l$ in $\mathbf{X}$, i.e.,

$$\mathbf{x}_l \leftarrow \frac{\mathbf{x}_l - \overline{x}_l}{std(\mathbf{x}_l)} \,, \tag{5.11}$$

and to the vector of property data $\tilde{\mathbf{y}}$, i.e.,

$$\tilde{\mathbf{y}} \leftarrow \frac{\tilde{\mathbf{y}} - \overline{\tilde{y}}}{std(\tilde{\mathbf{y}})} \,, \tag{5.12}$$

yielding the $N_c \times N_d$ matrix of scaled descriptor data $\mathbf{X}$ and the column-vector of scaled property data $\tilde{\mathbf{y}}$ of length N_c. Then, PLS regression proceeds by extracting few, orthogonal latent variables from $\mathbf{X}$ such that these are predictors of $\tilde{\mathbf{y}}$, but also model $\mathbf{X}$. This can be written as follows (Wold et al., 2001):

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{5.13}$$

$$\mathbf{T} = \mathbf{X}\mathbf{W} \tag{5.14}$$

$$\tilde{\mathbf{y}} = \mathbf{T}\mathbf{c}^{\mathrm{T}} + \mathbf{r} = \mathbf{X}\boldsymbol{\beta} + \mathbf{r} \tag{5.15}$$

$$\boldsymbol{\beta} = \mathbf{W}\mathbf{c}^{\mathrm{T}} \tag{5.16}$$

Eqn. (5.13) shows the decomposition of the descriptor data matrix into a score matrix $\mathbf{T} \in \mathbb{R}^{N_c \times N_{pls}}$, i.e., the matrix of N_{pls} orthogonal latent variables, and a loading matrix $\mathbf{P}^{\mathrm{T}} \in \mathbb{R}^{N_{pls} \times N_d}$, which is unlike in PCA non-orthogonal. $\mathbf{E}$ denotes the matrix of residuals. As can be seen from Eqn. (5.14), the score matrix $\mathbf{T}$ is the product of the original descriptor data and a *weight* matrix $\mathbf{W} \in \mathbb{R}^{N_d \times N_{pls}}$. This means that the latent variables are linear combinations of all N_d original descriptors. Eqn. (5.15) refers to the relationship between the property data vector and $\mathbf{T}$, where $\mathbf{c}^{\mathrm{T}} \in \mathbb{R}^{N_{pls}}$ is a column-vector of weights and $\mathbf{r}$ denotes the vector of residuals. Finally, the column-vector of regression coefficients $\boldsymbol{\beta} \in \mathbb{R}^{N_d}$ in Eqn. (5.15) is calculated as the product of the weight matrix and the weight vector (cf. Eqn. (5.16)). In this thesis, Matlab's (The MathWorks Inc., 2016) *plsregress* function which is based on the SIMPLS algorithm (de Jong, 1993) has been used to perform the PLS regression step.

The optimal number of PLS components N_{pls} has been selected based on 7-fold cross-validation (Cramer et al., 1988; Clark and Cramer, 1993; Wold et al., 2001; Eriksson et al., 2006b). To this end, the root-mean-square error (RMSE) for the model fit has been compared to the cross-validation RMSE as shown for the normal boiling point QSPR in Figure 5.8. Here, both fit RMSE and cross-validation RMSE decrease similarly in case of the first few PLS components. However, the addition of more than about seven components does not reduce the cross-validation RMSE any further, whereas the fit to the training data continues to improve. This is a clear sign of the onset of over-fitting (Geladi and Kowalski, 1986; Eriksson et al., 2006b). Hence, the optimal number of PLS components for the T_{boil}-QSPR has been set to seven.

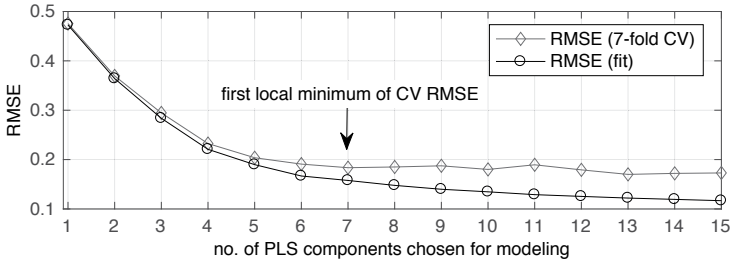


Figure 5.8: Selection of the optimal number of PLS components for the T_{boil}-QSPR model is based on a comparison of model fit RMSE and 7-fold cross-validation (CV) RMSE.

Obviously, descriptor data of new compounds need to be scaled in the same way the training data set $\mathbf{X}$ had been autoscaled. This shall be exemplified by considering the test set descriptor matrix $\mathbf{X}^{test}$. To this end, each column $\mathbf{x}_l^{test}$ of $\mathbf{X}^{test}$ needs to be scaled according to

$$\mathbf{x}_l^{test} \leftarrow \frac{\mathbf{x}_l^{test} - \overline{x}_l}{std(\mathbf{x}_l)} \ , \tag{5.17}$$

where the mean $\overline{x}_l$ and the standard deviation $std(\mathbf{x}_l)$ refer to the unscaled variant of $\mathbf{X}$, i.e., the descriptor matrix for the training set. Then, the vector of predicted property values $\mathbf{y}^{test}$ of the test set compounds can be calculated from (cf. Eqn. (5.15))

$$\mathbf{y}^{test} = std(\tilde{\mathbf{y}}) \cdot \left(\mathbf{X}^{test}\boldsymbol{\beta}\right) + \overline{\tilde{y}} \ , \tag{5.18}$$

where the mean $\overline{\tilde{y}}$ and the standard deviation $std(\tilde{\mathbf{y}})$ refer to the unscaled variant of $\tilde{\boldsymbol{y}}$, i.e., the property vector for the training set. The test sets are used to confirm the predictive

power of the models. An exemplary parity plot of $\mathbf{y}^{test}$ (predicted data) vs. $\tilde{\mathbf{y}}^{test}$ (data taken from DIPPR database) is shown for the T_{boil}-QSPR model in Figure 5.9. Average relative errors (AREs) and maximum relative errors (MREs) for the individual models will be given in the subsequent Section.

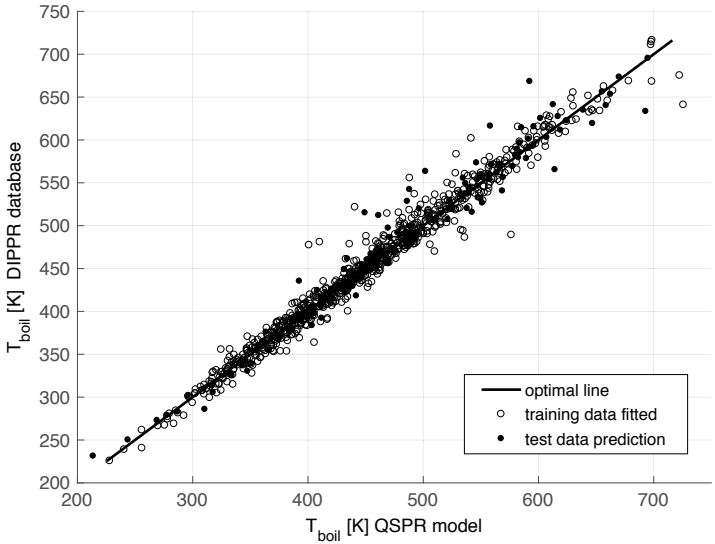


Figure 5.9: Parity plot for the T_{boil}-QSPR model (802 training compounds, 140 test compounds, 2.5% average relative prediction error, 12.3 K average absolute prediction error).

Finally, for applicability domain (AD) evaluation, $\mathbf{X}^{test}$ needs to be projected into the PC space of the training data $\mathbf{X}$. This can be written as follows (Wold et al., 1987):

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{5.19}$$

$$\mathbf{T}^{test} = \mathbf{X}^{test}\mathbf{P} \tag{5.20}$$

In Eqn. (5.19), the column-vectors of $\mathbf{T} \in \mathbb{R}^{N_c \times N_{pc}}$ denote the scores of $\mathbf{X}$, where N_{pc} is the minimum number of PCs required to explain 60% of the variation in $\mathbf{X}$. The *loading* matrix $\mathbf{P}$ then transforms $\mathbf{X}^{test}$ into the PC space of the training data by means of Eqn. (5.20), thus yielding $\mathbf{T}^{test} \in \mathbb{R}^{N_c^{test} \times N_{pc}}$, i.e., the score matrix of the test data. With the help of $\mathbf{T}$ and $\mathbf{T}^{test}$ it is possible to compute, for any compound i in either training or test set, the distance to the five nearest training neighbors $\delta_{5NTN,i}$ in the N_{pc}-dimensional PC space. More specifically, $\delta_{5NTN,i}$ refers to the average of the Euclidean distances to

the closest five compounds from the training set (Weaver and Gleeson, 2008). The $\delta_{5NTN,i}$ values of the training compounds themselves are stored in a vector $\boldsymbol{\delta}_{5NTN}$, and the 95% percentile of these values has been suggested to act as a normalization factor δ_{5NTN}^{norm} for AD evaluation (Weaver and Gleeson, 2008). Hence, any novel compound i is considered to be located inside the AD of a QSPR model built on the training data sets $\mathbf{X}$ / $\tilde{\mathbf{y}}$, if $\delta_{5NTN,i}/\delta_{5NTN}^{norm} \leq 1$. As can be seen from both Figure 5.10 and Table 5.4, some test set compounds are located outside the AD. This is a result of the modified SCA algorithm, which maximizes the diversity of both training and test sets. The chosen AD criterion is therefore rather restrictive.

Table 5.4: Characteristics of the training and test sets. $\delta_{5NTN,i}$ denotes the average distance of compound i to its five nearest training neighbors in the PC space. The normalization factor δ_{5NTN}^{norm} for AD calculation is the 95% percentile of $\boldsymbol{\delta}_{5NTN}$, i.e., the vector of $\delta_{5NTN,i}$ values of all training compounds. Data for LHV, ρ_m, T_{boil}, T_{melt}, H_{vap}, σ and μ have already been published in *Energy & Fuels* (Dahmen and Marquardt, 2016).

	size of **training** set (N_c)	size of **test** set (N_c^{test})	$\delta_{5NTN,i}$ of **training** set compounds		$\delta_{5NTN,i}$ of **test** set compounds		**normalization** factor δ_{5NTN}^{norm}
			mean	max.	mean	max.	
LHV	792	140	2.6	16.2	4.4	14.0	6.2
ρ_m	575	101	4.1	21.8	6.3	21.8	9.3
T_{boil}	802	140	2.5	11.0	4.2	19.3	5.9
T_{melt}	778	136	2.7	14.7	4.2	12.1	6.5
H_{vap}	409	71	3.5	18.1	5.5	12.6	9.1
σ	474	84	3.2	19.2	5.2	15.6	7.8
μ	390	69	3.7	14.1	4.9	13.2	9.0
T_{crit}	822	145	2.6	17.3	3.9	15.8	6.2
p_{crit}	789	139	2.6	18.9	4.7	14.1	6.4
V_{crit}	324	56	3.8	18.1	4.8	10.6	8.6

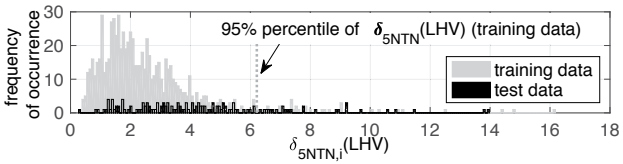


Figure 5.10: Histogram of $\delta_{5NTN,i}(LHV)$, i.e., the average distance of compound i to its five nearest training neighbors in the PC space $\mathbf{T}(LHV)$ / $\mathbf{T}^{test}(LHV)$ (cf. Eqns. (5.19)-(5.20)).

5.2 Comparison with GC-based prediction of thermophysical properties

The book of Poling et al. (2001) provides an excellent review of the most important group contribution methods (GCMs) for the estimation of thermophysical properties. Based on an extensive comparison of the different methods, Poling et al. (2001) argue that, with regard to pure-component constants, i.e., critical properties, normal boiling point and melting point, the newer models developed by Joback and Reid (1987) and Constantinou and Gani (1994) should be preferred, since the older models, e.g., the one proposed by Fedors (1982), do not appear to be as accurate. The model of Constantinou and Gani (1994) has since then been extended by Gani and co-workers (Marrero and Gani, 2001; Hukkerikar et al., 2012; Hukkerikar, 2013). In its most recent version (Hukkerikar et al., 2012; Hukkerikar, 2013), Gani's group contribution plus (GC^+) model represents the most comprehensive suite of group additivity based pure-component thermophysical property prediction in the publicly available scientific literature.

As can be seen from Table 5.5, Joback's method enables the direct prediction of only two fuel properties considered in this thesis (cf. Table 2.1), i.e., the normal boiling point T_{boil} and the melting point T_{melt}. Further properties, however, can be calculated if additional functional relationships are utilized. For instance, the lower heating value LHV [MJ/kg] of a $C_xH_yO_z$ species can be calculated based on the estimate for the enthalpy of formation at standard state H_f^0 [kJ/mol] as

$$HHV = 393.51 \cdot x + 142.915 \cdot y + H_f^0 \,, \tag{5.21}$$

$$M = 12.0107 \cdot x + 1.00794 \cdot y + 15.9994 \cdot z, \tag{5.22}$$

$$LHV = \frac{1}{M} \left(HHV - \frac{1}{2} \cdot y \cdot 43.99 \right) \,, \tag{5.23}$$

where HHV [kJ/mol] denotes the higher heating value, M is the molar mass [g/mol], x is the number of carbon atoms, y refers to the number of hydrogen atoms and z denotes the number of oxygen atoms (Lide, 2003). The following relationships for the enthalpy of vaporization H_{vap}, liquid density ρ_L, kinematic viscosity ν and surface tension σ have been chosen based on the recommendations made by Poling et al. (2001) and Gmehling et al. (2012).

The modified Rackett equation (Yamada and Gunn, 1973) correlates the liquid density ρ_L [kg/m^3] with critical property data via

$$\rho_L = \left[\frac{8.314 \cdot T_{crit}}{10^5 \cdot p_{crit}} \left(0.29056 - 0.08775 \cdot \omega\right)^{1+(1-298/T_{crit})^{(2/7)}} \cdot 1000 \right]^{-1} \cdot M \,, \tag{5.24}$$

Table 5.5: Prediction of the thermophysical fuel properties listed in Table 2.1 by means of two established GC models, i.e., Joback's method (Joback and Reid, 1987) and Gani's method (Hukkerikar et al., 2012; Hukkerikar, 2013), in combination with other empirical models (see text for details). A checkmark indicates that a GC model has been trained for the respective property. The molar mass computed from the sum formula (cf. Eqn. (5.22)) is used to convert between molar and mass bases (not explicitly shown).

	Joback's method (Joback and Reid, 1987)	**Gani's method** (Hukkerikar et al., 2012; Hukkerikar, 2013)
key fuel properties (cf. Table 2.1):		
boiling point T_{boil} at 1 atm [°C]	✓	✓
enthalpy of vaporization H_{vap} at 25°C [kJ/kg]	$H_{vap}(T_{crit}, \omega)$	✓
lower heating value LHV at 25°C [MJ/kg]	$LHV(H_f^0)$	$LHV(H_f^0)$
melting point T_{melt} at 1 atm [°C]	✓	✓
liquid density ρ_L at 25°C [kg/m^3]	$\rho_L(T_{crit}, p_{crit}, \omega)$	✓
kinematic viscosity ν at 25°C [mm^2/s]	$\nu(\mu, \rho_L)$	$\nu(\mu, \rho_L)$
surface tension σ at 25°C [mN/m]	$\sigma(T_{boil}, T_{crit}, p_{crit})$	✓
required for indirect prediction:		
critical temperature T_{crit} [K]	✓	✓
critical pressure p_{crit} [bar]	✓	✓
acentric factor ω [-]	$\omega(T_{boil}, T_{crit}, p_{crit})$	✓
dynamic viscosity μ [Pa·s]	✓	✓
std. enthalpy of formation H_f^0 [kJ/mol]	✓	✓

where T_{crit} denotes the critical temperature [K], p_{crit} the critical pressure [bar], ω the acentric factor [-] and M the molar mass [kg/kmol] (Poling et al., 2001). While T_{crit} and p_{crit} are readily available from Joback's method (cf. Table 5.5), the most accurate estimate for ω is typically obtained from

$$\omega = -\frac{\ln\left(\frac{p_{crit}}{1.01325}\right) + \left[-5.97616\chi + 1.29874\chi^{1.5} - 0.60394\chi^{2.5} - 1.06841\chi^{5}\right] / \frac{T_{boil}}{T_{crit}}}{\left[-5.03365\chi + 1.11505\chi^{1.5} - 5.41217\chi^{2.5} - 7.46628\chi^{5}\right] / \frac{T_{boil}}{T_{crit}}} \,, \quad (5.25)$$

where $\chi = 1 - (T_{boil}/T_{crit})$ (Poling et al., 2001).

The density estimate ρ_L then allows to convert the dynamic viscosity μ as calculated by Joback's method into the kinematic viscosity ν which is considered for the purpose of fuel screening (cf. Tables 2.1 and 5.5).

With the help of ω and T_{crit}, the enthalpy of vaporization H_{vap} [kJ/kg] at 298 K can be

estimated as (Poling et al., 2001)

$$H_{vap} = \frac{1}{M}\left[7.08\left(1-\frac{298}{T_{crit}}\right)^{0.354} + 10.95\cdot\omega\left(1-\frac{298}{T_{crit}}\right)^{0.456}\right] 8.314\cdot T_{crit}\ . \tag{5.26}$$

Finally, in order to compute the surface tension σ [mN/m] at 298 K, the Brock-Bird-Miller equation (Brock and Bird, 1955; Miller and Thodos, 1963; Gmehling et al., 2012) can be utilized:

$$\sigma = p_{crit}^{2/3}\cdot T_{crit}^{1/3}\cdot\left(0.1196\left[1+\frac{(T_{boil}/T_{crit})\cdot\ln\left(\frac{p_{crit}}{1.01325}\right)}{1-(T_{boil}/T_{crit})}\right]-0.279\right)\cdot\left(1-\frac{298}{T_{crit}}\right)^{11/9} \tag{5.27}$$

Gani's multi-level GC$^+$ models (Hukkerikar et al., 2012; Hukkerikar, 2013) have been trained on the CAPEC database, i.e., a property database established and extended by Gani and co-workers since 1998 (Nielsen et al., 2001). Gani and co-workers have argued that it is possible to improve the predictive capability and applicability range of their GC$^+$ models by taking all available data to train the models (Hukkerikar et al., 2012; Hukkerikar, 2013). Consequently, they have decided against setting aside a fraction of the available property data for the purpose of external validation. In general, however, the fitting error is an overly optimistic estimate for the true predictive power of a model (Hastie et al., 2009).

Table 5.6 summarizes the statistical measures for the QSPR models that have resulted from the workflow described above and compares these measures to those of GC-based prediction. In order to facilitate the comparison, Joback's model (Joback and Reid, 1987) and the additional functional relationships given in Eqns. (5.21)-(5.27) have been implemented in Matlab (The MathWorks Inc., 2016). In case of Gani's model (Hukkerikar et al., 2012; Hukkerikar, 2013), ICAS-ProPred (CAPEC, 2012) has been used to make the predictions. The lower heating value has been computed in Matlab from the ICAS-ProPred estimate for the enthalpy of formation H_f^0 (cf. Eqns. (5.21)-(5.23)). For any given property, identical test sets have been used to enable a fair comparison of the different models.

In case of the QSPR models, the average relative error (ARE) is below 10% for all properties and prediction accuracies are best for LHV, T_{boil}, T_{crit} and ρ_m. Similar AREs for the training and test sets suggest that all QSPR models have high predictive power (Golbraikh and Tropsha, 2002a). As can be seen from Table 5.6, the QSPR models perform similar to or better than Joback's and Gani's GCMs. In case of LHV, H_{vap}, p_{crit} and V_{crit}, differences between the three methods are generally marginal. The AREs for the T_{boil} and T_{crit} estimates from Joback's method are roughly two times larger than the corresponding

Table 5.6: Statistical measures for the QSPR models and for two established group contribution methods: For a given property, average relative errors (AREs) and maximum relative errors (MREs) have been calculated for identical test sets (values in brackets for inactive applicability domain). AREs and MREs for the LHV, ρ_m, T_{boil}, T_{melt}, H_{vap}, σ and μ QSPR models have already been published in *Energy & Fuels* (Dahmen and Marquardt, 2016). AREs and MREs for the T_{crit}, p_{crit} and V_{crit} QSPR models have already been published in *Energy & Fuels* (Dahmen and Marquardt, 2017).

	unit	**newly derived QSPR models**				**Joback's method** (Joback and Reid, 1987)		**Gani's method** (Hukkerikar et al., 2012; Hukkerikar, 2013)	
		fitting error on training set		prediction error on test set (w/o AD)		prediction error on test set		prediction error on test set	
		ARE [%]	MRE [%]	**ARE [%]**	MRE [%]	ARE [%]	MRE [%]	ARE [%]	MRE [%]
LHV	J/kmol	1.0	11.5	**1.0** (1.2)	8.9 (9.2)	1.8	10.2	1.7	12.4
ρ_m	kmol/m^3	1.6	12.6	**2.2** (2.5)	7.5 (18.0)	13.1	158.8	1.6	8.0
T_{boil}	K	2.0	17.9	**2.6** (2.5)	12.7 (12.7)	5.7	42.2	2.6	16.0
T_{melt}	K	9.5	77.2	**9.2** (10.3)	31.2 (41.5)	14.9	64.5	11.9	54.1
H_{vap}	J/kmol	5.3	54.5	**5.7** (6.0)	17.3 (22.4)	5.9	44.2	6.0	33.8
σ	N/m	5.9	86.8	**7.7** (7.7)	58.0 (58.0)	17.7	68.2	12.8[b]	76.2[b]
$\log_{10}(\mu)$	$\log_{10}$(Pa·s)	3.7	79.0	**5.0** (5.2)	58.2 (58.2)	7.0[a]	84.2[a]	8.3[c]	47.0[c]
T_{crit}	K	2.1	16.5	**2.5** (2.8)	15.5 (19.3)	5.7	42.4	2.8	15.0
p_{crit}	Pa	4.1	26.9	**5.1** (5.8)	29.5 (31.1)	5.1	28.9	5.0	35.6
V_{crit}	m^3/kmol	2.2	19.8	**4.0** (3.9)	19.0 (19.0)	3.4	14.6	3.4	14.1

[a] 4 molecules could not be evaluated due to missing parameters

[b] 11 molecules could not be evaluated due to missing parameters

[c] 5 molecules could not be evaluated due to missing parameters

AREs obtained with the QSPR and GC$^+$ methods. For T_{melt}, σ and μ, the QSPR models seem to provide somewhat better estimates than the GC methods. Moreover, insertion of critical property data estimates from Joback's method into the modified Rackett equation (cf. Eqn. (5.24)) to predict liquid density yields remarkably inaccurate results, while AREs for the corresponding QSPR and GC$^+$ models are small.

In a strict sense the conclusions drawn here apply to the specific test sets only and it is unclear to what extent compounds from the test sets have been used to train Joback's and/or Gani's GCMs. However, it is safe to say that the QSPRs have demonstrated high predictive power and their training and test sets comprise $C_xH_yO_z$ structures that are similar to those that are generated by the targeted procedure outlined in Chapter 3. Furthermore, the QSPR applicability domain concept allows to quantify this similarity. For these reasons, the QSPR models will be employed for computational property prediction as part of the virtual fuel screening in the subsequent Chapters.

Finally, it shall be mentioned that numerous QSPR models for the prediction of thermophysical properties can also be found in the literature. Katritzky and Fara (2005) and Katritzky et al. (2010) provide excellent reviews of such models. However, despite the existence of many guidelines on best practices for QSPR modeling (cf. Section 2.4.2), it is often difficult to judge the quality of a particular QSPR model based on the published data as technical details are often missing. An even greater problem is the fact that the information required to compute new descriptor data and to make new predictions is frequently incomplete. This has led to the paradox that although re-usability by others should be one of the prime uses of any published QSPR, poor transferability is considered rather the rule than the exception (Dearden et al., 2009; Cherkasov et al., 2014).

5.3 QSPR-based prediction of fuel auto-ignition quality

The QSPR modeling procedure has also been applied to the IQT ignition delay τ. Note that τ and the derived cetane number (DCN) represent physicochemical properties because, in addition to the mixture formation process, gas-phase chemical reactions strongly account for the time delay between fuel injection and start of combustion (cf. Figure 2.2). Total available data have been split into training and validation sets in the same way as described in Chapter 4 to enable a direct comparison between QSPR and GC models. PCA and PLS are particularly useful to deal with the highly unfavorable descriptors-to-observations ratio of 628 to 144. As can be seen from Table 5.7, applying Hotelling's T^2 statistic to the PCA model of the training data does not reveal any structural outliers.

An important difference between the GC model from Chapter 4 and the QSPR model results from the inability of the PLS algorithm to exploit the information provided by the

Table 5.7: QSPR model for the IQT ignition delay τ and the DCN.

no. of training compounds	144
no. of molecular descriptors after pretreatment	628
no. of PCs needed to describe 60% of variation in $\mathbf{X}(\tau)$	3
percentage of compounds ...	
... above 95.0% Hotelling's T^2 threshold in the PCA model of $\mathbf{X}(\tau)$	5.6%
... above 99.0% Hotelling's T^2 threshold in the PCA model of $\mathbf{X}(\tau)$	1.4%
... above 99.9% Hotelling's T^2 threshold in the PCA model of $\mathbf{X}(\tau)$	0.0%
no. of PLS components used to model $\tilde{\mathbf{y}}(\ln(\ln\tau))$ (determined by 7-fold CV)	8
statistical performance measures	
training data set	
average absolute error (AAE) in DCN space	4.1
maximum absolute error (MAE) in DCN space	17.4
repeated 5-fold cross-validation (LMO-CV) data sets (10,000 runs)	
average absolute error (AAE) in DCN space	6.8
maximum absolute error (MAE) in DCN space	68.0
test data set (i.e., the external validation data set)	
average absolute error (AAE) in DCN space	6.6
maximum absolute error (MAE) in DCN space	14.1

measurement uncertainty. However, similar to the GC model, the best PLS fit is obtained, if $\ln(\ln \tau)$ instead of $\ln \tau$ is modeled by a linear combination of descriptors (see discussion in Subsection 4.2.1). As can be seen from the parity plot in Figure 5.11, this yields a distribution of residuals that is approximately independent of the DCN.

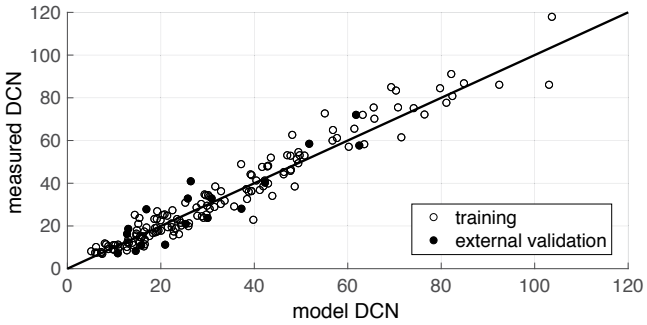


Figure 5.11: QSPR model for the IQT ignition delay τ and the derived cetane number (DCN). Note: $\ln \tau$ has been converted to DCN exclusively based on Eqn. (4.1).

As depicted in Figure 5.12, the GC model and the QSPR model perform similarly well on the external validation sets. This is also reflected in the statistical performance measures shown in Table 5.7 (compare to those of the GC model presented in Table 4.10). Two possible interpretations shall be given here: (i) The fact that the PLS algorithm is unable to extract a correlation between $\ln(\ln \tau)$ and 628 molecular descriptors, which is significantly more accurate than the one described by the group contribution method, suggests that the GC model indeed captures important relationships between molecular structure and the IQT ignition delay. (ii) The fact that the GC model does not appear to be considerably more accurate than the QSPR model is another piece of evidence supporting the claim that the GC model has not been substantially overfitted. Interestingly, both models fail to predict the low DCN of methyl tert-butyl ether (MTBE), whereas the predictions for all other compounds with experimental DCN<7, can be considered fairly accurate (cf. Figure 5.12, bottom). This observation underlines the special character of the ether group sitting next to a tertiary carbon atom (see discussion in Subsection 4.2.4).

The GC model, not the QSPR model, is used to predict DCN as part of the virtual fuel screening in the subsequent Chapters, because (i) the GC estimate can be considered more transparent because the rationale behind the GC model has been presented in detail in Section 4.2, (ii) the GC model does not seem to be inferior to the QSPR model in terms of predictive power, and (iii) the GC model takes into account the marked differences in measurement uncertainty of IQT-derived ignition delay and DCN.

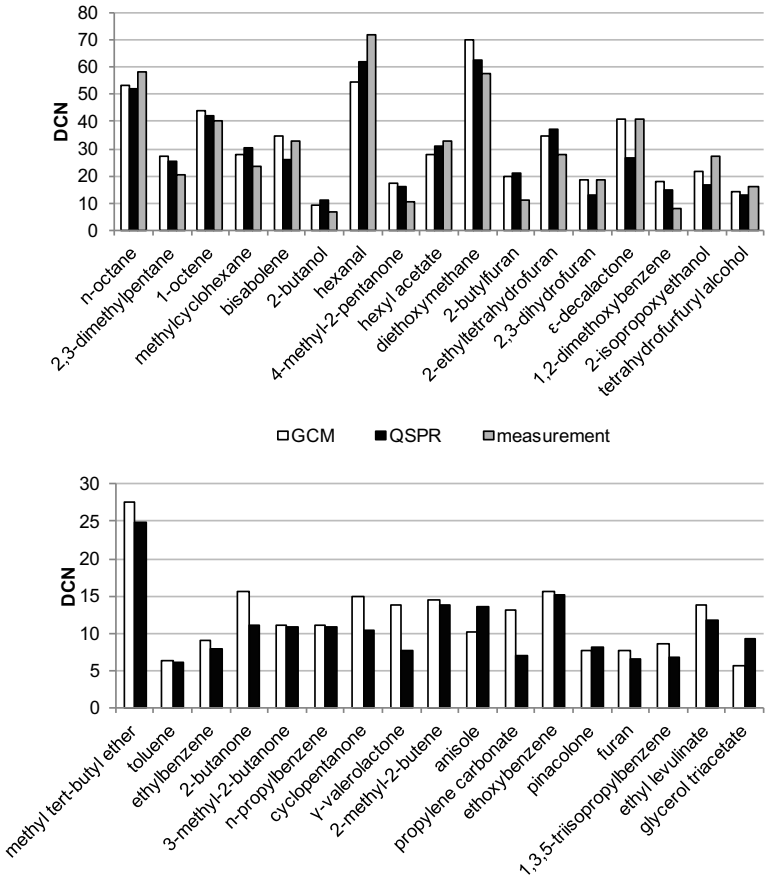


Figure 5.12: GC model vs. QSPR model: Predictions for the external validation set (top). Predictions for the compounds with experimental DCN<7 (cf. Figure 4.10) (bottom).

5.4 Conclusions and outlook

A generic workflow for QSPR modeling of physicochemical fuel properties based on PCA and PLS regression has been successfully applied to a range of thermophysical properties and the IQT ignition delay. The property models have been tailored for $C_xH_yO_z$ structures relevant to computational fuel screening by considering high-quality property data extracted from the DIPPR database of organic compounds. The models have been thoroughly validated by means of cross-validation and external validation. Moreover, an applicability domain concept has been implemented to quantify similarity between training compounds and target compounds. With regard to the thermophysical fuel properties from Table 2.1, the QSPR models perform similar to or somewhat better than established group contribution methods on the basis of the test sets investigated here. Similar prediction accuracies have been observed for QSPR-based and GC-based prediction of IQT ignition delay and DCN. These encouraging results suggest to apply this particular strategy for QSPR modeling to additional physicochemical properties relevant to fuel design. Given a sufficient amount of training data, application of the modeling workflow is straight-forward and does not require a substantial amount of time.

6 Model-based identification of biofuel candidates

In this Chapter, the generate-and-test variant of biofuel CAMD as depicted in Figure 1.3 is applied to two case studies. The aim of these case studies is to identify oxygenated fuel candidates for high-efficiency and low-emission spark-ignition and compression-ignition engines like those studied within TMFB (Janssen et al., 2011; Thewes et al., 2011a,b; Heuser et al., 2013b, 2014; Hoppe et al., 2016a). These fuels shall result from carbon- and energy-efficient chemo-catalytic refunctionalization of bio-derived platform molecules. In contrast to earlier biofuel CAMD studies (Hechinger et al., 2012b; Dahmen et al., 2012; Hechinger, 2014), the targeted approach for structure generation (cf. Chapter 3) and the improved model for fuel auto-ignition quality (cf. Chapter 4) allow to narrow down the range of potential fuel candidates to a smaller number of distinct molecular motifs. Fuel prototyping, i.e., the synthesis of a sufficiently pure, small sample for experimental investigation, constitutes a subsequent step which needs to be performed on manually selected compounds, but is not addressed explicitly in this thesis. The interested reader is referred to a recent publication (Hoppe et al., 2016b) where experimental results for both SI and CI one-cylinder research engines are presented for a range of TMFB fuels. The present Chapter has already been published in a slightly modified form in *Energy & Fuels* (Dahmen and Marquardt, 2016).

6.1 Candidates for spark-ignition (SI) engines

The first case study deals with the identification of fuel candidates for the spark-ignition (SI) engine. In order to achieve a high overall LHV efficiency (sugar/H_2 to fuel), those intermediates from Table 2.4 are selected as starting points for fuel synthesis that comprise at least 75% LHV efficiency from sugar to intermediate at highest-reported real yield. To compute the overall LHV efficiency, it is assumed that the chemical upgrading of intermediates to fuels occurs at (or close to) the maximum theoretical yield. The eleven intermediates selected for fuel production are given in Table 6.1.

The product spectrum shall be limited to molecules containing up to eight carbon atoms.

Table 6.1: Scenario definition for the generation of biofuel candidates for the **SI engine**.

intermediates	ethanol, lactic acid, 1-butanol, succinic acid, isobutanol, 2,3-butanediol, acetic acid, 1,3-propanediol, hydroxymethylfurfural, furfural, butyric acid	
structural constraints	max. of eight carbon atoms; five- and six-membered rings only; up to one aggregation by C-C coupling (aldol condensation) and up to one aggregation by C-O-C coupling (etherification or esterification); max. Joback T_{boil} (T_{melt}) of 500 °C (of 200 °C) for all molecules (intermediates and products) to be generated	
fuel specifications	oxygen content wt_{O_2}	$\geq$ 10 wt-%
	boiling point T_{boil}	$\geq$ 60 °C and $\leq$ 120 °C
	melting point T_{melt}	$\leq$ -20 °C
	surface tension σ	$\leq$ 30 mN/m
	viscosity ν	$\leq$ 2.0 mm^2/s
	enthalpy of vaporization H_{vap}	$\leq$ 60 kJ/kg(air) at λ=1
	derived cetane number DCN	$\leq$ 20

Since five- and six-membered rings are the most stable cyclic arrangements of heavy-atoms, other rings sizes are not considered here. Finally, the aggregations shall be limited to one C-C coupling and one C-O-C coupling.

For the purpose of fuel identification, the set of property constraints given in Table 6.1 is used. These constraints have evolved from definitions for tailor-made fuels established within TMFB (Kremer, 2011; Klankermayer et al., 2011; Thewes et al., 2011b; Victoria Villeda et al., 2012c; Hoppe et al., 2016a,b) and from the regular exchange between the author of this thesis and members of the TMFB Core Interaction Field "Fuel Design", most notably Florian Kremer, Benedikt Heuser, Fabian Hoppe and Stefan Pischinger (Institute for Combustion Engines, RWTH Aachen University) and Manuel Hechinger and Juan José Victoria Villeda (Institute for Process Systems Engineering at Aachener Verfahrenstechnik, RWTH Aachen University). With the aim of reducing soot emissions in direct-injection SI engines, fuel candidates containing at least 10 wt-% of oxygen are favored. The DCN shall be smaller than 20, which roughly corresponds to a RON of 85 and higher. The relationship between DCN and RON is based on a correlation between IQT ignition delay τ and RON derived by Perez and Boehman (2012), which is illustrated in Figure 6.1. The DCN for iso-octane is approximately 17 (Dooley et al., 2012b) and from Figure 6.1 it can be seen that the correlated RON is close to the true RON of 100. Likewise, the experimental DCN for methyl tert-butyl ether (MTBE) is <7 (cf. Appendix B), which is in good agreement with MTBE's RON of 116-118 (Rosell et al., 2006).

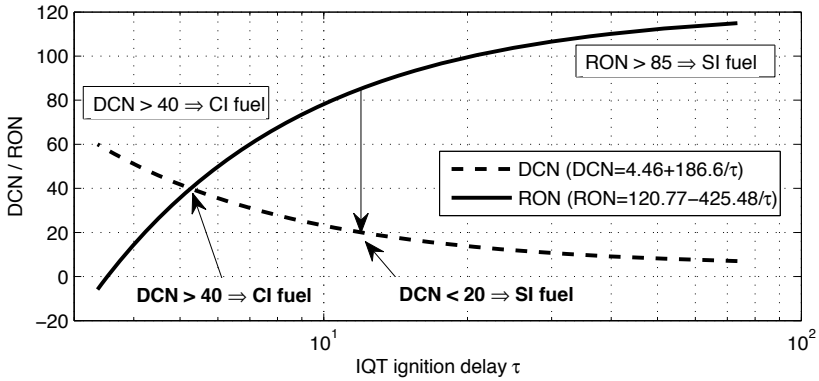
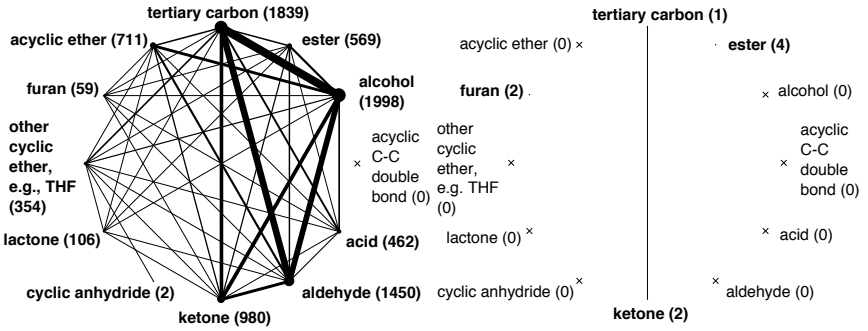


Figure 6.1: Relationship between IQT ignition delay τ, derived cetane number (DCN) and research octane number (RON) (correlations taken from ASTM D6890 (2011) and Perez and Boehman (2012)).

Similar to previous studies (Klankermayer et al., 2011; Hechinger et al., 2012b; Hechinger, 2014; Hoppe et al., 2016a,b), problems related to cold start/run and oil dilution shall be mitigated by limiting the boiling point to 120 °C and the enthalpy of vaporization to 60 kJ/kg(air). The enthalpy of vaporization in kJ/kg is related to 1 kg of air for a stoichiometric mixture in order to account for the fact that the mass of fuel to be injected depends on the fuel's atomic composition. Besides boiling point and enthalpy of vaporization, surface tension and kinematic viscosity are constrained, because droplet combustion and insufficient mixing of vaporized ethanol containing fuel and air have been linked to soot formation in direct-injection SI engines (Storch et al., 2015). Finally, to select only those fuels, which are liquids under typical ambient conditions, an upper bound is placed on the melting point and a lower bound is introduced for the boiling point (cf. Table 6.1).

Based on the scenario definition depicted in Table 6.1, the molecular structure generator suggests 3,215 potential products. Characteristics of these products are visualized in Figure 6.2 (a). Here, the diameter of the nodes denotes the number of molecules in the product spectrum containing one or multiple instances of a certain structural pattern. The exact number of molecules is given in parentheses. Moreover, the thickness of a line drawn between two nodes indicates the number of molecules containing both structural patterns.

The product spectrum covers many different functionalities. The most frequently occurring oxygen motifs are alcohol, aldehyde, ketone and ether. The tertiary carbon atom originates from iso-butanol or results from C-C coupling. Polyfunctional molecules consti-



(a) 3,215 molecules in total **(b)** 8 molecules in total

Figure 6.2: Pure-component fuel candidates for the SI engine: Product spectrum after structure generation (a) and after virtual fuel screening (b). The diameter of the nodes denotes the number of molecules containing a certain structural pattern (the exact number is given in parentheses). The thickness of a line indicates the number of molecules containing both structural patterns.

tute the vast majority. As can be seen from Figure 6.2 (b), only eight of 3,215 molecules are predicted to exhibit the desired properties. These oxygenates are monofunctional compounds. Figure 6.3 shows the detailed results from the virtual fuel screening. For each property filter the number of molecules satisfying the imposed constraint, the number of molecules not meeting the constraint and the number of molecules located outside of the applicability domain of the respective QSPR model are given. Interestingly, only 89 compounds cannot be evaluated due to limited model applicability. Obviously, the molecules generated by the targeted approach have sufficient similarity to the training molecules used in the derivation of the QSPRs in Chapter 5. The normal boiling point constraint (-98.1%) and the DCN constraint (-85.7%) act as the two main filters.

Table 6.2 summarizes key properties of the eight biofuel candidates. The structural relations between fuel candidates and intermediates are tracked by the molecular structure generator and can be found in the third last column of Table 6.2. The eight candidates can be grouped into furans, ketones and esters and will be discussed in the following.

2-Methylfuran and 2,5-dimethylfuran can be viewed as validation compounds demonstrating that the outlined computational approach is indeed capable of identifying fuel candidates. As discussed in Chapter 2, the favorable properties of these furanic fuels are well-known. Moreover, based on an optimized production set-up for furfural and hydroxymethylfurfural, quite high overall LHV efficiencies can be realized. The volumetric energy densities of 2-methylfuran and 2,5-dimethylfuran (29.0−29.3 MJ/l) are approximately 28%

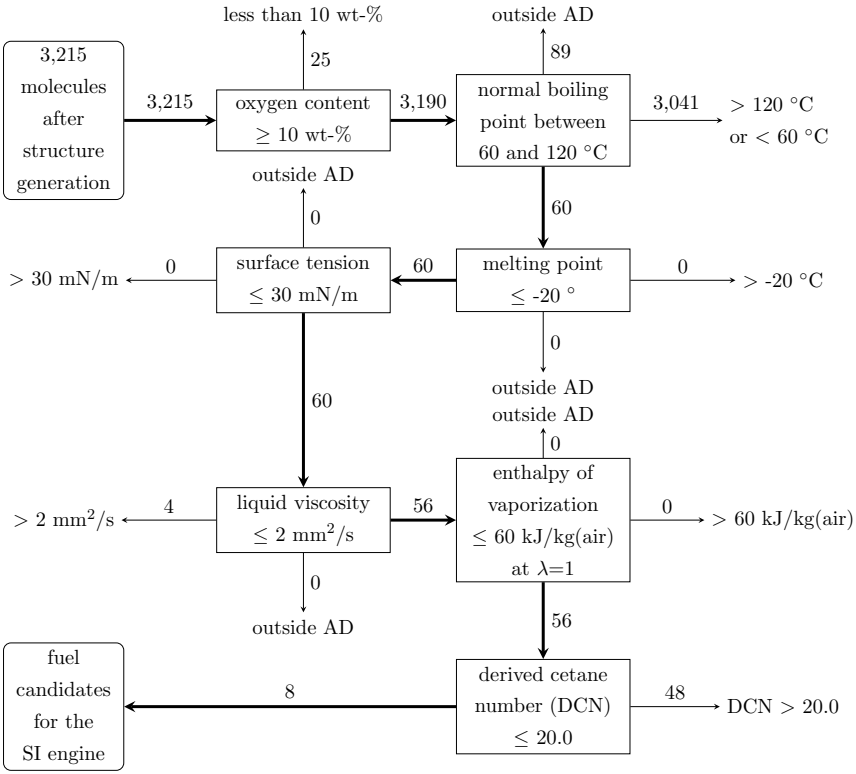


Figure 6.3: SI engine: Virtual fuel screening performed by means of QSPR and group contribution based property prediction (AD denotes applicability domain of respective QSPR model).

Table 6.2: Pure-component biofuel candidates for the **SI engine**.

fuel candidate	**boiling point [°C]**	**DCN**	**lower heating value [MJ/kg]**	**enthalpy of vaporization [kJ/kg(air)]**	**density [kg/m^3]**	**viscosity [mm^2/s]**	**intermediates**	**LHV eff. [%] (sugar/H_2 to fuel)**	**moles H_2 per mole of fuel**
 2-methylfuran	82.7	9.6	30.5	41.5	951	0.7	furfural	77	2
 2,5-dimethylfuran	111.2	9.8	32.1	35.7	914	0.9	hydroxymethyl furfural	85	3
 2-butanone	77.6	17.6	31.5	39.5	811	0.5	2,3-butanediol	86	0
 methyl isobutyl ketone	114.7	19.0	34.8	32.0	808	0.7	lactic acid, 1,3-propanediol	85 − 93	3 − 5
 ethyl acetate	90.4	13.7	23.2	47.5	920	0.4	ethanol, acetic acid	88	0

Continued on next page

Table 6.2: Pure-component biofuel candidates for the **SI engine** (continued).

fuel candidate	boiling point [°C]	DCN	lower heating value [MJ/kg]	enthalpy of vaporization [kJ/kg(air)]	density [kg/m^3]	viscosity [mm^2/s]	intermediates	LHV eff. [%] (sugar/H_2 to fuel)	moles H_2 per mole of fuel
ethyl propanoate	113.0	17.4	26.1	42.2	902	0.6	ethanol, lactic acid	95	1
propyl acetate	107.2	17.4	25.8	41.6	886	0.6	acetic acid, 1,3-propanediol, lactic acid	79 – 88	1 – 3
propanyl acetate	100.0	10.3	26.0	38.7	881	0.6	acetic acid, lactic acid	88	3

higher than that of ethanol, however about 9% lower than that of gasoline. Compared to RON95, engine efficiency can be increased with 2-methylfuran as a result of the higher octane number (Hoppe et al., 2016a,b). Thus, it is reasonable to assume that the two compact furans can provide a mileage close to that of gasoline given a fixed volume of fuel.

2-Butanone had been discovered by Hechinger (2014) in a previous round of computer-aided fuel design and since then the fuel has been studied in a highly-boosted direct-injection SI research engine (Hoppe et al., 2016a,b). Efficiency gains up to 20% at full load operation compared to RON95 gasoline can be achieved due to a higher effective compression ratio enabled by the extreme knock resistance of 2-butanone (RON 117, MON 107) (Hoppe et al., 2016b). Moreover, butanone's low boiling point, low viscosity and gasoline-like enthalpy of vaporization improve the mixture formation process compared to ethanol fuel, especially at low engine loads and cold boundary conditions (Hoppe et al., 2016a,b). As proposed by the molecular structure generator, 2-butanone can be produced from 2,3-butanediol (Török et al., 1996; Zhang et al., 2012). Direct fermentation of butanone has also been suggested, however, current yields are extremely low (Yoneda et al., 2014). Methyl isobutyl ketone (cf. Table 6.2) is another compact ketonic biofuel candidate for the SI engine. Its experimental DCN is 12.6 (cf. Appendix B). Presumably, it exhibits anti-knock properties similar to those of butanone, however, methyl isobutyl ketone has a higher boiling point, which is comparable to that of 1-butanol. However, in contrast to the alcohol-fuel, methyl isobutyl ketone comprises a more gasoline-like enthalpy of vaporization.

Small esters make the third group of pure-component biofuel candidates for the SI engine. Compared to the furanic or ketonic biofuels, their mass- and volume-related energy density is lower due to the higher oxygen-to-carbon ratio. Similar to ketones and furans, small esters exhibit enthalpies of vaporization similar to that of gasoline fuel. The experimental boiling points are somewhat lower than those predicted by the QSPR model, e.g., the true boiling point for ethyl acetate is 77 °C (Jenkins et al., 2013). In addition to these favorable mixture formation properties, small esters are also extremely knock-resistant (Dabbagh et al., 2013). For instance, a RON of 116 has been reported for ethyl acetate (Stickney et al., 2005). Ternary mixtures of small esters have been studied in HCCI engines (Contino et al., 2011a,b). Saturated esters have been found to be far more stable than the polyunsaturated esters in vegetable oils and biodiesel (Jenkins et al., 2013). Thus, oxidative instability should not be a major concern.

Given the fact that lots of research activity has centered around furanic biofuels in recent years (e.g., Zhong et al. (2010); Lange et al. (2010); Thewes et al. (2011a); Lange et al. (2012); Wang et al. (2013); Sudholt et al. (2015)), compact ketones and esters constitute interesting, possibly even superior alternatives. In contrast to the furans, the ketones and

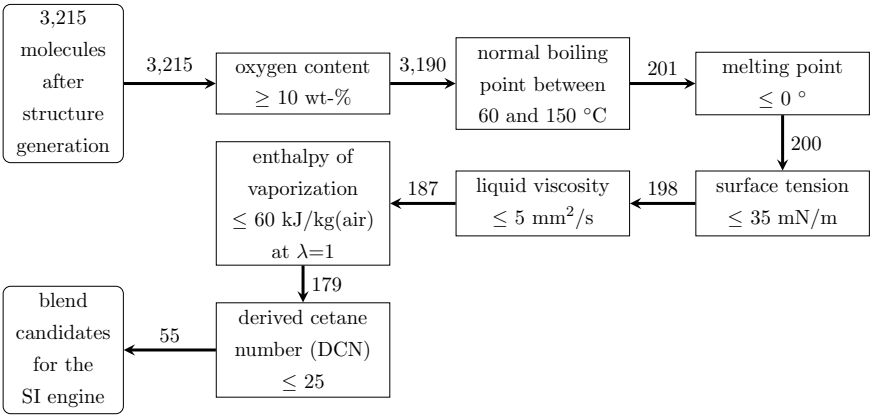
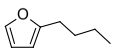
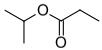


Figure 6.4: SI engine: Candidates for blending: Virtual fuel screening performed by means of QSPR and group contribution based property prediction.

esters can be produced from fermentation products with high overall LHV efficiencies. For instance, Jenkins et al. (2013) have recently esterified a variety of large-volume fermentation products (including acetic acid and lactic acid) with ethanol to yield potential fuel compounds.

The question arises, whether a broader range of components can be obtained for blending applications, if the property constraints are somewhat relaxed. For instance, raising the maximum tolerable boiling point from 120 °C to 150 °C yields additional molecules, which do not constitute pure-component fuel candidates. However, these species might be part of a blend in which lighter fractions support the in-cylinder mixture formation process. Similarly, one-at-a-time variation of all property constraints shown in Figure 6.3 can reveal how the number of molecules and the prevalent structural motifs depend on the chosen constraint values. Although the flexibility of the computational approach allows to perform such an analysis with very little effort, the interpretation of a sensitivity study goes beyond the scope of this case study. Therefore, the attention shall be restricted here to one particular scenario with relaxed property constraints, which is shown in Figure 6.4. In this scenario, the requirements on boiling point, melting point, viscosity, surface tension and DCN have been weakened. This yields a list of 55 candidates for blending applications that includes the eight pure-component fuel candidates discussed above. The full list of structures can be found in the Supporting Information of Dahmen and Marquardt (2016). Seven representative structures are depicted in Table 6.3. Structurally similar compounds from the list are given in the last column.

Table 6.3: SI engine: Selected candidates for blending.

candidate for blending	boiling point [°C]	DCN	lower heating value [MJ/kg]	viscosity [mm^2/s]	intermediates	LHV eff. [%] (sugar/H_2 to fuel)	moles H_2 per mole of fuel	similar candidates
 2,5-dimethyltetrahydrofuran	101.5	24.5	35.2	0.8	hydroxymethyl-furfural, lactic acid	86 − 93	5	2/3-MTHF, 2,3/2,4/3,4-dimethyltetrahydro-furan, ...
 2-butylfuran	139.4	21.2	34.8	1.3	furfural, lactic acid	82	5	2-(2-methylpropyl)furan, 2-propylfuran
 2-pentanol	121.3	14.6	34.9	4.2	furfural, acetic acid, lactic acid, ethanol	79 − 93	3 − 5	2-butanol, isobutanol, 3-methyl-1-butanol, 2-methyl-1-butanol
 3,4-dimethyl-2-pentanone	136.1	14.5	36.0	0.9	lactic acid, 2,3-butanediol, 1,3-propanediol	81 − 87	1 − 3	2-pentanone, 3-methyl-2-pentanone, 5-methyl-2-hexanone, ...
 propanyl propanoate	122.9	14.8	28.1	0.8	lactic acid	94	4	2-methylpropyl acetate, 2-butanyl acetate, butyl acetate, ...

Continued on next page

Table 6.3: SI engine: Selected candidates for blending (continued).

candidate for blending	boiling point [°C]	DCN	lower heating value [MJ/kg]	viscosity [mm^2/s]	intermediates	LHV eff. [%] (sugar/H_2 to fuel)	moles H_2 per mole of fuel	similar candidates
3-(propan-2-yloxy)butan-2-one	141.0	22.3	30.3	1.2	2,3-butanediol, lactic acid	88	2	
2-ethoxy-1-propanol	138.4	22.8	28.0	4.4	ethanol, lactic acid	94	2	

Note: The entire set of 55 blending candidates generated can be found in the Supporting Information of Dahmen and Marquardt (2016).

A range of tetrahydrofurans is proposed as a result of a slightly relaxed DCN constraint. The experimental DCN for 2-methyltetrahydrofuran (2-MTHF) is 21.3 (cf. Appendix B). The correlated RON is in good agreement with the reported RON of 86 (Yanowitz et al., 2011). Since the hydrogenation of furans significantly lowers their anti-knock quality, it does not seem reasonable to perform this synthesis step.

2-Butylfuran can be produced via aldol condensation of furfural and acetone as recently reported by Julis and Leitner (2012). If the side-chain length of a furanic structure is increased, the DCN increases as well (Sudholt et al., 2015). As a consequence, 2-butylfuran (experimental DCN of 13.1; cf. Sudholt et al. (2015)) is expected to be less knock-resistant than 2-methylfuran. Similar to the hydrogenation of the furan ring, extending the length of the side-chain will increase the lower heating value. However, since a high knock-resistance is typically more important, aldol condensation as a tool in furanic SI fuel synthesis does not seem plausible.

Isomers of butanol and pentanol as well as additional ketones are among the 55 candidates for blending applications (cf. Table 6.3). While the larger alcohols satisfy the constraint on the enthalpy of vaporization, the boiling point as well as the DCN increase compared to ethanol. Moreover, the viscosity of the larger alcohols is roughly six times higher than the viscosity of gasoline fuel. Thewes et al. (2011b) have suggested that the high viscosity of the alcohol-fuels further deteriorates the quality of mixture formation due to the larger droplet size.

With regard to furans, ketones and esters, the relaxation of property constraints gives rise to more variants. However, there is evidence that the small molecules, i.e., those satisfying the stricter constraints for pure-component fuels, should be favored: An increase in the molecular weight, as a general trend, will decrease fuel volatility and knock resistance. Furthermore, the synthesis of larger compounds is expected to require more steps.

Finally, the list of 55 candidates for blending applications includes polyfunctional compounds of two sorts: An ether group is combined either with a ketone group or with an alcohol group. The ether group allows to integrate another oxygen atom without strongly increasing the boiling point (Joback and Reid, 1987). Both ketone and alcohol groups raise the RON (cf. Chapter 4 and Figure 6.1). However, the effect of the ether group on fuel auto-ignition is difficult to predict (cf. discussion in Subsection 4.2.4). Given the similarity to diisopropylether, i.e., a widely used anti-knock agent for gasoline (Heese et al., 1999), the additional ketone group in 3-(propan-2-yloxy)butan-2-one (cf. Table 6.2) is not expected to change things for the worse. The RON for 2-ethoxy-1-propanol on the other hand, is difficult to predict, since a structurally similar compound is missing.

In summary, the case study has yielded compact furans, ketones and esters as candidates of first choice for the SI engine, if a pure-component biofuel is envisaged. The larger

alcohols, however, might constitute favorable components for blending applications. The anti-knock quality of the tetrahydrofurans and larger furans is lower than that of RON95 gasoline. Still, it is conceivable that these compounds might play a role as they can be used to raise the heating value of a biofuel blend, if another major component of the mixture is able to provide an acceptable overall RON.

In contrast to the targeted approach to molecular structure generation used here, Hechinger (2014) had performed an exhaustive search for pure-component fuel candidates for the SI engine. To this end, Hechinger (2014) combined the molecular structure generator Molgen (Gugisch et al., 2015) with QSPR models as described by Dahmen et al. (2012). The target properties and their ranges were somewhat different compared to the current study. Most importantly, the auto-ignition propensity had not been modeled by Hechinger (2014). His screening yielded 279 structures that were categorized into acyclic and cyclic ethers, furans, acetals, aldehydes, ketones and alcohols (Hechinger, 2014; Hoppe et al., 2016b). Based on the DCN group contribution model (cf. Chapter 4), neither the linear ethers/acetals nor the aldehydes do constitute fuels for the SI engine. Furans, ketones and alcohols have also been identified by the targeted approach used here. Since Hechinger (2014) had screened only for compounds with a lower heating value of 30 MJ/kg and higher, small esters could not be identified in his study. Highly-branched ethers like methyl-tert-butyl ether (MTBE) or ethyl-tert-butyl ether (ETBE) represent excellent fuels for turbocharged SI engines (cf. discussion in Section 2.1). Such species, however, do not result from the current set of platform molecules and/or transformation rules implemented in the structure generator. The same holds for non-furan aromatic species exhibiting low DCNs, e.g., toluene or anisole (cf. Figure 4.10). Although pyrolysis produces aromatics (Haveren et al., 2008; Marshall and Alaimo, 2010), it is the lack of selectivity of the reaction towards a desired molecule that makes these pathways unsuitable with regard to the outlined approach of selective refunctionalization of bio-based platform chemicals. Moreover, low volatility generally constrains the use of aromatics in SI engine combustion. While toluene (T_{boil} of 110.6 °C; cf. Przyjazny and Kokosa (2002)) satisfies the pure-component fuel volatility requirement (cf. Table 6.1), even slightly larger aromatic species, e.g., ethylbenzene (T_{boil} of 136.2 °C; cf. Przyjazny and Kokosa (2002)), xylenes (T_{boil} of 138.3−144.0 °C; cf. Przyjazny and Kokosa (2002)) or anisole (T_{boil} of 153.8 °C; cf. Dreisbach and Martin (1949)), most likely, will require the addition of lighter components, which support the in-cylinder mixture formation process. Finally, fuel aromatic content is closely linked to sooting propensity, since it is believed that the formation of the first benzene ring is the rate-controlling step in the production of particulate matter (Boot et al., 2008).

6.2 Candidates for compression-ignition (CI) engines

The second case study targets pure-component biofuels for the compression-ignition (CI) engine. Based on the refunctionalization rules given in Figure 3.2 and Appendix A, the number of generated molecular structures increases rapidly with the number of carbon atoms. While the proposed algorithm is capable of processing tens (or even hundreds) of thousands of molecules, this case study shall exemplify, that a large number of fuel candidates for the CI engine can be obtained, even if the number of intermediates chosen for fuel synthesis is as small as six and the maximum number of carbon atoms is set to ten. Furthermore, carbon-carbon coupling is restricted to chain-lengthening aldol condensation, i.e., aldol condensation extending a straight carbon chain without forming a tertiary carbon atom. The purpose is to keep the interpretation of the results reasonable, as inclusion of branched molecules will give many additional variants of the structures discussed below without changing the general conclusions. Moreover, branching is typically associated with decreasing CN (Heck et al., 1998; Santana et al., 2006).

Table 6.4: Scenario definition: Biofuel candidates for the **CI engine** based on six bio-derived intermediates.

intermediates	ethanol, 1,3-propanediol, 2,3-butanediol, lactic acid, succinic acid, furfural	
structural constraints	max. of ten carbon atoms; five- and six-membered rings only; up to one aggregation by C-C coupling (chain-lengthening aldol condensation only) and up to one aggregation by C-O-C coupling (etherification/esterification); max. Joback T_{boil} (T_{melt}) of 500 °C (of 200 °C) for all molecules (intermediates and products) to be generated	
fuel specifications	oxygen content wt_{O_2}	$\geq$ 10 wt-%
	boiling point T_{boil}	$\geq$ 60 °C and $\leq$ 250 °C
	melting point T_{melt}	$\leq$ -20 °C
	lower heating value LHV	$\geq$ 30 MJ/kg
	liquid density ρ_L	$\geq$ 700 kg/m^3
	surface tension σ	$\leq$ 30 mN/m
	kinematic viscosity ν	$\leq$ 4.5 mm^2/s
	derived cetane number DCN	$\geq$ 40

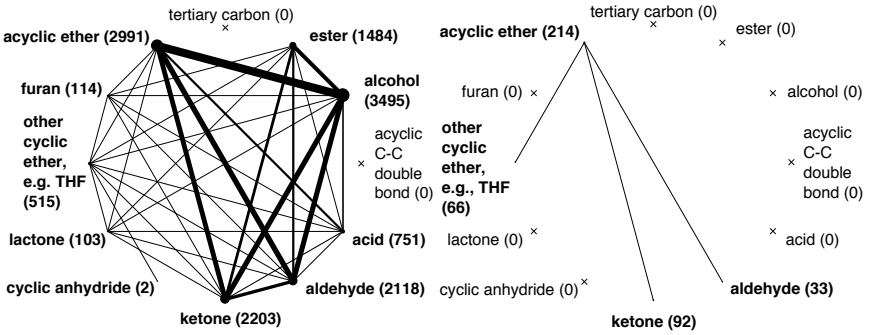
As can be seen from the scenario definition in Table 6.4, the intermediates ethanol, propanediol, butanediol, lactic acid and succinic acid have been chosen for fuel production, since these fermentation products comprise high LHV efficiencies at real yields. Moreover, compared to the other fermentation products mentioned in Table 2.3, high productivity

and high titer suggest that a large-volume production is feasible given state-of-the-art technology. In addition, furfural has been selected as a derivative of the C_5 sugar fraction.

Similar to the previous case study, the property constraints given in Table 6.4 have evolved from fuel definitions established within TMFB (Kremer, 2011; Janssen et al., 2011; Victoria Villeda et al., 2012c; Heuser et al., 2013a,b, 2014) and from the regular exchange between the author of this thesis and members of the TMFB Core Interaction Field "Fuel Design", most notably Florian Kremer, Andreas Janssen, Benedikt Heuser and Stefan Pischinger (Institute for Combustion Engines, RWTH Aachen University) and Manuel Hechinger and Juan José Victoria Villeda (Institute for Process Systems Engineering at Aachener Verfahrenstechnik, RWTH Aachen University).

Since CI engines can burn less volatile fuel, boiling points of up to 250 °C shall be tolerated here. This moderate upper bound, compared with the boiling range of fossil diesel fuel (180−350 °C), is thought to enhance mixture homogenization and to minimize the likelihood of wall impingement in low-temperature diesel combustion, where the fuel is injected early into relatively cool in-cylinder gases (Cheng et al., 2010; Fisher et al., 2010). Typical injection pressures in CI engines are higher than those used in direct-injection SI engines. Consequently, compared with the previous case study, the upper limit for the kinematic viscosity is raised to 4.5 mm^2/s, which is also the maximum allowed viscosity in the European diesel fuel standard EN 590 (2014). At least 10 wt-% of oxygen in the fuel molecule is desired to achieve low levels of engine-out soot emissions. Fuel and air shall auto-ignite under conditions typically attained in the cylinder of a diesel engine. Thus, the lower bound for the DCN is set to 40. The need for a long injection duration can cause higher NOx production and intensified formation of soot precursors in fuel rich cores, especially at high engine loads and speeds (Arcoumanis et al., 2008; Yao et al., 2010; Jung et al., 2011). To mitigate this problem, lower bounds on the heating value (30 MJ/kg) and the liquid density (700 kg/m^3) are introduced, since the injection duration correlates with the amount of fuel delivered.

Based on the scenario definition depicted in Table 6.4, 5,345 molecular structures are generated. Figure 6.5 (a) illustrates the nature of the product spectrum. The distribution of functionalities is similar to the one obtained in the previous case study, although a smaller set of intermediates has been used. If the physicochemical properties of these molecules are predicted by means of QSPR and GCM, 247 fuel candidates for the CI engine are obtained, which all fulfill the fuel specifications in Table 6.4. As can be seen from Figure 6.5 (b), these structures can be classified into acyclic and cyclic ethers, ketones, aldehydes and polyfunctional compounds. The polyfunctionality, however, is limited to the case, where an acyclic ether group is combined with either a cyclic ether group, a ketone group or an aldehyde group.



(a) 5,345 molecules in total **(b)** 247 molecules in total

Figure 6.5: Pure-component fuels for the **CI engine**: Product spectrum after structure generation (a) and after virtual fuel screening (b).

Figure 6.6 shows the detailed results from the virtual fuel screening. Except for the liquid density, all property constraints cause rejection of certain compounds and thus effectively reduce the number of molecules for further consideration. Compared with the previous case study, a larger fraction (∼18% instead of ∼3%) of the generated structures are located outside the applicability domain (AD) of the normal boiling point QSPR model (compare Figures 6.3 and 6.6). This can be explained by the fact that 95% of the structural outliers found here have either nine or ten carbon atoms, whereas in the generation step of the previous case study the maximum number of carbon atoms had been restricted to eight. If the final set of 247 fuel candidates for the CI engine is analyzed, seven types of molecules can be distinguished. These types are depicted in Table 6.5 and will be discussed in the following. The full list of molecular structures can be found in the Supporting Information of Dahmen and Marquardt (2016).

Type I structures are acyclic, straight chain ethers containing at least six carbon atoms. It is well-known that linear ethers like di-butylether exhibit very high auto-ignition propensities (Beatrice et al., 1996). At the same time, the boiling point of di-butylether (141 °C) is well below the boiling range of fossil diesel fuel (Heuser et al., 2014; Graziano et al., 2015). In addition, the low surface tension (22 mN/m) and the low viscosity (0.72 mPa·s) of di-butylether promote the break up of the spray during mixture formation (Heuser et al., 2013b, 2014). Although the air/fuel mixing time for di-butylether is very short (due to the high CN of ∼100), nearly soot-free combustion with very low NOx levels have been reported (Heuser et al., 2013b,a; García et al., 2016). Based on these findings, linear ethers (C_6 and larger) are considered prototype liquid biofuels for CI engines. They can be obtained from etherification of primary alcohols like propanol, butanol and pentanol.

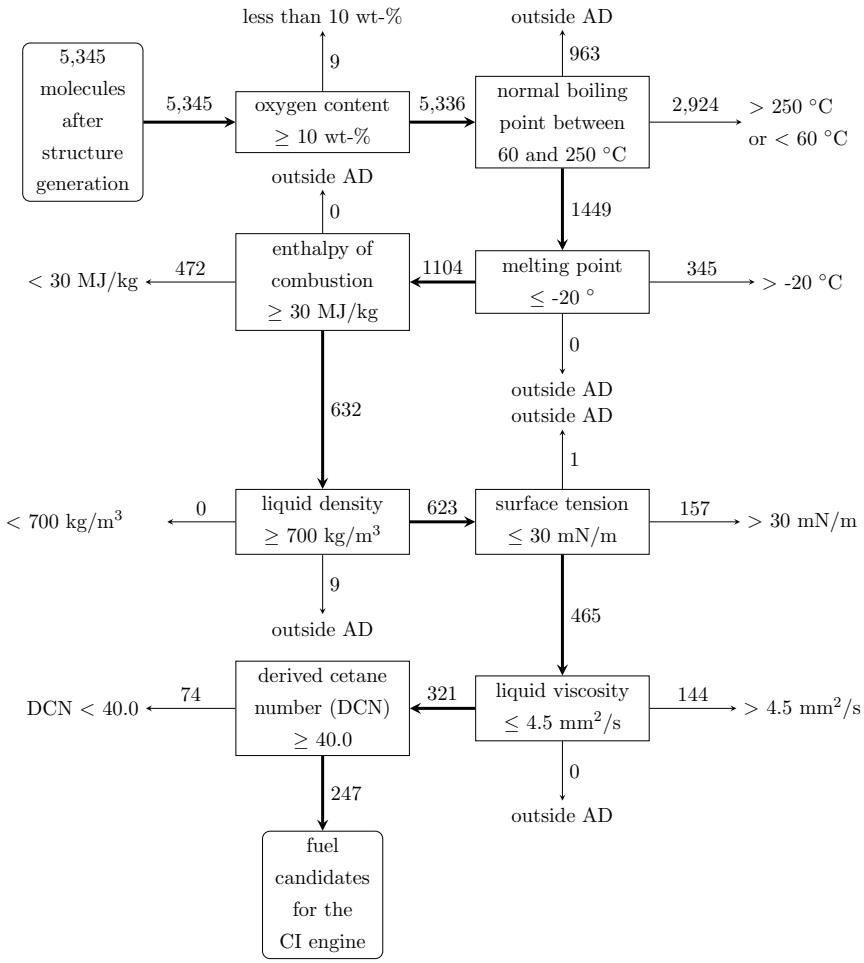
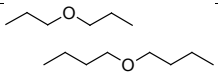
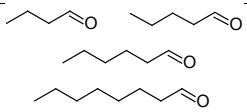
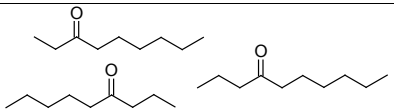
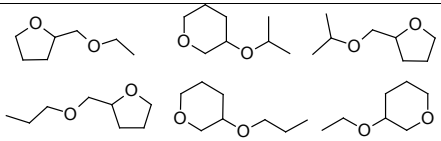
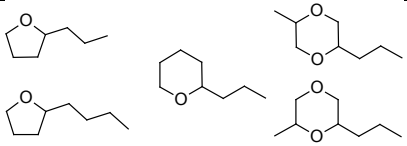
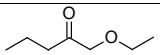
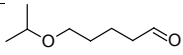


Figure 6.6: CI engine: Virtual fuel screening performed by means of QSPR and group contribution based property prediction (AD denotes applicability domain of respective QSPR model).

Table 6.5: Seven types of fuel candidates for the **CI engine**.

type	motif and exemplary candidates
I	 ethers (C_6 and larger)
II	 aldehydes (C_4 and larger)
III	 large ketones (C_9 and larger)
IV	 saturated cyclic ether plus ether side chain
V	 saturated cyclic ether plus alkane side chain
VI	 ether group plus ketone group
VII	 ether group plus aldehyde group

Note: The entire set of 247 fuel candidates can be found in the Supporting Information of Dahmen and Marquardt (2016).

Aldehydes with four or more carbon atoms satisfy the imposed property constraints. They are denoted type II molecules (cf. Table 6.5). Compared to their alkane counterparts, aldehydes are more prone to auto-ignition. For instance, the experimental DCN of butanal is 41.4 (cf. Appendix B). Similar to the linear ethers, the aldehydes have favorable mixture formation properties. However, unburnt fuel can result in massive engine-out aldehyde emissions. Since the formation of aldehydes during the incomplete combustion of other oxygenated fuel components is already a major concern (Magnusson and Nilsson, 2011), it is questionable whether aldehydes can represent potential biofuels, even if options for catalytic after-treatment are considered.

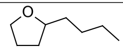
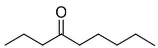
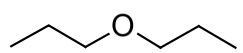
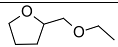
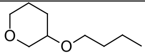
Although the ketone group retards auto-ignition (cf. Chapter 4), ketonic biofuels are considered a viable option for the CI engine, if the straight alkane backbone of the molecule is large enough (C_9 and larger) such that the DCN exceeds the threshold value of 40 (cf. type III structures in Table 6.5). Hence, bacterial methyl ketone synthesis targeting the C_{11} to C_{15} range has been proposed, recently (Goh et al., 2012). Chemo-catalytic synthesis starting from C_2-C_6 intermediates requires the formation of carbon-carbon bonds, e.g., via aldol condensation. However, large ketones have not been investigated in research engines yet. Ketonic CI biofuels would comprise high energy densities due to their low oxygen-to-carbon ratios. Boiling points in the lower range of fossil diesel fuel are expected.

If the furan motif, i.e., a structure relatively resistant to auto-ignition, is augmented with a linear ether side-chain of sufficient size (C_6 and larger), the predicted DCN exceeds the value of 40. However, such large furans are not included in Table 6.5, because they do not pass all of the AD filters. As the small furanic fuels, e.g., methylfuran and di-methylfuran, have been shown to posses a great tendency to form soot precursors (Sirignano et al., 2015), high levels of engine-out soot emissions might prohibit burning furanic fuels in the CI engine.

Type IV molecules from Table 6.5 describe saturated cyclic ethers with an additional ether side-chain of sufficient length (C_3 and larger). These compounds readily auto-ignite. For instance, Avantium (Gruter and De Jong, 2009), a spin-off of Royal Dutch Shell, reported a CN of 81-90 for ethyltetrahydrofurfuryl ether (De Jong, 2011; Lange et al., 2012). Moreover, the high oxygen content (> 20 wt-%) and the eliminated aromaticity are expected to result in less engine-out soot emissions. Predicted boiling points for these molecules range from 150 °C to 220 °C.

As can be seen from Figure 6.7, production of type IV molecules is considered to be based on etherification of alcohols. Primary alcohols are either fermentation products or can be derived from such. Furfural is converted into tetrahydrofurfuryl alcohol or tetrahydropyranol. In a production setting, some primary alcohols might also react to linear ethers (type I molecules). Similarly, some tetrahydrofurfurylalcohol might be converted to

Table 6.6: Selected pure-component biofuel candidates for the **CI engine**.

fuel candidate	**oxygen content [wt-%]**	**boiling point [°C]**	**DCN**	**lower heating value [MJ/kg]**	**density [kg/m^3]**	**intermediates**	**LHV eff. [%] (sugar/H_2 to fuel)**	**moles H_2 per mole of fuel**
fuel candidates from formation of carbon-carbon bonds								
 butyltetrahydrofuran (type V)	12.5	153	50.2	37.0	851	furfural, lactic acid	83	7
 4-nonanone (type III)	11.2	185	41.3	37.6	822	furfural, 2,3-butanediol, succinic acid	81 – 83	5 – 9
fuel candidates from formation of ether links								
 di-propylether (type I)	15.7	89	58.8	36.2	741	1,3-propanediol, lactic acid	78 – 92	2 – 6
 2-(ethoxymethyl)tetrahydrofuran (type IV)	24.6	156	60.0	30.4	921	furfural, ethanol	83	3
 2-butoxytetrahydropyran (type IV)	20.2	200	76.0	32.9	905	furfural, succinic acid, ethanol	83 – 85	3 – 8

Note: The entire set of 247 fuel candidates generated can be found in the Supporting Information of Dahmen and Marquardt (2016).

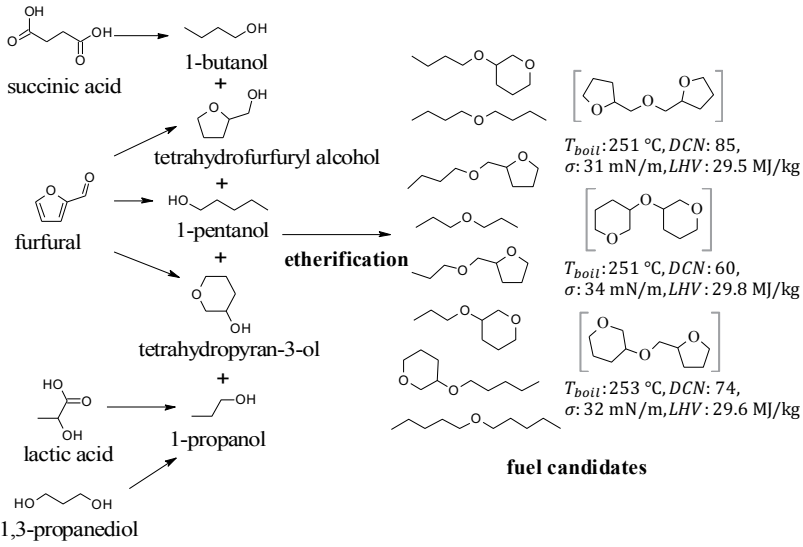


Figure 6.7: Potential pathways from acids, diols and furfural to fuel candidates for the CI engine (predicted properties for the candidates in brackets are slightly outside the desired range).

the compound depicted in the upper right of Figure 6.7. However, based on the predictions of the pure-component properties, it is likely that most practical mixtures of the fuel candidates depicted in Figure 6.7 would constitute feasible biofuels.

Type V molecules from Table 6.5 are cyclic saturated ethers with an alkane side-chain. Compared to their type IV counterparts, the DCN is somewhat lower due to the missing ether group in the side-chain. Butyltetrahydrofuran has been proposed as a lignocellulosic biofuel due to its favorable experimental DCN of 45.5 (Sudholt et al., 2015). Having a relatively low oxygen-to-carbon ratio, this compound has a high energy density (37.0 MJ/kg). Increasing the length of the side-chain is expected to raise both DCN and energy density even further. Butyltetrahydrofuran can be synthesized from furfural and acetone via aldol condensation (Julis and Leitner, 2012).

Molecular structures of type VI and VII are obtained, if the acyclic ether group appears together with either a ketone or an aldehyde group (cf. Table 6.5). Such polyfunctional compounds might pave the way to a variety of new fuel candidates and corresponding synthetic pathways. However, experimental data for structurally similar fuels are missing. Thus, the predicted properties, in particular the auto-ignition quality, must be experimentally validated in a first step.

Table 6.6 summarizes key properties of five exemplary fuel candidates belonging to the four most promising types, i.e., types I, III, IV and V. Compared to the SI engine, the range of potential biofuel molecules for the CI engine is significantly larger. This is primarily due to the less restrictive volatility constraint. The acyclic and cyclic ethers are probably the most interesting candidates, because (i) they do not require C−C coupling, (ii) a variety of them can be made from existing platform molecules, and (iii) the high oxygen content and the ether functional group should effectively reduce engine-out soot emissions.

Pragmatically, the size of the case study had been constrained by looking exclusively at straight-chain molecules with up to ten carbon atoms. It shall be noted, however, that branching can be used to achieve lower melting points since, as a general rule, the degree of molecular symmetry is positively correlated with the melting point of a pure compound (Brown and Brown, 2000). Similarly, branched esters are known to have lower cloud and pour points than their straight-chain counterparts (Lee et al., 1995; Knothe, 2005). There is no compelling reason against fuel candidates having more than ten carbon atoms. However, it is conceivable that the selective synthesis of a specific long-chained fuel molecule becomes more challenging as the number of coupling steps to be performed increases.

Interestingly, the case study suggests, that the production of CI fuels strictly requires coupling of intermediates, if existing carbon-carbon bonds in the chosen platform molecules shall not be altered. The coupling is essential for meeting auto-ignition requirements and can be achieved by the formation of carbon-carbon bonds or ether links. This observation even holds true for the production of renewable di-methylether, i.e., the smallest fuel molecule exhibiting a diesel-like CN. Obviously, the availability of linear ether or long-chained platform molecules would eliminate the coupling step. Such novel building blocks would therefore be of great interest for the synthesis of CI fuels.

6.3 Conclusions and outlook

In this Chapter, two case studies have illustrated the capabilities of the generate-and-test approach for computer-aided molecular design (CAMD) of tailor-made fuel components as outlined in this thesis (cf. Figure 1.3). To this end, desired ranges of key engine-related fuel properties, a set of bio-based platform chemicals and a set of carbon-skeleton preserving refunctionalization rules had to be specified for both spark-ignition (SI) and compression-ignition (CI) engines. The algorithm for targeted structure generation (cf. Chapter 3) has then been used to explore a spectrum of potential products, which has subsequently been narrowed down to a set of fuel candidates by means of a virtual fuel screening.

Compact furans, ketones and esters have emerged as first choice candidates for the SI

engine, since these molecules exhibit high volatilities, gasoline-like enthalpies of vaporization and high RONs. While the small ketonic biofuels are extremely knock-resistant, large ketones can achieve diesel-like CNs thanks to long, straight carbon backbones. Acyclic and cyclic ethers constitute viable options for volatile CI fuels, which can have very high DCNs and can be derived from a range of bio-derived intermediates, in principle. Based on the investigation of dibutylether, there is reasonable hope that such oxygenated fuels will enable nearly NOx- and soot-free combustion without the need for full mixture homogenization.

The large-scale, cost-effective and energy-efficient production of biofuel components requires further optimization of fermentation and chemo-catalytic pathways. Each collection of today's most promising platforms will continuously evolve as new molecules and pathways are proposed. With regard to the use of the fuel, a better understanding of the fundamental relationships between physicochemical fuel properties and the performance of a particular internal combustion engine (cf. discussion in Subsection 2.1.3) will allow to iteratively refine the definition of a tailored biofuel component. A particular strength of the outlined computational approach is its flexibility which allows to incorporate further physicochemical fuel properties, bio-derived platform chemicals and refunctionalization patterns.

7 Blend formulation by simultaneous product and pathway design

In an attempt to optimize both production and quality of the fuel, an integrated product and pathway design problem is posed on the molecular level in this Chapter. The complexity of this design problem is driven (i) by the rich variety of oxygenated fuel components that can be obtained from refunctionalization of bio-derived sugars in principle (cf. Chapters 3 and 6), (ii) by the numerous conversion pathways that connect these molecules with the biomass feedstock, and (iii) by the interactions between physicochemical fuel properties and the performance of an internal combustion engine (cf. Section 2.1). Hechinger et al. (2010) have suggested to tackle this problem by combining CAMD for de novo generation of biogenic fuel candidates exhibiting tailored properties with reaction network flux analysis (RNFA) (Voll and Marquardt, 2012b), i.e., an optimization-based methodology to identify the most promising routes for fuel production based on a network of competing reaction pathways. To this end, CAMD and RNFA have been applied consecutively or iteratively, however, have been mostly confined to the identification of pure-component fuels (Hechinger et al., 2010; Voll and Marquardt, 2012a,b; Hechinger et al., 2012b; Dahmen et al., 2012; Voll, 2014; Hechinger, 2014; Ulonska et al., 2016b; Dahmen and Marquardt, 2016). A first attempt at bringing the two approaches together to solve a simultaneous product and pathway design problem has been made by Zalfen (2014) in a student project supervised by the author of this thesis. Here, a linear optimization problem involving maximization of the production yield of a multicomponent fuel has been formulated by adding linear mixing rules for certain fuel properties to a simplified variant of the original RNFA model. The pure-component properties had been estimated with the models from Dahmen et al. (2012). Once the solution to the linear problem had been determined, the boiling curve of the mixture has been analyzed with the help of the UNIFAC group contribution model (Fredenslund et al., 1975; Hansen et al., 1991). Furthermore, similar to Yunus et al. (2014), the phase stability algorithm presented by Conte et al. (2011) has been applied to assess the miscibility of the multicomponent mixture by analyzing the miscibility of all binary pairs. The approach followed by Zalfen (2014), however, has multiple shortcomings. First, Zalfen (2014) did not calculate a batch distillation curve, which is typically used to

rate the volatility of a multicomponent fuel (ASTM D86, 1999), but instead considered the simpler case of an isobaric boiling process in a closed system based on repeated phase equilibria calculations. Second, as a consequence of the sequential analysis, constraints with regard to the fuel's boiling characteristics and phase stability could not be considered in the optimization problem, i.e., the actual design step. The integration of these aspects into the design problem has been touched upon only as part of the outlook on further work. And third, the original RNFA yield formulation is problematic in the context of mixture design because it does not take selectivity limitation of pathways into account. This issue will be discussed in detail below. It shall be noted that the addition of linear mixing rules to the original RNFA formulation has also been practiced by Victoria Villeda (2017) in the context of automatic reaction network generation and analysis. Here, the motivation has been to integrate streams otherwise considered waste into the product stream, thereby generating a biofuel mixture instead of a pure-component fuel. Again, the pure-component properties had been predicted with the help of the QSPR models published by Dahmen et al. (2012).

In the present Chapter, a new, more sophisticated attempt is made to formulate and solve a simultaneous product and pathway design problem. The resulting nonlinear program includes a local, necessary phase stability criterion, linear and nonlinear mixing rules, as well as constraints for a fuel's distillation curve. Furthermore, an improved pathway model takes both conversion and selectivity of each production pathway into account and the miscibility of the resulting multicomponent mixture is analyzed by means of global minimization of the so-called tangent plane distance function (McDonald and Floudas, 1995; Wasylkiewicz et al., 1996).

This Chapter has already been published in a slightly modified form in *Energy & Fuels* (Dahmen and Marquardt, 2017). The Chapter is structured as follows: Existing methodologies for mixture CAMD in the publicly available literature are briefly reviewed before the blend design framework for TMFB fuels is proposed. Three important aspects of the methodology are then described in detail. First, the concept of conversion pathway maps is introduced which is based on RNFA. Second, models for the prediction of mixture properties based on QSPR and GC estimates will be discussed. Third, the mathematical problem formulation is derived and the solution strategy is explained. The blend design framework is then applied to a case study in order to demonstrate the feasibility of the approach. The case study is dedicated to the identification of a 100%-renewable fuel mixture for the spark-ignition (SI) engine and respective production routes. In this scenario, the energy of fuel produced for a given amount of biomass shall be maximized. In order to meet this objective, renewable hydrogen from carbon-free energy sources such as wind or solar shall be provided. Finally, some conclusions and discussions are provided.

7.1 Existing methodologies for computer-aided mixture design

A comprehensive methodology for computer-aided mixture design and formulation of blended products has been proposed by Gani and co-workers (Karunanithi et al., 2005; Conte et al., 2011; Yunus et al., 2014). In their framework, a mixture design problem is formulated as a mixed-integer nonlinear program which is solved by means of a decomposition-based approach, i.e., the solution to the original problem is sought by considering a series of solvable subproblems and therefore by performing a sequential reduction of the number of blend candidates. This approach has been applied to a variety of case studies including the formulation of gasoline- and diesel-like fuels where the design objective has been to find optimal blend components for a so-called main ingredient, i.e., a fossil gasoline/diesel pseudo-compound that must constitute the main part of all mixtures under consideration (Yunus et al., 2011, 2012; Ariffin Kashinath et al., 2012; Yunus et al., 2014; Yunus, 2014; Phoon et al., 2015). In the aforementioned works, fixed cost factors ($/L) assigned to the individual blend components have been the only process-related criteria explicitly included in the optimization problem. Recently, Ng et al. (2015a) have developed a systematic methodology for mixture design in the context of integrated biorefineries that includes the identification of optimal conversion pathways via superstructure optimization once an optimal mixture has been determined based on CAMD techniques. Instead of such a two-step approach, Daoutidis and co-workers (Marvin et al., 2013; Daoutidis et al., 2013) have proposed a strategy that simultaneously identifies gasoline blends with desired properties and the corresponding chemistries by combining an automatic rule-based generator of reaction networks with an RNFA-like approach. The design problem is formulated as a mixed-integer linear problem typically having a large number of alternative solutions.

Similar to the works of Gani, Ng, Daoutidis and respective co-workers, the blend design problem is formulated as a nonlinear program (NLP) in this Chapter. In addition to the physicochemical fuel properties listed in Table 2.1, the Reid vapor pressure and the distillation profile of a fuel mixture are taken into account because these two properties constitute important fuel performance indicators for engine cold-start, vapor lock and oil dilution in spark-ignition (SI) engines (Kalghatgi, 2014a; ASTM D4814, 2016).

7.2 Blend design framework for tailor-made fuels

The blend design framework for TMFB fuels is depicted in Figure 7.1 and comprises four stages. In stage 1, a set of so-called palette compounds has to be selected. The term *palette compound* refers to any compound that could be part of a rationally formulated blend, in

principle. The set of palette compounds therefore defines the search space for the blend design. Palette compounds can either be assembled manually or generated automatically by the approach described in Chapter 3. Criteria for deciding whether a molecule should be included in the set of palette compounds include pure-component properties, yield data as well as additional process-related data if available.

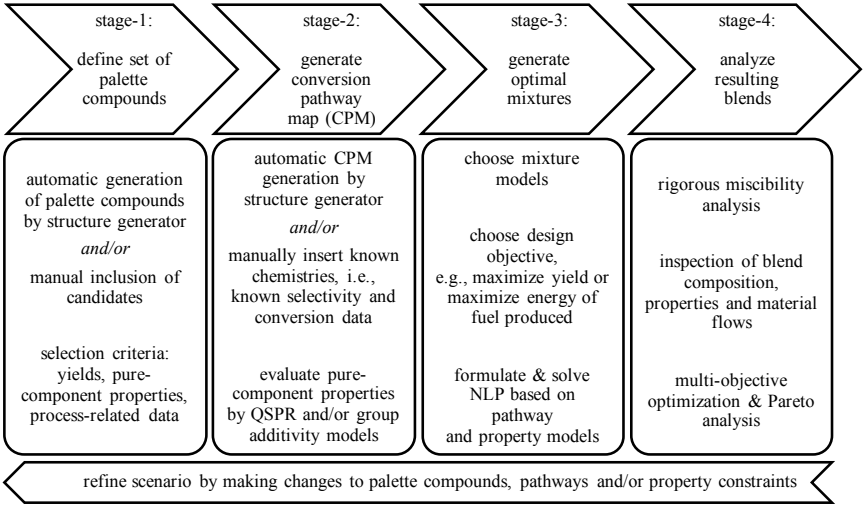


Figure 7.1: The blend design framework: Simultaneous product and pathway design.

Based on the set of palette compounds, a conversion pathway map (CPM) is created in stage 2. The CPM holds all conversion pathways to be considered in fuel production. Its basic structure could also be generated algorithmically since the generator of molecular structures described in Chapter 3 keeps track of substrate/product-relationships. This includes the case where a single product can be derived from multiple bio-derived platform chemicals as well as the case where multiple platform molecules have to be refunctionalized and joined into a single target molecule. Known performance measures of catalytic systems, i.e., selectivity and conversion data, have to be assigned manually to each pathway included in the CPM. QSPR and GC models from Chapters 4 and 5 are utilized to predict pure-component property data for palette compounds where experimental data are not available.

Stage 3 deals with setting up and solving the actual mixture design problem which is a nonlinear program (NLP). Different process-related objective functions, e.g., maximization of production yield or maximization of the energy of fuel produced for a fixed feed of biomass, can be considered here. Mathematical constraints include the pathway model,

i.e., mass-balances of all species together with selectivity data and conversion limits, as well as the property models to predict the blend's fuel properties. Due to nonlinearity in the property models, a sequential solution strategy is pursued.

Stage 4 is dedicated to a detailed analysis of the resulting mixtures. This includes the inspection of blend compositions, fuel properties and material flow diagrams. While a local, necessary phase stability criterion is included in the NLP formulation, a more rigorous stability analysis is performed *a posteriori* in stage 4 to check both necessary and sufficient conditions for miscibility. Multi-objective optimization and Pareto analysis can examine the trade-off between competing objectives, e.g., between maximization of yield and maximization of the fuel's knock resistance. After analyzing the proposed product and pathway designs, the modeler may choose to refine the scenario definition, for instance, by making changes to the set of palette compounds and/or fuel specifications or by including additional conversion pathways. In the following Sections, key elements of the blend design framework depicted in Figure 7.1 will be explained in more detail, starting with the concept of conversion pathway maps.

7.3 Pathway model

Figure 7.2 shows the graphical illustration of an exemplary conversion pathway map (CPM). The nodes of the CPM include biomass, hexoses, pentoses, platform molecules (shaded in gray) and palette compounds (encircled by dashed lines). By-products and intermediates are not considered here since the focus lies strictly on fuel production. Water, CO_2 and hydrogen are omitted in the graphical representation for the sake of simplicity.

7.3.1 Extension of reaction network flux analysis

The mathematical representation of the CPM is based on the methodology of reaction network flux analysis (RNFA) proposed by Voll and Marquardt (2012b). Analogous to RNFA, the stationary mole balances of a CPM comprising N_N nodes and N_P pathways can be formulated as

$$\mathbf{A} \cdot \mathbf{f} = \mathbf{b} \; . \tag{7.1}$$

Here, $\mathbf{A} \in \mathbb{R}^{N_N \times N_P}$ is the matrix of stoichiometric coefficients $\upsilon_{i,h}$, where the rows correspond to the $i \in \{1, ..., N_N\}$ nodes, i.e., reactants, and the columns correspond to the $h \in \{1, ..., N_P\}$ pathways, i.e., reactions. The vector $\mathbf{f}$ holds the molar fluxes over all pathways, whereas the vector $\mathbf{b}$ describes the molar product flows (Voll, 2014).

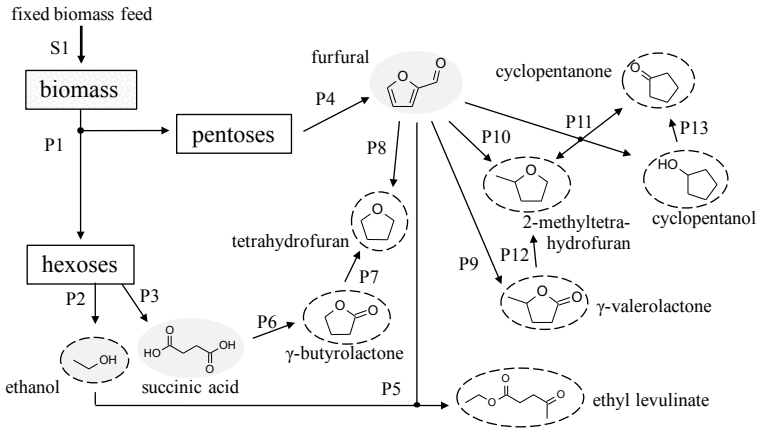


Figure 7.2: Graphical illustration of an exemplary conversion pathway map (CPM). A platform molecule (shaded in gray) can also be a palette compound (encircled by dashed lines), like in the case of ethanol. Note that other molecules, i.e., intermediates, by-products, water, CO_2 and hydrogen, are omitted in the graphical representation for the sake of simplicity.

In the original RNFA (Voll and Marquardt, 2012b), yield constraints are formulated to account for the fact that ideal reaction performance as described by Eqn. (7.1) is rarely achieved. These yield constraints act as conversion limits, i.e., selectivities S_h are assumed to be one and yields Y_h are effectively identical to conversions C_h, because of $Y_h = C_h \cdot S_h$. However, in practice, yields are more often constrained by low selectivity rather than by low conversion. In such cases, the yield constraints can lead to unrealistic results, especially in the context of mixture design where a single reactant can often be converted into a multitude of fuel components via independent pathways. To overcome this issue, both conversion C_h and selectivity S_h of each pathway h can be included in the model formulation. By introducing an artificial mismatch into the stoichiometric balance of a pathway h, selectivities S_h smaller than one can be accounted for. For instance, the pseudo-reaction for pathway $P6$ (cf. Figure 7.2)

$$\mathrm{C_4H_6O_4\ (succinic\ acid)\ } + 2\ \mathrm{H_2} \rightarrow S_{P6} \cdot (\ \mathrm{C_4H_6O_2\ (\gamma\text{-butyrolactone})} + 2\ \mathrm{H_2O}\ ) \quad (7.2)$$

recognizes the fact that unspecified by-products are created if $S_{P6} < 1$. Thus, in the model, the selectivity information is encoded in the matrix $\mathbf{A}$.

Regarding the formulation of pseudo-reactions it is therefore necessary to distinguish between co-products that constitute fuel components and unspecified byproducts and/or

waste. The fuel component co-product is considered a palette compound. For instance, pathway P11 in Figure 7.2 has one main product (cyclopentanone) and two co-products (cyclopentanol and 2-methyltetrahydorfuran). All three species constitute palette compounds and are balanced accordingly. Byproducts, most notably chemical products, and waste, however, are neglected in the pseudo-reactions for two reasons. First, the focus here lies strictly on fuel production and, as such, the analysis of an integrated bio-refinery that produces both fuel and chemicals is considered beyond the scope of this contribution, but offers an obvious extension of the suggested methodology. Second, whereas selectivity and conversion data for the main products of a conversion pathway can be extracted from the available literature in the great majority of cases, comprehensive (quantitative) information about the spectrum of byproducts is mostly lacking.

If the case of limited conversion shall be modeled, the following general form of a conversion constraint for a limiting reactant i in pathway h is used:

$$0 \geq \frac{-\upsilon_{i,h} \cdot f_h}{\sum_{k \neq h} \upsilon_{i,k} f_k} - C_h \ . \tag{7.3}$$

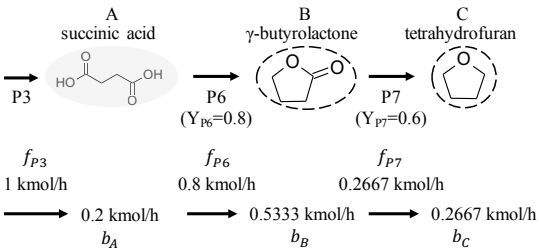


Figure 7.3: Original RNFA formulation with **yield** constraints (Voll and Marquardt, 2012b): Two-step conversion of succinic acid (species A) into γ-butyrolactone (species B) and tetrahydrofuran (species C).

The extension of RNFA from yield to selectivity and conversion limitation is best illustrated using a simple example, i.e., the two-step conversion of succinic acid (species A) into γ-butyolactone (species B) and tetrahydrofuran (species C) as depicted in Figure 7.3. A mole flow f_{P3} of 1 kmol/h succinic acid is supplied to this sample network. Note that both γ-butyrolactone and tetrahydrofuran are considered palette compounds. The pathway yields Y_{P6} and Y_{P7} shall be 0.8 and 0.6, respectively. The stoichiometric matrix **A** for this example reads

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \tag{7.4}$$

and the RNFA yield constraints (Voll, 2014) for pathways P6 and P7 are formulated as

$$Y_{P6} \geq \frac{\upsilon_{B,P6} \cdot f_{P6}}{\sum_{k \neq P6} \upsilon_{A,k} \cdot f_k} = \frac{1 \cdot f_{P6}}{1 \cdot f_{P3} + 0 \cdot f_{P7}} = \frac{f_{P6}}{f_{P3}} \qquad \Rightarrow f_{P6} \leq Y_{P6} \cdot f_{P3} \ , \tag{7.5}$$

$$Y_{P7} \geq \frac{\upsilon_{C,P7} \cdot f_{P7}}{\sum_{k \neq P7} \upsilon_{B,k} \cdot f_k} = \frac{1 \cdot f_{P7}}{0 \cdot f_{P3} + 1 \cdot f_{P6}} = \frac{f_{P7}}{f_{P6}} \qquad \Rightarrow f_{P7} \leq Y_{P7} \cdot f_{P6} \ . \tag{7.6}$$

Considering the simple objective of maximizing the mole flow of a blend composed of B and C, the linear program can then be written as

$$\max_{\mathbf{f} \geq 0, \mathbf{b} \geq 0} \quad (b_B + b_C) \tag{7.7}$$

$$\text{s.t.} \quad \mathbf{A} \cdot \mathbf{f} = \mathbf{b} \tag{7.8}$$

$$f_{P3} = 1 \tag{7.9}$$

$$f_{P6} - Y_{P6} \cdot f_{P3} \leq 0 \tag{7.10}$$

$$f_{P7} - Y_{P7} \cdot f_{P6} \leq 0 \quad . \tag{7.11}$$

The solution to this problem, i.e., vectors $\mathbf{f}$ and $\mathbf{b}$ yielding the maximum product flow of 0.8 kmol/h (composed of 0.5333 kmol/h of γ-butyolactone and 0.2667 kmol/h of tetrahydrofuran), is shown in Figure 7.3. Thus, the result is a binary blend composed of γ-butyolactone and tetrahydrofuran with the molar ratio of 2:1. Note that there is also unconverted succinic acid leaving the network (0.2 kmol/h).

Separately accounting for selectivity S_h and conversion C_h allows to investigate the scenario where yield is actually limited due to selectivity and not due to conversion, i.e., S_{P6}=0.8, S_{P7}=0.6, C_{P6}=1.0 and C_{P7}=1.0. Then the stoichiometric matrix $\mathbf{A}$ needs to be modified to

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0.8 & -1 \\ 0 & 0 & 0.6 \end{bmatrix} \tag{7.12}$$

and the optimization problem is rewritten as

$$\max_{\mathbf{f} \geq 0, \mathbf{b} \geq 0} \quad (b_B + b_C) \tag{7.13}$$

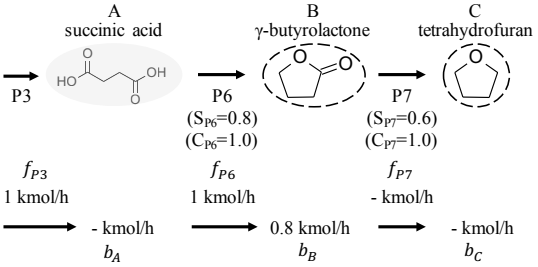


Figure 7.4: Modified RNFA formulation with **selectivity** constraints: Two-step conversion of succinic acid (species A) into γ-butyrolactone (species B) and tetrahydrofuran (species C).

$$\text{s.t.} \quad \mathbf{A} \cdot \mathbf{f} = \mathbf{b} \tag{7.14}$$

$$f_{P3} = 1 \quad . \tag{7.15}$$

The solution to this scenario is shown in Figure 7.4. Instead of a binary mixture, pure γ-butyolactone is produced here. Moreover, there is no succinic acid left.

Finally, it shall be assumed that there is also a conversion limitation for pathway P6, i.e., $C_{P6} = 0.9$. Adapting the general form of a conversion constraint (cf. Eqn. (7.3)) to this example yields

$$0 \geq \frac{-\upsilon_{A,P6} \cdot f_{P6}}{\sum_{k \neq P6} \upsilon_{A,k} f_k} - C_{P6} = \frac{f_{P6}}{f_{P3}} - C_{P6} \quad \Rightarrow f_{P6} \leq C_{P6} \cdot f_{P3} \; . \tag{7.16}$$

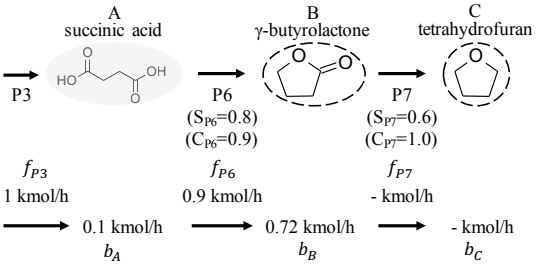


Figure 7.5: Modified RNFA formulation with **selectivity and conversion** constraints: Two-step conversion of succinic acid (species A) into γ-butyrolactone (species B) and tetrahydrofuran (species C).

Note that Eqn. (7.16) gives a constraint that is identical to the one stated in Eqn. (7.5),

thus demonstrating that the original RNFA yield constraints act as conversion constraints. The optimization problem is reformulated as

$$\max_{\mathbf{f}\geq 0, \mathbf{b}\geq 0} \quad (b_B + b_C) \tag{7.17}$$

$$\text{s.t.} \quad \mathbf{A}\cdot\mathbf{f} = \mathbf{b} \tag{7.18}$$

$$f_{P6} - C_{P6}\cdot f_{P3} \leq 0 \tag{7.19}$$

$$f_{P3} = 1 \quad . \tag{7.20}$$

Figure 7.5 shows the solution to this problem. As expected, unconverted succinic acid (0.1 kmol/h) leaves the network. In combination with the selectivity constraints for pathways P6 and P7, the conversion limitation leads to a product flow of only 0.72 kmol/h (pure γ-butyrolactone).

7.3.2 Biomass supply

It is assumed that a constant flow of biomass is supplied to the CPM via a pathway S1 to be included in the stoichiometric matrix **A**. Although being omitted in the graphical illustration of the CPM (cf. Figure 7.2), a hydrogen supply pathway is added to matrix **A**, thus allowing to balance the hydrogen demand associated with the production of a specific fuel blend. Analogously, a CO_2 mole balance is included to assess the amount of CO_2 generated from the biomass.

The supplied biomass is fractionated and depolymerized into hexoses and pentoses in the first conversion step (cf. Figure 7.2, pathway P1). A variety of processing concepts, e.g., Organosolv (Zhao et al., 2009), OrganoCat (Grande et al., 2015) or mechanocatalytic strategies (Käldström et al., 2014), have been proposed for this purpose. Although hemicellulose is a polymer of different sugars, its most abundant building block is xylan, i.e., a polymer of xylose (Huber et al., 2006). Therefore, hemicellulose is often pragmatically treated as a polymer of the monomer unit $C_5H_8O_4$, as can be seen from Table 7.1.

Table 7.1: Fractions of lignocellulosic biomass and corresponding monomer units (Petrus and Noordermeer, 2006).

biomass fraction	monomer unit	monomer molar weight [g/mol]
cellulose	$C_6H_{10}O_5$	162.14
hemicellulose	$C_5H_8O_4$	132.12
lignin	$C_{10}H_{12}O_3$	180.20

For a fixed, hypothetical biomass composition of 50 wt-% cellulose (polymer of glucose), 30 wt-% hemicellulose (polymer of xylose) and 20 wt-% lignin, 1 kmol of biomass monomers

($\sim$ 154.71 kg of biomass) can theoretically yield 0.477 kmol of hexoses ($C_6H_{12}O_6$) and 0.351 kmol of pentoses ($C_5H_{10}O_5$) after hydrolysis. Therefore, one can formulate the pseudo-reaction

$$\text{biomass} \rightarrow S_{P1,C_6} \cdot 0.477 \text{ hexoses } (C_6H_{12}O_6) + S_{P1,C_5} \cdot 0.351 \text{ pentoses } (C_5H_{10}O_5) \ , \tag{7.21}$$

which, in case of $S_{P1,C_6} = S_{P1,C_5} = 1$, would represent a perfect recovery of hexoses and pentoses. Note that the composition of beechwood is similar to the 50/30/20 composition used here (Couhert et al., 2009). The selectivity constants S_{P1,C_6} and S_{P1,C_5} can be used to adjust the yields of hexoses and pentoses depending on the performance of a practical fractionation and depolymerization process. Pragmatically, it is assumed that $S_{P1,C_6} = S_{P1,C_5} = 0.9$ based on a recent overview of different depolymerization approaches published by Luterbacher et al. (2014). Obviously, instead of a single conversion pathway P1, multiple fractionation and depolymerization pathways could be included in one CPM to account for the fact that different processing concepts might yield different shares of C_5 and C_6 sugars.

7.3.3 Calculation of process performance measures

In order to compute process-related performance indicators, e.g., hydrogen demand, mass of fuel produced or lower heating value (LHV) energy efficiency, the entries of the solution vector **b** referring to the amounts of palette compounds produced need to be investigated.

The total molar flow of fuel produced is defined as

$$\dot{n}_{fuel} = \sum_{i \in PC} b_i \ , \tag{7.22}$$

where $PC \in \{1, ..., n_C\}$ denotes the set of palette compounds included in the CPM. Similarly, the total mass flow of fuel produced can be derived from

$$\dot{m}_{fuel} = \sum_{i \in PC} b_i \cdot M_i \ , \tag{7.23}$$

where M_i denotes the molar weight of palette compound i. Then, mole (z_i) and mass (ξ_i) fractions of all blend components can be calculated from:

$$z_i = \frac{b_i}{\dot{n}_{fuel}} \qquad \forall i \in PC \tag{7.24}$$

$$\xi_i = \frac{b_i \cdot M_i}{\dot{m}_{fuel}} \qquad \forall i \in PC \tag{7.25}$$

The total energy flow of fuel produced is defined as

$$\dot{E}_{fuel} = \sum_{i \in PC} b_i \cdot M_i \cdot LHV_i \ , \tag{7.26}$$

where LHV_i is the lower heating value of blend component i.

The stoichiometric hydrogen demand for fuel production is computed based on f_{S2}, i.e., the molar flow of hydrogen supplied to the network. To this end, the constraint $b_{H_2} = 0$ is added making sure that there is no residual hydrogen leaving the network. It is convenient to normalize the mass-based hydrogen demand by division with the mass of fuel produced, i.e.,

$$m_{H_2} \frac{[\text{kg H}_2]}{[\text{kg fuel}]} = \frac{f_{S2} \cdot M_{H_2}}{\dot{m}_{fuel}} \ , \tag{7.27}$$

where M_{H_2} denotes the molecular weight of hydrogen (2.02 kg/kmol). Similarly, the mole-based hydrogen consumption can be stated as

$$n_{H_2} \frac{[\text{mol H}_2]}{[\text{mol fuel}]} = \frac{f_{S2}}{\dot{n}_{fuel}} \ . \tag{7.28}$$

The amount of CO_2 generated from the carbohydrate fraction of the biomass is related to the energy of fuel produced, i.e.,

$$m_{CO_2} \frac{[\text{g CO}_2]}{[\text{MJ fuel}]} = \frac{b_{CO_2} \cdot M_{CO_2}}{\dot{E}_{fuel}} \cdot 1000 \ , \tag{7.29}$$

where b_{CO_2} is the molar product flow of CO_2 leaving the network and M_{CO_2} denotes the molecular weight of carbon dioxide (44.01 kg/kmol).

Given a fixed, hypothetical supply of biomass of 1 kmol/h ($\sim$ 154.71 kg/h), the LHV efficiency (η_{LHV}) can be computed from

$$\eta_{LHV} = \frac{\dot{E}_{fuel}}{154.71 \cdot LHV_{biomass} + f_{S2} \cdot M_{H_2} \cdot LHV_{H_2}} \ , \tag{7.30}$$

where LHV_{H_2} denotes the lower heating value of hydrogen (120 MJ/kg). Based on the dry biomass composition of 50 wt-% cellulose (monomer $C_6H_{10}O_5$), 30 wt-% hemicellulose (monomer $C_5H_8O_4$) and 20 wt-% lignin (monomer $C_{10}H_{12}O_3$), the lower heating value $LHV_{biomass}$ in Eqn. (7.30) is estimated as 18.6 MJ/kg with the help of a regression model proposed by Demirbas et al. (1997). Typical lower heating values for wood reported in the literature range from 18.6 to 18.9 MJ/kg depending on type and source (Huber et al., 2006; Ptasinski et al., 2007).

7.4 Fuel property models

Predictive models are required to evaluate pure-component and mixture properties relevant to fuel design. The QSPR and GC models from Chapters 4 and 5 are used to predict the primary properties, i.e., those pure-component properties which are predicted solely based on the molecular structure. In contrast, secondary properties are derived on the basis of primary properties (and/or secondary properties) and additional relationships. It shall be noted that the outlined approach has been designed to be fully predictive. This means that the required component-specific input data can be estimated solely on the basis of the two-dimensional molecular graph with the given set of methods. This way, the blend design approach can also cope with novel compounds which have not been characterized experimentally yet.

Activity coefficients are required to describe the non-ideal behavior of mixtures with regard to Reid vapor pressure and distillation profile, two important measures for fuel volatility (Kalghatgi, 2014a; ASTM D4814, 2016). For the present contribution, the modified UNIFAC (Dortmund) model (Weidlich and Gmehling, 1987) has been implemented in Matlab (The MathWorks Inc., 2016). UNIFAC has been chosen because it is a widely-used group contribution method that is fully predictive and applicable to the variety of molecular structures considered here. Moreover, it is of limited computational complexity to be used in an optimization-based approach. Different variants of the UNIFAC approach have been used extensively in the context of CAMD and mixture design before (Gani and Brignole, 1983; Naser and Fournier, 1991; Van Dyk and Nieuwoudt, 2000; Conte et al., 2011; Yunus et al., 2014; Benavides et al., 2015). Only recently, first CAMD and mixture design approaches based on SAFT (Pereira et al., 2011; Lampe et al., 2015; Burger et al., 2015) and COSMO-RS (Scheffczyk et al., 2017; Austin et al., 2017) have been presented. In contrast to the original UNIFAC model developed by Fredenslund et al. (1975), the modified UNIFAC (Dortmund) model proposed by Gmehling and co-workers features temperature-dependent group interaction parameters and an empirically modified combinatorial part that is thought to offer a better description of asymmetric systems (Lohmann et al., 2001; Gmehling et al., 2012). In quantitative comparisons with the original UNIFAC formulation, modified UNIFAC (Dortmund) has demonstrated significantly improved results in different areas, including the prediction of vapor-liquid equilibria (VLE) and liquid-liquid equilibria (LLE) mole fractions (Lohmann et al., 2001; Gmehling et al., 2012). It is therefore used throughout this Chapter.

Linear blending rules are the simplest mixing rules and they are often used to describe fuel properties due to the lack of better alternatives (Mueller et al., 2012; Yunus et al., 2014). The error caused by the assumption of linear blending depends on the property and mixture under investigation. For instance, the error in case of the specific volume is

typically less than 3.5%, whereas for the liquid viscosity of a mixture the linear rule can only be expected to yield the correct order of magnitude (Gmehling et al., 2012). The following linear and nonlinear mixing rules for the liquid density ρ_L, the surface tension σ and the kinematic viscosity ν have been compiled based on the compendium of property modeling published by Gmehling et al. (2012). The chosen models are considered to be among the best in class under the premise that they are fully predictive and applicable to the diversity of molecular structures considered here.

In order to approximate the liquid density ρ_L [kg/m^3] of a mixture, a linear mixing rule is applied to the specific volume v_L [m^3/kg] where the contribution of the excess volume is neglected (Gmehling et al., 2012):

$$v_L = \sum_{i \in PC} \xi_i v_{L,i} \qquad \Rightarrow \quad \rho_L = \left(\sum_{i \in PC} \frac{\xi_i}{\rho_{L,i}} \right)^{-1} \tag{7.31}$$

According to Gmehling et al. (2012), this simplified density calculation typically leads to small errors and Eqn. (7.31) should be preferred over a linear mixing rule for the density ($\rho_L = \sum_{i \in PC} \xi_i \rho_{L_i}$).

The surface tension σ of a blend is computed with the help of a Parachor-based mixing rule (Gmehling et al., 2012), i.e.,

$$\sigma = \left(\rho_m \sum_{i \in PC} \sum_{j \in PC} z_i z_j \frac{(\rho_{m,i})^{-1} \sigma_i^{1/4} + (\rho_{m,j})^{-1} \sigma_j^{1/4}}{2} \right)^4 \qquad \text{[mN/m]} \ , \tag{7.32}$$

where ρ_m [mol/cm^3] is the molar density of the blend.

The blend's kinematic viscosity ν is calculated by assuming linear mixing for the dynamic viscosity μ (Gmehling et al., 2012), i.e.,

$$\nu = \frac{1}{\rho_L} \sum_{i \in PC} z_i \mu_i \ . \tag{7.33}$$

Some authors (Balabin et al., 2007; Kar et al., 2009) have proposed to employ the Clausius-Clapeyron equation to derive an estimate for the latent heat of vaporization H_{vap} of fuel mixtures. Assuming the ideal gas law for the vapor phase and further assuming the gas volume to be much larger than the liquid volume, i.e., $v_V >> v_L$, the Clausius-Clapeyron equation is written as (Gmehling et al., 2012)

$$H_{vap} = \frac{RT^2}{p^S} \frac{dp^S}{dT} \ , \tag{7.34}$$

where the vapor pressure of the mixture p^S, i.e., the bubble point pressure, at a given temperature T can be calculated from (Stichlmair and Fair, 1998)

$$p^S(T) = \sum_{i \in PC} z_i \cdot \gamma_i(T, \mathbf{z}) \cdot p_i^S(T) \ . \tag{7.35}$$

In Eqn. (7.35), p_i^S and γ_i denote the vapor pressure and activity coefficient of component i, respectively. Eqn. (7.35) can be differentiated to compute $\frac{dp^S}{dT}$ in Eqn. (7.34). However, strictly speaking, the Clausius-Clapeyron equation is valid for the phase transition of a pure compound at a constant temperature only. Chupka et al. (2015) have recently shown that the application of the Clausius-Clapeyron equation to ethanol-hydrocarbon mixtures will significantly underestimate the true heat of vaporization. This can be explained by the fact that the vapor pressure is predominantly influenced by the most volatile component, whereas the entire mixture is often wide-boiling (Chupka et al., 2015). As an alternative approach, Chupka et al. (2015) have suggested to employ the linear mixing rule

$$H_{vap} = \sum_{i \in PC} \xi_i \cdot H_{vap,i} \ , \tag{7.36}$$

which is based on the enthalpies of vaporization H_{vap_i} of the individual constituents of the blend. Such estimates were found to compare favorably with experimentally determined heats of vaporization, although the enthalpy of mixing is not accounted for by this approach (Chen and Stone, 2011; Chupka et al., 2015). Based on these findings, Eqn. (7.36) will be used to obtain an estimate for the heat of vaporization of a mixture in this Chapter, whereas the enthalpy of vaporization of a pure compound, i.e., $H_{vap,i}$, is computed from the Clausius-Clapyron equation (cf. Eqn. (7.34)). To this end, the vapor pressure p_i^S of a pure compound i at temperature T is calculated by means of the extended Antoine equation (Stichlmair and Fair, 1998; Gmehling et al., 2012), i.e.,

$$\ln\left(p_i^S(T)\right) = C_{ant,1,i} + \frac{C_{ant,2,i}}{T + C_{ant,3,i}} + C_{ant,4,i} \cdot T + C_{ant,5,i} \cdot \ln(T) + C_{ant,6,i} \cdot T^{C_{ant,7,i}} \ , \tag{7.37}$$

if parameter values $C_{ant,1,i}$, $C_{ant,2,i}$, $C_{ant,3,i}$, $C_{ant,4,i}$, $C_{ant,5,i}$, $C_{ant,6,i}$ and $C_{ant,7,i}$ are available. Alternatively, the vapor pressure is considered a secondary property and the Hoffmann-Florin (HF) equation (Hoffmann and Florin, 1943; Gmehling et al., 2012) is applied as part of a fully predictive approach (cf. Appendix E):

$$p_i^S(T) = f_{\mathrm{HF}}(T_{crit,i}, p_{crit,i}, T_{boil,i}, T) \qquad [\mathrm{kPa}] \tag{7.38}$$

The material specific inputs to Eqn. (7.38), i.e., the critical temperature $T_{crit,i}$, the critical pressure $p_{crit,i}$ and the normal boiling point $T_{boil,i}$, are primary properties that are predicted by means of the QSPRs from Chapter 5.

The lower heating value LHV and the derived cetane number DCN of the blend are calculated based on the assumption of linear mixing as well, i.e.,

$$LHV = \sum_{i \in PC} \xi_i \cdot LHV_i \ , \tag{7.39}$$

$$DCN = \sum_{i \in PC} z_i \cdot DCN_i \ . \tag{7.40}$$

Note that in Eqn. (7.40), the dimensionless DCN is weighted by mole fractions due to the fact that the mixture octane number (ON) of a gasoline surrogate model composed of n-heptane/iso-octane/toluene (Knop et al., 2014) as well as the DCN of mixtures of furanic species and n-heptane (Sudholt et al., 2015) have been observed to follow linear-by-mole blending rules. Nonlinear synergistic/antagonistic blending effects, however, have been reported for ON in case of ethanol and toluene (Foong et al., 2014; Dryer et al., 2014). Although cetane number of diesel fuel is generally believed to closely follow a linear blending rule (Kalghatgi, 2005), in principle, nonlinear effects are plausible from a kinetics point of view, especially if fuels with different chemistries are investigated.

7.4.1 Distillation curve model

Fuel volatility is important for in-cylinder mixture formation. In both spark-ignition (SI) and compression-ignition (CI) engines, too heavy components can cause engine oil dilution (Fisher et al., 2010; Thewes et al., 2011b; ASTM D4814, 2016). Especially in SI engines, high fuel volatility is critical in obtaining a well-mixed, ignitable fuel/air mixture. Engine cold-start and low ambient temperatures place particular demands on the fuel in this respect (Yanowitz et al., 2011; Thewes et al., 2011b; Hoppe et al., 2016a). Although the CI engine is generally capable of burning less volatile fuel (Dec, 2009), high volatilities seen in some biofuels and certain diesel blends have been linked to better mixture homogenization in low-temperature combustion concepts (Park et al., 2009; De Ojeda et al., 2011; Heuser et al., 2013a,b; Wang et al., 2014; Liu et al., 2014; Hoppe et al., 2016a). It shall be noted that the flash point is a critical volatility-related parameter of diesel fuel for safety reasons, as only at temperatures above the flash point (typically at least 55 °C) an ignitable mixture of gaseous fuel and air is formed (Kalghatgi, 2014a; EN 590, 2014; ASTM D975, 2016). In case of gasoline, the vapor above the liquid fuel is typically too rich to be flammable as a result of gasoline's high volatility (Kalghatgi, 2014a).

The standard way of measuring fuel volatility is to perform a batch distillation according to ASTM D86 (1999). This yields a so-called distillation curve, i.e., a (graphical) relationship between the temperature of the fuel vapor and the volume fraction distilled. Although ASTM D86 (1999) is the de facto standard for the determination of distillation curves of real fuels, the testing method suffers from design drawbacks which make it impossible to interpret the reported temperature as a true thermodynamic state point (Bruno, 2006; Mueller et al., 2012). In the 2000s, this has motivated Bruno and co-workers (Bruno, 2006; Smith and Bruno, 2007) to develop an improved experimental setup which they have called Advanced Distillation Curve (ADC) method. Subsequently, Huber et al. (2008) as well as Backhaus and Rothamer (2012) have demonstrated that results generated from simple open distillation (batch distillation) models (Stichlmair and Fair, 1998), which are combined with either equations of state or UNIFAC, closely follow experimental ADC data for binary mixtures of n-decane and n-tetradecane.

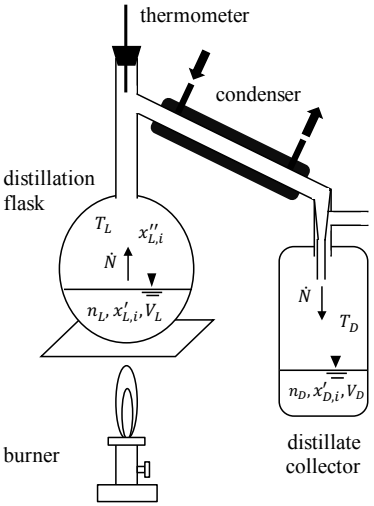


Figure 7.6: Conceptual model of an idealized batch distillation process.

For the purpose of blend design, such a model is integrated into the optimization-based approach. The idea is that the modeler can impose constraints on the volatility characteristics of the blended products to be formulated. Figure 7.6 represents a sketch of the considered idealized batch distillation process. The model equations are

$$\frac{dn_L}{dt} = -\dot{N} \ , \tag{7.41}$$

$$\frac{dn_D}{dt} = \dot{N} \ , \tag{7.42}$$

$$\frac{dn_{L,i}}{dt} = -\dot{N} x''_{L,i} \qquad \forall i = 1, ..., n_C - 1 \ , \tag{7.43}$$

$$\frac{dn_{D,i}}{dt} = \dot{N} x''_{L,i} \qquad \forall i = 1, ..., n_C - 1 \ , \tag{7.44}$$

$$\frac{dV_D}{dt} = \frac{\dot{N}}{\rho_m(T_D, x''_{L,i})} \ , \tag{7.45}$$

$$x''_{L,i} = \frac{x'_{L,i} \cdot \gamma_i(T_L, \mathbf{x'_L}) \cdot p_i^S(T_L)}{p} \qquad \forall i = 1, ..., n_C \ , \tag{7.46}$$

$$x'_{L,i} = \frac{n_{L,i}}{n_L} \qquad \forall i = 1, ..., n_C \ , \tag{7.47}$$

$$x'_{D,i} = \frac{n_{D,i}}{n_D} \qquad \forall i = 1, ..., n_C \ , \tag{7.48}$$

$$n_D = \sum_i n_{D,i} \ , \tag{7.49}$$

$$n_L = \sum_i n_{L,i} \ , \tag{7.50}$$

$$p = \sum_i x'_{L,i} \cdot \gamma_i(T_L, \mathbf{x'_L}) \cdot p_i^S(T_L) \ , \tag{7.51}$$

$$p = \sum_i x'_{D,i} \cdot \gamma_i(T_D, \mathbf{x'_D}) \cdot p_i^S(T_D) \ . \tag{7.52}$$

Eqns. (7.41) and (7.42) represent the total mole balances for the liquid (L) in the distillation flask and for the distillate (D) in the distillate receiver, whereas Eqns. (7.43) and (7.44) are the component mole balances for the liquid and the distillate, respectively (Stichlmair and Fair, 1998). At each point in time, the liquid is assumed to be at its bubble point (Huber et al., 2008). Mathematically, the vapor-liquid equilibrium can be written as Eqn. (7.46), where $x''_{L,i}$ denote the mole fractions of the vapor which is assumed to be in equilibrium with the liquid (Stichlmair and Fair, 1998). In Eqn. (7.46), the notation $\mathbf{x'_L} = [x'_{L,1}, x'_{L,2}, ..., x'_{L,n_C}]^{\mathrm{T}}$ is used to indicate that the mole fractions of all species are needed to calculate the activity coefficient γ_i of species i.

It is further assumed that the vapor stream $\dot{N}$ is immediately transferred to and totally condensed in the distillate receiver (Huber et al., 2008). The change in the distillate volume V_D is computed based on the liquid molar density of the condensate at the bubble point temperature T_D of the distillate in the distillate receiver (Huber et al., 2008). This allows to derive a differential equation for the volume of the distillate (cf. Eqn. (7.45)). The molar fractions of the liquid and the distillate are defined by Eqns. (7.47) and (7.48), whereas Eqns. (7.49) and (7.50) represent the closure conditions for the liquid and the distillate, respectively. The bubble point temperatures T_L and T_D have to be calculated by iteratively solving Eqns. (7.51) and (7.52) (Stichlmair and Fair, 1998). Beginning at $t = 0$ and setting $\dot{N} = 1$, the entire liquid is distilled at $t = 1$. The distillation is assumed

to take place at sea level atmospheric pressure ($p = 1.01325$ bar).

Again, the vapor pressure p_i^S of a pure compound i at temperature T (cf. Eqns. (7.46),(7.51)-(7.52)) is calculated by means of the extended Antoine equation (cf. Eqn. (7.37)). Alternatively, the Hoffmann-Florin (HF) equation (cf. Eqn. (7.38)) is applied as part of a fully predictive approach, if Antoine parameter values are not available.

The liquid molar density ρ_m in Eqn. (7.45) is described as

$$\rho_m(T, x''_{L,i}) = \left(\sum_i \frac{x''_{L,i}}{\rho_{m,i}(T)} \right)^{-1} , \tag{7.53}$$

where the liquid molar density $\rho_{m,i}$ of species i is computed from the DIPPR 105 equation (AIChE, 2012)

$$\rho_{m,i}(T) = \frac{C_{dip,1,i}}{C_{dip,2,i}^{1+(1-T/C_{dip,3,i})^{C_{dip,4,i}}}} , \tag{7.54}$$

if parameter values $C_{dip,1,i}$, $C_{dip,2,i}$, $C_{dip,3,i}$ and $C_{dip,4,i}$ are available. Alternatively, the COSTALD method (Hankinson and Thomson, 1979; Gmehling et al., 2012) is applied as part of a fully predictive approach (cf. Appendix E):

$$\rho_{m,i}(T) = f_{\text{COSTALD}}(T_{crit,i}, V_{crit,i}, p_{crit,i}, T_{boil,i}, T) \qquad [\text{kmol/m}^3] \tag{7.55}$$

The four material-specific inputs to Eqn. (7.55), i.e., critical temperature $T_{crit,i}$, critical volume $V_{crit,i}$, critical pressure $p_{crit,i}$ and the normal boiling point $T_{boil,i}$ are primary properties that are predicted by means of the QSPRs from Chapter 5.

Once the solution to Eqn. (7.45) is available, the evolution of the volume fraction distilled V_f is calculated based on the final volume of distillate $V_D(t = 1)$, i.e.,

$$V_f(t) = \frac{V_D(t)}{V_D(t = 1)} . \tag{7.56}$$

Having built a similar model, Huber et al. (2008) have observed a horizontal offset between experimental and computed data when plotting the temperature T_D over the volume fraction V_f. This offset is thought to correspond to the fluid transit delay, i.e., the time it takes for the vapor formed above the liquid in the kettle to travel upwards, to condense in the condenser and to be eventually collected in the distillate receiver. Huber et al. (2008) have proposed to correct for this delay by shifting the computed distillation profile leftwards by a constant volume increment V_{shift} to be found by manual adjustment.

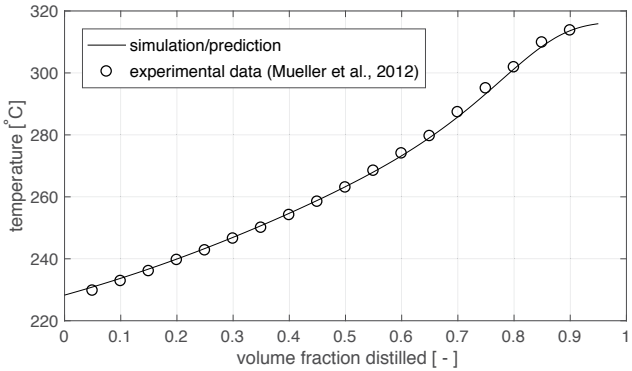


Figure 7.7: Distillation curve model applied to a diesel fuel surrogate proposed by Mueller et al. (2012) and compared with experimental data generated by the Advanced Distillation Curve (ADC) method. Experimental data are also taken from Mueller et al. (2012). Activity coefficients have been calculated by modified UNIFAC (Dortmund) (Gmehling et al., 1993, 1998; Constantinescu and Gmehling, 2016). Note: A volume shift V_{shift} of -0.05 has been applied to compensate for the fluid transit delay (Huber et al., 2008) .

Figure 7.7 compares computed and measured ADC data for a diesel fuel surrogate proposed by Mueller et al. (2012). The surrogate is composed of n-hexadecane, n-octadecane, iso-cetane, n-butylcyclohexane, trans-decalin, 1,2,4-trimethylbenzene, tetralin and 1-methylnaphthalene. If the volume increment V_{shift} is set to -0.05, the simulated distillation curve shows a nice agreement with the measured ADC data (cf. Figure 7.7).

7.4.2 Reid vapor pressure model

The Reid vapor pressure (RVP) as defined in ASTM D323 (1999) is another important measure for the volatility of gasoline fuels. RVP is highly relevant for the ability of performing engine start when the engine is cold relative to its design operating temperature (ASTM D4814, 2016). It is also important to prevent vapor-lock, i.e., the formation of fuel vapor bubbles in the fuel delivery system which leads to abnormal fuel flow to the combustion chamber (ASTM D4814, 2016). Conceptually, the Reid vapor pressure is the pressure p^S_{Reid} in a closed system in which the biphasic fuel is stored at a constant temperature of 37.8°C with a constant vapor-to-liquid volume ratio V_V/V_L of four (ASTM D323, 1999). Hatzioannidis et al. (1998) have shown that UNIFAC is generally capable of predicting the RVP of simulated gasoline blended with a variety of oxygenated additives and fuel extenders. More recently, Backhaus and Rothamer (2012) have employed UNIFAC for

RVP prediction as part of a design approach for drop-in gasoline fuels satisfying certain volatility constraints.

The Reid vapor pressure test can be modeled by the flash equations (Stichlmair and Fair, 1998):

$$n = n_L + n_V \tag{7.57}$$

$$n z_i = n_L x_i' + n_V x_i'' \tag{7.58}$$

$$x_i'' = K_i x_i' \tag{7.59}$$

In Eqn. (7.57), n is the total moles, whereas n_L and n_V are the moles of the liquid and the vapor phase, respectively. The overall mole fractions z_i are determined from Eqn. (7.58) based on the mole fractions for the liquid phase x_i' and for the vapor phase x_i''. The vapor-liquid equilibrium is described by the equilibrium ratio K_i in Eqn. (7.59). Inserting Eqns. (7.57) and (7.59) into Eqn. (7.58) and performing rearrangement gives (Stichlmair and Fair, 1998)

$$x_i' = z_i / \left[1 + (K_i - 1)\, n_V / n\right] \ . \tag{7.60}$$

The volume of the gas phase is determined from the ideal gas law, i.e., $V_V = n_V RT / p_{Reid}^S$, and the volume of the liquid phase is given by $V_L = n_L / \rho_m$, where the molar density of the liquid ρ_m is again calculated from Eqn. (7.53). The vapor-to-liquid volume ratio of four then leads to the following ratio between n_V and n:

$$4 \stackrel{!}{=} \frac{V_V}{V_L} = \frac{n_V RT \rho_m}{(n - n_V) p_{Reid}^S} \quad \Leftrightarrow \quad \frac{n_V}{n} = \frac{4 p_{Reid}^S}{RT\rho_m + 4 p_{Reid}^S} \tag{7.61}$$

Inserting Eqn. (7.61) into Eqn. (7.60), rewriting the K-factor $K_i = x_i''/x_i' = \gamma_i(\mathbf{x}') \cdot p_i^S / p_{Reid}^S$ (Gmehling et al., 2012), and adding the closure condition, one obtains the system of nonlinear equations

$$x_i' = z_i / \left(1 + \frac{4 p_{Reid}^S}{RT\rho_m + 4 p_{Reid}^S} \left(\frac{\gamma_i(\mathbf{x}') \cdot p_i^S}{p_{Reid}^S} - 1 \right)\right) \qquad \forall i = 1, ..., n_C \ , \tag{7.62}$$

$$1 = \sum_i x_i' \ , \tag{7.63}$$

which can be solved for x_i' and p_{Reid}^S, given a temperature of 37.8°C and the overall blend composition $\mathbf{z}$. Note that due to the fact that the ratio n_V/n is always small (cf. Eqn. (7.61)), the bubble point pressure p_{bp} of a mixture with composition $\mathbf{z}$ at 37.8°C, i.e.,

$p_{bp} = \sum_i z_i \cdot \gamma_i(\mathbf{z}) \cdot p_i^S$, is sometimes taken as an approximation of the Reid vapor pressure p_{Reid}^S (Yunus et al., 2014). Furthermore, the liquid mole fractions x_i' are similar to the total mole fractions z_i. Hence, the bubble point pressure p_{bp} and the z_i are used as initial values for p_{Reid}^S and x_i', respectively.

7.4.3 Phase stability criteria

The formulated fuel shall form one liquid phase under typical ambient conditions. To this end, the Gibbs energy change of mixing Δg as a function of mixture composition $\mathbf{z}$ and temperature T is considered here. Δg can be written as

$$\Delta g(T, \mathbf{z}) = RT \sum_{i \in PC} \left(z_i \cdot \ln(z_i) + z_i \cdot \ln(\gamma_i(T, \mathbf{z})) \right) \ , \tag{7.64}$$

where R denotes the ideal gas constant and activity coefficients $\gamma_i(T, \mathbf{z})$ describe the contribution of the excess Gibbs energy (Gmehling et al., 2012).

In case of a binary mixture composed of components 1 and 2,

$$\frac{\partial^2 \Delta g}{\partial z_1^2} > 0 \tag{7.65}$$

can be considered a local, necessary phase stability criterion for a blend with compositions z_1 and $z_2 = 1 - z_1$ (Bausa and Marquardt, 2000; Mitsos et al., 2009). However, if a miscibility gap does exist, it extends beyond the region where $\frac{\partial^2 \Delta g}{\partial z_1^2} < 0$ and the so-called tangent line criterion must be applied to reveal the gap's full dimension, i.e., the locally unstable region plus the metastable region (Baker et al., 1982; Segura et al., 2000; Conte et al., 2011).

Similarly, in the multicomponent case, the Hessian matrix $\mathbf{H}$ of the Gibbs energy change of mixing with respect to the $n_C - 1$ independent mole fractions can be analyzed to derive a local, necessary phase stability criterion (Van Dongen et al., 1983; Wasylkiewicz et al., 1996; Segura et al., 2000):

$$\lambda_i \left(\mathbf{H}(\Delta g(\mathbf{z})) \right) > 0 \qquad\qquad i = 1, ..., n_C - 1 \tag{7.66}$$

In Eqn. (7.66), λ_i denotes the i-th eigenvalue of $\mathbf{H} \in \mathbb{R}^{(n_C-1)\times(n_C-1)}$. Thus, if $\mathbf{H}$ is not positive definite, the mixture is unstable (Van Dongen et al., 1983). Furthermore, the minimization of the so-called tangent plane distance function f_{tpd} constitutes the more rigorous, global phase stability check (Wasylkiewicz et al., 1996):

$$\min_{0\leq x'_i\leq 1} \quad f_{tpd}(\mathbf{x}') = \sum_i x'_i \left[\ln x'_i + \ln \gamma_i(T,\mathbf{x}') - \ln z_i - \ln \gamma_i(T,\mathbf{z})\right] \tag{7.67}$$

$$\text{s.t.} \quad 1 = \sum_i x'_i \tag{7.68}$$

Assuming that $\mathbf{x}'^*$ is the global minimizer of Eqn. (7.67) such that the closure condition (cf. Eqn. (7.68)) holds, the necessary and sufficient condition for phase stability is $f_{tpd}(\mathbf{x}'^*) \geq 0$ (McDonald and Floudas, 1995; Wasylkiewicz et al., 1996). Due to the strong nonlinear behavior of Eqn. (7.67) and due to the existence of many local minima, several approaches to global optimization have been applied to the tangent plane stability problem in combination with equations of state, NRTL and UNIQUAC (McDonald and Floudas, 1995; Wasylkiewicz et al., 1996; Yushan and Zhihong, 1999; Harding and Floudas, 2000; Mitsos and Barton, 2007). In the present contribution, the solver *globalsearch* from Matlab's Global Optimization toolbox (The MathWorks Inc., 2016) is used to solve the problem defined by Eqns. (7.67) and (7.68). *Globalsearch* is an implementation of the OQNLP algorithm (Ugray et al., 2007), i.e., a multistart heuristic algorithm combining scatter search with local NLP solvers. Ugray et al. (2007) have reported that the OQNLP algorithm is capable of finding global solutions to all 142 continuous-variable NLP test problems compiled by Floudas et al. (1999). However, it should be noted that due to its heuristic nature the algorithm cannot guarantee that the global minimum of the tangent plane distance function f_{tpd} is identified.

7.5 Problem formulation and solution strategy

After pathway and property models have been defined, the blend design problem can now be formulated as a nonlinear program (NLP). Ultimately, the purpose of fuel production is the provision of energy for transport applications. Therefore, the objective is chosen as the maximization of the energy of fuel produced (in terms of lower heating value) for a given feed of biomass. Then, the problem can be stated as follows:

objective function:

$$\min_{\mathbf{f}\geq\mathbf{0},\mathbf{b}\geq\mathbf{0}} \quad -\sum_{j\in PC} b_j M_j LHV_j \qquad \text{(maximize energy of fuel produced)} \tag{7.69}$$

$$\text{s.t.} \tag{7.70}$$

pathway model:

$$\text{mass balance} \quad \mathbf{A} \cdot \mathbf{f} = \mathbf{b} \tag{7.71}$$

$$\text{biomass supply} \quad f_{S1} = 1 \tag{7.72}$$

$$\text{no residual } H_2 \quad b_{H_2} = 0 \tag{7.73}$$

definitions:

$$\text{mole fractions} \quad z_i = b_i / \left(\sum_{j \in PC} b_j \right) \qquad \forall i \in PC \tag{7.74}$$

$$\text{molar mass} \quad M = \sum_{j \in PC} z_j M_j \tag{7.75}$$

$$\text{mass fractions} \quad \xi_i = z_i M_i / M \qquad \forall i \in PC \tag{7.76}$$

$$\text{mass density} \quad \rho_L = 1 / \left(\sum_{j \in PC} \frac{\xi_j}{\rho_{L,j}} \right) \tag{7.77}$$

$$\text{molar density} \quad \rho_m = \frac{\rho_L}{M} \tag{7.78}$$

property constraints:

$$\text{derived cetane number} \quad DCN^{\text{lb}} \leq \sum_{j \in PC} z_j DCN_j \leq DCN^{\text{ub}} \tag{7.79}$$

$$\text{lower heating value} \quad LHV^{\text{lb}} \leq \sum_{j \in PC} \xi_j LHV_j \leq LHV^{\text{ub}} \tag{7.80}$$

$$\text{enthalpy of vaporization} \quad H_{vap}^{\text{lb}} \leq \sum_{j \in PC} \xi_j H_{vap,j} \leq H_{vap}^{\text{ub}} \tag{7.81}$$

$$\text{oxygen content} \quad wt_{O_2}^{\text{lb}} \leq \sum_{j \in PC} \xi_j wt_{O_2,j} \leq wt_{O_2}^{\text{ub}} \tag{7.82}$$

$$\text{liquid density} \quad \rho_L^{\text{lb}} \leq \rho_L \leq \rho_L^{\text{ub}} \tag{7.83}$$

$$\text{kinematic viscosity} \quad \nu^{\text{lb}} \leq \frac{1}{\rho_L} \sum_{j \in PC} z_j \mu_j \leq \nu^{\text{ub}} \tag{7.84}$$

$$\text{surface tension} \quad \sigma^{\text{lb}} \leq \left(\frac{\rho_L}{M} \sum_{i \in PC} \sum_{j \in PC} z_i z_j \frac{(\rho_{m,i})^{-1} \sigma_i^{1/4} + (\rho_{m,j})^{-1} \sigma_j^{1/4}}{2} \right)^4 \leq \sigma^{\text{ub}} \tag{7.85}$$

$$\text{approximated RVP} \quad p_{bp}^{\text{lb}} \leq \sum_{j \in PC} z_j \cdot \gamma_j(37.8°\text{C}, \mathbf{z}) \cdot p_j^S(37.8°\text{C}) \leq p_{bp}^{\text{ub}} \tag{7.86}$$

$$\text{dist. (10 mol-\% evap.)} \quad T10(m)^{\text{lb}} \leq \text{Dist.-Model}_{\text{T10(m)}}(\mathbf{z}) \leq T10(m)^{\text{ub}} \tag{7.87}$$

$$\text{dist. (50 mol-\% evap.)} \quad T50(m)^{\text{lb}} \leq \text{Dist.-Model}_{\text{T50(m)}}(\mathbf{z}) \leq T50(m)^{\text{ub}} \tag{7.88}$$

$$\text{dist. (90 mol-\% evap.)} \quad T90(m)^{\text{lb}} \leq \text{Dist.-Model}_{\text{T90(m)}}(\mathbf{z}) \leq T90(m)^{\text{ub}} \tag{7.89}$$

stability constraint:

$$\text{local, necessary condition} \quad \lambda_i \left(\mathbf{H} \left(\Delta g(298\text{K}, \mathbf{z}) \right) \right) \geq 0 \qquad i = 1, ..., n_C - 1 \tag{7.90}$$

The optimization variables are the flows $\mathbf{f}$ and $\mathbf{b}$. Thus, the task of the optimizer is to find a feasible combination of $\mathbf{f}$ and $\mathbf{b}$ that maximizes the energy of fuel produced (cf. Eqn. (7.69)). The property constraints are formulated on the basis of the n_C palette compounds. This means that, in principle, the solution to the optimization problem could be a blend of n_C components. In practical cases, however, the solution is typically composed of few species only, as the mole fractions of the other palette compounds become zero. Likewise, most entries of $\mathbf{f}$ and $\mathbf{b}$ are zero at the solution.

Constraints based on the pathway model (cf. Eqns. (7.71)-(7.73)) are followed by a set of definitions (cf. Eqns. (7.74)-(7.78)) which are given here for the sake of clarity, although, in principle, these could also be integrated into the subsequently listed property and stability constraints (cf. Eqns. (7.79)-(7.90)). Note that the pure-component properties, e.g., the derived cetane number, can be calculated *a priori*, whereas those properties depending on the composition, namely, activity coefficient, Reid vapor pressure, distillation profile and Gibbs energy of mixing, have to be recomputed in each step taken by the NLP solver. The efforts associated with the evaluation of those properties scale with the number of palette compounds. This means that also the total computational time for solving the blend design problem (cf. Eqns. (7.69)-(7.90)) primarily depends on the number of palette compounds.

Two measures have been taken to speed up the calculations. First, inside the optimization problem, the Reid vapor pressure (RVP) is approximated by the bubble point pressure p_{bp} (cf. Eqn. (7.86)). The correct RVP p^S_{Reid} is always computed from the system of nonlinear Eqns. (7.62)-(7.63), once the solution to the optimization problem has been obtained. From experience with the case study below, the approximated RVP (p_{bp}) is typically between zero and ten kPa higher than the correct RVP (p^S_{Reid}). Second, within the NLP the distillation curve is considered on a molar basis (instead of volume). This way, the idealized batch distillation model (cf. Eqns. (7.41)-(7.52)) can be stripped down to (Stichlmair and Fair, 1998)

$$\frac{dn_L}{dt} = -\dot{N} = -1 \tag{7.91}$$

$$\frac{dx'_{L,i}}{dt} = \frac{1}{n_L}\left(x'_{L,i} - \frac{x'_{L,i}\cdot\gamma_i(T_L,\mathbf{x'_L})\cdot p_i^S(T_L)}{p}\right) \qquad \forall i = 1,...,n_C-1 \tag{7.92}$$

$$1 = \sum_i x'_{L,i} \tag{7.93}$$

$$p = \sum_i x'_{L,i}\cdot\gamma_i(T_L,\mathbf{x'_L})\cdot p_i^S(T_L)\ , \tag{7.94}$$

where the initial values are set to $x'_{L,i} = z_i$ (z_i is the blend composition defined in Eqn. (7.74)) and $n_L = 1$. The characteristic $T10$ temperature (cf. Eqn. (7.87)), i.e., the tem-

perature where 10% of the fuel have been evaporated, is then simply $T_L(t = 0.1)$. The solution of this smaller DAE system (cf. Eqns. (7.91)-(7.94)) is approximated with the explicit Euler method (Ascher and Petzold, 1998) based on a fixed grid. Compared with the error-controlled, variable step length solvers *ode*45 and *ode*15*s* built into Matlab (The MathWorks Inc., 2016), the computational overhead as well as the number of activity coefficient calculations are significantly reduced, thereby greatly improving performance. The approximation error correlates with the chosen step length, i.e., the size of the mole increment (1 mol-% has been used here), and is assessed by once running the error-controlled solver *a posteriori*. Although the distillation constraints are defined on a molar basis in order to reduce the computational effort, the blend's volumetric distillation curve is computed with Eqns. (7.41)-(7.52) *a posteriori*, because the volume basis is the more frequently used variant. Therefore, it will be distinguished between $T10(m)/T50(m)/T90(m)$ and $T10(v)/T50(v)/T90(v)$ in the following, where (m) refers to mole and (v) refers to volume.

The local, necessary criterion for phase stability is integrated into the NLP formulation in an attempt to circumvent parts of miscibility gaps (cf. Eqn. (7.90)). The Hessian $\mathbf{H}$ is approximated with finite differences (Nocedal and Wright, 2006) and the eigenvalues are computed with Matlab's (The MathWorks Inc., 2016) *eig* routine which is based on the QR algorithm (Francis, 1961; Watkins, 2002). In order to avoid both the locally unstable and the metastable regions, a bilevel optimization problem would have to be formulated and solved. This is not done here for the sake of simplicity and performance. Instead, each solution to Eqns. (7.69)-(7.90) is checked *a posteriori* by minimization of the tangent plane distance function, i.e., by solving a second but smaller NLP (cf. Eqns. (7.67)-(7.68)). This way blends which turn out to be immiscible are discarded from the set of solutions.

Note that the optimization problem defined by Eqns. (7.69)-(7.90) represents a general formulation where each property is constrained by a lower and by an upper bound. However, depending on the modeler's intentions, the formulation can easily be modified for the design task at hand. Such modification might include removal of certain bounds. For instance, in the case study below, no upper bound is considered for the oxygen content and the lower heating value is not constrained at all. Instead of maximization of a process-related criteria like mass or energy of fuel produced, specific blend properties can be optimized as well. For instance, smaller values for the DCN correlate with higher research octane numbers (RONs) (cf. Figure 6.1) and a higher RON allows to raise the effective compression ratio and thereby the efficiency of a spark-ignition engine (cf. discussion in Subsection 2.1.1). Thus, the objective function

$$\min_{\mathbf{f}\geq\mathbf{0},\mathbf{b}\geq\mathbf{0}} \quad \sum_{j\in PC} z_j DCN_j \qquad \text{(minimize blend DCN)} \tag{7.95}$$

may be chosen. Note that the inequality constraint

$$\sum_{j\in PC} b_j M_j \geq \dot{m}_{fuel,min} \tag{7.96}$$

can be added simultaneously to ensure that a certain minimum mass flow of fuel $\dot{m}_{fuel,min}$ is still produced. Alternatively, if no more than $n_{H_2,max}$ [$\mathrm{mol}_{H_2}/\mathrm{mol}_{fuel}$] of external hydrogen shall be used to upgrade the carbohydrates, the inequality constraint

$$\frac{f_{S2}}{\left(\sum_{j\in PC} b_j\right)} \leq n_{H_2,max} \tag{7.97}$$

may be used.

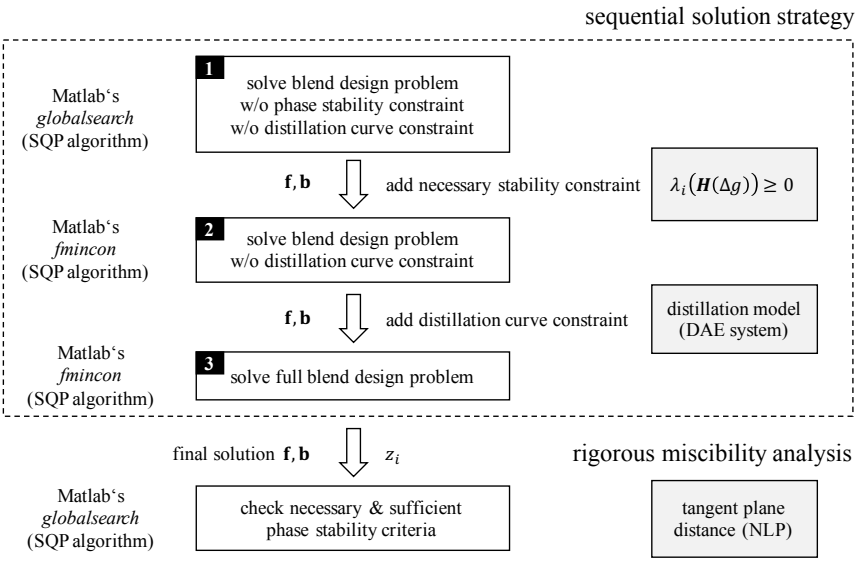


Figure 7.8: Sequential solution strategy and *a posteriori* rigorous miscibility check. The molar flows $\mathbf{f}$ and $\mathbf{b}$ resulting from steps 1 and 2 are used as initial values for the respective subsequent problems. The optimal blend composition z_i is determined in step 3.

Although the complete problem described by Eqns. (7.69)-(7.90) could be approached directly, it has been observed that if initial values for the full problem are derived systematically by first solving two simpler subproblems, the overall time required to find the solution

can often be reduced. This sequential solution strategy is depicted in Figure 7.8. The first subproblem is derived from ignoring both the distillation profile constraints (cf. Eqns. (7.87)-(7.89)) and the phase stability constraint (cf. Eqn. (7.90)) (cf. Figure 7.8, step 1). This nonlinear problem is small enough to be efficiently handled by *Globalsearch* and the global solver is employed to decrease the risk of getting stuck at a local optimum that is not the global optimum. Once solution vectors $\mathbf{f}$ and $\mathbf{b}$ have been obtained, these are taken as initial guesses for the next subproblem that includes the phase stability constraint, but still lacks the distillation model (cf. Figure 7.8, step 2). Since the computation of the Hessian of Δg makes this subproblem significantly more computationally demanding, a local NLP solver is used here. If the initial values do not violate the phase stability constraint, completion of step 2 is extremely fast as the solution to subproblem 2 then is identical to the initial guesses. Finally, the results from step 2 are used to initialize the full problem (cf. Figure 7.8, step 3). Solving the full problem is by far the most computationally demanding step because of the distillation model, however, only if the supplied initial values violate the distillation profile constraints. With the proposed sequential strategy (cf. Figure 7.8), the total computational effort required to solve the blend design task therefore strongly depends on whether the path to the solution is somehow affected by the distillation curve constraints. As can be seen from Figure 7.8, the necessary and sufficient phase stability criteria are always checked *a posteriori*. Total computational demand for solving the blend design task and for performing the rigorous miscibility check scales with the number of palette species under consideration. Some specific numbers on computational times will be given for the case study below.

If the CPM contains numerous palette compounds, blends of high complexity may arise. Here high complexity refers to a large number of components to be included in the final fuel. Trace amounts of a component i can easily be removed by adding a constraint $b_i = 0$ to the general problem formulation (cf. Eqn. (7.69)-(7.90)). However, typically there is a multitude of blends with similar objective values and these blends can differ substantially with respect to composition and utilization of production pathways. Being able to select from a range of promising blends rather than having to rely on a single best solution may prove particularly valuable in this context. To obtain blends with a specified number (n_C) of components, subproblems can be extracted from the CPM and solved in a systematic fashion. The enumeration-based strategy proposed in Figure 7.9 can be readily applied to generate binary, ternary and quaternary mixtures, if the total number of palette compounds to choose from ($n_{C,tot}$) is in the range of 20-30. Although in principle the strategy can also be applied to multicomponent mixtures with $n_C > 4$, it quickly approaches its limits because of the combinatorial explosion encountered in step A (cf. Figure 7.9). The reason is that there are generally $\binom{n_{C,tot}}{n_C} = \frac{n_{C,tot}!}{(n_{C,tot}-n_C)!\cdot n_C!}$ ways

of choosing n_C palette compounds from a CPM that comprises a total number of $n_{C,tot}$ palette compounds. Different subproblem definition strategies have to be developed to cope with such situations.

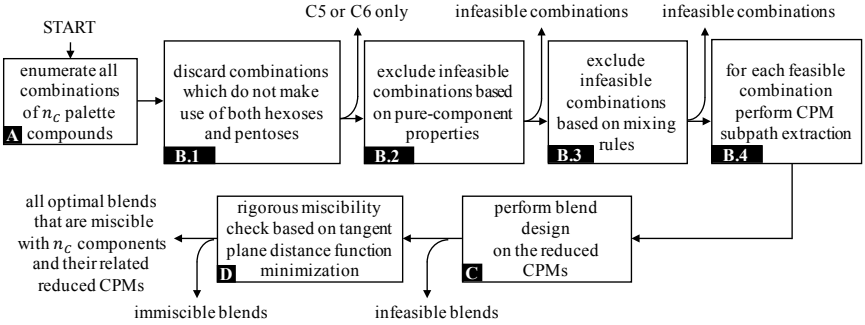


Figure 7.9: Workflow for the generation of all blends consisting of n_C components.

As can be seen from Figure 7.9, the workflow proposed here starts with the enumeration of all possible combinations of palette compounds (step A). Subsequently, the CPM is analyzed to discard those combinations that do not utilize both hexoses and pentoses (step B.1), because conversion of a single sugar fraction strongly limits the amount of fuel obtained. In step B.2, the feasibility of each combination is judged on the basis of pure-component properties. For instance, if a binary blend shall have a mixture DCN smaller than 15, at least one of the two components needs to exhibit a DCN < 15. In step B.3, a first optimization problem is formulated based on the mixture property models. The objective of this problem is to find a feasible point, i.e., a specific composition that satisfies all imposed property constraints. Here it is still ignored whether this composition can actually be achieved in a process. In order to reduce the computational burden for the final design step, only those species and pathways are considered which eventually lead towards the palette compounds under investigation. An example for such CPM subpath extraction (step B.4) is depicted in Figure 7.10. Finally, the actual blend design and the rigorous miscibility check (steps C and D) are carried out by means of the sequential approach described before (cf. Figure 7.8).

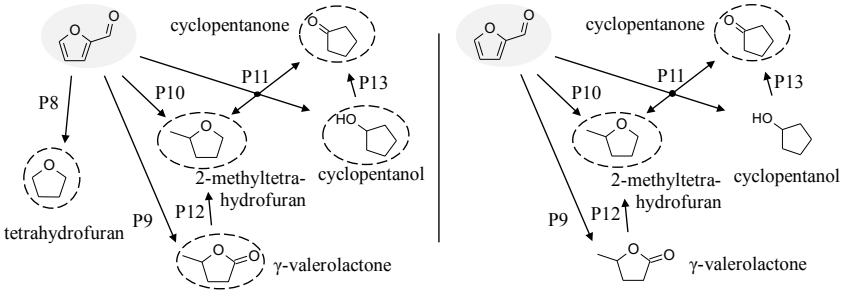


Figure 7.10: Example for CPM subpath extraction (step B.4 in Figure 7.9): Based on the original CPM on the left, the smaller scenario on the right, where only 2-methyltetrahydrofuran and cyclopentanone are considered palette compounds (indicated by dashed circles), contains exclusively those pathways and species that eventually lead towards the species of interest.

7.6 Case study: Blends for the spark-ignition (SI) engine

A case study is dedicated to the identification of biofuels with tailored properties and their respective optimal production pathways.

7.6.1 Scenario description

The target combustion system is a boosted direct injection spark-ignition engine, like those studied within TMFB previously (Thewes et al., 2011a,b; Hoppe et al., 2016a). The objective is to maximize the energy of fuel produced (in terms of lower heating value) given a fixed feed of biomass (cf. Eqn. (7.69)). This can be achieved particularly well by considering production routes that (a) have high yields and (b) are capable of integrating major amounts of additional hydrogen. Since the amount of biomass that can be grown for the purpose of fuel production is limited, biomass hydrogenation by means of water electrolysis utilizing carbon-free energy sources such as wind or solar is seen as an opportunity to leverage the amount and energy of liquid hydrocarbons that can be produced from the biomass feedstock (Shinnar and Citro, 2006; Agrawal et al., 2007; Muradov and Veziroğlu, 2008; Connolly et al., 2014). Therefore, in this case study, the supply of hydrogen is not constrained, but instead hydrogenation is considered to be an enabler for directing as many carbon as possible from the carbohydrates towards the fuel, thereby maximizing the energy of fuel produced.

As can be seen from Figure 7.11, 12 bio-based intermediates (shaded in gray) which can be produced from hexoses and pentoses in large volumes, e.g., ethanol, acetic acid

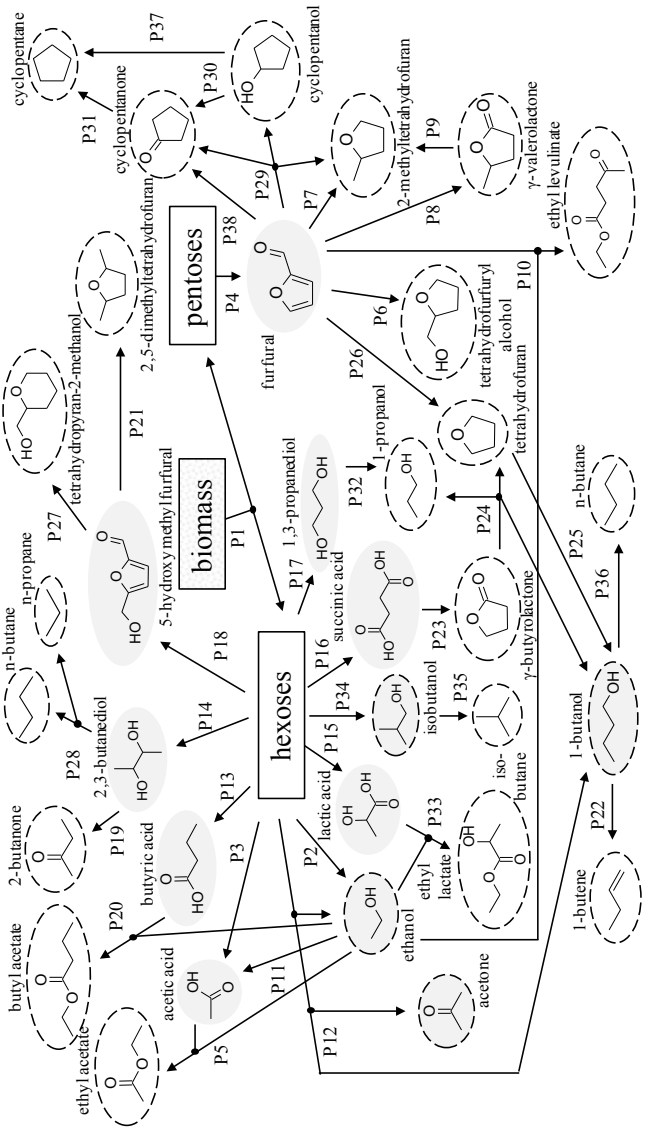


Figure 7.11: Conversion pathway map (CPM) for the case study (bio-derived platform chemicals in gray; palette compounds marked with dashed circles). Selectivity data and references for all pathways can be found in Appendix E.

and furfural, are considered. These platforms have been chosen because they comprise ≥75% LHV energy efficiencies, i.e., LHV of the intermediate per LHV of glucose/xylose, at highest-reported production yields (cf. Table 2.4). Because knock-resistance and fuel volatility are key for achieving high engine efficiencies and low engine-out emission levels, two criteria have been applied to select the palette compounds: (i) The normal boiling point of each blend component shall be well below the final boiling point of a typical gasoline (∼225 °C). (ii) No blend component shall exhibit a diesel-like auto-ignition tendency. Therefore, DCN of all blend components must be smaller than ∼30. Selectivity data for all pathways have been collected from the open literature. Data and references can be found in Appendix E. For the sake of simplicity, conversion-limitation is neglected in this case study, either because the catalytic system enables close-to-full conversion (most cases), or because recycling of unconverted reactants is assumed. In total, 24 potential fuel components and 38 conversion pathways are considered (cf. Figure 7.11).

A limitation of the case study is that furans could not have been included because the UNIFAC group interaction parameters for furans are available exclusively to sponsors of the UNIFAC consortium (Constantinescu and Gmehling, 2016; DDBST Dortmund Data Bank Software & Separation Technology GmbH, 2017) at this time. The furan-derivative γ-valerolactone has been included in the case study because of its high knock-resistance (cf. Table 2.2), although its use as a fuel component is viewed with some criticism due to γ-valerolactone's potential for abuse as a drug (Yanowitz et al., 2011). With a normal boiling point of 207-208 °C (Horváth et al., 2008), γ-valerolactone is also the least volatile compound in the case study.

The desired fuel properties are summarized in Table 7.2. The property constraints are based on the SI engine case study presented in Section 6.1. Again, the expectation is that a fuel with a DCN≤20 exhibits a knock-resistance sufficiently high to justify further investigation as a candidate fuel for the spark-ignition engine. This may include experimental RON/MON determination by means of ASTM D2699 (2013) and ASTM D2700 (2014).

Reid vapor pressure (RVP), T10(m), T50(m) are constrained to mitigate problems with vapor lock, cold-starting and during the warm-up driving phase (Kalghatgi, 2014a). Moreover, T90(m) is constrained to avoid lube oil dilution and excessive deposit formation (Kalghatgi, 2014a). The upper and lower bounds for RVP, T10(m), T50(m) and T90(m) were chosen such that all distillation and vapor pressure classes of US gasoline fuel as defined in ASTM D4814 (2016) are covered. The minimum Reid vapor pressure, which is important for engine cold-starting (Kalghatgi, 2014a), has been taken from the European gasoline fuel standard EN 228 (2014), since the US standard does not explicitly include such a constraint.

For this case study, pure-component property data for LHV, ρ_L, σ and μ have been

Table 7.2: Specifications for a multicomponent biofuel with tailored properties for combustion in a boosted direct-injection spark-ignition (SI) engine.

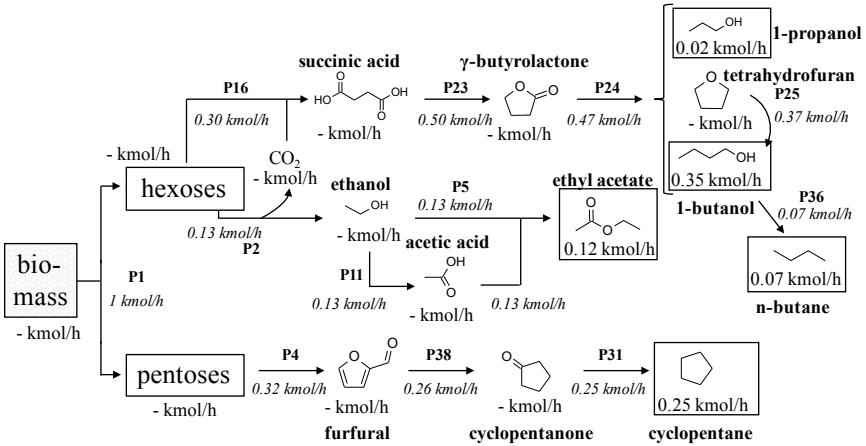
fuel property	lower bound	upper bound
derived cetane number DCN [-]	–	20
oxygen content wt_{O_2} [wt-%]	10	–
enthalpy of vaporization H_{vap} [kJ/kg$_{\text{air},\lambda=1}$]	–	60
surface tension σ [mN/m]	–	30
kinematic viscosity ν [mm^2/s]	0.5	2.0
approximated Reid vapor pressure p_{bp} [kPa]	45	100
distillation profile[a]		
temperature at 10% mole fraction evp. $T10(m)$ [°C]	45	70
temperature at 50% mole fraction evp. $T50(m)$ [°C]	65	125
temperature at 90% mole fraction evp. T90(m) [°C]	65	190

[a] idealized open batch distillation (cf. Figure 7.6)

retrieved from the DIPPR 801 database (AIChE, 2012) by means of a software tool developed by Manuel Hechinger and Luise Schulze Langenhorst (Hechinger and Schulze Langenhorst, 2012). Only data that has been classified as "accepted" by DIPPR staff and that is thought to have an error of $\leq$10% have been queried (cf. discussion in Section 5.1). This covers approximately 74% of the required data for LHV, ρ_L, σ and μ. The remaining fraction has been predicted with the QSPR models from Chapter 5. The predictions can be found in Appendix E. In case of vapor pressure and molar density, parameters for the extended Antoine equation (Eqn. (7.37)) and the DIPPR 105 equation (Eqn. (7.54)) have been retrieved from the Aspen Physical Property System pure-component databank (Aspen Technology, 2011) where such data were available, thus limiting the use of the fully predictive approach, i.e., critical data QSPR plus Hoffmann-Florin equation and/or COSTALD method, to the remaining instances (cf. Appendix E). Pure-component enthalpy of vaporization H_{vap} has been calculated from the vapor pressure curve via the Clausius-Clapeyron equation (cf. Eqn. (7.34)). Modified UNIFAC (Dortmund) parameters relevant to the species in the case study have been taken from different literature sources (Gmehling et al., 1993, 1998; Constantinescu and Gmehling, 2016). Experimental DCNs for eleven compounds were taken from Appendix B, whereas the other 13 DCNs had to be predicted with the help of the group contribution model from Chapter 4. The predictions can be found in Appendix E. Pragmatically, a value of 7 has been assigned to those compounds which do not ignite in the IQT as a consequence of the measurement limit, i.e., DCN $\approx$ 7.2 (Haas et al., 2011).

7.6.2 Results

Maximization of the energy of fuel produced yields a five-components blend (referred to as blend A in the following) which is rich in 1-butanol (44 mol-%), cyclopentane (31 mol-%) and ethyl acetate (15 mol-%). This design features a high degree of hydrogenation (4.6 mol_{H_2} per mol_{fuel}) and yields 0.75 MJ_{fuel} per $MJ_{biomass}$. The corresponding material flow diagram is depicted in Figure 7.12. The C_5 sugars are converted into cyclopentane via pathways P4, P38 and P31, whereas the majority of the C_6 sugars is directed to succinic acid fermentation. The remaining C_6 sugars are fed to ethanol fermentation. Ethanol, in this case, does not constitute a fuel component, but acts as a substrate for the production of acetic acid. The intermediates acetic acid and ethanol are finally converted into the fuel component ethyl acetate.



energy of fuel produced MJ_{fuel} / $MJ_{biomass}$	mass of fuel produced kg_{fuel} / $kg_{biomass}$	hydrogen demand mol_{H2} / mol_{fuel}	LHV energy efficiency MJ_{fuel} / ($MJ_{biomass}$+MJ_{H2})	CO_2 from biomass g / MJ_{fuel}	derived cetane number (DCN)
0.75	0.39	4.6	0.57	0.0	18.8

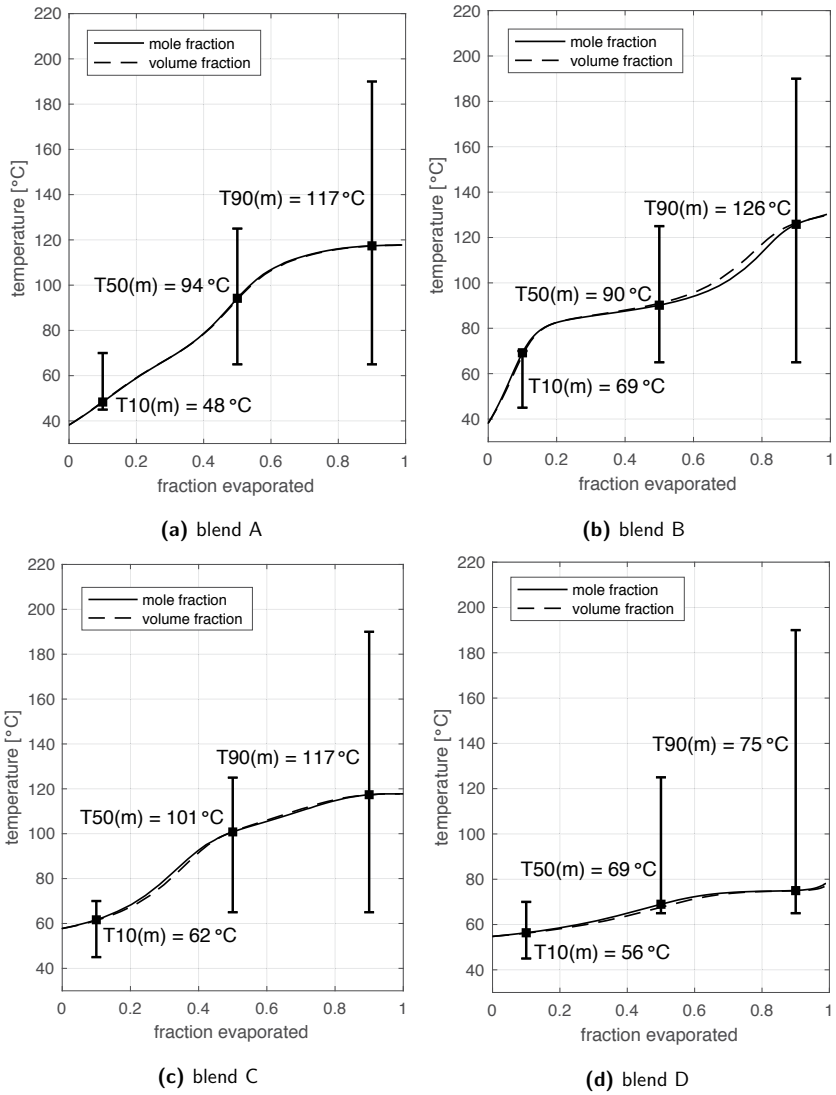
Figure 7.12: Material flow diagram for the design that maximizes the energy of fuel produced under the constrains stated in Table 7.2 (**blend A**). The resulting five-components blend is composed of 44 mol-% 1-butanol, 31 mol-% cyclopentane, 15 mol-% ethyl acetate, 8 mol-% n-butane and 3 mol-% 1-propanol. For each conversion pathway, the mole flow rate of the main reactant is given. Beneath each species or pseudo-compound, the product flow rate, i.e., the amount of substance leaving the fuel production plant in a given amount of time, is stated.

As can be seen from Figure 7.12, the carbon dioxide released in ethanol fermentation is fully re-used in the succinic acid fermentation step. While the CO_2 has zero enthalpy of

Table 7.3: Predicted properties of the designs discussed in this case study.

	blend composition [mol-%]	energy of fuel produced [MJ_{fuel}/ $MJ_{biomass}$]	mol H_2 per mol fuel	η_{LHV} [MJ_{fuel}/ ($MJ_{biomass}$ +MJ_{H_2})]	DCN (cRON /cMON)	LHV [MJ /kg]	ρ_L [kg /m^3]	H_{vap} [kJ /kg_{air}] (λ=1)	σ [mN /m]	ν [mm^2 /s]	p^S_{Reid}/ p_{bp} [kPa]	T10 (v)[a] [°C]	T50 (v)[a] [°C]	T90 (v)[a] [°C]
A	1-butanol (44%), cyclopentane (31%), ethyl acetate (15%), n-butane (8%), 1-propanol (3%)	0.75	4.6	0.57	18.8 (92/85)	35.5	779	47	22	1.8	96 / 100	49	94	117
B	ethanol (42%), cyclopentanone (26%), 2-butanone (15%), 1-butene (7%), 1-butanol (6%), cyclopentane (3%), 1-propanol (<1%)	0.61	1.7	0.54	9.0 (114/104)	31.6	818	60	24	1.2	92 / 100	68	91	126
C	1-butanol (45%), ethanol (29%), cyclopentane (27%)	0.73	3.8	0.56	14.5 (103/94)	34.9	781	57	23	2.0	51 / 51	61	101	117
D	ethanol (58%), cyclopentane (27%), 2,5-DMTHF (15%)	0.65	2.2	0.56	15.0 (102/93)	34.3	785	54	22	1.1	54 / 54	56	68	75

[a] computed with Matlab's (The MathWorks Inc., 2016) error-controlled, variable step-length solver *ode15s*



(a) blend A **(b)** blend B

(c) blend C **(d)** blend D

Figure 7.13: Predicted distillation curves for the blends discussed in this case study. The curves have been computed with Matlab's (The MathWorks Inc., 2016) error-controlled, variable step-length solver *ode15s*. The bars show the constraints for T10(m), T50(m) and T90(m) (cf. Table 7.2). Note that V_{shift} has been set to zero in an attempt to simulate an idealized batch distillation process.

combustion, the re-use of carbon allows for higher degrees of hydrogenation in subsequent steps. Thus, co-processing of ethanol and succinic acid can be seen as an elegant way for chemical hydrogen storage in the case of SI fuels where a certain oxygen content is mandatory. It shall be noted that an analysis of the general trade-off between hydrogen demand [$\mathrm{mol}_{H_2}/\mathrm{mol}_{fuel}$] and energy of fuel produced [$\mathrm{MJ}_{fuel}/\mathrm{MJ}_{biomass}$] can be found in Appendix E. The discussion there includes a blend that has been designed with the premise of zero hydrogen demand.

In case of blend A, succinic acid is converted into a mixture of alcohols and n-butane via pathways P23−P25 and P36 (cf. Figure 7.12). Cyclopentane and n-butane are responsible for the blend's excellent front-end volatility, i.e., low T10 and high Reid vapor pressure. Table 7.3 shows the estimated blend properties and Figure 7.13 depicts the distillation curve. Based on the pure-component DCN values, the alcohols and the ethyl acetate are considered the most knock-resistant species in the mixture. The equations reported by Perez and Boehman (2012) can be applied to calculate a quantitative estimate for RON/MON from the DCN (cf. Figure 6.1). These will be referred to as correlated RON (cRON) and correlated MON (cMON) in the following to distinguish them from true RON/MON. Since the accuracy of the IQT-RON/MON correlations proposed by Perez and Boehman (2012) has not been systematically assessed for a wider range of fuels, cRON and cMON should be generally taken with caution. For instance, the cRON for iso-octane is only 96.7 (based on the DCN of 17 reported by Dooley et al. (2012b)), whereas its true RON is 100.

The cRON calculated for blend A is 92 (cf. Table 7.3). While RON91 constitutes the minimum fuel quality required for most modern passenger cars equipped with knock sensors, future fuels will probably require higher RONs because of the strive for higher engine efficiencies (Hoppe et al., 2016a,b). In this context, the model-based design approach can be used to analyze the trade-off between DCN as a simplified measure for knock-resistance and the energy of fuel produced. To this end, a sequence of design problems is solved with decreasing upper bounds for the DCN. The resulting trade-off is visualized in Figure 7.14.

It can be seen that the energy of fuel produced in MJ_{fuel} per $\mathrm{MJ}_{biomass}$ decreases as DCNs get smaller. However, for DCNs above 12, the correlation is generally weak. While there is no clear trend for mass and LHV energy efficiencies, hydrogen consumption follows the trend seen for energy of fuel produced. The reduced opportunity for hydrogenation in case of low DCN is accompanied by significant CO_2 emissions, because the emitted carbon can no longer be directed towards the fuel without violating the property constraints.

Figure 7.15 shows the material flow diagram for blend B, i.e., the blend with a DCN of 9 (cf. Figure 7.14) which translates into a cRON of 114 and a cMON of 104. The

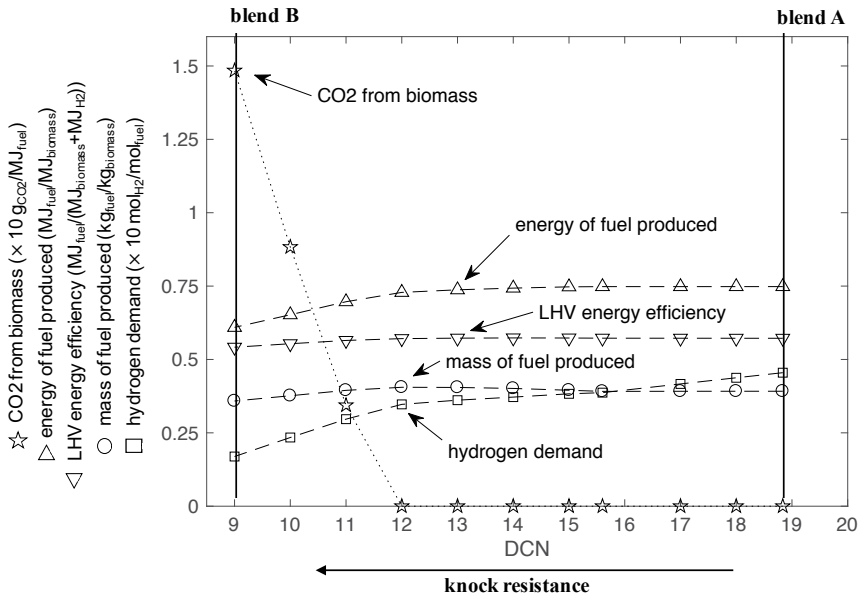
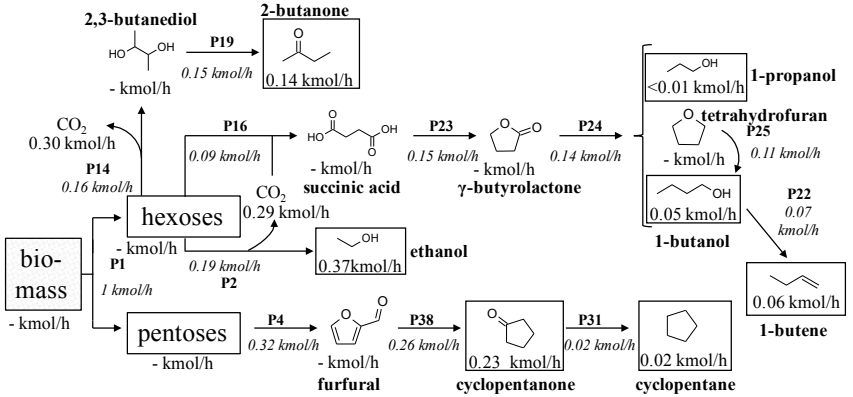


Figure 7.14: Trade-off between derived cetane number (DCN) and energy of fuel produced [$\mathrm{MJ}_{fuel}/\mathrm{MJ}_{biomass}$]. The blend design problem has been solved multiple times, whereby the upper bound for the DCN has been lowered stepwise (20, 19, 18, 17 etc.). While the points in the diagram represent properties of optimal designs, the dashed and dotted lines are for illustration purposes only. A feasible blend can be generated for DCN$\leq$8, however, its LHV efficiency is as low as 0.40 $\mathrm{MJ}_{fuel}/(\mathrm{MJ}_{biomass}+\mathrm{MJ}_{H_2})$ as significant portions of the biomass are left unconverted. DCN$\leq$7 makes the problem infeasible.

presumed high knock-resistance of this seven-components blend primarily results from the high fractions of ethanol, cyclopentanone and 2-butanone. The latter compound has an experimental RON/MON of 117/107 (Hoppe et al., 2016b). The production of 2-butanone, however, necessitates the inclusion of a separate 2,3-butanediol fermentation step in the design (cf. Figure 7.15). In contrast to the previous design, carbon dioxide released in ethanol fermentation is mostly emitted here (compare Figures 7.12 and 7.15). The low hydrogen consumption in the production of blend B (1.7 $\mathrm{mol}_{H_2}/\mathrm{mol}_{fuel}$) is accompanied by a lower energy of fuel produced (0.61 $\mathrm{MJ}_{fuel}/\mathrm{MJ}_{biomass}$). However, the LHV energy efficiency is only slightly lower compared to the previous design. The estimated blend properties can be found in Table 7.3.

From the trade-off analysis in Figure 7.14 it appears that the blend with a DCN of 12



energy of fuel produced $MJ_{fuel}/MJ_{biomass}$	mass of fuel produced $kg_{fuel}/kg_{biomass}$	hydrogen demand mol_{H2} / mol_{fuel}	LHV energy efficiency $MJ_{fuel}/(MJ_{biomass}+MJ_{H2})$	CO_2 from biomass g / MJ_{fuel}	derived cetane number (DCN)
0.61	0.36	1.7	0.54	14.84	9.0

Figure 7.15: Material flow diagram for the design that maximizes the energy of fuel produced if DCN is constrained to $\leq$9 (**blend B**). The resulting seven-components blend is composed of 42 mol-% ethanol, 26 mol-% cyclopentanone, 15 mol-% 2-butanone, 7 mol-% 1-butene, 6 mol-% 1-butanol, 3 mol-% cyclopentane and <1 mol-% 1-propanol.

would represent a nice compromise between knock-resistance on the one hand and energy of fuel produced and CO_2 emissions on the other hand. However, it should be kept in mind that the the inverse relationship between DCN and RON/MON holds only approximately. Moreover, linear mixing is assumed here and pure-component DCNs of species that do not ignite in the IQT (e.g., ethanol, cyclopentanone) are assumed to be 7. In principle, the uncertainties of the outlined design approach could be reduced significantly if experimental RON and MON values were available for the pure compounds. It is also possible to derive nonlinear RON/MON mixing rules, once blending effects can be quantified based on comprehensive experimental RON/MON mixture data for the compounds of interest (AlRamadan et al., 2016).

The complexities of blends A and B are quite high, meaning that both blend components and conversion pathways are numerous. The number of conversion pathways in a design can be taken as a first rough estimate of process complexity. Moreover, it is often assumed that the number of processing steps correlates with the capital cost of a production plant (El-Halwagi, 2012; Ulonska et al., 2016a). With the aim of arriving at less complex designs, all optimal binary and ternary blends, which can be formulated based on the conversion pathway map depicted in Figure 7.11, have been generated. To this

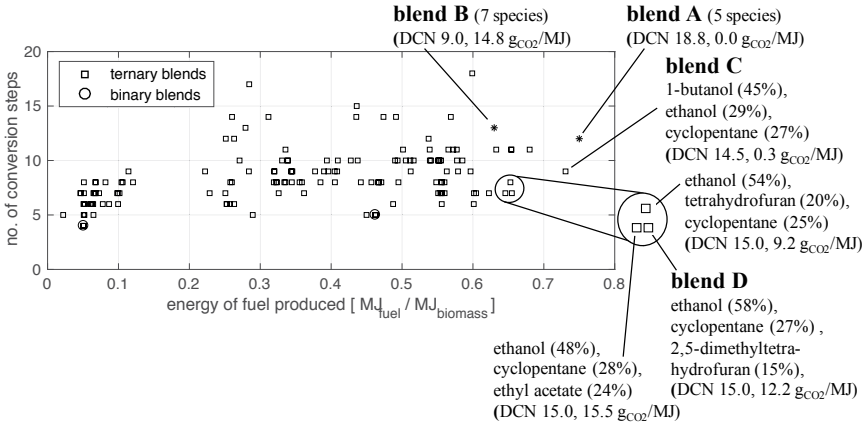
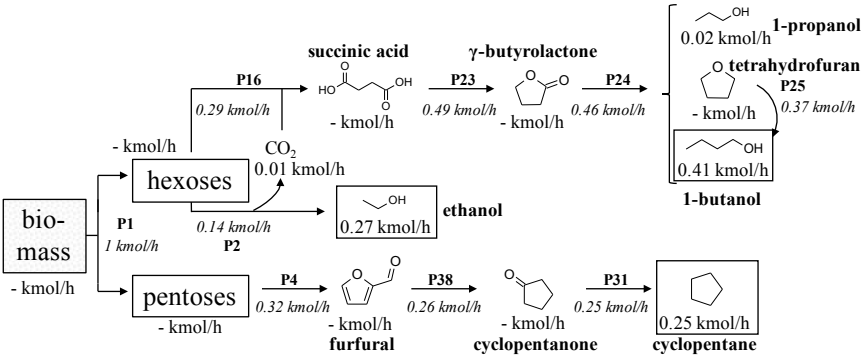


Figure 7.16: Number of active conversion steps and energy of fuel produced for all optimal binary (2 circles) and ternary (149 squares) blends. For comparison, the graph also includes the more complex blends A and B (2 stars).

end, the strategy proposed in Figure 7.9 has been employed. The DCN constraint has now been tightened to DCN$\leq$15, which corresponds to a cRON/cMON of $\geq$102/93, in order to place particular focus on fuels especially well-suited for highly-boosted, direct-injection SI engines (Kalghatgi, 2005; Hoppe et al., 2016a,b). Figure 7.16 shows the number of active conversion steps and the energy of fuel produced for all optimal binary and ternary blends. The ideal design would be located in the lower right part of the graph. It can be noted that there is a ternary blend (called blend C in the following) which is slightly inferior to blend A in terms of energy of fuel produced. However, this blend is less complex. Indeed, the material flow diagram for blend C can be seen as a simplified version of design A (compare Figures 7.12 and 7.17). Instead of acetic acid fermentation and ethyl acetate production, ethanol is directly used as a fuel component here. Furthermore, the conversion of 1-butanol to n-butane has been eliminated which somewhat deteriorates the front-end volatility, as can be seen from Figure 7.13. Still, all fuel properties are well within the specified bounds (compare Tables 7.2 and 7.3). With a cRON of 103, blend C is expected to offer greater knock-resistance than blend A. At the same time, blend C retains a key property of blend A, i.e., the almost complete avoidance of carbon dioxide emissions that is enabled by concurrent ethanol and succinic acid fermentation (cf. Figure 7.17).

If the process complexity shall be reduced even further, Figure 7.16 shows that blend D represents a nice trade-off between the energy of fuel produced and the number of conversion steps. As can be seen from the material flow diagram in Figure 7.18, the majority of

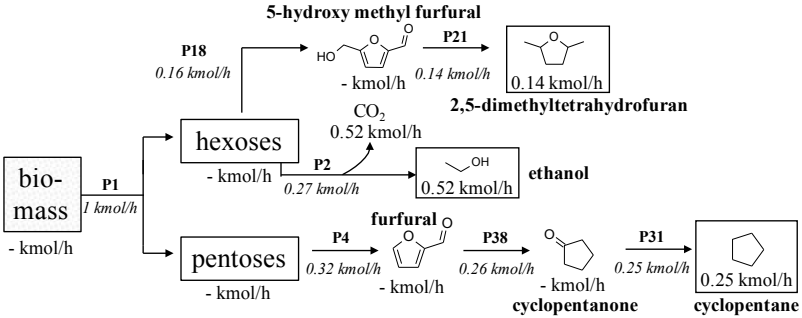


energy of fuel produced $MJ_{fuel}/MJ_{biomass}$	mass of fuel produced $kg_{fuel}/kg_{biomass}$	hydrogen demand mol_{H2}/mol_{fuel}	LHV energy efficiency $MJ_{fuel}/(MJ_{biomass}+MJ_{H2})$	CO_2 from biomass g/MJ_{fuel}	derived cetane number (DCN)
0.73	0.39	3.8	0.56	0.28	14.5

Figure 7.17: Material flow diagram for the optimal ternary blend of 1-butanol (45 mol-%), ethanol (29 mol-%) and cyclopentane (27 mol-%) (**blend C**, cf. Figure 7.16).

C_6 sugars is still directed towards ethanol fermentation in this design. However, a significant fraction is also converted chemo-catalytically into 2,5-dimethyltetrahydrofuran (2,5-DMTHF), a compound with a high heating value. Knock-resistance is provided primarily by ethanol, and cyclopentane is required to satisfy the front-end volatility requirements. Two additional ternary blends are located in the vicinity of blend D in Figure 7.16 and their main components are ethanol and cyclopentane as well. This suggests that the third component of blend D, i.e., 2,5-DMTHF, can well be exchanged for tetrahydrofuran or ethyl acetate, if the proportions of ethanol and cyclopentane are adjusted accordingly (cf. Figure 7.16).

In summary, the case study has demonstrated the capabilities of the model-based approach for integrated product and pathway design of novel biofuel blends. The succinic acid pathway constitutes an elegant option for re-using CO_2 that is produced elsewhere in the plant, e.g., in concurrent ethanol fermentation. This way more carbon reaches the fuel and, together with hydrogen produced from renewable electricity, this allows to increase the fuel energy available for combustion given a fixed input of biomass. The case study has also revealed the promising role of cyclopentanone production from furfural in the context of SI fuel synthesis. Cyclopentanone is considered an ideal fuel for highly-boosted direct injection SI engines due to its extremely low auto-ignition reactivity and hence extreme knock-resistance (Yang and Dec, 2013). Further hydrogenation of cyclopentanone yields cyclopentane, a volatile compound which is expected to mitigate cold-start and cold-run



energy of fuel produced MJ_{fuel} / $MJ_{biomass}$	mass of fuel produced kg_{fuel} / $kg_{biomass}$	hydrogen demand mol_{H2} / mol_{fuel}	LHV energy efficiency $MJ_{fuel}/(MJ_{biomass}+MJ_{H2})$	CO_2 from biomass g / MJ_{fuel}	derived cetane number (DCN)
0.65	0.36	2.2	0.56	12.16	15.0

Figure 7.18: Material flow diagram for the optimal ternary blend of ethanol (58 mol-%), cyclopentane (27 mol-%) and 2,5-dimethyltetrahydrofuran (15 mol-%) (**blend D**, cf. Figure 7.16).

problems typically observed in oxygenated fuels. The analysis of selected ternary blends has shown that, as a general rule, it is beneficial to generate and analyze a range of designs with similar properties, as these designs can differ considerably with respect to fuel components and production pathways employed. A subsequent, manual inspection from the perspective of early conceptual process design can then be used to discriminate among the alternatives. At the latest when the number of designs shrinks to a few, experimental validation of predicted blend properties should be performed. Finally, Figure 7.16 has also revealed the importance of performing combined product and pathway design. For though all binary and ternary blends plotted in this graph are considered fuels with tailored properties, only few of these are attractive from a process point of view, which means that the respective designs feature both high energy of fuel produced and limited complexity.

Based on the Matlab (The MathWorks Inc., 2016) implementation of the sequential solution strategy and the rigorous miscibility check (cf. Figure 7.8), the following computational times have been measured on a PC equipped with an Intel® Core™ i5-6360U CPU during the work on the case study: Depending on the chosen fuel property constraints, the total computational time required to solve the blend design problem ranges from approximately 4 minutes, if the distillation and phase stability constraints never become active during the optimization, to about 1 hour. Likewise, optimization of a single ternary blend extracted from the case study's CPM by means of the approach depicted in Figure 7.9 takes between 10 seconds and a few minutes. Given these numbers, there is still some room to

increase the size of the case study in the future, as the available sets of interesting fuel components and promising conversion pathways continuously evolve. Generally, a larger number of options is expected to increase both the number and the quality of interesting product/pathway combinations for further investigation.

7.7 Conclusions and outlook

In this Chapter, an optimization-based approach has been presented that facilitates the rational formulation of biofuel blends with tailored properties. The blends are designed such that a process-related performance measure, i.e., the energy of fuel produced, is optimized. This has been achieved by stating and solving a nonlinear program that couples reaction flux analysis and mixture property prediction based on QSPR and group contribution modeling. Because feedstock cost is thought to dominate total biofuel production cost, selectivity data taken from the literature were assigned to the conversion pathways considered in this study. It was demonstrated that a multitude of product/pathway designs can be generated automatically, if subproblems are defined and solved in a systematic fashion. Such a strategy then yields a set of alternatives for further investigation by means of process network flux analysis (PNFA) (Ulonska et al., 2016a) and conceptual process design (CPD) (Upadhye et al., 2011; Sen et al., 2012; Han et al., 2015). PFNA and CPD are ultimately needed to bridge the gap from the mass- and energy-based molecular level analysis considered here to a process level analysis that deals with the economics of fuel production by taking into account aspects like separation of byproducts and solvents and energy recovery from waste streams.

A weakness of the computational approach is the fact that reliable liquid-liquid equilibria prediction is generally considered to be very difficult (Gmehling et al., 2012). Luckily, experimental validation of miscibility is quite straightforward. Removal of the phase stability criteria is conceivable, if experimental knowledge can rule out the presence of a miscibility gap. As a general rule, measured pure-component property data should be used wherever possible to improve the accuracy of mixture property prediction. However, experimental data is often incomplete in the early screening phase. This underscores the value of the fully predictive approach taken here.

The blend design methodology is based on data and models which both are subject to uncertainty. For instance, pathway yield data often stem from lab-scale experiments and thus may change as a consequence of scale-up and transition from batch to continuous production. In case of RNFA, a comprehensive sensitivity study based on Monte Carlo analysis has revealed that differentiation of fuel candidates can become difficult or even impossible, if simultaneously occurring deviations of $\pm 15\%$ in all model parameters are

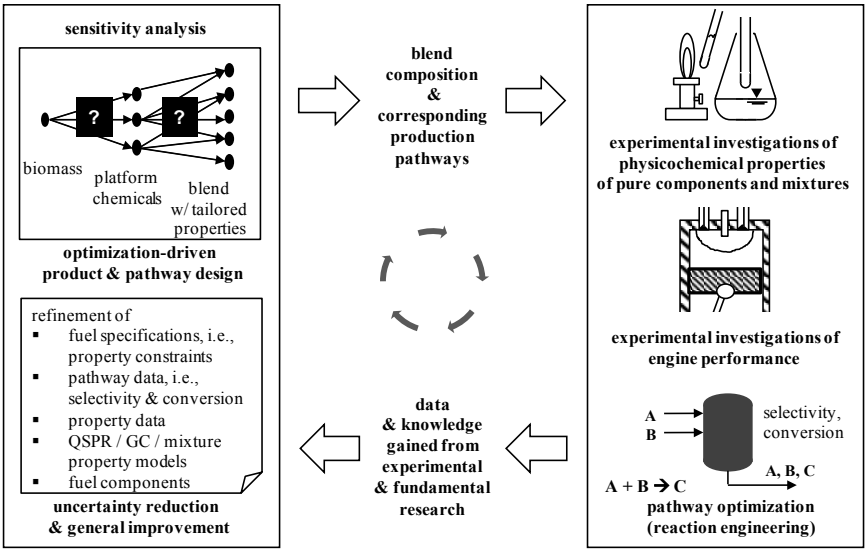


Figure 7.19: Iterative strategy combining computational blend design (left) with experimental and fundamental investigations of blends, engine performance and production pathways (right).

assumed (Ulonska et al., 2016b). For a practical blend design problem, logical next steps could be: The blend's physicochemical properties should be measured in order to confirm the fuel's quality. If measured and predicted fuel properties are too far apart, accuracy of pure-component property data should be improved (by acquiring measurement data instead of using predictions). Relaxation or tightening of property constraints may be tried, if the solution lies on a constraint and the divergence of measured and predicted data is likely due to model inaccuracy, e.g., underestimation of DCN due to nonlinear mixing behavior. It is also conceivable to improve property models based on the characteristics of the specific mixture under investigation. One-at-a-time sensitivity analysis (Ulonska et al., 2016b) with regard to the selectivities of all pathways involved in the blend's production can be performed to evaluate how the optimum of the objective function and the blend's composition will be affected. As depicted in Figure 7.19, an iterative strategy, where the model-based design approach is combined with experimental and fundamental investigations, should be pursued aiming at gradual uncertainty reduction, model improvement and pathway debottlenecking. This is even more important since the fuel definition itself, i.e., the properties to be modeled and their target ranges, is continuously evolving as the fuel/engine interaction becomes more understood.

A potential problem that has not been addressed here is chemical reactivity among the blend components. For instance, if it is known that two components A and B tend to react with each other, precautions must be taken to avoid that both of them appear in the same blend. This could be achieved by stating a mixed-integer nonlinear program which, however, is significantly more challenging to solve. If the issue of reactivity is limited to few components, the CPM might also be decomposed into subproblems manually. In any case, reactivity, including the issue of long-term fuel stability, is probably difficult to judge in advance, thus requiring specialized testing as a part of experimental fuel quality validation. To mitigate fuel oxidation stability issues, the formulation of certain oxygenated blends will likely require the use of stabilizers like butylated hydroxytoluene.

Throughout this thesis, lignin has been treated as a source for process power and heat. Therefore only carbon from the C_5 and C_6 sugars has been directed towards the fuel. However, lignin could also be added as a pseudo-component to the CPM, as both lignin combustion generating CO_2 and lignin gasification yielding syngas bear the potential to improve the carbon efficiency and/or energy of fuel produced.

Perspectively, network generator tools like RING (Marvin et al., 2013) or ReNeGen (Victoria Villeda, 2017) might be used to generate alternative pathways and associated pathway performance data for integration into the formulation of the blend design problem. It shall also be mentioned that surrogate models for fossil fuels, i.e., gasoline and diesel fuel, can easily be integrated into the framework outlined above. For instance, the eight components diesel surrogate proposed by Mueller et al. (2012) has been rationally formulated to match the DCN, the liquid density, the distillation curve and the distribution of key functional groups of a real diesel fuel. The property models used here can be applied readily to mixtures of oxygenates and these eight hydrocarbon model compounds, thereby giving the opportunity to design biofuels for blending applications with fossil fuels.

8 Conclusions and outlook

Lignocellulosic biofuels are considered to play an important role in a low carbon energy future. However, at least with regard to light duty vehicles, biofuels are now expected to face strong competition from electricity and potentially hydrogen (Fulton et al., 2015). For the internal combustion engine to remain a prominent drive technology in this segment of the transport sector, engines and vehicles need to become as clean as possible and at the same time as efficient as possible. In this context, the shift from the fossil to the renewable carbon resource and, most importantly, the different nature of the biomass feedstock offer the opportunity to redesign the entire value chain and therefore to tailor the fuel to the engine and vice versa (Marquardt et al., 2010; Janssen et al., 2011; Victoria Villeda et al., 2012a, 2015; Hoppe et al., 2016b; Leitner et al., 2017). This has also been the central motivation behind the Cluster of Excellence "Tailor-Made Fuels from Biomass" (TMFB) at RWTH Aachen University since its initiation in 2007. Here, the fuel's molecular structure (or its molecular composition in the case of mixtures) constitutes the single most important design degree of freedom in the optimization of fuel production and combustion.

To systematically screen the molecular search space for viable fuel components, methods from computer-aided molecular design (CAMD) had first been applied as part of a model-based fuel design approach by Hechinger et al. (2010) at the Institute for Process Systems Engineering (Prof. Dr.-Ing. Wolfgang Marquardt) at Aachener Verfahrenstechnik. The present thesis represents a continuation of the earlier works performed by the Marquardt group (Hechinger et al., 2010; Hechinger and Marquardt, 2010; Hechinger et al., 2012a,b; Dahmen et al., 2012; Victoria Villeda et al., 2012a; Voll and Marquardt, 2012a,b; Hechinger, 2014; Victoria Villeda et al., 2015; Victoria Villeda, 2017) and has emerged from the research performed by the author during his time as a member of both TMFB and the Marquardt group (from 2011 to early 2017). The thesis covers four main contributions:

- Based on a comprehensive literature survey, the basic principles of model-based fuel design have been presented in Chapter 2. Particular focus has been placed on criteria and methods for the (model-based) assessment of the suitability of bio-derived components as tailor-made fuels for advanced internal combustion engines. Moreover, platform molecules derived via hydrolysis and aqueous-phase processing of carbohydrates of lignocellulosic biomass have been evaluated with regard to their potential

in carbon- and energy-efficient biofuel production. The findings of Chapter 2 therefore constitute the theoretical foundation for the methodological contributions and computational investigations described in the subsequent Chapters.

- An efficient algorithm for molecular structure generation that resembles carbon- and energy-efficient chemo-catalytic refunctionalization of bio-derived platform chemicals has been outlined. This approach links CAMD products to pre-defined building blocks and therefore allows to direct experimental investigations specifically towards those molecular entities that are thought to exhibit both the desired properties and structural affinities with the available substrates (cf. Chapter 3). The structure generator has subsequently been coupled with predictive quantitative structure-property relationship (QSPR) and group contribution (GC) models to identify oxygenated biofuel candidates for spark-ignition (SI) and compression-ignition (CI) engines. Compared to earlier biofuel CAMD studies (Dahmen et al., 2012; Hechinger, 2014), the targeted approach has allowed to narrow down the range of potential fuel candidates to a much smaller number of most promising molecular motifs. Whereas compact furans, ketones and esters have emerged as first choice candidates for the SI engine, acyclic and cyclic ethers represent attractive options for the CI engine because they combine high cetane numbers with high volatilities (cf. Chapter 6).

- High-quality ignition delay data from a single, well-defined experimental set-up, i.e., the Ignition Quality Tester (IQT), have been collected for a wide range of oxygenated hydrocarbon species and have been used to establish a predictive quantitative relationship between a fuel's molecular structure and its auto-ignition propensity by means of a group contribution approach (cf. Chapter 4). The model represents an easy to use estimator of the IQT ignition delay and corresponding derived cetane number (DCN) from the two-dimensional molecular graph. It can be applied in situations where a sufficiently large sample for experimental testing is not yet available. Furthermore, the unique molecular diversity of the underlying database has allowed to integrate the auto-ignition model into the CAMD approach where DCN filters are used to decide whether a molecule is considered a candidate for spark-ignition (DCN<20) or compression-ignition (DCN>40) engines (cf. Chapter 6).

- A computational framework for simultaneous product and pathway design of biofuel mixtures and their corresponding production routes has been presented in Chapter 7. To this end, the optimization-driven approach of reaction network flux analysis (RNFA) has been extended and combined with mixture property modeling based on QSPR and GC. The resulting nonlinear program (NLP) includes constraints for fuel oxygen content, heating value, density, viscosity, surface tension, enthalpy of

vaporization, Reid vapor pressure and derived cetane number. Furthermore, a local, necessary phase stability criterion and a batch distillation model that allows to formulate constraints for the blend's distillation curve have been integrated into the NLP. The additional degrees of freedom offered by a multicomponent fuel have been utilized to optimize a production-related performance measure, i.e., the energy of fuel produced for a fixed amount of biomass. The analysis of a case study has highlighted the importance of performing combined product and pathway design, because only few blends are predicted to exhibit both the desired fuel properties and attractive process-related properties. High levels of hydrogenation could be achieved by re-using CO_2, which is generated in ethanol and/or 2,3-butanediol fermentation, to produce additional fuel components via the CO_2-consuming succinic acid fermentation pathway.

Discussions of the limitations and perspectives of individual parts of the model-based approach have been provided in the final Sections of the respective Chapters and shall not be repeated in detail here. Nevertheless, four aspects will be particularly important to advance model-based fuel design:

- Although the prediction accuracy of the DCN model proposed here is sufficient to distinguish gasoline-like (DCN ∼15) fuels from diesel-like (DCN ∼50) fuels, DCN estimates in the intermediate range (∼20 to ∼40) are still difficult to interpret. The analysis in Chapter 4, however, suggests that a substantial improvement in model accuracy can only be achieved, if the size of the database is increased. Since the number of pure compounds that can be bought at a reasonable price is inherently limited, the experimental investigation of specifically designed mixtures of few components might prove useful for generating further DCN training data, if no significant deviations from the linear mixing rule are encountered. Recently, such a strategy has been applied successfully by Abdul Jameel et al. (2016) who have trained a multiple linear regression model for the DCN of hydrocarbon species by considering both DCN data of pure compounds and DCN data of blends with known composition.

- Perspectively, predictive models for other important fuel properties developed by colleagues in TMFB, e.g., for the threshold sooting index (TSI) (Graziano et al., 2016), for the wear scar diameter (Masuch et al., 2011; Weinebeck and Murrenhoff, 2013), or even for novel characteristic fuel numbers like the oxidation potential number (Graziano et al., 2016), should be integrated into the generate-and-test approach, such that the number of fuel candidates can be further narrowed down. At the same time, the definition of a tailor-made fuel, i.e., the list of physicochemical properties

and their desired ranges, should be continuously updated based on new insights into the fuel/engine interaction.

- With regard to the virtual fuel screening described in Chapter 6, sensitivity studies should be performed on the most important property constraints, e.g., the DCN threshold, in order to assess how uncertainties in property estimates and fuel definition do influence the results. Likewise sensitivity analysis can help to identify critical pathways, properties and constraints in the context of blend design (cf. discussion at the end of Chapter 7).

- The blend design methodology outlined in Chapter 7 relies on selectivity and conversion as the only assessment criteria for pathway performance. By analogy with RNFA, it is assumed that the need for separation of intermediates, (side-)products and solvents will not significantly influence the design decision, because biofuel production cost is thought to be dominated by the feedstock cost. However, this can only be a first step, since feasibility and energy demand of inescapable separation steps will certainly constitute important factors, when multiple design alternatives with similarly high energy of fuel produced ($\mathrm{LHV}_{fuel}/\mathrm{LHV}_{biomass}$) need to be further elaborated and discriminated. To bridge the gap between RNFA and conceptual process design, Ulonska et al. (2016a) have extended RNFA to process network flux analysis (PNFA). In the PNFA methodology, the energy demand of reactions and separations is assessed with the help of shortcut models such that the cumulative energy demand of a biomass conversion process can be optimized. An attempt should be made to transfer the core elements of the PNFA methodology to the blend design problem as described in Chapter 7. This would allow to move from the molecular level considered here towards an integrated design of biofuels and their production processes.

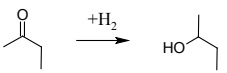
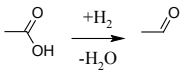
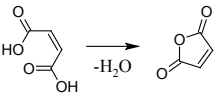
The core value of model-based fuel design is to translate experimental data and (empirical) knowledge into a computational tool that facilitates a systematic exploration of the search space and that enables rational decision making on the basis of evaluated alternatives. However, the results from such an approach can only be as good as its inputs, i.e., the many assumptions made, the chosen boundary conditions and evaluation criteria, and the experimental data used to train the models. Therefore, an iterative cycle between the computational investigations and the experimental and fundamental research as being performed within TMFB constitutes the most promising way forward towards a reduction of uncertainties and towards an improvement of the overall fuel design methodology.

Appendices

A Transformation rules for molecular structure generation

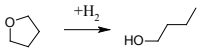
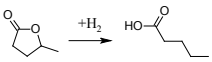
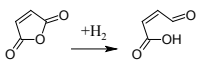
The transformation rules shown in Tables A.1 to A.4 have been implemented in the molecular structure generator (cf. Chapter 3) and have been used in the case studies presented in Chapter 6. The transformations are sorted by target functional groups and have been compiled in collaboration by the author of this thesis and Juan José Victoria Villeda (Institute for Process Systems Engineering at Aachener Verfahrenstechnik, RWTH Aachen University) based on the reviews of catalytic strategies for the production of biofuels prepared by Corma et al. (2007), Alonso et al. (2010), Serrano-Ruiz et al. (2011), Serrano-Ruiz and Dumesic (2011), Lange et al. (2010, 2012) and Climent et al. (2014). The works of Julis and Leitner (2012) and Luska et al. (2014) have motivated the inclusion of the carbon-carbon coupling scheme based on aldol condensation (cf. Tables A.2 and A.3). The rules shown in Tables A.1 to A.4 have already been published in *Energy & Fuels* (Dahmen and Marquardt (2016); Supporting Information).

Table A.1: Refunctionalization: Single-molecule transformations classified by target structural groups.

=O group	
 example: butanone → 2-butanol	**hydrogenation of ketones/aldehydes to alcohols** 1. C^1 is a carbon carrying an =O group 2. C^1 is not part of an oxygen coupling ($C^1{-}O{-}C^2$), i.e., it is not part of an acyclic/cyclic ester or anhydride 3. replace =O group with −OH group at C^1 [$+\mathbf{H_2}$]
COOH group	
 example: formic acid → ethanal	**reduction of carboxylic acid to aldehyde** 1. C^1 is a carbon member of a COOH group 2. transform C^1 to carry an =O group instead of being part of a COOH group [$+\mathbf{H_2}$, $\mathbf{-H_2O}$]
 example: maleic acid → maleic anhydride	**intramolecular condensation of dicarboxylic acid to cyclic anhydride** 1. C^1 and C^2 are two carbons in the same molecule, that a. are parts of COOH groups b. are not identical ($C^1{\neq}C^2$) c. have at least k carbons/oxygens inbetween (no rings smaller than k+3) d. have a maximum of l carbons/oxygens inbetween (no rings larger then l+3) e. have only one existing path between each other (not part of an existing ring) OR have two existing paths between each other, but one path is a subset of the other path (allows for the formation of two rings attached to each other) 2. take away COOH groups from C^1 and C^2 3. make a C-O-C coupling between C^1 and C^2, i.e., $C^1{-}O{-}C^2$ (ring formation) 4. place =O groups at C^1 and C^2 [$\mathbf{-H_2O}$]

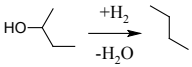
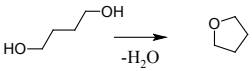
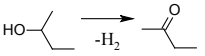
Continued on next page

Table A.1: Refunctionalization: Single-molecule transformations classified by target structural groups (continued).

cyclic −O− group (ester, ether, anhydride)	
 example: tetrahydrofuran → 1-butanol	**cyclic ether ring-opening yielding alcohols/diols** 1. C^1 and C^2 are two carbons in the same molecule, that a. are connected via a C-O-C coupling, i.e., C^1-O-C^2 b. do not have =O groups (it is not a cyclic ester or anhydride, but a cyclic ether) c. C^1−O−C^2 is not a member of an aromatic ring (it is not a furan) 2. open the ring, i.e., take away the C-O-C coupling 3. produce up to three new molecules: a. place −OH group at C^2, if no −OH group is already there [+$\mathbf{H_2}$] b. place −OH group at C^1, if no −OH group is already there [+$\mathbf{H_2}$] c. place −OH group at both C^1 and C^2, if no −OH groups are already there [+$\mathbf{H_2O}$]
 example: γ-valerolactone → butyric acid	**cyclic ester ring-opening yielding an acid** 1. C^1 and C^2 are two carbons in the same molecule, that a. are connected via a C-O-C coupling, i.e., C^1-O-C^2 b. C^1 carries an =O group, but C^2 does not carry an =O group (cyclic ester) 2. open the ring, i.e., take away the C-O-C coupling 3. remove the =O group from C^1 4. make C^1 part of a COOH group [+$\mathbf{H_2}$]
 example: cleavage of maleic anhydride	**cyclic anhydride ring-opening** 1. C^1 and C^2 are two carbons in the same molecule, that a. are connected via a C-O-C coupling, i.e., C^1-O-C^2 b. both C^1 and C^2 carry an =O group (cyclic anhydride) 2. open the ring, i.e., take away the C-O-C coupling 3. remove the =O groups from C^1 and C^2 4. produce two new molecules: a. make C^1 part of a COOH group, add =O group at C^2 [+$\mathbf{H_2}$] b. make C^2 part of a COOH group, add =O group at C^1 [+$\mathbf{H_2}$]

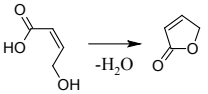
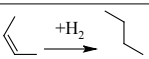
Continued on next page

Table A.1: Refunctionalization: Single-molecule transformations classified by target structural groups (continued).

−OH group	
 example: 2-butanol → butane	**reduction of alcohol to alkane** 1. C^1 is a carbon carrying an −OH group 2. remove the −OH group from C^1 [**+H$_2$,-H$_2$O**]
 example: 1,4-butanediol → tetrahydrofuran	**cyclodehydration of diols to cyclic ether (self-etherification)** 1. C^1 and C^2 are two carbons in the same molecule, that a. both carry an −OH group b. are not identical ($C^1 \neq C^2$) c. have at least k carbons/oxygens inbetween (no rings smaller than k+3) d. have a maximum of l carbons/oxygens inbetween (no rings larger then l+3) e. have only one existing path between each other (not part of an existing ring) OR have two existing paths between each other, but one path is a subset of the other path (allows for the formation of two rings attached to each other) 2. remove the −OH groups from C^1 and C^2 3. make a C-O-C coupling between C^1 and C^2, i.e., C^1−O−C^2 [**-H$_2$O**]
 example: 2-butanol → butanone	**dehydrogenation of alcohols to aldehydes/ketones** 1. C^1 is a carbon that carries an −OH group 2. C^1 carries at least one hydrogen atom 3. C^1 has no double bond 4. replace −OH group by =O group [**-H$_2$**]

Continued on next page

Table A.1: Refunctionalization: Single-molecule transformations classified by target structural groups (continued).

−OH group together with COOH group	
 example: formation of a lactone	**self-esterification (of carboxylic acid and alcohol groups)** 1. C^1 and C^2 are two carbons in the same molecule and a. C^1 and C^2 are not identical ($C^1 \neq C^2$) b. C^1 carries an −OH group c. C^2 is part of a COOH group d. C^1 and C^2 have at least k carbons/oxygens inbetween (no rings smaller than k+3) e. C^1 and C^2 have a maximum of l carbons/oxygens inbetween (no rings larger than l+3) f. C^1 and C^2 have only one existing path between each other (not part of an existing ring) OR have two existing paths between each other, but one path is a subset of the other path (allows for formation of two rings attached to each other) 2. remove the −OH group from C^1 and remove the acid group from C^2 3. make a C-O-C coupling between C^1 and C^2, i.e., C^1-O-C^2 4. introduce a =O group at C^2 to complete the cyclic ester [**-H_2O**]
C=C group	
 example: 2-butene → butane	**hydrogenation of carbon-carbon double bond** 1. C^1 and C^2 are connected by a C=C double bond ($C^1=C^2$) 2. $C^1=C^2$ is not a member of an aromatic ring 3. replace this C=C double-bond by a C−C single-bond, i.e. C^1-C^2 [+**H_2**]

Continued on next page

Table A.1: Refunctionalization: Single-molecule transformations classified by target structural groups (continued).

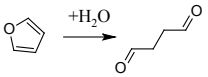
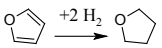
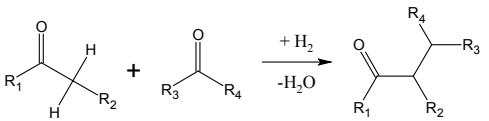
furan group	
 example: cleavage of furan	**furan ring-opening yielding aldehydes/ketones** 1. furan ring, i.e., C^1=C^2–C^3=C^4 and a C-O-C coupling of C^1 and C^4 2. no –OH groups attached to C^1 and/or C^4 3. change C^1=C^2 to C^1–C^2 4. change C^3=C^4 to C^3–C^4 5. remove C-O-C coupling of C^1 and C^4 6. place =O groups at C^1 and C^2 [+$\mathbf{H_2O}$]
 example: furan → tetrahydrofuran	**hydrogenation of furan yielding tetrahydrofuran** 1. furan ring, i.e., C^1=C^2–C^3=C^4 and a C-O-C coupling of C^1 and C^4 2. remove double-bonds C^1=C^2 and C^3=C^4 [+**2** $\mathbf{H_2}$]

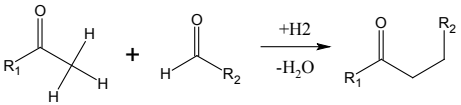
Table A.2: Aggregation: C–C coupling of two molecules via aldol condensation.



1. C^1 is a carbon in molecule M1, that
a. carries an =O group
b. is no member of a ring
c. has a single-bonded carbon neighbor C^3 carrying at least two hydrogen atoms
d. C^3 is not bonded to an oxygen atom
2. C^2 is a carbon in molecule M2
a. carrying an =O group
b. is no member of a ring
3. remove =O group from C^2
4. make a C–C bond between C^2 and C^3 by introducing two C/C identities involving a newly created auxiliary chain consisting of two carbons [+$\mathbf{H_2}$ -$\mathbf{H_2O}$]

Table A.3: Aggregation: C−C coupling of two molecules via **chain-lengthening** aldol condensation.

chain-lengthening aldol condensation



1. C^1 is a carbon in molecule M1, that

a. carries an =O group

b. is no member of a ring

c. has a single-bonded carbon neighbor C^3 carrying three hydrogen atoms

2. C^2 is a carbon in molecule M2

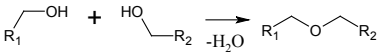
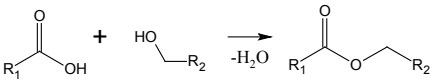
a. carrying an aldehyde group, i.e., an =O group at the end of the chain

b. is no member of a ring

3. remove the =O group from C^2

4. make a C−C bond between C^2 and C^3 by introducing two C/C identities involving a newly created auxiliary chain consisting of two carbons [$+\mathbf{H}_2$ $-\mathbf{H}_2\mathbf{O}$]

Table A.4: Aggregation: C−O−C coupling of two molecules via etherification/esterification.

etherification of alcohols

1. C^1 is a carbon carrying an −OH group in molecule M1
2. C^2 is a carbon carrying an −OH group in molecule M2
3. remove −OH group from C^1
4. remove −OH group from C^2
5. make a C−O−C coupling of C^1 and C^2 [**-H_2O**]
note: the alcohols do not have to be terminal alcohols
esterification of alcohol and carboxylic acid

1. C^1 is a carbon carrying an −OH group in molecule M1
2. carbon C^2 is a part of a COOH group in molecule M2
3. remove −OH group from C^1
4. transform COOH group at C^2 to =O group at C^2
5. make a C−O−C coupling of C^1 and C^2 [**-H_2O**]
note: the alcohol does not have to be a terminal alcohol

B Experimental ignition delay data

All data collected for the purpose of modeling are summarized in Table B.1. The allocation of a measurement to the training and external validation sets is depicted in the last column. The contents of Table B.1 have already been published in *Energy & Fuels* (Dahmen and Marquardt (2015); Supporting Information).

Table B.1: Ignition delay data collected for the purpose of model building.

compound	source	supplier and purity	stabilizer	delay [ms]	std.dev. [ms]	DCN[1]	peroxide status[2]	set
n-alkanes								
n-hexane	Yanowitz et al. (2014)	-	-	4.30[a]	-	47.9	-	T
n-heptane	ASG	-	-	3.86	0.05	52.8	-	T
n-octane	Yanowitz et al. (2014)	-	-	3.47[a]	-	58.2	-	V
n-nonane	Yanowitz et al. (2014)	-	-	3.31[a]	-	60.9	-	T
n-decane	Yanowitz et al. (2014)	-	-	3.10[b]	-	65.5	-	T
n-dodecane	Yanowitz et al. (2014)	-	-	2.85[b]	-	72.9	-	T
n-tetradecane	Yanowitz et al. (2014)	-	-	2.56[b]	-	85.1	-	T
n-hexadecane	ASG	Alfa 99%	none	2.34	0.05	98.5	<1 meq O/kg	T
iso-alkanes								
2-methylbutane	ASG	Aldrich ≥ 99%	none	10.04	0.25	24.0	-	T
2-methylpentane	Yanowitz et al. (2014)	-	-	6.21[a]	-	34.5	-	T
3-ethylpentane	Yanowitz et al. (2014)	-	-	6.30[a]	-	34.1	-	T
2-methylheptane	Yanowitz et al. (2014)	-	-	3.88[a]	-	52.6	-	T
3-methylheptane	Santana et al. (2006)	-	-	4.60[a]	-	45.0	-	T
2-methylhexane	Sarathy et al. (2011)	-	-	4.78[a]	-	43.5	-	T
2-methylheptadecane	Yanowitz et al. (2014)	-	-	2.45[b]	-	91.0	-	T
2-methyloctadecane	Yanowitz et al. (2014)	-	-	2.27[b]	-	104.0	-	T
2,2-dimethylbutane	Murphy et al. (2004)	-	-	9.82[b]	-	24.4	-	T
2,3-dimethylpentane	Yanowitz et al. (2014)	-	-	11.52[b]	-	22.0	-	V
2,4-dimethylpentane	Bogin Jr et al. (2011)	-	-	7.80[b]	-	28.6	-	T
2,2,3-trimethylbutane	Yanowitz et al. (2014)	-	-	29.61[b]	-	12.9	-	T
2,2,4-trimethylpentane	Dooley et al. (2012b)	-	-	17.69[b]	-	17.0	-	T
2,6-dimethyloctane	Yanowitz et al. (2014)	-	-	3.95[a]	-	51.7	-	T
2,2,4,4,6,8,8-heptamethylnonane	ASG	Alfa ≥ 97%	none	24.65	1.28	14.2	<1 meq O/kg	T
2,6,10-trimethyldodecane	Smagala et al. (2013); Yanowitz et al. (2014)	-	-	3.49[a]	-	58.0	-	T
alkenes								
1-hexene	Yanowitz et al. (2014)	-	-	9.04[b]	-	25.8	-	T
1-heptene	Murphy et al. (2004)	-	-	6.69[b]	-	32.0	-	T
1-octene	Yanowitz et al. (2014)	-	-	5.25[a]	-	40.0	-	V

Continued on next page

Table B.1: Ignition delay data collected for the purpose of model building (continued).

compound	source	supplier and purity	stabilizer	delay [ms]	std.dev. [ms]	DCN[1]	peroxide status[2]	set
3-octene	Hellier et al. (2013); Yanowitz et al. (2014)	-	-	5.92[a]	-	36.0[3]	-	T
1-nonene	Murphy et al. (2004)	-	-	4.01[a]	-	51.0	-	T
1-decene	Yanowitz et al. (2014)	-	-	4.18[a]	-	49.1	-	T
1-dodecene	Yanowitz et al. (2014)	-	-	3.57[a]	-	56.8	-	T
1-hexadecene	Yanowitz et al. (2014)	-	-	2.77[b]	-	75.9	-	T
1,9-decadiene	Yanowitz et al. (2014)	-	-	5.11[a]	-	41.0	-	T
2-methyl-1,3-butadiene	ASG	Alfa 99%	0.02% TBC	29.38	7.12	13.0	2 meq O/kg	T
2,5-dimethyl-2,4-hexadiene	ASG	Aldrich 96%	none	13.50	4.19	19.9	2 meq O/kg	T
cyclo-alkanes								
cyclohexane	ASG	Aldrich $\geq$ 99%	none	13.39	1.54	20.0	-	T
cyclooctane	Yanowitz et al. (2014)	-	-	11.28[b]	-	22.3	-	T
methylcyclopentane	Yanowitz et al. (2014)	-	-	17.33[b]	-	17.2	-	T
methylcyclohexane	Yanowitz et al. (2014)	-	-	9.82[b]	-	24.4	-	V
1,3,5-trimethylcyclohexane	Yanowitz et al. (2014)	-	-	7.14[b]	-	30.5	-	T
ethylcyclohexane	Yanowitz et al. (2014)	-	-	5.95[a]	-	35.8	-	T
butylcyclohexane	Yanowitz et al. (2014)	-	-	4.33[a]	-	47.6	-	T
decahydronaphthalene	Santana et al. (2006)	-	-	6.43[a]	-	33.0	-	T
1,2,4-trimethylcyclohexane	ASG	Aldrich 97%	none	9.80	0.14	24.4	<1 meq O/kg	T
1,3,5-triisopropylcyclohexane	Yanowitz et al. (2014)	-	-	9.30[b]	-	25.3	-	T
cyclo-alkenes								
cyclohexene	ASG	Aldrich 99%	none	15.89	0.41	18.1	-	T
1,5-cyclooctadiene	ASG	Aldrich 99%	50-150ppm TBC	9.07	0.12	25.7	5 meq O/kg*	T
α-pinene	ASG	Aldrich 98%	none	14.74	0.50	18.9	31 meq O/kg*	T
β-pinene	ASG	Aldrich 99%	none	12.87	0.57	20.5	10 meq O/kg*	T
limonene	ASG	Alfa 97%	yes, unspecified	14.77	0.30	18.9	8 meq O/kg*	T
γ-terpinene	ASG	Aldrich $\geq$ 97%	none	13.14	1.75	20.3	4 meq O/kg*	T
bisabolene	Yanowitz et al. (2014)	-	-	6.64[b]	-	32.2	-	V

Continued on next page

Table B.1: Ignition delay data collected for the purpose of model building (continued).

compound	source	supplier and purity	stabilizer	delay [ms]	std.dev. [ms]	DCN[1]	peroxide status[2]	set
alcohols								
1-heptanol	Yanowitz et al. (2014)	-	-	7.65[b]	-	29.0	-	T
2-ethyl-1-hexanol	Yanowitz et al. (2014)	-	-	10.40[b]	-	23.5	-	T
3-octanol	Murphy et al. (2004)	-	-	9.41[b]	-	25.1	-	T
2-nonanol	Yanowitz et al. (2014)	-	-	5.31[b]	-	39.6	-	T
1-butanol	Haas et al. (2011)	-	-	34.29[b]	-	12.0	-	T
2-butanol	Haas et al. (2011)	-	-	75.36[b]	-	8.5	-	V
isobutanol	Haas et al. (2011)	-	-	75.36[b]	-	8.5	-	T
1-octanol	ASG	Roth $\geq$ 99%	none	6.33	0.15	33.9	-	T
3-methyl-1-butanol	ASG	Roth $\geq$ 98%	none	28.62	1.22	13.1	-	T
2-methyl-2-butanol	ASG	Aldrich 99%	none	33.71	8.64	12.1	-	T
nerolidol	ASG	Alfa $\geq$ 97%	none	14.31	0.61	19.2	6 meq O/kg*	T
cyclohexanol	ASG	Aldrich 99%	none	19.13	7.96	16.3	-	T
cyclopentanol	ASG	Alfa 99%	none	52.86	28.59	9.8	-	T
geraniol	ASG	Aldrich $\geq$ 97%	none	18.69	2.44	16.5	-	T
3,7-dimethyl-1-octanol	Yanowitz et al. (2014)	-	-	7.54[b]	-	29.3	-	T
β-citronellol	Yanowitz et al. (2014)	-	-	9.14[b]	-	25.6	-	T
linalool	ASG	Aldrich $\geq$ 97%	none	29.66	1.28	12.9	18 meq O/kg*	T
aldehydes								
octanal	ASG	Aldrich 99%	none	2.29	0.06	102.5	-	T
hexanal	ASG	Alfa 98%	none	2.78	0.06	75.2	-	V
butanal	ASG	Aldrich $\geq$ 99%	none	5.09	0.08	41.1	-	T
isobutyraldehyde	ASG	Aldrich 98%	none	12.28	0.28	21.1	-	T
pentanal	ASG	Alfa 97%	none	3.23	0.08	62.2	-	T
3-cyclohexene-1-carboxaldehyde	ASG	Alfa 98%	none	8.01	0.22	28.1	5 meq O/kg*	T
ketones								
3-pentanone	ASG	Alfa 99%	none	14.04	10.85	19.5	22 meq O/kg*	T
2-heptanone	Murphy et al. (2004)	-	-	7.30[b]	-	30.0	-	T
2-octanone	ASG	Alfa 98%	none	5.80	0.10	36.6	-	T
3-octanone	Yanowitz et al. (2014)	-	-	5.92[a]	-	36.0	-	T
2-nonanone	ASG	Alfa 98%	none	4.48	0.09	46.1	9 meq O/kg*	T

Continued on next page

Table B.1: Ignition delay data collected for the purpose of model building (continued).

compound	source	supplier and purity	stabilizer	delay [ms]	std.dev. [ms]	DCN[1]	peroxide status[2]	set
4-methyl-2-pentanone	ASG	Alfa 99%	none	30.89	6.09	12.6	-	V
2,4-dimethyl-3-pentanone	ASG	Aldrich 98%	none	20.18	10.30	15.8	-	T
cyclohexanone	ASG	Aldrich $\geq$ 99%	none	47.09	36.95	10.4	-	T
cycloheptanone	ASG	Alfa 99%	none	11.13	0.26	22.5	-	T
3,3,5-trimethylcyclohexanone	ASG	Aldrich 98%	none	35.03	8.66	11.9	<1 meq O/kg	T
menthone	ASG	Alfa 98%	none	12.77	0.23	20.6	41 meq O/kg*	T
methyl valerate	Yanowitz et al. (2014)	-	-	27.88[b]	-	13.3	-	T
methyl hexanoate	ASG	Aldrich 99%	none	9.32	0.24	25.3	-	T
methyl heptanoate	Yanowitz et al. (2014)	-	-	6.27[a]	-	34.2	-	T
propyl valerate	Yanowitz et al. (2014)	-	-	12.69[b]	-	20.7	-	T
butyl butanoate	Murphy et al. (2004)	-	-	16.33[b]	-	17.8	-	T
butyl valerate	Yanowitz et al. (2014)	-	-	10.40[b]	-	23.5	-	T
pentyl valerate	Yanowitz et al. (2014)	-	-	7.72[b]	-	28.8	-	T
hexyl acetate	ASG	Aldrich 99%	none	6.63	0.14	32.2	-	V
ethyl valerate	Yanowitz et al. (2014)	-	-	15.15[b]	-	18.6	-	T
ethyl hexanoate	ASG	Aldrich $\geq$ 99%	none	8.24	0.21	27.5	-	T
methyl decanoate	Yanowitz et al. (2014)	-	-	3.87[a]	-	52.7	-	T
methyl laurate	Yanowitz et al. (2014)	-	-	3.07[b]	-	66.3	-	T
methyl tetradecanoate	ASG	Alfa 99%	none	2.77	0.06	75.8	<1 meq O/kg	T
menthyl acetate	ASG	Alfa 98%	none	14.02	0.25	19.5	<1 meq O/kg	T
2-methyl-9-decenoate	Yanowitz et al. (2014)	-	-	5.51[a]	-	38.3	-	T
methyl linolenate	Yanowitz et al. (2014)	-	-	5.73[a]	-	37.0	-	T
methyl linoleate	Yanowitz et al. (2014)	-	-	4.73[a]	-	43.9	-	T
vinyl laurate	ASG	Aldrich 99%	none	2.74	0.05	77.0	4 meq O/kg*	T
acyclic ethers								
1-methoxyhexane	Yanowitz et al. (2014)	-	-	2.32[b]	-	99.8	-	T
dibutyl ether	ASG	-	-	2.16	0.04	115.4	-	T
diisopropyl ether	ASG	Aldrich 99%	BHT or HYD	10.31	1.14	23.6	-	T
diisobutyl ether	Yanowitz et al. (2014)	-	-	3.38[a]	-	59.7	-	T
diisoamyl ether	Yanowitz et al. (2014)	-	-	2.37[b]	-	96.3	-	T

Continued on next page

Table B.1: Ignition delay data collected for the purpose of model building (continued).

compound	source	supplier and purity	stabilizer	delay [ms]	std.dev. [ms]	DCN[1]	peroxide status[2]	set
diethylene glycol dimethyl ether	ASG	Aldrich ≥ 99%	none	1.65	0.03	315.7	-	T
1,2-dimethoxyethane	ASG	Aldrich ≥ 99%	0.01% DTB	2.64	0.04	81.0	<1 meq O/kg	T
diethoxymethane	ASG	Aldrich > 99%	50-150 ppm BHT	3.53	0.04	57.3	1 meq O/kg	V
2,2-dimethoxypropane	ASG	Aldrich 98%	none	6.88	0.29	31.3	<1 meq O/kg	T
cyclopentyl methyl ether	ASG	Aldrich > 99%	50 ppm BHT	4.35	0.08	47.3	<1 meq O/kg	T
acetaldehyde diethyl acetal	ASG	Aldrich 99%	none	3.76	0.04	54.1	-	T
dodecyl vinyl ether	ASG	Aldrich 98%	none	2.30	0.04	101.7	<1 meq O/kg	T
1,4-cyclohexanedimethanol divinyl ether	ASG	Aldrich 98%	none	3.30	0.09	61.1	<1 meq O/kg	T
furans								
2-methylfuran	ASG	-	-	63.53	16.85	9.1	-	T
furfural	ASG	Alfa 98%	none	25.65	3.92	13.9	-	T
2-ethylfuran	Sudholt et al. (2015)	-	-	48.67[b]	-	10.2	-	T
2-butylfuran	Sudholt et al. (2015)	-	-	28.73[b]	-	13.1	-	V
2,5-dimethylfuran	Sudholt et al. (2015)	-	-	42.02[b]	-	10.9	-	T
tetrahydrofurans								
2-methyltetrahydrofuran	ASG	-	-	12.12	1.42	21.3	-	T
tetrahydrofuran	ASG	Aldrich > 99%	250 ppm BHT	11.59	2.68	21.9	-	T
tetrahydrofurfuryl acetate	ASG	Alfa 98%	none	15.71	0.30	18.2	<1 meq O/kg	T
2-ethyltetrahydrofuran	Sudholt et al. (2015)	-	-	7.99[b]	-	28.1	-	V
2-butyltetrahydrofuran	Sudholt et al. (2015)	-	-	4.55[a]	-	45.5	-	T
other cyclic ethers								
eucalyptol	ASG	Alfa 99%	none	14.85	1.22	18.8	5 meq O/kg*	T
rose oxide	Yanowitz et al. (2014)	-	-	7.30[b]	-	30.0	-	T
3,4-dihydro-2H-pyran	ASG	Aldrich 97%	none	10.12	0.70	23.9	-	T
tetrahydropyran	ASG	Alfa ≥ 98%	none	5.52	0.10	38.2	-	T

Continued on next page

Table B.1: Ignition delay data collected for the purpose of model building (continued).

compound	source	supplier and purity	stabilizer	delay [ms]	std.dev. [ms]	DCN[1]	peroxide status[2]	set
2,3-dihydrofuran	Sudholt et al. (2015)	-	-	13.42[b]	-	20.0	-	V
2,5-dihydrofuran	Sudholt et al. (2015)	-	-	20.63[b]	-	15.6	-	T
lactones								
γ-undecanolactone	ASG	Alfa 99%	none	3.88	0.17	52.6	23 meq O/kg*	T
whiskey lactone	ASG	Aldrich 98%	none	8.38	0.16	27.2	2 meq O/kg	T
ϵ-decalactone	ASG	Aldrich 99%	none	5.18	0.05	40.5	<1 meq O/kg	V
δ-undecalactone	ASG	Aldrich $\geq$ 97%	none	4.23	0.11	48.6	2 meq O/kg	T
benzene compounds								
n-butylbenzene	ASG	Alfa 99%	none	29.79	4.53	12.9	-	T
dibenzylether	ASG	Aldrich $\geq$ 98%	none	84.12	8.47	8.1	-	T
1,2-dimethoxybenzene	ASG	Alfa 99%	none	53.22	8.53	9.8	-	V
benzaldehyde dimethyl acetal	ASG	Alfa 99%	none	46.85	15.69	10.4	2 meq O/kg	T
α-methyl-trans-cinnamaldehyde	ASG	Aldrich 98%	none	14.51	2.52	19.1	2 meq O/kg*	T
1,2,4-trimethylbenzene	Yanowitz et al. (2014)	-	-	67.14[b]	-	8.9	-	T
1,2,3,4-tetrahydronaphthalene	Yanowitz et al. (2014)	-	-	67.14[b]	-	8.9	-	T
4-methoxybenzaldehyde	ASG	Alfa 98%	none	9.04	0.95	25.8	<1 meq O/kg	T
polyfunctionals								
1-butoxy-2-propanol	Yanowitz et al. (2014)	-	-	5.90[a]	-	36.1	-	T
dipropylene glycol monomethyl ether	Yanowitz et al. (2014)	-	-	4.73[a]	-	43.9	-	T
tripropylene glycol monomethyl ether	Yanowitz et al. (2014)	-	-	2.64[b]	-	81.3	-	T
triethylene glycol monomethyl ether	Yanowitz et al. (2014)	-	-	2.65[b]	-	80.7	-	T
diethylene glycol monomethyl ether	ASG	Alfa 98%	none	5.51	0.11	38.3	-	T

Continued on next page

Table B.1: Ignition delay data collected for the purpose of model building (continued).

compound	source	supplier and purity	stabilizer	delay [ms]	std.dev. [ms]	DCN[1]	peroxide status[2]	set
2-isopropoxyethanol	ASG	Aldrich 99%	none	8.13	0.25	27.8	-	V
2-methoxyethanol	ASG	Aldrich 99%	50 ppm BHT	20.98	0.50	15.5	-	T
3-methoxy-3-methyl-1-butanol	ASG	Alfa $\geq$ 98%	none	50.94	21.39	10.0	<1 meq O/kg	T
1-methoxy-2-propanol	ASG	Aldrich 99%	none	18.97	0.38	16.3	<1 meq O/kg	T
furfuryl alcohol	ASG	Aldrich 99%	none	55.11	7.40	9.7	-	T
ethyl acetoacetate	ASG	Alfa 99%	none	26.44	1.40	13.7	<1 meq O/kg	T
tetrahydrofurfuryl alcohol	Sudholt et al. (2015)	-	-	16.17[b]	-	17.9	-	V
propylene glycol monomethyl ether acetate	ASG	Alfa 99%	50 ppm BHT	10.07	0.18	24.0	<1 meq O/kg	T
ethylene glycol vinyl ether	ASG	Aldrich 97%	none	25.30	3.07	14.0	<1 meq O/kg	T
4-hydroxy-4-methyl-2-pentanone	ASG	Aldrich 99%	none	37.19	12.86	11.5	1 meq O/kg*	T
butyl levulinate	Yanowitz et al. (2014)	-	-	23.93[b]	-	14.4	-	T

[1] according to ASTM D6890 (2011): if $3.1 \leq \tau \leq 6.5$ then $\mathrm{DCN} = 4.46 + 186.6/\tau$ else $\mathrm{DCN} = 83.99(\tau - 1.512)^{-0.658} + 3.547$

[2] according to ISO 3960 (2010)

[a] calculated from $\tau = 186.6/(\mathrm{DCN} - 4.46)$

[b] calculated from $\tau = [(\mathrm{DCN} - 3.547)/83.99]^{(1/-0.658)} + 1.512$

* value reported refers to peroxide level before column chromatography

abbreviations: ignition delay (τ); standard deviation of the ignition delay (std.dev.); butyl-hydroxy-toluene (BHT); hydroquinone (HYD); 4-tert-butylcatechol (TBC); 2,6-di-tert-butyl-4-methylphenol (DTB); Sigma-Aldrich Corporation (Aldrich); ASG Analytik-Service GmbH (ASG); Alfa Aesar (Alfa); Carl Roth GmbH + Co. KG (Roth); training set (T); external validation set (V)

C GC model for the derived cetane number

The contents of Appendix C have already been published in *Energy & Fuels* (Dahmen and Marquardt (2015); Supporting Information).

C.1 Group decomposition and descriptor data

Group decomposition and descriptor data for all compounds from the training and validation sets can be found in Table C.1.

Table C.1: Group decomposition and descriptors for compounds in training and external validation sets.

	d_{nAB}	d_{nCCDB}	$\ln\left(d_{pS}\right)$	d_{nQC}	g_{-CH3}	$\ln\left(g_{-CH2-(non-ring)}+1\right)$	g_{-OH}	$g_{-O-(non-ring)}$	$g_{-O-(ring)}$	$g_{>C=O(non-ring)}$	$g_{>C=O(ring)}$	$g_{O=CH-}$	$g_{-COO-(non-ring)}$
n-hexane			-1.450		2	1.609							
n-heptane			-2.358		2	1.792							
n-nonane			-4.273		2	2.079							
n-decane			-5.277		2	2.197							
n-dodecane			-7.358		2	2.398							
n-tetradecane			-9.494		2	2.565							
n-hexadecane			-11.612		2	2.708							
2-methylbutane			-0.553		3	0.693							
2-methylpentane			-1.414		3	1.099							
3-ethylpentane			-2.309		3	1.386							
2-methylheptane			-3.237		3	1.609							
3-methylheptane			-3.237		3	1.609							
2-methylhexane			-2.309		3	1.386							
2-methylheptadecane			-13.479		3	2.708							
2-methyloctadecane			-14.410		3	2.773							
2,2-dimethylbutane			-1.272	1	4	0.693							
2,4-dimethylpentane			-2.261		4	0.693							
2,2,3-trimethylbutane			-2.094	1	5								
2,2,4-trimethylpentane			-2.981	1	5	0.693							
2,6-dimethyloctane			-5.103		4	1.609							
2,2,4,6,8,8-heptamethylnonane			-9.902	3	9	1.386							
2,6,10-trimethyldodecane			-10.148		5	2.079							
1-hexene		1	-1.316		1	1.386							

Continued on next page

Table C.1: Group decomposition and descriptors for compounds in training and external validation sets (continued).

	d_{nAB}	d_{nCCDB}	$\ln\left(d_{pS}\right)$	d_{nQC}	g_{-CH3}	$\ln\left(g_{-CH2-(non-ring)}+1\right)$	g_{-OH}	$g_{-O-(non-ring)}$	$g_{-O-(ring)}$	$g_{>C=O(non-ring)}$	$g_{>C=O(ring)}$	$g_{O=CH-}$	$g_{-COO-(non-ring)}$
1-heptene		1	-2.214		1	1.609							
3-octene		1	-3.432		2	1.609							
1-nonene		1	-4.111		1	1.946							
1-decene		1	-5.106		1	2.079							
1-dodecene		1	-7.174		1	2.303							
1-hexadecene		1	-11.430		1	2.639							
1,9-decadiene		2	-4.938			1.946							
2-methyl-1,3-butadiene		2	-0.320		1								
2,5-dimethyl-2,4-hexadiene		2	-3.472		4								
cyclohexane			-2.345										
cyclooctane			-4.427										
methylcyclopentane			-1.992		1								
1,3,5-trimethylcyclohexane			-4.375		3								
ethylcyclohexane			-3.941		1	0.693							
butylcyclohexane			-5.867		1	1.386							
decahydronaphthalene			-6.217										
1,2,4-trimethylcyclohexane			-4.375		3								
1,3,5-triisopropylcyclohexane			-9.894		6								
cyclohexene		1	-2.324										
1,5-cyclooctadiene		2	-4.398										
α-pinene		1	-5.738	1	3								
β-pinene		1	-5.523	1	2								
limonene		2	-5.798		2								
γ-terpinene		2	-6.405		3								

Continued on next page

Table C.1: Group decomposition and descriptors for compounds in training and external validation sets (continued).

	d_{nAB}	d_{nCCDB}	$\ln\left(d_{pS}\right)$	d_{nQC}	g_{-CH3}	$\ln\left(g_{-CH2-(non-ring)}+1\right)$	g_{-OH}	$g_{-O-(non-ring)}$	$g_{-O-(ring)}$	$g_{>C=O(non-ring)}$	$g_{>C=O(ring)}$	$g_{O=CH-}$	$g_{-COO-(non-ring)}$
1-heptanol			-7.397		1	1.946	1						
2-ethyl-1-hexanol			-8.473		2	1.792	1						
3-octanol			-8.473		2	1.792	1						
2-nonanol			-9.679		2	1.946	1						
1-butanol			-3.922		1	1.386	1						
isobutanol			-3.850		2	0.693	1						
1-octanol			-8.601		1	2.079	1						
3-methyl-1-butanol			-4.966		2	1.099	1						
2-methyl-2-butanol			-4.700		3	0.693	1						
nerolidol		3	-15.919		4	1.609	1						
cyclohexanol			-6.954				1						
cyclopentanol			-5.684				1						
geraniol		2	-11.036		3	1.386	1						
3,7-dimethyl-1-octanol			-10.740		3	1.792	1						
β-citronellol		1	-10.892		3	1.609	1						
linalool		2	-10.268		3	1.099	1						
octanal			-5.689		1	1.946						1	
butanal			-1.657		1	1.099						1	
isobutyraldehyde			-1.617		2							1	
pentanal			-2.617		1	1.386						1	
3-cyclohexene-1-carboxaldehyde		1	-5.284									1	
3-pentanone			-2.794		2	1.099				1			
2-heptanone			-4.793		2	1.609				1			

Continued on next page

Table C.1: Group decomposition and descriptors for compounds in training and external validation sets (continued).

	d_{nAB}	d_{nCCDB}	$\ln(d_{p^S})$	d_{nQC}	g_{-CH3}	$\ln(g_{-CH2-(non-ring)}+1)$	g_{-OH}	$g_{-O-(non-ring)}$	$g_{-O-(ring)}$	$g_{>C=O(non-ring)}$	$g_{>C=O(ring)}$	$g_{O=CH-}$	$g_{-COO-(non-ring)}$
2-octanone			-5.838		2	1.792				1			
3-octanone			-5.838		2	1.792				1			
2-nonanone			-6.909		2	1.946				1			
2,4-dimethyl-3-pentanone			-4.629		4					1			
cyclohexanone			-5.243								1		
cycloheptanone			-6.346								1		
3,3,5-trimethylcyclohexanone			-7.562	1	3						1		
menthone			-8.558		3						1		
methyl valerate			-3.974		2	1.386							1
methyl hexanoate			-4.997		2	1.609							1
methyl heptanoate			-6.049		2	1.792							1
propyl valerate			-6.049		2	1.792							1
butyl butanoate			-6.049		2	1.792							1
butyl valerate			-7.126		2	1.946							1
pentyl valerate			-8.223		2	2.079							1
ethyl valerate			-4.997		2	1.609							1
ethyl hexanoate			-6.049		2	1.792							1
methyl decanoate			-9.333		2	2.197							1
methyl laurate			-11.557		2	2.398							1
methyl tetradecanoate			-13.713		2	2.565							1
menthyl acetate			-10.056		4								1
2-methyl-9-decenoate		1	-9.135		1	2.079							1
methyl linolenate		3	-17.792		2	2.398							1

Continued on next page

Table C.1: Group decomposition and descriptors for compounds in training and external validation sets (continued).

	d_{nAB}	d_{nCCDB}	$\ln\left(d_{pS}\right)$	d_{nQC}	g_{-CH3}	$\ln\left(g_{-CH2-(non-ring)}+1\right)$	g_{-OH}	$g_{-O-(non-ring)}$	$g_{-O-(ring)}$	$g_{>C=O(non-ring)}$	$g_{>C=O(ring)}$	$g_{O=CH-}$	$g_{-COO-(non-ring)}$
methyl linoleate		2	-17.678		2	2.565							1
vinyl laurate		1	-12.456		1	2.398							1
1-methoxyhexane			-3.348		2	1.792		1					
dibutyl ether			-4.340		2	1.946		1					
diisopropyl ether			-2.291		4			1					
diisobutyl ether			-4.189		4	1.099		1					
diisoamyl ether			-6.207		4	1.609		1					
diethylene glycol dimethyl ether			-4.481		2	1.609		3					
1,2-dimethoxyethane			-1.478		2	1.099		2					
2,2-dimethoxypropane			-2.196		4			2					
cyclopentyl methyl ether			-2.936		1			1					
acetaldehyde diethyl acetal			-3.336		3	1.099		2					
dodecyl vinyl ether		1	-10.550		1	2.485		1					
1,4-cyclohexanedimethanol divinyl ether		2	-9.580			1.099		2					
2-methylfuran	5	2	-2.586		1				1				
furfural	5	2	-4.942						1			1	
2-ethylfuran	5	2	-3.538		1	0.693			1				
2,5-dimethylfuran	5	2	-3.733		2				1				
2-methyltetrahydrofuran			-2.211		1				1				
tetrahydrofuran			-1.520						1				
tetrahydrofurfuryl acetate			-6.917		1	0.693			1				1
2-butyltetrahydrofuran			-5.078		1	1.386			1				
eucalyptol			-6.901		3				1				

Continued on next page

Table C.1: Group decomposition and descriptors for compounds in training and external validation sets (continued).

	d_{nAB}	d_{nCCDB}	$\ln\left(d_p{}^s\right)$	d_{nQC}	g_{-CH3}	$\ln\left(g_{-CH2-(non-ring)}+1\right)$	g_{-OH}	$g_{-O-(non-ring)}$	$g_{-O-(ring)}$	$g_{>C=O(non-ring)}$	$g_{>C=O(ring)}$	$g_{O=CH-}$	$g_{-COO-(non-ring)}$
rose oxide		1	-6.927		3				1				
3,4-dihydro-2H-pyran		1	-2.557						1				
tetrahydropyran			-2.578						1				
2,5-dihydrofuran		1	-1.496						1				
γ-undecanolactone			-11.593		1	1.946			1		1		
whiskey lactone			-9.035		2	1.386			1		1		
δ-undecalactone			-11.638		1	1.792			1		1		
n-butylbenzene	6	3	-6.308		1	1.386							
dibenzylether	12	6	-12.691			1.099		1					
benzaldehyde dimethyl acetal	6	3	-7.481		2			2					
α-methyl-trans-cinnamaldehyde	6	4	-8.946		1							1	
1,2,4-trimethylbenzene	6	3	-5.689		3								
1,2,3,4-tetrahydronaphthalene	6	3	-7.162										
4-methoxybenzaldehyde	6	3	-8.235		1			1				1	
1-butoxy-2-propanol			-8.631		2	1.609	1	1					
dipropylene glycol monomethyl ether			-9.900		3	1.099	1	2					
tripropylene glycol monomethyl ether			-14.915		4	1.386	1	3					
triethylene glycol monomethyl ether			-11.675		1	1.946	1	3					
diethylene glycol monomethyl ether			-7.680		1	1.609	1	2					
2-methoxyethanol			-3.993		1	1.099	1	1					
3-methoxy-3-methyl-1-butanol			-7.083		3	1.099	1	1					
1-methoxy-2-propanol			-5.058		2	0.693	1	1					
furfuryl alcohol	5	2	-7.831			0.693	1		1				

Continued on next page

Table C.1: Group decomposition and descriptors for compounds in training and external validation sets (continued).

	d_{nAB}	d_{nCCDB}	$\ln\left(d_{pS}\right)$	d_{nQC}	g_{-CH3}	$\ln\left(g_{-CH2-(non-ring)}+1\right)$	g_{-OH}	$g_{-O-(non-ring)}$	$g_{-O-(ring)}$	$g_{>C=O(non-ring)}$	$g_{>C=O(ring)}$	$g_{O=CH-}$	$g_{-COO-(non-ring)}$
ethyl acetoacetate			-6.675		2	1.099				1			1
propylene glycol monomethyl ether acetate			-4.996		3	0.693		1					1
ethylene glycol vinyl ether		1	-4.961			1.099	1	1					
4-hydroxy-4-methyl-2-pentanone			-8.890		3	0.693	1			1			
butyl levulinate			-10.081		2	1.792				1			1
n-octane			-3.299		2	1.946							
2,3-dimethylpentane			-2.261		4	0.693							
1-octene		1	-3.146		1	1.792							
methylcyclohexane			-3.020		1								
bisabolene		3	-11.566		4	1.099							
2-butanol			-3.850		2	0.693	1						
hexanal			-3.609		1	1.609						1	
4-methyl-2-pentanone			-3.709		3	0.693				1			
hexyl acetate			-6.049		2	1.792							1
diethoxymethane			-2.422		2	1.386		2					
2-butylfuran	5	2	-5.529		1	1.386			1				
2-ethyltetrahydrofuran			-3.137		1	0.693			1				
2,3-dihydrofuran		1	-1.496						1				
ϵ-decalactone			-10.593		1	1.386			1		1		
1,2-dimethoxybenzene	6	3	-6.742		2			2					
2-isopropoxyethanol			-6.225		2	1.099	1	1					
tetrahydrofurfuryl alcohol			-7.238			0.693	1		1				

C.2 Comparison of measurement data with results of GC model

Measurements and predictions for all compounds from the training and validation sets can be found in Table C.2. The predictions have been computed with the DCN group contribution model from Chapter 4.

Table C.2: Measurements vs. predictions.

	$\ln \tau$		τ		DCN[1]	
	meas.	pred.	meas.	pred.	meas.	pred.
n-hexane	1.46	1.60	4.30	4.94	47.9	42.3
n-heptane	1.35	1.46	3.86	4.31	52.8	47.7
n-nonane	1.20	1.24	3.31	3.46	60.9	58.3
n-decane	1.13	1.15	3.10	3.16	64.7	63.6
n-dodecane	1.05	0.99	2.85	2.69	69.9	73.9
n-tetradecane	0.94	0.86	2.56	2.35	77.4	83.8
n-hexadecane	0.85	0.74	2.34	2.11	84.2	93.0
2-methylbutane	2.31	2.17	10.07	8.78	23.0	25.7
2-methylpentane	1.83	1.89	6.21	6.62	34.5	32.6
3-ethylpentane	1.84	1.69	6.30	5.41	34.1	38.9
2-methylheptane	1.35	1.53	3.88	4.62	52.6	44.9
3-methylheptane	1.53	1.53	4.60	4.62	45.0	44.9
2-methylhexane	1.56	1.69	4.78	5.41	43.5	38.9
2-methylheptadecane	0.90	0.71	2.45	2.04	80.5	96.1
2-methyloctadecane	0.82	0.67	2.27	1.95	86.5	100.1
2,2-dimethylbutane	2.28	2.66	9.82	14.28	23.5	17.5
2,4-dimethylpentane	2.05	2.09	7.80	8.08	28.4	27.6
2,2,3-trimethylbutane	3.39	3.15	29.61	23.39	10.8	12.4
2,2,4-trimethylpentane	2.87	2.56	17.69	12.90	15.0	18.9
2,6-dimethyloctane	1.37	1.46	3.95	4.30	51.7	47.8
2,2,4,4,6,8,8-heptamethylnonane	3.20	2.71	24.53	15.04	12.1	16.9
2,6,10-trimethyldodecane	1.25	1.06	3.49	2.90	58.0	68.8
1-hexene	2.20	1.87	9.04	6.46	25.1	33.4
1-heptene	1.90	1.69	6.69	5.43	32.3	38.8
3-octene	1.78	1.67	5.92	5.30	36.0	39.7
1-nonene	1.39	1.42	4.01	4.15	51.0	49.5
1-decene	1.43	1.31	4.18	3.71	49.1	54.7
1-dodecene	1.27	1.12	3.57	3.08	56.8	65.1
1-hexadecene	1.02	0.84	2.77	2.32	71.9	85.0
1,9-decadiene	1.63	1.50	5.11	4.49	41.0	46.0
2-methyl-1,3-butadiene	3.33	3.14	27.94	23.10	11.1	12.5
2,5-dimethyl-2,4-hexadiene	2.54	3.08	12.68	21.69	19.2	13.1
cyclohexane	2.58	2.05	13.20	7.78	18.6	28.4
cyclooctane	2.42	1.85	11.28	6.37	21.0	33.7
methylcyclopentane	2.85	2.18	17.33	8.88	15.2	25.5
1,3,5-trimethylcyclohexane	1.97	2.12	7.14	8.37	30.6	26.8
ethylcyclohexane	1.78	1.68	5.95	5.37	35.8	39.2
butylcyclohexane	1.46	1.30	4.33	3.65	47.6	55.5
decahydronaphthalene	1.86	1.70	6.43	5.45	33.5	38.7
1,2,4-trimethylcyclohexane	2.28	2.12	9.78	8.37	23.5	26.8
1,3,5-triisopropylcyclohexane	2.23	1.85	9.30	6.38	24.5	33.7
cyclohexene	2.77	2.36	15.96	10.63	16.2	22.0
1,5-cyclooctadiene	2.21	2.46	9.12	11.66	24.9	20.5

Continued on next page

Table C.2: Measurements vs. predictions (continued).

	$\ln \tau$		τ		DCN[1]	
	meas.	pred.	meas.	pred.	meas.	pred.
α-pinene	2.69	2.77	14.73	15.98	17.1	16.1
β-pinene	2.55	2.68	12.81	14.55	19.0	17.3
limonene	2.69	2.51	14.73	12.28	17.1	19.7
γ-terpinene	2.57	2.55	13.07	12.76	18.7	19.1
1-heptanol	2.03	2.18	7.65	8.84	28.9	25.6
2-ethyl-1-hexanol	2.34	2.24	10.40	9.42	22.4	24.3
3-octanol	2.24	2.24	9.41	9.42	24.3	24.3
2-nonanol	1.67	2.04	5.31	7.67	39.6	28.8
1-butanol	3.53	2.95	34.29	19.20	9.9	14.2
isobutanol	4.32	3.66	75.36	38.87	6.9	9.3
1-octanol	1.85	1.99	6.36	7.31	33.8	30.0
3-methyl-1-butanol	3.35	3.14	28.50	23.21	11.0	12.5
2-methyl-2-butanol	3.47	3.67	32.14	39.31	10.3	9.2
nerolidol	2.66	2.71	14.30	14.99	17.5	16.9
cyclohexanol	2.85	3.39	17.29	29.64	15.3	10.8
cyclopentanol	3.72	3.61	41.26	36.88	9.0	9.5
geraniol	2.92	3.02	18.54	20.42	14.5	13.6
3,7-dimethyl-1-octanol	2.02	2.10	7.54	8.15	29.2	27.4
β-citronellol	2.21	2.50	9.14	12.22	24.9	19.7
linalool	3.39	3.36	29.67	28.67	10.8	11.0
octanal	0.83	1.10	2.29	3.00	85.8	66.7
butanal	1.63	1.64	5.10	5.16	41.0	40.6
isobutyraldehyde	2.51	2.24	12.30	9.35	19.6	24.4
pentanal	1.17	1.46	3.22	4.31	62.4	47.8
3-cyclohexene-1-carboxaldehyde	2.08	1.96	8.00	7.12	27.8	30.7
3-pentanone	2.44	2.43	11.47	11.38	20.7	20.9
2-heptanone	1.99	1.95	7.30	7.04	30.0	31.0
2-octanone	1.76	1.77	5.81	5.90	36.6	36.1
3-octanone	1.78	1.77	5.92	5.90	36.0	36.1
2-nonanone	1.50	1.62	4.48	5.07	46.1	41.3
2,4-dimethyl-3-pentanone	2.88	3.16	17.81	23.60	14.9	12.4
cyclohexanone	3.49	2.72	32.79	15.16	10.2	16.8
cycloheptanone	2.41	2.58	11.13	13.13	21.2	18.7
3,3,5-trimethylcyclohexanone	3.51	3.36	33.45	28.92	10.0	10.9
menthone	2.55	2.64	12.81	14.06	19.0	17.7
methyl valerate	3.33	2.52	27.88	12.40	11.2	19.5
methyl hexanoate	2.23	2.27	9.30	9.67	24.5	23.7
methyl heptanoate	1.84	2.06	6.27	7.87	34.2	28.2
propyl valerate	2.54	2.06	12.69	7.87	19.2	28.2
butyl butanoate	2.79	2.06	16.33	7.87	15.9	28.2
butyl valerate	2.34	1.89	10.40	6.59	22.4	32.8
pentyl valerate	2.04	1.73	7.72	5.64	28.6	37.5
ethyl valerate	2.72	2.27	15.15	9.67	16.8	23.7
ethyl hexanoate	2.11	2.06	8.25	7.87	27.1	28.2
methyl decanoate	1.35	1.59	3.87	4.92	52.7	42.4
methyl laurate	1.12	1.36	3.07	3.90	65.3	52.3
methyl tetradecanoate	1.02	1.18	2.77	3.24	71.7	62.0

Continued on next page

Table C.2: Measurements vs. predictions (continued).

	$\ln \tau$		τ		DCN[1]	
	meas.	pred.	meas.	pred.	meas.	pred.
menthyl acetate	2.64	2.84	14.01	17.18	17.8	15.3
2-methyl-9-decenoate	1.71	1.82	5.51	6.18	38.3	34.7
methyl linolenate	1.75	1.53	5.73	4.60	37.0	45.0
methyl linoleate	1.55	1.28	4.73	3.60	43.9	56.2
vinyl laurate	1.01	1.43	2.75	4.19	72.4	49.0
1-methoxyhexane	0.84	1.13	2.32	3.08	84.7	65.0
dibutyl ether	0.77	1.03	2.16	2.81	90.9	70.8
diisopropyl ether	2.33	1.99	10.28	7.32	22.6	29.9
diisobutyl ether	1.22	1.40	3.38	4.03	59.7	50.7
diisoamyl ether	0.86	1.12	2.37	3.06	83.1	65.5
diethylene glycol dimethyl ether	0.50	0.73	1.65	2.07	117.6	94.6
1,2-dimethoxyethane	0.97	1.18	2.64	3.25	75.2	61.9
2,2-dimethoxypropane	1.93	1.62	6.89	5.04	31.5	41.5
cyclopentyl methyl ether	1.47	1.69	4.35	5.40	47.4	39.0
acetaldehyde diethyl acetal	1.32	1.12	3.74	3.08	54.3	65.1
dodecyl vinyl ether	0.83	0.74	2.29	2.09	85.8	93.8
1,4-cyclohexanedimethanol divinyl ether	1.19	0.96	3.29	2.60	61.2	76.1
2-methylfuran	4.10	4.03	60.34	56.46	7.6	7.8
furfural	3.23	3.30	25.28	27.10	11.8	11.3
2-ethylfuran	3.89	3.26	48.67	26.09	8.3	11.6
2,5-dimethylfuran	3.74	3.99	42.02	53.92	8.9	7.9
2-methyltetrahydrofuran	2.49	2.25	12.06	9.53	19.9	24.0
tetrahydrofuran	2.43	2.23	11.36	9.30	20.9	24.5
tetrahydrofurfuryl acetate	2.75	2.56	15.64	12.99	16.4	18.8
2-butyltetrahydrofuran	1.51	1.41	4.55	4.08	45.5	50.2
eucalyptol	2.69	1.96	14.73	7.09	17.1	30.8
rose oxide	1.99	2.25	7.30	9.49	30.0	24.1
3,4-dihydro-2H-pyran	2.31	2.44	10.07	11.45	23.0	20.8
tetrahydropyran	1.71	2.12	5.53	8.30	38.2	26.9
2,5-dihydrofuran	3.03	2.57	20.63	13.05	13.5	18.8
γ-undecanolactone	1.35	1.36	3.86	3.91	52.8	52.2
whiskey lactone	2.13	1.85	8.41	6.36	26.6	33.8
δ-undecalactone	1.44	1.41	4.22	4.10	48.7	49.9
n-butylbenzene	3.38	2.83	29.37	17.03	10.8	15.4
dibenzylether	4.43	3.83	83.93	46.27	6.7	8.5
benzaldehyde dimethyl acetal	3.73	2.55	41.68	12.80	8.9	19.0
α-methyl-trans-cinnamaldehyde	2.66	3.83	14.30	46.23	17.5	8.5
1,2,4-trimethylbenzene	4.21	4.45	67.14	85.87	7.2	6.6
1,2,3,4-tetrahydronaphthalene	4.21	3.62	67.14	37.31	7.2	9.5

Continued on next page

Table C.2: Measurements vs. predictions (continued).

	$\ln \tau$		τ		DCN[1]	
	meas.	pred.	meas.	pred.	meas.	pred.
4-methoxybenzaldehyde	2.20	2.79	9.03	16.28	25.1	15.9
1-butoxy-2-propanol	1.77	1.88	5.90	6.55	36.1	32.9
dipropylene glycol monomethyl ether	1.55	1.69	4.73	5.41	43.9	39.0
tripropylene glycol monomethyl ether	0.97	1.04	2.64	2.83	75.2	70.3
triethylene glycol monomethyl ether	0.97	0.93	2.65	2.54	74.9	77.8
diethylene glycol monomethyl ether	1.71	1.52	5.53	4.59	38.2	45.2
2-methoxyethanol	3.04	2.55	20.91	12.81	13.4	19.0
3-methoxy-3-methyl-1-butanol	3.81	2.40	45.15	10.99	8.6	21.4
1-methoxy-2-propanol	2.94	2.79	18.92	16.27	14.3	15.9
furfuryl alcohol	4.00	5.23	54.60	187.06	7.9	5.5
ethyl acetoacetate	3.27	3.40	26.31	29.98	11.6	10.7
propylene glycol monomethyl ether acetate	2.31	2.39	10.07	10.91	23.0	21.6
ethylene glycol vinyl ether	3.22	2.68	25.03	14.52	11.9	17.3
4-hydroxy-4-methyl-2-pentanone	3.54	4.30	34.47	73.94	9.9	7.0
butyl levulinate	3.18	2.44	23.93	11.44	12.3	20.8
n-octane	1.24	1.35	3.47	3.84	58.2	53.1
2,3-dimethylpentane	2.44	2.09	11.52	8.08	20.7	27.6
1-octene	1.66	1.55	5.25	4.70	40.0	44.2
methylcyclohexane	2.28	2.08	9.82	7.97	23.5	27.9
bisabolene	1.89	1.83	6.64	6.23	32.6	34.4
2-butanol	4.32	3.66	75.36	38.87	6.9	9.3
hexanal	1.02	1.32	2.77	3.74	71.7	54.4
4-methyl-2-pentanone	3.41	2.68	30.27	14.58	10.6	17.3
hexyl acetate	1.89	2.06	6.62	7.87	32.6	28.2
diethoxymethane	1.26	1.05	3.53	2.86	57.4	69.8
2-butylfuran	3.36	2.51	28.73	12.25	11.0	19.7
2-ethyltetrahydrofuran	2.08	1.83	7.99	6.20	27.8	34.5
2,3-dihydrofuran	2.60	2.57	13.42	13.05	18.4	18.8
ϵ-decalactone	1.64	1.64	5.16	5.14	40.7	40.7
1,2-dimethoxybenzene	3.95	2.64	51.94	14.07	8.1	17.7
2-isopropoxyethanol	2.09	2.39	8.08	10.92	27.5	21.6
tetrahydrofurfuryl alcohol	2.78	2.96	16.17	19.21	16.0	14.2

[1] DCN computed as $\mathrm{DCN} = 4.46 + 186.6/\tau$

Note: The measured $\ln \tau$ is the mean of the 32 individual *logarithmic* delays measured for a compound. This mean value is used to compute τ and the DCN. In contrast, τ given in Table B.1 represents the mean of the 32 *non-logarithmic* delays.

C.3 Sensitivity equations and covariance matrix

In order to compute the confidence interval $[y_i - \Delta y_i,\, y_i + \Delta y_i]$ for the predicted logarithmic ignition delay $\ln(\tau_i) = y_i$ from

$$\Delta y_i = t_{144-14}^{1-(\alpha/2)} \left(\mathbf{s}_i^{\mathrm{T}} \mathbf{V}_{\mathbf{F}} \left(\boldsymbol{\Theta}^*, \mathbf{u} \right) \mathbf{s}_i \right)^{1/2} \quad , \tag{C.1}$$

the covariance matrix $\mathbf{V}_{\mathbf{F}}(\boldsymbol{\Theta}^*, \mathbf{u})$ (cf. Table C.3) is needed as well as the sensitivity vector

$$\mathbf{s}_i^{\mathrm{T}}(\mathbf{u}_i, \boldsymbol{\Theta}^*) = \left[\frac{\partial y_i(\mathbf{u}_i, \boldsymbol{\Theta}^*)}{\partial \Theta_1^*}, .., \frac{\partial y_i(\mathbf{u}_i, \boldsymbol{\Theta}^*)}{\partial \Theta_{14}^*} \right] \tag{C.2}$$

for the model inputs $\mathbf{u}_i^{\mathrm{T}} = [\mathbf{g}_i, \mathbf{d}_i]^{\mathrm{T}}$. Given the model structure y_i (cf. Table 4.9) the entries of the vector $\mathbf{s}_i^{\mathrm{T}}$ can be written as:

$$s_{i,1}\left(\mathbf{g}_i, \mathbf{d}_i\right) = y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.3}$$

$$s_{i,2}\left(\mathbf{g}_i, \mathbf{d}_i\right) = d_{nAB,i} \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.4}$$

$$s_{i,3}\left(\mathbf{g}_i, \mathbf{d}_i\right) = d_{nCCDB,i} \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.5}$$

$$s_{i,4}\left(\mathbf{g}_i, \mathbf{d}_i\right) = \ln(d_{p^S,i}) \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.6}$$

$$s_{i,5}\left(\mathbf{g}_i, \mathbf{d}_i\right) = d_{nQC,i} \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.7}$$

$$s_{i,6}\left(\mathbf{g}_i, \mathbf{d}_i\right) = g_{-CH3,i} \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.8}$$

$$s_{i,7}\left(\mathbf{g}_i, \mathbf{d}_i\right) = \ln(g_{-CH2-(non-ring),i} + 1) \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.9}$$

$$s_{i,8}\left(\mathbf{g}_i, \mathbf{d}_i\right) = g_{-OH,i} \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.10}$$

$$s_{i,9}\left(\mathbf{g}_i, \mathbf{d}_i\right) = g_{-O-(non-ring),i} \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.11}$$

$$s_{i,10}\left(\mathbf{g}_i, \mathbf{d}_i\right) = g_{-O-(ring),i} \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.12}$$

$$s_{i,11}\left(\mathbf{g}_i, \mathbf{d}_i\right) = g_{>C=O(non-ring),i} \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.13}$$

$$s_{i,12}\left(\mathbf{g}_i, \mathbf{d}_i\right) = g_{>C=O(ring),i} \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.14}$$

$$s_{i,13}\left(\mathbf{g}_i, \mathbf{d}_i\right) = g_{O=CH-,i} \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.15}$$

$$s_{i,14}\left(\mathbf{g}_i, \mathbf{d}_i\right) = g_{-COO-(non-ring),i} \cdot y_i\left(\mathbf{g}_i, \mathbf{d}_i\right) \tag{C.16}$$

A typical value for α is 0.05 corresponding to the 95% confidence interval. Then, Student's t-distribution yields $t_{144-14}^{1-(0.05/2)}$=1.9784.

Table C.3: Estimate of the covariance matrix $\mathbf{V_F}(\mathbf{\Theta}^*, \mathbf{u})$ at the estimated parameter values $\mathbf{\Theta}^*$.

	P	D_{nAB}	D_{nCCDB}	D_{pS}	D_{nQC}	G_{-CH3}	$G_{-CH2-(non-ring)}$	G_{-OH}
P	1.1348e-03	1.0671e-04	-3.1321e-04	1.1425e-05	1.4281e-04	-2.2082e-04	-2.1150e-04	-9.6642e-05
D_{nAB}	1.0671e-04	3.2391e-04	-1.4696e-04	-5.6006e-06	3.9843e-06	-1.7138e-05	-2.0602e-05	-2.8408e-05
D_{nCCDB}	-3.1321e-04	-1.4696e-04	4.4433e-04	6.2121e-05	-9.2702e-05	1.1207e-04	1.0553e-04	2.7187e-04
D_{pS}	1.1425e-05	-5.6006e-06	6.2121e-05	2.9219e-05	-2.3509e-05	2.5236e-05	3.1056e-05	1.1517e-04
D_{nQC}	1.4281e-04	3.9843e-06	-9.2702e-05	-2.3509e-05	2.1026e-03	-2.0777e-04	7.9546e-06	5.0204e-05
G_{-CH3}	-2.2082e-04	-1.7138e-05	1.1207e-04	2.5236e-05	-2.0777e-04	1.2019e-04	3.6162e-05	9.8557e-05
$G_{-CH2-(non-ring)}$	-2.1150e-04	-2.0602e-05	1.0553e-04	3.1056e-05	7.9546e-06	3.6162e-05	2.4143e-04	3.0830e-06
G_{-OH}	-9.6642e-05	-2.8408e-05	2.7187e-04	1.1517e-04	5.0204e-05	9.8557e-05	3.0830e-06	1.4874e-03
$G_{-O-(non-ring)}$	-1.6558e-04	-5.4478e-05	7.5185e-05	-1.8874e-06	9.1794e-05	-1.8617e-06	1.9430e-05	-1.4268e-04
$G_{-O-(ring)}$	-4.9419e-04	-8.9543e-05	2.1149e-04	3.7234e-05	-9.2820e-05	1.3812e-04	1.2003e-04	2.1197e-04
$G_{>C=O(non-ring)}$	-2.4162e-04	-6.5467e-05	2.4109e-04	4.6296e-05	8.8568e-05	6.3265e-05	-8.6291e-05	4.7506e-04
$G_{>C=O(ring)}$	-3.8764e-04	-8.6365e-05	5.1483e-04	1.6130e-04	7.1142e-05	1.5024e-04	2.6265e-04	8.1243e-04
$G_{O=CH-}$	-6.4898e-04	-1.1896e-04	1.9848e-04	4.6373e-06	4.6871e-05	1.0294e-04	7.6135e-05	2.3391e-04
$G_{-COO-(non-ring)}$	-1.0746e-04	-3.6827e-05	2.9933e-04	1.1337e-04	1.4231e-04	6.5443e-05	4.2145e-05	7.3286e-04

Continued on next page

Table C.3: Estimate of the covariance matrix $\mathbf{V_F}(\mathbf{\Theta}^*, \mathbf{u})$ at the estimated parameter values $\mathbf{\Theta}^*$ (continued).

	$G_{-O-(non-ring)}$	$G_{-O-(ring)}$	$G_{>C=O(non-ring)}$	$G_{>C=O(ring)}$	$G_{O=CH-}$	$G_{-COO-(non-ring)}$
P	-1.6558e-04	-4.9419e-04	-2.4162e-04	-3.8764e-04	-6.4898e-04	-1.0746e-04
D_{nAB}	-5.4478e-05	-8.9543e-05	-6.5467e-05	-8.6365e-05	-1.1896e-04	-3.6827e-05
D_{nCCDB}	7.5185e-05	2.1149e-04	2.4109e-04	5.1483e-04	1.9848e-04	2.9933e-04
D_{p^S}	-1.8874e-06	3.7234e-05	4.6296e-05	1.6130e-04	4.6373e-06	1.1337e-04
D_{nQC}	9.1794e-05	-9.2820e-05	8.8568e-05	7.1142e-05	4.6871e-05	1.4231e-04
G_{-CH3}	-1.8617e-06	1.3812e-04	6.3265e-05	1.5024e-04	1.0294e-04	6.5443e-05
$G_{-CH2-(non-ring)}$	1.9430e-05	1.2003e-04	-8.6291e-05	2.6265e-04	7.6135e-05	4.2145e-05
G_{-OH}	-1.4268e-04	2.1197e-04	4.7506e-04	8.1243e-04	2.3391e-04	7.3286e-04
$G_{-O-(non-ring)}$	4.2468e-04	9.7916e-05	1.1748e-04	1.1624e-04	1.3232e-04	4.1829e-05
$G_{-O-(ring)}$	9.7916e-05	1.7442e-03	2.3741e-04	-1.1500e-05	3.1257e-04	3.6364e-05
$G_{>C=O(non-ring)}$	1.1748e-04	2.3741e-04	2.7727e-03	4.7904e-04	2.9766e-04	3.6908e-04
$G_{>C=O(ring)}$	1.1624e-04	-1.1500e-05	4.7904e-04	3.0598e-03	3.6994e-04	9.0404e-04
$G_{O=CH-}$	1.3232e-04	3.1257e-04	2.9766e-04	3.6994e-04	2.7744e-03	2.4415e-04
$G_{-COO-(non-ring)}$	4.1829e-05	3.6364e-05	3.6908e-04	9.0404e-04	2.4415e-04	1.3381e-03

D Derivation of QSPR models

Table D.1 gives an overview about the nature of the outliers that have been excluded in the course of QSPR modeling in Chapter 5. Figure D.1 shows the bisection approach that is used to control the similarity threshold $\overline{\delta}$ in the modified SCA algorithm (cf. Figure 5.6) introduced in Chapter 5.

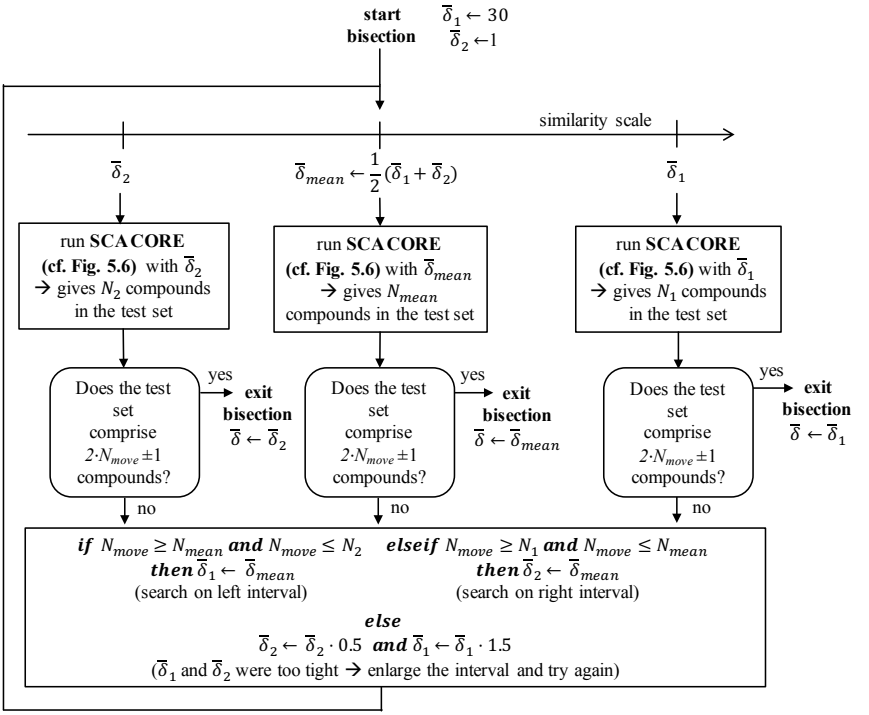


Figure D.1: Bisection approach used to control the similarity threshold $\overline{\delta}$ in the modified SCA algorithm (cf. Figure 5.6).

Table D.1: Compounds excluded from the modeling data sets due to unusually high/low property values and/or Hotelling's T^2 statistic.

data set	**property outliers**	**excluded due to T^2 statistic**
liquid viscosity μ	DL-α-tocopherol; 2-nonylphenol; glycerol; 3-ethylphenol	1 large alkane; 9 large (aromatic) esters
enthalpy of vaporization H_{vap}	bis(2-ethylhexyl) phthalate; DL-α-tocopherol; 1,2,3-butanetriol	4 large aromatic esters
surface tension σ	–	1 large phenol; 1 large ether, 1 large alkene, 11 large (aromatic) esters, 1 large aromatic species
melting point T_{melt}	1,4-benzenedicarboxylic acid	11 large alkanes; 4 large alkenes; 1 large phenol, 21 large (aromatic) esters; 2 large polycyclic alkanes
liquid molar density ρ_m	–	16 large (aromatic) esters, 2 large aromatic species; 6 large alkenes; 1 large alkane; 2 large ethers; 4 large acids
boiling point T_{boil}	–	13 large alkanes; 2 large polycyclic alkanes; 1 large phenol; 19 large (aromatic) esters
lower heating value LHV	1-tetracontene; n-hexatriacontane; glycerol dioleate; glycerol trioleate	2 large polycyclic alkanes; 11 large alkanes; 4 large alkenes; 1 large phenol; 17 large (aromatic) esters
critical temperature T_{crit}	–	1 large phenol; 2 large polycyclic alkanes; 6 large alkenes; 13 large alkanes; 1 large ether; 17 large (aromatic) esters
critical pressure p_{crit}	1,3-propylene glycol; 1,2,3-benzenetriol; 2-oxacyclobutanone; glycerol; 1,3-benzenediol; succinic anhydride; maleic anhydride	6 large alkenes; 3 large alkanes; 19 large (aromatic) esters
critical volume V_{crit}	–	5 large alkenes; 2 large esters

E Rational formulation of biofuel mixtures

The contents of Appendix E have already been published in *Energy & Fuels* (Dahmen and Marquardt (2017); Supporting Information).

E.1 Application of Hoffmann-Florin equation

The Hoffmann-Florin equation (Hoffmann and Florin, 1943; Gmehling et al., 2012) can be used to compute the temperature-dependent vapor pressure p_i^S for a pure compound i at temperature T from QSPR estimates for critical temperature $T_{crit,i}$ [K], critical pressure $p_{crit,i}$ [bar] and the normal boiling point $T_{boil,i}$ [K] according to the following relations:

$$f_{T_{boil,i}} = \frac{1}{T_{boil,i}} - 7.9151 \cdot 10^{-3} + 2.6726 \cdot 10^{-3} \cdot \log_{10}(T_{boil,i}) - 0.8625 \cdot 10^{-6} \cdot T_{boil,i} \quad \text{(E.1)}$$

$$f_{T_{crit,i}} = \frac{1}{T_{crit,i}} - 7.9151 \cdot 10^{-3} + 2.6726 \cdot 10^{-3} \cdot \log_{10}(T_{crit,i}) - 0.8625 \cdot 10^{-6} \cdot T_{crit,i} \quad \text{(E.2)}$$

$$f_{T_i} = \frac{1}{T} - 7.9151 \cdot 10^{-3} + 2.6726 \cdot 10^{-3} \cdot \log_{10}(T) - 0.8625 \cdot 10^{-6} \cdot T \quad \text{(E.3)}$$

$$\chi_{1,i} = 11.5261 - \ln\left(\frac{1.01325}{p_{crit,i}}\right) \cdot \frac{f_{T_{boil,i}}}{f_{T_{boil,i}} - f_{T_{crit,i}}} \quad \text{(E.4)}$$

$$\chi_{2,i} = \ln\left(\frac{1.01325}{p_{crit,i}}\right) / \left(f_{T_{boil,i}} - f_{T_{crit,i}}\right) \quad \text{(E.5)}$$

$$p_i^S(T) = \exp\left(\chi_{1,i} + \chi_{2,i} \cdot f_{T_i}\right) \cdot 1000 \qquad \text{[kPa]} \quad \text{(E.6)}$$

E.2 Application of COSTALD method

The COSTALD method (Hankinson and Thomson, 1979; Gmehling et al., 2012) can be used to compute the temperature-dependent liquid molar density $\rho_{m,i}$ of a pure compound i at temperature T from QSPR estimates for critical temperature $T_{crit,i}$ [K], critical volume $V_{crit,i}$ [cm^3/mol] and the acentric factor ω_i [-] according to the following relations:

$$\chi = \frac{T}{T_{crit,i}} \tag{E.7}$$

$$V_R^{(0)} = 1 + C_{cost,a}\,(1-\chi)^{1/3} + C_{cost,b}\,(1-\chi)^{2/3} + C_{cost,c}\,(1-\chi) + C_{cost,d}\,(1-\chi)^{4/3} \tag{E.8}$$

with $C_{cost,a}$=-1.52816, $C_{cost,b}$=1.43907

$C_{cost,c}$=-0.81446 and $C_{cost,d}$=0.190454

$$V_R^{(\delta)} = \left(C_{cost,e} + C_{cost,f}\cdot\chi + C_{cost,g}\cdot(\chi)^2 + C_{cost,h}\cdot(\chi)^3\right) / (\chi - 1.00001) \tag{E.9}$$

with $C_{cost,e}$=-0.296123, $C_{cost,f}$=0.386914

$C_{cost,g}$=-0.0427458 and $C_{cost,h}$=-0.0480645

$$\rho_{m,i}(T) = \left(V_{c,i}\cdot V_R^{(0)}\cdot(1-\omega_i V_R^{(\delta)})\right)^{-1}\cdot 1000 \qquad [\mathrm{kmol/m^3}] \tag{E.10}$$

The acentric factor ω_i in Eqn. (E.10) is computed from Eqn. (5.25).

E.3 Trade-off analysis: Hydrogen demand vs. energy of fuel produced

Figure E.1 shows the trade-off between the hydrogen demand [$\mathrm{mol}_{H_2}/\mathrm{mol}_{fuel}$] and the energy of fuel produced [$\mathrm{MJ}_{fuel}/\mathrm{MJ}_{biomass}$] for the case study from Chapter 7. The CO_2 emissions increase as the hydrogen demand is lowered as a consequence of the fact that the necessary deoxygenation can be achieved either via removal of CO_2 or via removal of water. The latter option, however, often requires hydrogenation. Note that blend A from Chapter 7 corresponds to the design with the maximum utilization of external hydrogen.

Blend E in Figure E.1 refers to a design that does not rely on any supply of external hydrogen. Instead, the hydrogen that is needed to upgrade the furfural is produced internally via concurrent butyric acid fermentation, as can be seen from Figure E.2. Consequently, the main component of this four-components blend is butyl acetate (53 mol-%). Table E.1 shows the properties of blend E. As can be seen from Figure E.3, the distillation curve touches the upper bounds for both T10(m) and T50(m) (cf. Table 7.2) and the heaviest blend component, i.e., ethyl levulinate (12%), is responsible for the blend's high final boiling point (205 °C). Note that there are residual pentoses and residual furfural in case of blend E (cf. Figure E.2). Obviously, these cannot be converted into fuel on the basis of the property constraints stated in Table 7.2. This deteriorates the LHV energy efficiency compared to the design with external hydrogen supply (cf. Figure E.1).

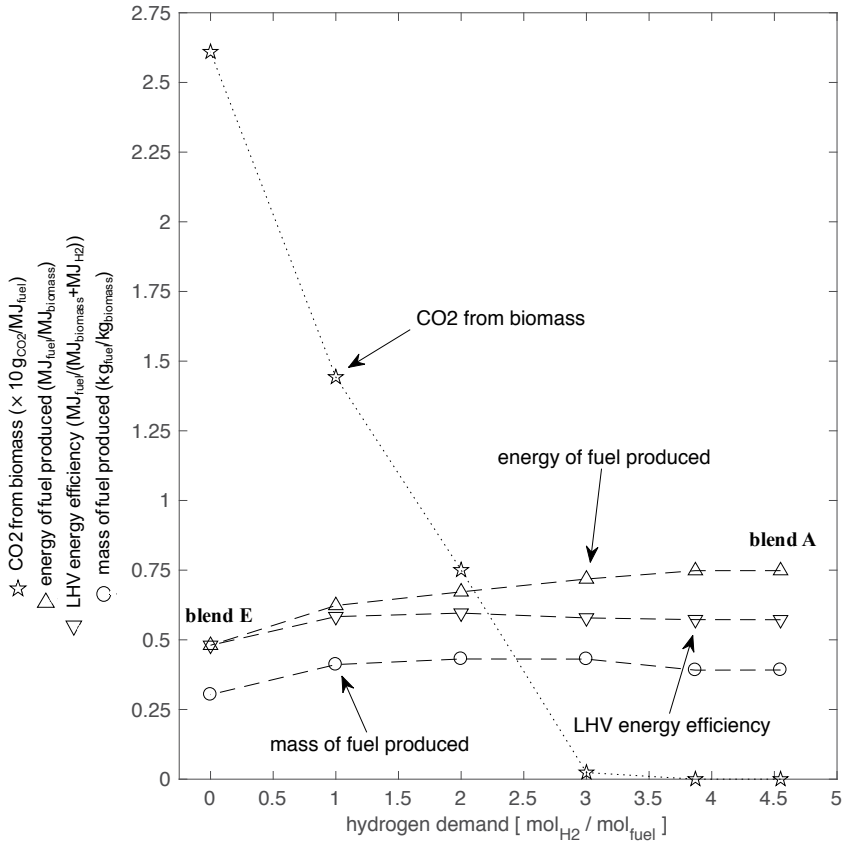
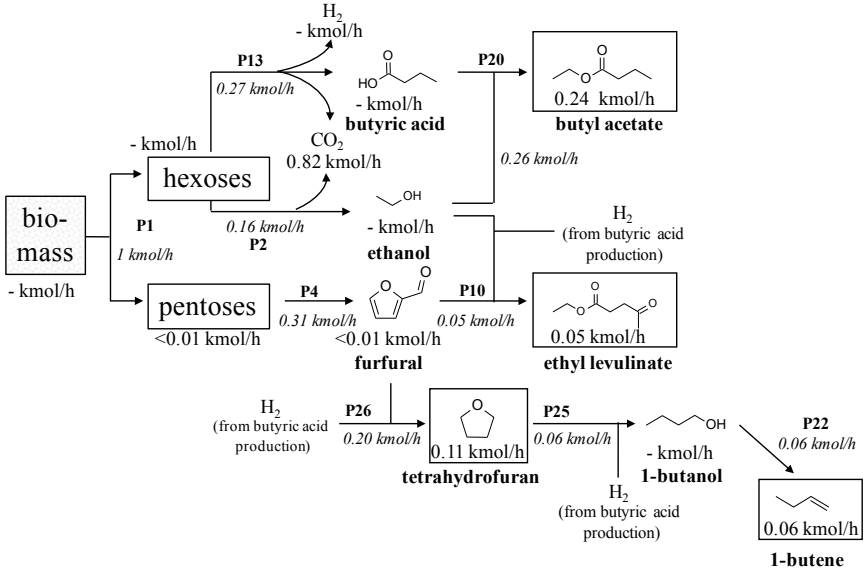


Figure E.1: Trade-off between hydrogen demand [mol$_{H_2}$/mol$_{fuel}$] and energy of fuel produced [MJ$_{fuel}$/MJ$_{biomass}$]. The blend design problem (maximization of the energy of fuel produced) has been solved multiple times, whereby the upper bound for the hydrogen demand has been lowered stepwise (5, 4, 3, 2, 1, 0). While the points in the diagram represent properties of optimal designs, the dashed and dotted lines are for illustration purposes only.



energy of fuel produced $MJ_{fuel}/MJ_{biomass}$	mass of fuel produced $kg_{fuel}/kg_{biomass}$	hydrogen demand mol_{H2}/mol_{fuel}	LHV energy efficiency $MJ_{fuel}/(MJ_{biomass}+MJ_{H2})$	CO_2 from biomass g/MJ_{fuel}	derived cetane number (DCN)
0.48	0.30	0.0	0.48	26.09	19.8

Figure E.2: Material flow diagram for the design that does not require external supply of hydrogen (cf. Figure E.1, $\mathbf{blend\ E}$). Note that butyric acid fermentation is used to provide the hydrogen needed to upgrade the furfural. The resulting four-components blend is composed of 53 mol-% butyl acetate, 23 mol-% tetrahydrofuran, 13 mol-% 1-butene and 12 mol-% ethyl levulinate.

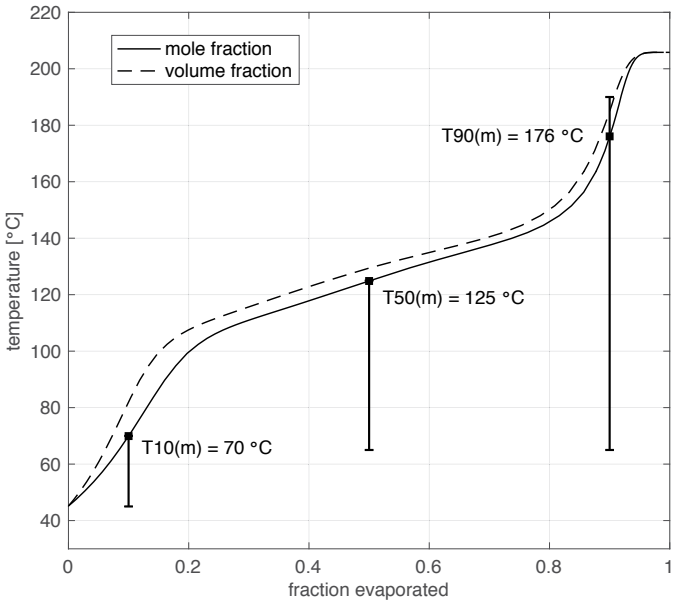


Figure E.3: Predicted distillation curve for $\mathbf{blend\ E}$ (cf. Figure E.1). The curve has been computed with Matlab's (The MathWorks Inc., 2016) error-controlled, variable step-length solver *ode15s*. The bars show the constraints for T10(m), T50(m) and T90(m). Note that V_{shift} has been set to zero in an attempt to simulate an idealized batch distillation process.

Table E.1: Predicted properties of **blend E** (cf. Figure E.1).

blend composition [mol-%]	butyl acetate (53%), tetrahydrofuran (23%), 1-butene (13%), ethyl levulinate (12%)
energy of fuel produced [MJ_{fuel}/ $MJ_{biomass}$]	0.48
hydrogen demand [mol H_2 per mol fuel]	0.0
LHV energy efficiency η_{LHV} [MJ_{fuel}/ ($MJ_{biomass}$ +MJ_{H_2})]	0.48
derived cetane number DCN (cRON /cMON)	19.8 (90/82)
heating value LHV [MJ /kg]	29.5
liquid density ρ_L [kg /m^3]	866
heat of vaporization H_{vap} [kJ /kg_{air}] (λ=1)	41
surface tension σ [mN /m]	25
viscosity ν [mm^2 /s]	0.8
Reid vapor pressure p^S_{Reid}/ p_{bp} [kPa]	77 / 82
distillation profile T10 (v)[a] [°C]	82
distillation profile T50 (v)[a] [°C]	129
distillation profile T90 (v)[a] [°C]	185

[a] computed with Matlab's (The MathWorks Inc., 2016) error-controlled, variable step-length solver *ode15s*

E.4 Data and references for the case study

The conversion pathways used in the case study in Chapter 7 are shown in Table E.2. The resulting stoichiometric matrix can be found in Table E.3. The property data estimates are given in Table E.4.

Table E.2: Description of conversion pathways included in the SI engine case study.

pathway	description	reference(s)
P1	fixed biomass composition: 0.5 kg cellulose, 0.3 kg hemicellulose and 0.2 kg lignin per kg biomass ⇒ 0.4293 mole hexoses and 0.3159 mole pentoses per mole biomass monomers (assuming 90% selectivity for both pentoses and hexoses)	assumption based on Luterbacher et al. (2014)
P2	ethanol fermentation yield of 0.50 g/g on glucose ⇒ 1.96 mol ethanol and 1.96 mol CO_2 per mol of glucose	Straathof (2014)
P3	acetic acid fermentation yield of 0.8 g/g on glucose ⇒ 2.40 mol acetic acid per mol of glucose	Straathof (2014)
P4	furfural production from xylose (81% selectivity)	Gürbüz et al. (2013)
P5	esterification of acetic acid with ethanol (93.2% selectivity)	Bamoharram et al. (2006)
P6	tetrahydrofurfuryl production from furfural (100% selectivity)	Bhogeswararao and Srinivas (2015)
P7	2-methyltetrahydrofuran production from furfural via levulinic acid (92% selectivity)	Geilen et al. (2010); Gupta et al. (2015)
P8	γ-valerolactone production from furfural via levulinic acid (100% selectivity)	Wright and Palkovits (2012); Gupta et al. (2015)
P9	2-methyltetrahydrofuran production from γ-valerolactone (97% selectivity)	Geilen et al. (2011)
P10	ethyl levulinate production from ethanol and furfural (100% selectivity)	Cirujano et al. (2015)
P11	acetic acid fermentation on ethanol at 98% molar yield	Straathof (2014)
P12	acetone-butanol-ethanol (ABE) fermentation yield 0.38 g/g on glucose ⇒ 0.16 mole acetone, 0.50 mol butanol and 0.48 mol ethanol per mol of glucose	Jang et al. (2013)
P13	butyric acid fermentation yield 0.46 g/g on glucose ⇒ 0.94 mol butyric acid, 1.88 mol CO_2 and 1.88 mol H_2 per mol of glucose	Jiang et al. (2011); Straathof (2014)

Continued on next page

Table E.2: Description of conversion pathways included in the SI engine case study (continued).

pathway	description	reference(s)
P14	2,3-butanediol fermentation yield 0.48 g/g on glucose ⇒ 0.96 mol 2,3-butanediol and 1.92 mol CO_2 per mol of glucose	Ma et al. (2009); Straathof (2014)
P15	lactic acid fermentation close to ideal yield of 1 g/g on glucose ⇒ 2 mol lactic acid per mol of glucose	Straathof (2014)
P16	succinic acid fermentation yield 1.1 g/g on glucose ⇒ 1.68 mol succinic acid per mol of glucose; consumes 0.86 mol of CO_2 per mole of glucose	Vemuri et al. (2002); Straathof (2014)
P17	1,3-propanediol fermentation yield 0.51 g/g on glucose ⇒ 1.21 mol 1,3-propanediol and 1.21 mol CO_2 per mol of glucose	Nakamura and Whited (2003); Straathof (2014)
P18	hydroxymethylfurfural production from glucose (85.3% selectivity)	Fan et al. (2011)
P19	dehydration of 2,3-butanediol to 2-butanone (90% selectivity)	Multer et al. (2013); Goulas and Toste (2016)
P20	esterification of butyric acid with ethanol (95.1% selectivity)	Bamoharram et al. (2006)
P21	hydrogenation of hydroxymethylfurfural yielding 2,5-dimethyltetrahydrofuran (97.4% selectivity)	Kong et al. (2015)
P22	dehydration of 1-butanol to 1-butene (94% selectivity)	Wright (2015)
P23	hydrogenation of succinic acid to γ-butyrolactone (94% selectivity)	You et al. (2015)
P24	hydrogenation of γ-butyrolactone to tetrahydrofuran (79.2% selectivity), 1-butanol (10.5% selectivity) and 1-propanol (4.8% selectivity)	Shao et al. (2014)
P25	hydrogenolysis of tetrahydrofuran to 1-butanol (99% selectivity)	Koso et al. (2011)
P26	decarbonylation of furfural to furan (90% selectivity) and subsequent hydrogenation of furan to tetrahydrofuran (95% selectivity)	Wojcik (1948); Corma et al. (2007); Ulonska et al. (2016b)

Continued on next page

Table E.2: Description of conversion pathways included in the SI engine case study (continued).

pathway	description	reference(s)
P27	hydrogenation of HMF to 2,5-THF-dimethanol (99% selectivity), subsequent ring opening yielding 1,2,6-hexanetriol (97% selectivity) and final conversion to tetrahydropyran-2-methanol (99% selectivity)	Buntara et al. (2011); van Putten et al. (2013)
P28	hydrodeoxygenation of 2,3-butanediol to n-propane (9.3% selectivity) and n-butane (87% selectivity) and CO_2 (3.5% selectivity)	Li and Huber (2010)
P29	conversion of furfural to cyclopentanol (36.23% selectivity), cyclopentanone (40.23% selectivity) and 2-methyltetrahydrofuran (9.44% selectivity)	Hronec and Fulajtarová (2012)
P30	dehydrogenative oxidation of cyclopentanol to cyclopentanone (100% selectivity)	Kawahara et al. (2012)
P31	hydrogenation of cyclopentanone to cyclopentane (98% selectivity)	Török et al. (2000)
P32	conversion of 1,3-propanediol to 1-propanol (99.5% selectivity)	Zhu et al. (2012)
P33	conversion of lactic acid and ethanol to ethyl lactate (94.9% selectivity)	Zhang et al. (2015b)
P34	isobutanol fermentation yield 0.35 g/g on glucose $\Rightarrow$ 0.85 mol isobutanol and 1.7 mol CO_2 per mol of glucose	Atsumi et al. (2008); Straathof (2014)
P35	deoxygenation of isobutanol to isobutane ($\sim$95% selectivity)	Glebov et al. (1985)
P36	deoxygenation of 1-butanol to n-butane ($\sim$95% selectivity)	Glebov et al. (1985)
P37	deoxygenation of cyclopentanol to cyclopentane ($\sim$95% selectivity)	Glebov et al. (1985)
P38	hydrogenation of furfural to cyclopentanone (99% selectivity)	Zhang et al. (2016)

Table E.3: Stoichiometric matrix **A** for the SI engine case study.

	S1	S2	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18
biomass	1		-1																	
hexoses			0.4293	-1	-1									-1	-1	-1	-1	-1	-1	-1
pentoses			0.3159			-1														
hydrogen		1						-3	-4	-2	-2	-1			1.88					
carbon dioxide				1.96											1.88	1.92		-0.86	1.21	
ethanol				1.96			-1					-1	-1	0.48						
acetic acid					2.4		-1						0.98							
furfural						0.81		-1	-1	-1		-1								
ethyl acetate							0.932													
tetrahydrofurfuryl alcohol								1												
γ-valerolactone										1	-1									
2-methyltetrahydrofuran									0.92		0.97									
ethyl levulinate												1								
acetone														0.16						
1-butanol														0.5						
butyric acid															0.94					
2,3-butanediol																0.96				
5-hydroxymethylfurfural																				0.853
lactic acid																	2			
succinic acid																		1.68		
1,3-propanediol																			1.21	
2-butanone																				
butyl acetate																				
2,5-dimethyltetrahydrofuran																				
1-butene																				
γ-butyrolactone																				
tetrahydrofuran																				
1-propanol																				
tetrahydropyran-2-methanol																				
cyclopentanol																				
cyclopentanone																				
cyclopentane																				
ethyl lactate																				
propane																				
butane																				
isobutanol																				
isobutane																				

Continued on next page

Table E.3: Stoichiometric matrix $\mathbf{A}$ for the SI engine case study (continued).

	P19	P20	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30	P31	P32	P33	P34	P35	P36	P37	P38
biomass																				
hexoses																-1				
pentoses																				
hydrogen			-5		-2	-2	-1	-2	-4	-2	-4	1	-2	-1			-1	-1	-1	-3
carbon dioxide										0.035						1.70				
ethanol		-1													-1					
acetic acid																				
furfural								-1			-1									-1
ethyl acetate																				
tetrahydrofurfuryl alcohol																				
γ-valerolactone																				
2-methyltetrahydrofuran											0.0944									
ethyl levulinate																				
acetone																				
1-butanol				-1		0.105	0.99											-1		
butyric acid		-1																		
2,3-butanediol	-1									-1										
5-hydroxymethylfurfural			-1						-1											
lactic acid															-1					
succinic acid					-1															
1,3-propanediol														-1						
2-butanone	0.9																			
butyl acetate		0.951																		
2,5-dimethyltetrahydrofuran			0.974																	
1-butene				0.94																
γ-butyrolactone					0.94	-1														
tetrahydrofuran						0.792	-1	0.86												
1-propanol						0.048								0.995						
tetrahydropyran-2-methanol									0.95											
cyclopentanol											0.3623	-1							-1	
cyclopentanone											0.4023	1	-1							0.99
cyclopentane													0.98						0.95	
ethyl lactate															0.949					
propane										0.093										
butane										0.87								0.95		
isobutanol																0.85	-1			
isobutane																	0.95			

Table E.4: Property data estimates for the SI engine case study. Except for the derived cetane number (DCN), all estimates have been made with the QSPR models from Chapter 5. The DCN has been predicted with the group contribution model from Chapter 4.

	Only data that could not be retrieved from the DIPPR database (AIChE, 2012) or from the DCN database (cf. Appendix B) have been estimated.					Data have been estimated exclusively for those compounds where the Hoffmann-Florin and COSTALD equations need to be applied.			
palette compound	lower heating value [MJ/kg] at 298 K	derived cetane number	liquid density [kg/m^3] at 298 K	surface tension [N/m] at 298 K	liquid viscosity [Pa·s] at 298 K	normal boiling point [K]	critical temperature [K]	critical pressure [bar]	critical volume [cm^3/mol]
ethanol									
ethyl acetate		17.4							
tetrahydrofurfuryl alcohol					0.0062				
γ-valerolactone	24.7								
2-methyltetrahydrofuran	33.9		872	0.024	0.0005	358	559	42.0	291
ethyl levulinate									
acetone		12.4							
1-butanol									
2-butanone				0.023	0.0004				
butyl acetate		21.0							
2,5-dimethyltetrahydrofuran	35.2	24.5	845	0.024	0.0006	375	578	36.8	353
1-butene		22.7			0.0002				
γ-butyrolactone		16.3							
tetrahydrofuran									
1-propanol		13.2							
tetrahydropyran-2-methanol	28.9	18.0	1055	0.035	0.0103	466	671	41.0	335
cyclopentanol	33.3		927	0.030	0.0058	423	624	47.4	304
cyclopentanone									
cyclopentane		26.6		0.021					
ethyl lactate		6.7							
propane		25.0			0.0001				
butane		30.6							
isobutanol					0.0027				
isobutane		19.1		0.008					

Bibliography

Abbasi, T. and Abbasi, S. (2011). Renewable hydrogen: Prospects and challenges. *Renewable and Sustainable Energy Reviews*, 15(6):3034–3040.

Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS regression). *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):97–106.

Abdul Jameel, A. G., Naser, N., Emwas, A.-H., Dooley, S., and Sarathy, S. M. (2016). Predicting fuel ignition quality using 1H NMR spectroscopy and multiple linear regression. *Energy & Fuels*, 30(11):9819–9835.

Achenie, L., Venkatasubramanian, V., and Gani, R. (2003). *Computer Aided Molecular Design: Theory and Practice*. Elsevier, Amsterdam, The Netherlands.

Agrawal, R., Singh, N. R., Ribeiro, F. H., and Delgass, W. N. (2007). Sustainable fuel for the transportation sector. *Proceedings of the National Academy of Sciences*, 104(12):4828–4833.

Ahmed, A., Goteng, G., Shankar, V. S., Al-Qurashi, K., Roberts, W. L., and Sarathy, S. M. (2015). A computational methodology for formulating gasoline surrogate fuels with accurate physical and chemical kinetic properties. *Fuel*, 143:290–300.

AIChE (2012). DIPPR 801 Database version 6.0.3. Design Institute for Physical Properties (DIPPR), American Institute of Chemical Engineers (AIChE), USA. `http://www.aiche.org/dippr`.

Albahri, T. A. (2003). Structural group contribution method for predicting the octane number of pure hydrocarbon liquids. *Industrial & Engineering Chemistry Research*, 42(3):657–662.

Allard, L. N., Hole, N. J., Webster, G. D., Ryan, T. W., Ott, D., Beregszazy, A., Fairbridge, C. W., Cooley, J., Mitchell, K., Richardson, E. K., Elliot, N. G., and Rickeard, D. J. (1997). Diesel fuel ignition quality as determined in the Ignition Quality Tester (IQT) – Part II. Technical report, SAE Technical Paper 971636.

Allard, L. N., Webster, G. D., Hole, N. J., Ryan, T. W., Ott, D., and Fairbridge, C. W. (1996). Diesel fuel ignition quality as determined in the Ignition Quality Tester (IQT). Technical report, SAE Technical Paper 961182.

Alnajjar, M., Cannella, B., Dettman, H., Fairbridge, C., Franz, J., Gallant, T., Gieleciak, R., Hager, D., Lay, C., Lewis, S., Ratcliff, M., Sluder, S., Storey, J., Yin, H., and Zigler, B. (2010). CRC Report No. FACE-1 - chemical and physical properties of the fuels for advanced combustion engines (FACE) research diesel fuels. Technical report, Coordinating Research Council, Inc., Alpharetta, Georgia, United States.

Alonso, D. M., Bond, J. Q., and Dumesic, J. A. (2010). Catalytic conversion of biomass to biofuels. *Green Chemistry*, 12(9):1493–1513.

Alonso, D. M., Wettstein, S. G., and Dumesic, J. A. (2013). γ-Valerolactone, a sustainable platform molecule derived from lignocellulosic biomass. *Green Chemistry*, 15(3):584–595.

AlRamadan, A. S., Sarathy, S. M., Khurshid, M., and Badra, J. (2016). A blending rule for octane numbers of PRFs and TPRFs with ethanol. *Fuel*, 180:175–186.

Anastas, P. and Eghbali, N. (2010). Green chemistry: Principles and practice. *Chemical Society Reviews*, 39(1):301–312.

Arcoumanis, C., Bae, C., Crookes, R., and Kinoshita, E. (2008). The potential of di-methyl ether (DME) as an alternative fuel for compression-ignition engines: A review. *Fuel*, 87(7):1014–1030.

Ariffin Kashinath, S. A., Abdul Manan, Z., Hashim, H., and Wan Alwi, S. R. (2012). Design of green diesel from biofuels using computer aided technique. *Computers & Chemical Engineering*, 41:88–92.

Ascher, U. M. and Petzold, L. R. (1998). *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA.

Aspen Technology, I. (2011). Aspen Properties V7.3 (25.0.4987). Aspen Technology, Inc., Cambridge, Massachusetts, USA. `www.aspentech.com`.

ASTM D2699 (2013). ASTM D 2699 "Standard Test Method for Research Octane Number of Spark-Ignition Engine Fuel", American Society for Testing and Materials (ASTM), West Conshohocken, Pennsylvania, USA.

ASTM D2700 (2014). ASTM D 2700 "Standard Test Method for Motor Octane Number of Spark-Ignition Engine Fuel", American Society for Testing and Materials (ASTM), West Conshohocken, Pennsylvania, USA.

ASTM D323 (1999). ASTM D 323 "Standard Test Method for Vapor Pressure of Petroleum Products (Reid Method)", American Society for Testing and Materials (ASTM), West Conshohocken, Pennsylvania, USA.

ASTM D4814 (2016). ASTM D 4814 "Standard Specification for Automotive Spark-Ignition Engine Fuel", American Society for Testing and Materials (ASTM), West Conshohocken, Pennsylvania, USA.

ASTM D613 (2015). ASTM D 613 "Standard Test Method for Cetane Number of Diesel Fuel Oil", American Society for Testing and Materials (ASTM), West Conshohocken, Pennsylvania, USA.

ASTM D6890 (2011). ASTM D 6890 "Standard Test Method for Determination of Ignition Delay and Derived Cetane Number (DCN) of Diesel Fuel Oils by Combustion in a Constant Volume Chamber", American Society for Testing and Materials (ASTM), West Conshohocken, Pennsylvania, USA.

ASTM D7170 (2011). ASTM D 7170 "Standard Test Method for Determination of Derived Cetane Number (DCN) of Diesel Fuel Oils - Fixed Range Injection Period, Constant Volume Combustion Chamber Method", American Society for Testing and Materials (ASTM), West Conshohocken, Pennsylvania, USA.

ASTM D86 (1999). ASTM D 86 "Standard Test Method for Distillation of Petroleum Products and Liquid Fuels at Atmospheric Pressure", American Society for Testing and Materials (ASTM), West Conshohocken, Pennsylvania, USA.

ASTM D975 (2016). ASTM D 975 "Standard Specification for Diesel Fuel Oils", American Society for Testing and Materials (ASTM), West Conshohocken, Pennsylvania, USA.

Atsumi, S., Hanai, T., and Liao, J. C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*, 451(7174):86–89.

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.

Azadi, P., Inderwildi, O. R., Farnood, R., and King, D. A. (2013). Liquid fuels, hydrogen and chemicals from lignin: A critical review. *Renewable and Sustainable Energy Reviews*, 21:506–523.

Backhaus, J. and Rothamer, D. (2012). Design methodology for volatility characteristics of bio-derived drop-in gasoline fuels. In *Proceedings of the Spring Technical Meeting of the Central States Section of the Combustion Institute, 22.-24.04.2012, Dayton, Ohio, USA*.

Bader, R. F. and Bayles, D. (2000). Properties of atoms in molecules: Group additivity. *The Journal of Physical Chemistry A*, 104(23):5579–5589.

Baker, L. E., Pierce, A. C., and Luks, K. D. (1982). Gibbs energy analysis of phase equilibria. *Society of Petroleum Engineers Journal*, 22(05):731–742.

Balabin, R. M., Syunyaev, R. Z., and Karpov, S. A. (2007). Molar enthalpy of vaporization of ethanol–gasoline mixtures and their colloid state. *Fuel*, 86(3):323–327.

Bamoharram, F. F., Heravi, M. M., Roshani, M., Jahangir, M., and Gharib, A. (2006). Preyssler catalyst, [NaP5W30O110]14-: A green, efficient and reusable catalyst for esterification of salicylic acid with aliphatic and benzylic alcohols. *Applied Catalysis A: General*, 302(1):42–47.

Bard, Y. (1974). *Nonlinear Parameter Estimation*. Academic Press, New York, New York, USA.

Barrientos, E. J., Lapuerta, M., and Boehman, A. L. (2013). Group additivity in soot formation for the example of C-5 oxygenated hydrocarbon fuels. *Combustion and Flame*, 160(8):1484–1498.

Baumgardner, M. E., Vaughn, T. L., Lakshminarayanan, A., Olsen, D. B., Ratcliff, M. A., McCormick, R. L., and Marchese, A. J. (2015). Combustion of ligno-cellulosic biomass based oxygenated components in a compression ignition engine. *Energy & Fuels*, 29(11):7317–7326.

Baumlin, S., Broust, F., Bazer-Bachi, F., Bourdeaux, T., Herbinet, O., Toutie Ndiaye, F., Ferrer, M., and Lédé, J. (2006). Production of hydrogen by lignins fast pyrolysis. *International Journal of Hydrogen Energy*, 31(15):2179–2192.

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.

Beatrice, C., Bertoli, C., D'Alessio, J., Del Giacomo, N., Lazzaro, M., and Massoli, P. (1996). Experimental characterization of combustion behaviour of new diesel fuels for low emission engines. *Combustion Science and Technology*, 120(1-6):335–355.

Benavides, P. T., Gebreslassie, B. H., and Diwekar, U. M. (2015). Optimal design of adsorbents for NORM removal from produced water in natural gas fracking. Part 2: CAMD for adsorption of radium and barium. *Chemical Engineering Science*, 137:977–985.

Benson, S. W. and Buss, J. H. (1958). Additivity rules for the estimation of molecular properties. Thermodynamic properties. *The Journal of Chemical Physics*, 29(3):546–572.

Bergez-Lacoste, M., Thiebaud-Roux, S., De Caro, P., Fabre, J.-F., Gerbaud, V., and Mouloungui, Z. (2014). From chemical platform molecules to new biosolvents: Design engineering as a substitution methodology. *Biofuels, Bioproducts and Biorefining*, 8(3):438–451.

Bhardwaj, O. P., Lüers, B., Kolbeck, A. F., Koerfer, T., Kremer, F., Pischinger, S., von Berg, A., and Roth, G. (2013). Tailor made biofuels: Effect of fuel properties on the soot microstructure and consequences on particle filter regeneration. In *Proceedings of the ASME 2013 Internal Combustion Engine Division Fall Technical Conference ICEF2013, 13.-16.10.2013, Dearborn, Michigan, USA*. American Society of Mechanical Engineers. Paper No. ICEF2013-19165.

Bhogeswararao, S. and Srinivas, D. (2015). Catalytic conversion of furfural to industrial chemicals over supported Pt and Pd catalysts. *Journal of Catalysis*, 327:65–77.

BioMetics, Inc. (2002). Final technical report: Commercialization of the biofine technology for levulinic acid production from paper sludge. Technical report. BioMetics, Inc., Waltham, Massachusetts, USA.

Bluhm, K., Heger, S., Seiler, T.-B., Hallare, A. V., Schäffer, A., and Hollert, H. (2012). Toxicological and ecotoxicological potencies of biofuels used for the transport sector - A literature review. *Energy & Environmental Science*, 5(6):7381–7392.

Bluhm, K., Seiler, T.-B., Anders, N., Klankermayer, J., Schaeffer, A., and Hollert, H. (2016). Acute embryo toxicity and teratogenicity of three potential biofuels also used as flavor or solvent. *Science of the Total Environment*, 566-567:786–795.

Bogin Jr, G. E., DeFilippo, A., Chen, J., Chin, G., Luecke, J., Ratcliff, M. A., Zigler, B. T., and Dean, A. M. (2011). Numerical and experimental investigation of n-heptane autoignition in the Ignition Quality Tester (IQT). *Energy & Fuels*, 25(12):5562–5572.

Bogin Jr, G. E., Osecky, E., Chen, J.-Y., Ratcliff, M. A., Luecke, J., Zigler, B. T., and Dean, A. M. (2014). Experiments and computational fluid dynamics modeling analysis of large

n-alkane ignition kinetics in the Ignition Quality Tester. *Energy & Fuels*, 28(7):4781–4794.

Bogin Jr, G. E., Osecky, E., Ratcliff, M. A., Luecke, J., He, X., Zigler, B. T., and Dean, A. M. (2013). Ignition Quality Tester (IQT) investigation of the negative temperature coefficient region of alkane autoignition. *Energy & Fuels*, 27(3):1632–1642.

Boot, M., Frijters, P., Luijten, C., Somers, B., Baert, R., Donkerbroek, A., Klein-Douwel, R. J., and Dam, N. (2008). Cyclic oxygenates: A new class of second-generation biofuels for diesel engines? *Energy & Fuels*, 23(4):1808–1817.

Bowden, J., Johnston, A., and Russell, J. (1974). *Octane-Cetane Relationship*. US Army Fuels and Lubricants Research Laboratory, Southwest Research Institute San Antonio, Texas, USA. Final Report AFLRL No. 33, National Technical Information Service, U.S. Department of Commerce.

Bozell, J. J. and Petersen, G. R. (2010). Technology development for the production of biobased products from biorefinery carbohydrates - the US Department of Energy's "top 10" revisited. *Green Chemistry*, 12(4):539–554.

Bradley, D. (2009). Combustion and the design of future engine fuels. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 223(12):2751–2765.

Bro, R. and Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, 6(9):2812–2831.

Broadbelt, L. J., Stark, S. M., and Klein, M. T. (1994). Computer generated pyrolysis modeling: On-the-fly generation of species, reactions, and rates. *Industrial & Engineering Chemistry Research*, 33(4):790–799.

Brock, J. R. and Bird, R. B. (1955). Surface tension and the principle of corresponding states. *AIChE Journal*, 1(2):174–177.

Brown, R. and Brown, R. (2000). Melting point and molecular symmetry. *Journal of Chemical Education*, 77(6):724–731.

Bruggink, A., Schoevaart, R., and Kieboom, T. (2003). Concepts of nature in organic synthesis: Cascade catalysis and multistep conversions in concert. *Organic Process Research & Development*, 7(5):622–640.

Bruno, T. J. (2006). Improvements in the measurement of distillation curves. 1. A composition-explicit approach. *Industrial & Engineering Chemistry Research*, 45(12):4371–4380.

Buntara, T., Noel, S., Phua, P. H., Melián-Cabrera, I., de Vries, J. G., and Heeres, H. J. (2011). Caprolactam from renewable resources: Catalytic conversion of 5-hydroxymethylfurfural into caprolactone. *Angewandte Chemie International Edition*, 50(31):7083–7087.

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.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York, New York, USA.

Buxton, A., Livingston, A., and Pistikopoulos, E. (1997). Reaction path synthesis for environmental impact minimization. *Computers & Chemical Engineering*, 21:959–964.

Cai, L., Sudholt, A., Lee, D. J., Egolfopoulos, F. N., Pitsch, H., Westbrook, C. K., and Sarathy, S. M. (2014). Chemical kinetic study of a novel lignocellulosic biofuel: Di-n-butyl ether oxidation in a laminar flow reactor and flames. *Combustion and Flame*, 161(3):798–809.

Cai, L., Uygun, Y., Togbé, C., Pitsch, H., Olivier, H., Dagaut, P., and Sarathy, S. M. (2015). An experimental and modeling study of n-octanol combustion. *Proceedings of the Combustion Institute*, 35(1):419–427.

Cao, N., Du, J., Gong, C., and Tsao, G. (1996). Simultaneous production and recovery of fumaric acid from immobilized *Rhizopus oryzae* with a rotary biofilm contactor and an adsorption column. *Applied and Environmental Microbiology*, 62(8):2926–2931.

CAPEC (2012). ICAS ProPred version 4.3. Department of Chemical Engineering, Technical University of Denmark. `http://www.capec.kt.dtu.dk`.

Carlini, C., Patrono, P., Galletti, A. M. R., and Sbrana, G. (2004). Heterogeneous catalysts based on vanadyl phosphate for fructose dehydration to 5-hydroxymethyl-2-furaldehyde. *Applied Catalysis A: General*, 275(1):111–118.

Carney, J. G., Cunningham, P., and Bhagwan, U. (1999). Confidence and prediction intervals for neural network ensembles. In *Proceedings of the International Joint Conference on Neural Networks 1999 IJCNN'99, 10.-16.07.1999, Washington DC, USA*, volume 2, pages 1215–1218. IEEE.

Chang, C., Ma, X., and Cen, P. (2006). Kinetics of levulinic acid formation from glucose decomposition at high temperature. *Chinese Journal of Chemical Engineering*, 14(5):708–712.

Chemmangattuvalappil, N. G. and Eden, M. R. (2013). A novel methodology for property-based molecular design using multiple topological indices. *Industrial & Engineering Chemistry Research*, 52(22):7090–7103.

Chemmangattuvalappil, N. G. and Ng, D. K. (2013). A systematic methodology for optimal product design in an integrated biorefinery. *Computer Aided Chemical Engineering*, 32:91–96.

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.

Cheng, A., Fisher, B. T., Martin, G. C., and Mueller, C. J. (2010). Effects of fuel volatility on early direct-injection, low-temperature combustion in an optical diesel engine. *Energy & Fuels*, 24(3):1538–1551.

Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y. C., Todeschini, R., Consonni, V., Kuz'min, V. E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., and Tropsha, A. (2014). QSAR Modeling: Where have you been? Where are you going to? *Journal of Medicinal Chemistry*, 57(12):4977–5010.

Cherubini, F. and Stromman, A. H. (2010). Production of biofuels and biochemicals from lignocellulosic biomass: Estimation of maximum theoretical yields and efficiencies using matrix algebra. *Energy & Fuels*, 24(4):2657–2666.

Chheda, J. N., Román-Leshkov, Y., and Dumesic, J. A. (2007). Production of 5-hydroxymethylfurfural and furfural by dehydration of biomass-derived mono-and polysaccharides. *Green Chemistry*, 9(4):342–350.

Chiang, L. H., Russell, E. L., and Braatz, R. D. (2000). Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 50(2):243–252.

Cholakov, G. S. (2011). Towards computer aided design of fuels and lubricants. *Journal of the University of Chemical Technology and Metallurgy*, 46(3):217–236.

Christensen, E., Williams, A., Paul, S., Burton, S., and McCormick, R. L. (2011a). Properties and performance of levulinate esters as diesel blend components. *Energy & Fuels*, 25(11):5422–5428.

Christensen, E., Yanowitz, J., Ratcliff, M., and McCormick, R. L. (2011b). Renewable oxygenate blending effects on gasoline properties. *Energy & Fuels*, 25(10):4723–4733.

Chupka, G. M., Christensen, E., Fouts, L., Alleman, T. L., Ratcliff, M. A., and McCormick, R. L. (2015). Heat of vaporization measurements for ethanol blends up to 50 volume percent in several hydrocarbon blendstocks and implications for knock in SI engines. *SAE International Journal of Fuels and Lubricants*, 8(2):251–263.

Churi, N. and Achenie, L. E. (1996). Novel mathematical programming model for computer aided molecular design. *Industrial & Engineering Chemistry Research*, 35(10):3788–3794.

Cirujano, F., Corma, A., and i Xamena, F. L. (2015). Conversion of levulinic acid into chemicals: Synthesis of biomass derived levulinate esters over Zr-containing MOFs. *Chemical Engineering Science*, 124:52–60.

Clark, M. and Cramer, R. D. (1993). The probability of chance correlation using partial least squares (PLS). *Quantitative Structure-Activity Relationships*, 12(2):137–145.

Climent, M. J., Corma, A., and Iborra, S. (2014). Conversion of biomass platform molecules into fuel additives and liquid hydrocarbon fuels. *Green Chemistry*, 16(2):516–547.

Connolly, D., Mathiesen, B. V., and Ridjan, I. (2014). A comparison between renewable transport fuels that can supplement or replace biofuels in a 100% renewable energy system. *Energy*, 73:110–125.

Constable, D. J., Curzons, A. D., and Cunningham, V. L. (2002). Metrics to green chemistry - Which are the best? *Green Chemistry*, 4(6):521–527.

Constantinescu, D. and Gmehling, J. (2016). Further development of modified UNIFAC (Dortmund): Revision and extension 6. *Journal of Chemical & Engineering Data*, 61(8):2738–2748.

Constantinou, L., Bagherpour, K., Gani, R., Klein, J., and Wu, D. (1996). Computer aided product design: Problem formulations, methodology and applications. *Computers & Chemical Engineering*, 20(6):685–702.

Constantinou, L. and Gani, R. (1994). New group contribution method for estimating properties of pure compounds. *AIChE Journal*, 40(10):1697–1710.

Conte, E., Gani, R., and Ng, K. M. (2011). Design of formulated products: A systematic methodology. *AIChE Journal*, 57(9):2431–2449.

Contino, F., Foucher, F., Mounaïm-Rousselle, C., and Jeanmart, H. (2011a). Combustion characteristics of tricomponent fuel blends of ethyl acetate, ethyl propionate, and ethyl butyrate in homogeneous charge compression ignition (HCCI). *Energy & Fuels*, 25(4):1497–1503.

Contino, F., Foucher, F., Mounaïm-Rousselle, C., and Jeanmart, H. (2011b). Experimental characterization of ethyl acetate, ethyl propionate, and ethyl butanoate in a homogeneous charge compression ignition engine. *Energy & Fuels*, 25(3):998–1003.

Cookson, D. J. and Smith, B. E. (1990). Calculation of jet and diesel fuel properties using carbon-13 NMR spectroscopy. *Energy & Fuels*, 4(2):152–156.

Cookson, D. J., Smith, B. E., and Johnston, R. R. (1993). Relationships between diesel fuel ignition quality indicators and composition. *Fuel*, 72(5):661–664.

Corey, E. and Wipke, W. T. (1969). Computer-assisted design of complex organic syntheses. *Science*, 166(3902):178–192.

Corma, A., Iborra, S., and Velty, A. (2007). Chemical routes for the transformation of biomass into chemicals. *Chemical Reviews*, 107(6):2411–2502.

Cortright, R., Davda, R., and Dumesic, J. (2002). Hydrogen from catalytic reforming of biomass-derived hydrocarbons in liquid water. *Nature*, 418(6901):964–967.

Couhert, C., Commandre, J.-M., and Salvador, S. (2009). Is it possible to predict gas yields of any biomass after rapid pyrolysis at high temperature from its composition in cellulose, hemicellulose and lignin? *Fuel*, 88(3):408–417.

Cramer, R. D. (1993). Partial least squares (PLS): Its strengths and limitations. *Perspectives in Drug Discovery and Design*, 1(2):269–278.

Cramer, R. D., Bunce, J. D., Patterson, D. E., and Frank, I. E. (1988). Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quantitative Structure-Activity Relationships*, 7(1):18–25.

Creton, B., Dartiguelongue, C., de Bruin, T., and Toulhoat, H. (2010). Prediction of the cetane number of diesel compounds using the quantitative structure property relationship. *Energy & Fuels*, 24(10):5396–5403.

Cronin, M. T. and Schultz, T. W. (2003). Pitfalls in QSAR. *Journal of Molecular Structure (Theochem)*, 622(1):39–51.

Dabbagh, H., Ghobadi, F., Ehsani, M., and Moradmand, M. (2013). The influence of ester additives on the properties of gasoline. *Fuel*, 104:216–223.

Dahmen, M., Hechinger, M., Victoria Villeda, J. J., and Marquardt, W. (2012). Towards model-based identification of biofuels for compression ignition engines. *SAE International Journal of Fuels and Lubricants*, 5(3):990–1003. SAE 2012 International Powertrains, Fuels & Lubricants Meeting, 18.-20.09.2012, Malmö, Sweden.

Dahmen, M., Klankermayer, J., Marquardt, W., and Leitner, W. (2013a). Fuel design for tailor-made fuels from biomass - From biofuel production to fuel conversion in the combustion engine. In *GeCatS Infoday - Selective catalytic transformation of biogenic feedstocks (DECHEMA), 22.11.2013, Frankfurt am Main, Germany.*

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.

Dahmen, M., Victoria Villeda, J. J., and Marquardt, W. (2013b). Refunctionalization of bio-based platform chemicals into novel biofuels: A computational approach. In *3rd International Conference on Sustainable Chemical Product and Process Engineering, 27.-30.05.2013, Dalian, China.*

Daoutidis, P., Marvin, W. A., Rangarajan, S., and Torres, A. I. (2013). Engineering biomass conversion processes: A systems perspective. *AIChE Journal*, 59(1):3–18.

DDBST Dortmund Data Bank Software & Separation Technology GmbH (2017). The UNIFAC Consortium. `http://www.unifac.org` accessed on: 02/23/2017.

De Jong, E. (2011). YXY building blocks: Biorefinery approach towards fuels and plastic applications. In *World Biofuels Markets, Biorefinery Platforms, 24.03.2011, Rotterdam, The Netherlands.*

de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263.

De Ojeda, W., Bulicz, T., Han, X., Zheng, M., and Cornforth, F. (2011). Impact of fuel properties on diesel low temperature combustion. *SAE International Journal of Engines*, 4(1):188–201.

Dearden, J., Cronin, M., and Kaiser, K. (2009). How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research*, 20(3-4):241–266.

Dec, J. E. (2009). Advanced compression-ignition engines - Understanding the in-cylinder processes. *Proceedings of the Combustion Institute*, 32(2):2727–2742.

DeFries, T. H., Kastrup, R. V., and Indritz, D. (1987). Prediction of cetane number by group additivity and carbon-13 nuclear magnetic resonance. *Industrial & Engineering Chemistry Research*, 26(2):188–193.

Demirbas, A., Gullu, D., Caglar, A., and Akdeniz, F. (1997). Estimation of calorific values of fuels from lignocellulosics. *Energy Sources*, 19(8):765–770.

Donaldson, J. R. and Schnabel, R. B. (1987). Computational experience with confidence regions and confidence intervals for nonlinear least squares. *Technometrics*, 29(1):67–82.

Donkerbroek, A., Boot, M., Luijten, C., Dam, N., and Ter Meulen, J. (2011). Flame lift-off length and soot production of oxygenated fuels in relation with ignition delay in a DI heavy-duty diesel engine. *Combustion and Flame*, 158(3):525–538.

Dooley, S., Won, S. H., Chaos, M., Heyne, J., Ju, Y., Dryer, F. L., Kumar, K., Sung, C.-J., Wang, H., Oehlschlaeger, M. A., Santoro, R. J., and Litzinger, T. A. (2010). A jet fuel surrogate formulated by real fuel properties. *Combustion and Flame*, 157(12):2333–2339.

Dooley, S., Won, S. H., Heyne, J., Farouk, T. I., Ju, Y., Dryer, F. L., Kumar, K., Hui, X., Sung, C.-J., Wang, H., Oehlschlaeger, M., Iyer, V., Iyer, S., L. T., Santoro, R. J., Malewicki, T., and Brezinsky, K. (2012a). The experimental evaluation of a methodology for surrogate fuel formulation to emulate gas phase combustion kinetic phenomena. *Combustion and Flame*, 159(4):1444–1466.

Dooley, S., Won, S. H., Jahangirian, S., Ju, Y., Dryer, F. L., Wang, H., and Oehlschlaeger, M. A. (2012b). The combustion kinetics of a synthetic paraffinic jet aviation fuel and a fundamentally formulated, experimentally validated surrogate fuel. *Combustion and Flame*, 159(10):3014–3020.

Dortmund Data Bank Software & Separation Technology GmbH (2014a). DDB Online Property Estimation by the Joback Method for dioctyl carbonate

(C17H34O3). http://ddbonline.ddbst.com/OnlinePropertyEstimation/OnlinePropertyEstimation.exe?Calculate=Calculate&complist=32322 (accessed on 15.08.2014).

Dortmund Data Bank Software & Separation Technology GmbH (2014b). DDB Online Property Estimation by the Joback Method for l-galactonic acid gamma-lactone (C6H10O6). http://ddbonline.ddbst.com/OnlinePropertyEstimation/OnlinePropertyEstimation.exe?Calculate=Calculate&complist=20551 (accessed on 15.08.2014).

Dortmund Data Bank Software & Separation Technology GmbH (2014c). DDB Online Property Estimation by the Joback Method for tetrafluorosuccinic anhydride (C4F4O3). http://ddbonline.ddbst.com/OnlinePropertyEstimation/OnlinePropertyEstimation.exe?Calculate=Calculate&complist=38739 (accessed on 15.08.2014).

Dreisbach, R. and Martin, R. (1949). Physical data on some organic compounds. *Industrial & Engineering Chemistry*, 41(12):2875–2878.

Dryer, F. L. (2015). Chemical kinetic and combustion characteristics of transportation fuels. *Proceedings of the Combustion Institute*, 35(1):117–144.

Dryer, F. L., Jahangirian, S., Dooley, S., Won, S. H., Heyne, J., Iyer, V. R., Litzinger, T. A., and Santoro, R. J. (2014). Emulating the combustion behavior of real jet aviation fuels by surrogate mixtures of hydrocarbon fluid blends: Implications for science and engineering. *Energy & Fuels*, 28(5):3474–3485.

Duvedi, A. P. and Achenie, L. E. (1996). Designing environmentally safe refrigerants using mathematical programming. *Chemical Engineering Science*, 51(15):3727–3739.

El-Halwagi, M. M. (2012). *Sustainable Design Through Process Integration: Fundamentals and Applications to Industrial Pollution Prevention, Resource Conservation, and Profitability Enhancement*. Butterworth-Heinemann, Elsevier, Oxford, United Kingdom.

EN 228 (2014). DIN EN 228 "Automotive fuels - Unleaded petrol - Requirements and test methods; German version", DIN Deutsches Institut fuer Normung e.V., Berlin, Beuth Verlag GmbH, Berlin, Germany.

EN 590 (2014). DIN EN 590 "Automotive fuels - Diesel - Requirements and test methods; German version", DIN Deutsches Institut fuer Normung e.V., Berlin, Beuth Verlag GmbH, Berlin, Germany.

Engel, C. A. R., Straathof, A. J., Zijlmans, T. W., van Gulik, W. M., and van der Wielen, L. A. (2008). Fumaric acid production by fermentation. *Applied Microbiology and Biotechnology*, 78(3):379–389.

Ericksen, D., Wilding, W. V., Oscarson, J. L., and Rowley, R. L. (2002). Use of the DIPPR database for development of QSPR correlations: Normal boiling point. *Journal of Chemical & Engineering Data*, 47(5):1293–1302.

Eriksson, L., Andersson, P. L., Johansson, E., and Tysklind, M. (2006a). Megavariate analysis of environmental QSAR data. Part I - A basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD). *Molecular Diversity*, 10(2):169–186.

Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T., McDowell, R. M., and Gramatica, P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. *Environmental Health Perspectives*, 111(10):1361–1375.

Eriksson, L., Johansson, E., Müller, M., and Wold, S. (2000). On the selection of the training set in environmental QSAR analysis when compounds are clustered. *Journal of Chemometrics*, 14(5-6):599–616.

Eriksson, L., Kettaneh-Wold, N., Trygg, J., Wikström, C., and Wold, S. (2006b). *Multi- and Megavariate Data Analysis: Part I: Basic Principles and Applications*. Umetrics AB, Umea, Sweden.

Ezeji, T. C., Qureshi, N., and Blaschek, H. P. (2007). Bioproduction of butanol from biomass: From genes to bioreactors. *Current Opinion in Biotechnology*, 18(3):220–227.

Fan, C., Guan, H., Zhang, H., Wang, J., Wang, S., and Wang, X. (2011). Conversion of fructose and glucose into 5-hydroxymethylfurfural catalyzed by a solid heteropolyacid salt. *Biomass and Bioenergy*, 35(7):2659–2665.

Fedors, R. (1982). A relationship between chemical structure and the critical temperature. *Chemical Engineering Communications*, 16(1-6):149–151.

Fehér, C., Kriván, E., Eller, Z., Hancsók, J., and Skoda-Földes, R. (2014). *Oligomerization of Chemical and Biological Compounds*, chapter The use of ionic liquids in the oligomerization of alkenes, pages 31–68. InTech.

Figueras, J. (1993). Morgan revisited. *Journal of Chemical Information and Computer Sciences*, 33(5):717–718.

Fisher, B. T., Knothe, G., and Mueller, C. J. (2010). Liquid-phase penetration under unsteady in-cylinder conditions: Soy-and cuphea-derived biodiesel fuels versus conventional diesel. *Energy & Fuels*, 24(9):5163–5180.

Floudas, C. A., Gümüş, Z. H., Adjiman, C. S., Schweiger, C. A., Esposito, W. R., Klepeis, J. L., Pardalos, P. M., Meyer, C. A., and Harding, S. T. (1999). *Handbook of Test Problems in Local and Global Optimization*. Kluwer Academic Publishers, Boston, Massachusetts, USA.

Foong, T. M., Morganti, K. J., Brear, M. J., da Silva, G., Yang, Y., and Dryer, F. L. (2014). The octane numbers of ethanol blended with gasoline and its surrogates. *Fuel*, 115:727–739.

Franceschini, G. and Macchietto, S. (2008). Novel anticorrelation criteria for model-based experiment design: Theory and formulations. *AIChE Journal*, 54(4):1009–1024.

Francis, J. G. (1961). The QR transformation - A unitary analogue to the LR transformation - Part 1. *The Computer Journal*, 4(3):265–271.

Fredenslund, A., Jones, R. L., and Prausnitz, J. M. (1975). Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal*, 21(6):1086–1099.

Fulton, L. M., Lynd, L. R., Körner, A., Greene, N., and Tonachel, L. R. (2015). The need for biofuels as part of a low carbon energy future. *Biofuels, Bioproducts and Biorefining*, 9(5):476–483.

Gallezot, P. (2012). Conversion of biomass to selected chemical products. *Chemical Society Reviews*, 41(4):1538–1558.

Gani, R. (2004a). Chemical product design: Challenges and opportunities. *Computers & Chemical Engineering*, 28(12):2441–2457.

Gani, R. (2004b). Computer-aided methods and tools for chemical product design. *Chemical Engineering Research and Design*, 82(11):1494–1504.

Gani, R. and Brignole, E. (1983). Molecular design of solvents for liquid extraction based on UNIFAC. *Fluid Phase Equilibria*, 13:331–340.

Gani, R. and González, H. E. (2009). Letter to the editor. *AIChE Journal*, 55(6):1626–1627.

Gani, R., Harper, P. M., and Hostrup, M. (2005). Automatic creation of missing groups through connectivity index for pure-component property prediction. *Industrial and Engineering Chemistry Research*, 44(18):7262–7269.

Gani, R. and Ng, K. M. (2015). Product design – Molecules, devices, functional products, and formulated products. *Computers & Chemical Engineering*, 81:70–79.

Gani, R., Nielsen, B., and Fredenslund, A. (1991). A group contribution approach to computer-aided molecular design. *AIChE Journal*, 37(9):1318–1332.

García, A., Monsalve-Serrano, J., Heuser, B., Jakob, M., Kremer, F., and Pischinger, S. (2016). Influence of fuel properties on fundamental spray characteristics and soot emissions using different tailor-made fuels from biomass. *Energy Conversion and Management*, 108:243–254.

Geilen, F., Engendahl, B., Harwardt, A., Marquardt, W., Klankermayer, J., and Leitner, W. (2010). Selective and flexible transformation of biomass-derived platform chemicals by a multifunctional catalytic system. *Angewandte Chemie*, 122(32):5642–5646.

Geilen, F. M., Engendahl, B., Hoelscher, M., Klankermayer, J., and Leitner, W. (2011). Selective homogeneous hydrogenation of biogenic carboxylic acids with [Ru(TriPhos)H]+: A mechanistic study. *Journal of the American Chemical Society*, 133(36):14349–14358.

Geladi, P. and Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, 185:1–17.

Gerbaud, V., Dos Santos, M. T., Pandya, N., and Aubry, J. (2017). Computer aided framework for designing bio-based commodity molecules with enhanced properties. *Chemical Engineering Science*, 159:177–193.

Ghosh, P. (2008). Predicting the effect of cetane improvers on diesel fuels. *Energy & Fuels*, 22(2):1073–1079.

Ghosh, P. and Jaffe, S. B. (2006). Detailed composition-based model for predicting the cetane number of diesel fuels. *Industrial & Engineering Chemistry Research*, 45(1):346–351.

Glebov, L., Mikaya, A., Yatsenko, A., Zaikin, V., Kliger, G., and Loktev, S. (1985). Effective gas-phase deoxygenation of alcohols and ketones on iron catalyst. *Tetrahedron Letters*, 26(28):3373–3376.

Global Bioenergies (2017). Press Release "First production in history of fully renewable ETBE". `http://www.global-bioenergies.com/first-production-in-history-of-fully-renewable-etbe/?lang=en` (accessed on 22.04.2017).

Gmehling, J., Kolbe, B., Kleiber, M., and Rarey, J. (2012). *Chemical Thermodynamics for Process Simulation*. Wiley-VCH, Weinheim, Germany.

Gmehling, J., Li, J., and Schiller, M. (1993). A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Industrial & Engineering Chemistry Research*, 32(1):178–193.

Gmehling, J., Lohmann, J., Jakob, A., Li, J., and Joh, R. (1998). A modified UNIFAC (Dortmund) model. 3. Revision and extension. *Industrial & Engineering Chemistry Research*, 37(12):4876–4882.

Goh, E.-B., Baidoo, E. E., Keasling, J. D., and Beller, H. R. (2012). Engineering of bacterial methyl ketone synthesis for biofuels. *Applied and Environmental Microbiology*, 78(1):70–80.

Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y.-D., Lee, K.-H., and Tropsha, A. (2003). Rational selection of training and test sets for the development of validated QSAR models. *Journal of Computer-Aided Molecular Design*, 17(2-4):241–253.

Golbraikh, A. and Tropsha, A. (2002a). Beware of q2! *Journal of Molecular Graphics and Modelling*, 20(4):269–276.

Golbraikh, A. and Tropsha, A. (2002b). Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Journal of Computer-Aided Molecular Design*, 16(5):357–369.

Gonzalez, H. E., Abildskov, J., Gani, R., Rousseaux, P., and Le Bert, B. (2007). A method for prediction of UNIFAC group interaction parameters. *AIChE Journal*, 53(6):1620–1632.

Goulas, K. A. and Toste, F. D. (2016). Combining microbial production with chemical upgrading. *Current Opinion in Biotechnology*, 38:47–53.

Graboski, M. S. and McCormick, R. L. (1998). Combustion of fat and vegetable oil derived fuels in diesel engines. *Progress in Energy and Combustion Science*, 24(2):125–164.

Gramatica, P. (2007). Principles of QSAR models validation: Internal and external. *QSAR & Combinatorial Science*, 26(5):694–701.

Gramatica, P. (2014). External evaluation of QSAR models, in addition to cross-validation: Verification of predictive capability on totally new chemicals. *Molecular Informatics*, 33(4):311–314.

Grande, P. M., Viell, J., Theyssen, N., Marquardt, W., de María, P. D., and Leitner, W. (2015). Fractionation of lignocellulosic biomass using the OrganoCat process. *Green Chemistry*, 17(6):3533–3539.

Graziano, B., Heuser, B., Kremer, F., Pischinger, S., and Rohs, H. (2016). The oxidation potential number: An index to evaluate inherent soot reduction in DI diesel spray plumes. *SAE International Journal of Engines*, 9(1):222–236.

Graziano, B., Kremer, F., Heufer, K. A., Rohs, H., and Pischinger, S. (2015). On the potential of oxygenated fuels as an additional degree of freedom in the mixture formation in direct injection diesel engines. *SAE International Journal of Fuels and Lubricants*, 8(1):62–79.

Green, E. M. (2011). Fermentative production of butanol - The industrial perspective. *Current Opinion in Biotechnology*, 22(3):337–343.

Griffiths, J., Halford-Maw, P., and Mohamed, C. (1997). Spontaneous ignition delays as a diagnostic of the propensity of alkanes to cause engine knock. *Combustion and Flame*, 111(4):327–337.

Groendyk, M. A. and Rothamer, D. (2015). Effects of fuel physical properties on autoignition characteristics in a heavy duty compression ignition engine. *SAE International Journal of Fuels and Lubricants*, 8(1):200–213.

Gruter, G. and De Jong, E. (2009). Furanics: Novel fuel options from carbohydrates. *Biofuels Technology*, 1:11–17.

Gugisch, R., Kerber, A., Kohnert, A., Laue, R., Meringer, M., Rücker, C., and Wassermann, A. (2015). MOLGEN 5.0, A molecular structure generator. In Basak, S. C., Restrepo, G., and Villaveces, J. L., editors, *Advances in Mathematical Chemistry and Applications*, volume 1, pages 113–138. Bentham Science Publishers Ltd., Elsevier, Amsterdam, The Netherlands.

Gulder, O. L. and Glavincevski, B. (1986). Prediction of cetane number of diesel fuels from carbon type structural composition determined by proton NMR spectroscopy. *Industrial & Engineering Chemistry Product Research and Development*, 25(2):153–156.

Gupta, K., Tyagi, D., Dwivedi, A. D., Mobin, S. M., and Singh, S. K. (2015). Catalytic transformation of bio-derived furans to valuable ketoacids and diketones by water-soluble ruthenium catalysts. *Green Chemistry*, 17(9):4618–4627.

Gürbüz, E. I., Gallo, J. M. R., Alonso, D. M., Wettstein, S. G., Lim, W. Y., and Dumesic, J. A. (2013). Conversion of hemicellulose into furfural using solid acid catalysts in γ-valerolactone. *Angewandte Chemie International Edition*, 52(4):1270–1274.

Haas, F. M., Ramcharan, A., and Dryer, F. L. (2011). Relative reactivities of the isomeric butanols and ethanol in an Ignition Quality Tester. *Energy & Fuels*, 25(9):3909–3916.

Han, J., Sen, S. M., Luterbacher, J. S., Alonso, D. M., Dumesic, J. A., and Maravelias, C. T. (2015). Process systems engineering studies for the synthesis of catalytic biomass-to-fuels strategies. *Computers & Chemical Engineering*, 81:57–69.

Hankinson, R. W. and Thomson, G. H. (1979). A new correlation for saturated densities of liquids and their mixtures. *AIChE Journal*, 25(4):653–663.

Hansen, H. K., Rasmussen, P., Fredenslund, A., Schiller, M., and Gmehling, J. (1991). Vapor-liquid equilibria by UNIFAC group contribution. 5. Revision and extension. *Industrial & Engineering Chemistry Research*, 30(10):2352–2355.

Harding, S. and Floudas, C. (2000). Phase stability with cubic equations of state: Global optimization approach. *AIChE Journal*, 46(7):1422–1440.

Harnisch, F., Blei, I., dos Santos, T. R., Möller, M., Nilges, P., Eilts, P., and Schröder, U. (2013). From the test-tube to the test-engine: Assessing the suitability of prospective liquid biofuel compounds. *RSC Advances*, 3(25):9594–9605.

Harper, P. M. and Gani, R. (2000). A multi-step and multi-level approach for computer aided molecular design. *Computers & Chemical Engineering*, 24(2):677–683.

Harper, P. M., Gani, R., Kolar, P., and Ishikawa, T. (1999). Computer-aided molecular design with combined molecular modeling and group contribution. *Fluid Phase Equilibria*, 158:337–347.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction - Second Edition*. Springer, New York, New York, USA.

Hatzimanikatis, V., Li, C., Ionita, J. A., Henry, C. S., Jankowski, M. D., and Broadbelt, L. J. (2005). Exploring the diversity of complex metabolic networks. *Bioinformatics*, 21(8):1603–1609.

Hatzioannidis, I., Voutsas, E. C., Lois, E., and Tassios, D. P. (1998). Measurement and prediction of Reid vapor pressure of gasoline in the presence of additives. *Journal of Chemical & Engineering Data*, 43(3):386–392.

Haveren, J. v., Scott, E. L., and Sanders, J. (2008). Bulk chemicals from biomass. *Biofuels, Bioproducts and Biorefining*, 2(1):41–57.

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, 44(1):1–12.

Hawkins, D. M., Basak, S. C., and Mills, D. (2003). Assessing model fit by cross-validation. *Journal of Chemical Information and Computer Sciences*, 43(2):579–586.

Hayashi, Y. (2016). Pot economy and one-pot synthesis. *Chemical Science*, 7(2):866–880.

Hechinger, M. (2014). *Model-based Identification of Promising Biofuel Candidates for Spark-Ignited Engines*. PhD thesis, RWTH Aachen University, AVT - Process Systems Engineering, Fortschritt-Berichte VDI, Reihe 3 (Verfahrenstechnik), Nr. 940, VDI-Verlag, Duesseldorf, Germany.

Hechinger, M., Dahmen, M., Victoria Villeda, J. J., and Marquardt, W. (2012a). Rigorose Generierung und modellbasierte Auswahl von Biokraftstoff-Kandidaten. *Chemie Ingenieur Technik*, 84(8):1390–1391. Special Issue: ProcessNet-Jahrestagung 2012 und 30. DECHEMA-Jahrestagung der Biotechnologen, 10.-13.09.2012, Karlsruhe, Germany.

Hechinger, M., Dahmen, M., Victoria Villeda, J. J., and Marquardt, W. (2012b). Rigorous generation and model-based selection of future biofuel candidates. In Karimi, I. A. and Srinivasan, R., editors, *11th International Symposium on Process Systems Engineering - PSE2012 (15.-19.07.2012, Singapore)*, volume 31 of *Computer Aided Chemical Engineering*, pages 1341 – 1345. Elsevier.

Hechinger, M., Leonhard, K., and Marquardt, W. (2012c). What is wrong with QSPR models based on three-dimensional descriptors? *Journal of Chemical Information and Modeling*, 52(8):1984–1993.

Hechinger, M. and Marquardt, W. (2010). Targeted QSPR for the prediction of the laminar burning velocity of biofuels. *Computers & Chemical Engineering*, 34(9):1507–1514.

Hechinger, M. and Schulze Langenhorst, L. (2012). DIPPR database extraction tool for Matlab. Aachener Verfahrenstechnik - Process Systems Engineering, RWTH Aachen University, Aachen, Germany.

Hechinger, M., Voll, A., and Marquardt, W. (2010). Towards an integrated design of biofuels and their production pathways. *Computers & Chemical Engineering*, 34(12):1909–1918.

Heck, S. M., Pritchard, H. O., and Griffiths, J. F. (1998). Cetane number vs. structure in paraffin hydrocarbons. *Journal of the Chemical Society, Faraday Transactions*, 94(12):1725–1727.

Heese, F. P., Dry, M. E., and Möller, K. P. (1999). Single stage synthesis of diisopropyl ether - An alternative octane enhancer for lead-free petrol. *Catalysis Today*, 49(1):327–335.

Hellier, P., Ladommatos, N., Allan, R., Filip, S., and Rogerson, J. (2013). The importance of double bond position and cis-trans isomerisation in diesel combustion and emissions. *Fuel*, 105:477–489.

Hemdal, S., Andersson, M., Dahlander, P., Ochoterena, R., and Denbratt, I. (2011). In-cylinder soot imaging and emissions of stratified combustion in a spark-ignited spray-guided direct-injection gasoline engine. *International Journal of Engine Research*, 12(6):549–563.

Herbinet, O., Pitz, W. J., and Westbrook, C. K. (2008). Detailed chemical kinetic oxidation mechanism for a biodiesel surrogate. *Combustion and Flame*, 154(3):507–528.

Heuser, B., Jakob, M., Kremer, F., Pischinger, S., Kerschgens, B., and Pitsch, H. (2013a). Tailor-made fuels from biomass: Influence of molecular structures on the exhaust gas emissions of compression ignition engines. Technical report. SAE Technical Paper 2013-36-0571.

Heuser, B., Kremer, F., Pischinger, S., and Klankermayer, J. (2013b). Optimization of diesel combustion and emissions with tailor-made fuels from biomass. *SAE International Journal of Fuels and Lubricants*, 6(3):922–934.

Heuser, B., Kremer, F., Pischinger, S., Rohs, H., Holderbaum, B., and Körfer, T. (2016). An experimental investigation of dual-fuel combustion in a light duty diesel engine by in-cylinder blending of ethanol and diesel. *SAE International Journal of Engines*, 9(1):11–25.

Heuser, B., Laible, T., Jakob, M., Kremer, F., and Pischinger, S. (2014). C8-oxygenates for clean diesel combustion. Technical report. SAE Technical Paper 2014-01-1253.

Hoffmann, W. and Florin, F. (1943). Zweckmäßige Darstellung von Dampfdruckkurven. *Verfahrenstechnik, Z. VDI-Beiheft*, (2):47–51.

Holladay, J. D., Hu, J., King, D. L., and Wang, Y. (2009). An overview of hydrogen production technologies. *Catalysis Today*, 139(4):244–260.

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. Special Issue on the 12th International Congress of Engine Combustion Processes, 12.-13.03.2015, Ludwigsburg, Germany.

Horváth, I. T., Mehdi, H., Fábos, V., Boda, L., and Mika, L. T. (2008). γ-Valerolactone - A sustainable liquid for energy and carbon-based chemicals. *Green Chemistry*, 10(2):238–242.

Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics*, 2(3):211–228.

Hronec, M. and Fulajtarová, K. (2012). Selective transformation of furfural to cyclopentanone. *Catalysis Communications*, 24:100–104.

Huber, G. W., Iborra, S., and Corma, A. (2006). Synthesis of transportation fuels from biomass: Chemistry, catalysts, and engineering. *Chemical Reviews*, 106(9):4044–4098.

Huber, M., Smith, B., Ott, L., and Bruno, T. (2008). Surrogate mixture model for the thermophysical properties of synthetic aviation fuel S-8: Explicit application of the advanced distillation curve. *Energy & Fuels*, 22(2):1104–1114.

Hui, X., Kumar, K., Sung, C.-J., Edwards, T., and Gardner, D. (2012). Experimental studies on the combustion characteristics of alternative jet fuels. *Fuel*, 98:176–182.

Hukkerikar, A. S. (2013). *Development of Pure Component Property Models for Chemical Product-Process Design and Analysis*. PhD thesis, Computer Aided Process-Product Engineering Center, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Lyngby, Denmark.

Hukkerikar, A. S., Sarup, B., Ten Kate, A., Abildskov, J., Sin, G., and Gani, R. (2012). Group-contribution plus (GC^+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilibria*, 321:25–43.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.

ISO 3960 (2010). DIN EN ISO 3960 "Animal and vegetable fats and oils - Determination of peroxide value - Iodometric (visual) endpoint determination - Deutsche Fassung", DIN Deutsches Institut fuer Normung e.V., Berlin, Beuth Verlag GmbH, Berlin, Germany.

Jäger, G. and Büchs, J. (2012). Biocatalytic conversion of lignocellulose to platform chemicals. *Biotechnology Journal*, 7(9):1122–1136.

Jang, Y.-S., Kim, B., Shin, J. H., Choi, Y. J., Choi, S., Song, C. W., Lee, J., Park, H. G., and Lee, S. Y. (2012). Bio-based production of C2–C6 platform chemicals. *Biotechnology and Bioengineering*, 109(10):2437–2459.

Jang, Y.-S., Malaviya, A., and Lee, S. Y. (2013). Acetone–butanol–ethanol production with high productivity using *Clostridium acetobutylicum* BKM19. *Biotechnology and Bioengineering*, 110(6):1646–1653.

Janssen, A. J., Kremer, F. W., Baron, J. H., Muether, M., Pischinger, S., and Klankermayer, J. (2011). Tailor-made fuels from biomass for homogeneous low-temperature diesel combustion. *Energy & Fuels*, 25(10):4734–4744.

Jaworska, J., Nikolova-Jeliazkova, N., and Aldenberg, T. (2005). QSAR applicabilty domain estimation by projection of the training set descriptor space: A review. *Alternatives to Laboratory Animals: ATLA*, 33(5):445–459.

Jenkins, R. W., Munro, M., Nash, S., and Chuck, C. J. (2013). Potential renewable oxygenated biofuels for the aviation and road transport sectors. *Fuel*, 103:593–599.

Jeuland, N., Montagne, X., and Gautrot, X. (2004). Potentiality of ethanol as a fuel for dedicated engine. *Oil & Gas Science and Technology*, 59(6):559–570.

Jiang, L., Wang, J., Liang, S., Cai, J., Xu, Z., Cen, P., Yang, S., and Li, S. (2011). Enhanced butyric acid tolerance and bioproduction by *Clostridium tyrobutyricum* immobilized in a fibrous bed bioreactor. *Biotechnology and Bioengineering*, 108(1):31–40.

Jin, C., Yao, M., Liu, H., Lee, C.-f. F., and Ji, J. (2011). Progress in the production and application of n-butanol as a biofuel. *Renewable and Sustainable Energy Reviews*, 15(8):4080–4106.

Joback, K. G. (1989). *Designing Molecules Possessing Desired Physical Property Values*. PhD thesis, Department of Chemical Engineering, Massachusetts Institute of Technology, Boston, Massachusetts, USA.

Joback, K. G. and Reid, R. C. (1987). Estimation of pure-component properties from group-contributions. *Chemical Engineering Communications*, 57(1-6):233–243.

Jolliffe, I. T. (2002). *Principal Component Analysis - Second Edition*. Springer, New York, New York, USA.

Julis, J. and Leitner, W. (2012). Synthesis of 1-octanol and 1,1-dioctyl ether from biomass-derived platform chemicals. *Angewandte Chemie International Edition*, 51(34):8615–8619.

Jung, D. W., Jeong, J. H., Lim, O. T., Pyo, Y. D., Lee, Y. J., and Iida, N. (2011). Influence of pilot injection on combustion characteristics and emissions in a DI diesel engine fueled with diesel and DME. Technical report. SAE Technical Paper 2011-01-1958.

Käldström, M., Meine, N., Farès, C., Rinaldi, R., and Schüth, F. (2014). Fractionation of water-soluble lignocelluloseinto C5/C6 sugars and sulfur-free lignins. *Green Chemistry*, 16(5):2454–2462.

Kalghatgi, G. (2014a). *Fuel / Engine Interactions*. SAE International, Warrendale, Pennsylvania, USA.

Kalghatgi, G. (2015). Developments in internal combustion engines and implications for combustion science and future transport fuels. *Proceedings of the Combustion Institute*, 35(1):101–115.

Kalghatgi, G. T. (2005). Auto-ignition quality of practical fuels and implications for fuel requirements of future SI and HCCI engines. Technical report. SAE Technical Paper 2005-01-0239.

Kalghatgi, G. T. (2014b). The outlook for fuels for internal combustion engines. *International Journal of Engine Research*, 15(4):383–398.

Kapdan, I. K. and Kargi, F. (2006). Bio-hydrogen production from waste materials. *Enzyme and Microbial Technology*, 38(5):569–582.

Kar, K., Last, T., Haywood, C., and Raine, R. (2009). Measurement of vapor pressures and enthalpies of vaporization of gasoline and ethanol blends and their effects on mixture preparation in an SI engine. *SAE International Journal of Fuels and Lubricants*, 1(1):132–144.

Karunanithi, A. T., Achenie, L. E., and Gani, R. (2005). A new decomposition-based computer-aided molecular/mixture design methodology for the design of optimal solvents and solvent mixtures. *Industrial & Engineering Chemistry Research*, 44(13):4785–4797.

Katritzky, A. R. and Fara, D. C. (2005). How chemical structure determines physical, chemical, and technological properties: An overview illustrating the potential of quantitative structure-property relationships for fuels science. *Energy & Fuels*, 19(3):922–935.

Katritzky, A. R., Karelson, M., and Petrukhin, R. (2005). Codessa Pro - QSPR/QSAR software. University of Florida, Gainesville, Florida, USA. http://www.codessa-pro.com.

Katritzky, A. R., Kuanar, M., Slavov, S., Hall, C. D., Karelson, M., Kahn, I., and Dobchev, D. A. (2010). Quantitative correlation of physical and chemical properties with chemical structure: Utility for prediction. *Chemical Reviews*, 110(10):5714–5789.

Katritzky, A. R., Lobanov, V. S., and Karelson, M. (1995). QSPR: The correlation and quantitative prediction of chemical and physical properties from structure. *Chemical Society Reviews*, 24(4):279–287.

Kawahara, R., Fujita, K.-i., and Yamaguchi, R. (2012). Cooperative catalysis by iridium complexes with a bipyridonate ligand: Versatile dehydrogenative oxidation of alcohols and reversible dehydrogenation-hydrogenation between 2-propanol and acetone. *Angewandte Chemie International Edition*, 51(51):12790–12794.

Kim, J.-H., Block, D. E., and Mills, D. A. (2010). Simultaneous consumption of pentose and hexose sugars: An optimal microbial phenotype for efficient fermentation of lignocellulosic biomass. *Applied Microbiology and Biotechnology*, 88(5):1077–1085.

Klankermayer, J., Dahmen, M., Victoria Villeda, J. J., Marquardt, W., and Leitner, W. (2013). Tailor-made fuels from biomass - Recent progress in model-based fuel design and fuel prototyping. In *Book of Extended Abstracts of the 1st TMFB International Conference, 18.-20.06.2013, Aachen, Germany.*

Klankermayer, J., Thewes, M., and Muether, M. (2011). From production to combustion - An example for the fuel design process. In *Book of Extended Abstracts of the 4th TMFB International Workshop of the Cluster of Excellence Tailor-Made Fuels from Biomass, 24.-26.05.2011, Aachen, Germany.*

Klein-Marcuschamer, D. and Blanch, H. W. (2013). Survival of the fittest: An economic perspective on the production of novel biofuels. *AIChE Journal*, 59(12):4454–4460.

Klincewicz, K. and Reid, R. (1984). Estimation of critical properties with group contribution methods. *AIChE Journal*, 30(1):137–142.

Knop, V., Loos, M., Pera, C., and Jeuland, N. (2014). A linear-by-mole blending rule for octane numbers of n-heptane/iso-octane/toluene mixtures. *Fuel*, 115:666–673.

Knothe, G. (2005). Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. *Fuel Processing Technology*, 86(10):1059–1070.

Knotts, T. A., Wilding, W. V., Oscarson, J. L., and Rowley, R. L. (2001). Use of the DIPPR database for development of QSPR correlations: Surface tension. *Journal of Chemical & Engineering Data*, 46(5):1007–1012.

Kohse-Höinghaus, K., Oßwald, P., Cool, T. A., Kasper, T., Hansen, N., Qi, F., Westbrook, C. K., and Westmoreland, P. R. (2010). Biofuel combustion chemistry: From ethanol to biodiesel. *Angewandte Chemie International Edition*, 49(21):3572–3597.

Kong, X., Zheng, R., Zhu, Y., Ding, G., Zhu, Y., and Li, Y.-W. (2015). Rational design of Ni-based catalysts derived from hydrotalcite for selective hydrogenation of 5-hydroxymethylfurfural. *Green Chemistry*, 17(4):2504–2514.

Koso, S., Nakagawa, Y., and Tomishige, K. (2011). Mechanism of the hydrogenolysis of ethers over silica-supported rhodium catalyst modified with rhenium oxide. *Journal of Catalysis*, 280(2):221–229.

Kremer, F. (2011). Fuel numbers – TP-CIF-1-6. Internal Presentation, Core Interaction Field (CIF), 16.05.2011, Cluster of Excellence "Tailor-Made Fuels from Biomass", Institute for Combustion Engines, RWTH Aachen University, Aachen, Germany.

Kumar, S., Chauhan, M. K., and Varun (2013). Numerical modeling of compression ignition engine: A review. *Renewable and Sustainable Energy Reviews*, 19:517–530.

Lakó, J., Hancsók, J., Yuzhakova, T., Marton, G., Utasi, A., and Rédey, Á. (2008). Biomass - A source of chemicals and energy for sustainable development. *Environmental Engineering and Management Journal*, 7(5):499–509.

Lampe, M., Stavrou, M., Schilling, J., Sauer, E., Gross, J., and Bardow, A. (2015). Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT. *Computers & Chemical Engineering*, 81:278–287.

Lange, J.-P. (2007). Lignocellulose conversion: An introduction to chemistry, process and economics. *Biofuels, Bioproducts and Biorefining*, 1(1):39–48.

Lange, J.-P., Price, R., Ayoub, P. M., Louis, J., Petrus, L., Clarke, L., and Gosselink, H. (2010). Valeric biofuels: A platform of cellulosic transportation fuels. *Angewandte Chemie International Edition*, 49(26):4479–4483.

Lange, J.-P., van der Heide, E., van Buijtenen, J., and Price, R. (2012). Furfural - A promising platform for lignocellulosic biofuels. *ChemSusChem*, 5(1):150–166.

Lapidus, A., Smolenskii, E., Bavykin, V., Myshenkova, T., and Kondratev, L. (2008). Models for the calculation and prediction of the octane and cetane numbers of individual hydrocarbons. *Petroleum Chemistry*, 48(4):277–286.

Larsen, U., Johansen, T., and Schramm, J. (2009). Ethanol as a fuel for road transportation. Technical report. International Energy Agency - Advanced Motor Fuels Agreement, Technical University of Denmark, Lyngby, Denmark.

Lee, C. S., Park, S. W., and Kwon, S. I. (2005). An experimental study on the atomization and combustion characteristics of biodiesel-blended fuels. *Energy & Fuels*, 19(5):2201–2208.

Lee, I., Johnson, L. A., and Hammond, E. G. (1995). Use of branched-chain esters to reduce the crystallization temperature of biodiesel. *Journal of the American Oil Chemists Society*, 72(10):1155–1160.

Lee, S.-w., Tanaka, D., Kusaka, J., and Daisho, Y. (2002). Effects of diesel fuel characteristics on spray and combustion in a diesel engine. *JSAE Review*, 23(4):407–414.

Leitner, W., Klankermayer, J., Pischinger, S., Pitsch, H., and Kohse-Höinghaus, K. (2017). Synthese, motorische Verbrennung, Emissionen: Chemische Aspekte des Kraftstoffdesigns. *Angewandte Chemie*, 129(20):5457–5457.

Leonard, J. T. and Roy, K. (2006). On selection of training and test sets for the development of predictive QSAR models. *QSAR & Combinatorial Science*, 25(3):235–251.

Li, C.-J. and Trost, B. M. (2008). Green chemistry for chemical synthesis. *Proceedings of the National Academy of Sciences*, 105(36):13197–13202.

Li, N. and Huber, G. W. (2010). Aqueous-phase hydrodeoxygenation of sorbitol with Pt/SiO2–Al2O3: Identification of reaction intermediates. *Journal of Catalysis*, 270(1):48–59.

Lide, D. R. (2003). *CRC Handbook of Chemistry and Physics*. CRC Press Inc., Boca Raton, Florida, USA.

Liu, H., Bi, X., Huo, M., Lee, C. F., and Yao, M. (2012). Soot emissions of various oxygenated biofuels in conventional diesel combustion and low-temperature combustion conditions. *Energy & Fuels*, 26(3):1900–1911.

Liu, H., Zheng, Z., Yue, L., Kong, L., and Yao, M. (2014). Effects of fuel volatility on combustion and emissions over a wide range of EGR rates in a diesel engine. Technical report. SAE Technical Paper 2014-01-2659.

Lohmann, J., Joh, R., and Gmehling, J. (2001). From UNIFAC to modified UNIFAC (Dortmund). *Industrial & Engineering Chemistry Research*, 40(3):957–964.

Luska, K. L., Julis, J., Stavitski, E., Zakharov, D. N., Adams, A., and Leitner, W. (2014). Bifunctional nanoparticle–SILP catalysts (NPs@SILP) for the selective deoxygenation of biomass substrates. *Chemical Science*, 5(12):4895–4905.

Luterbacher, J., Alonso, D. M., and Dumesic, J. (2014). Targeted chemical upgrading of lignocellulosic biomass to platform molecules. *Green Chemistry*, 16(12):4816–4838.

Lynch, M. D., Gill, R. T., and Lipscomb, T. E. (2013). Methods for producing 3-hydroxypropionic acid and other products. US Patent Application Publication No. 2013/0071893 A1.

Ma, C., Wang, A., Qin, J., Li, L., Ai, X., Jiang, T., Tang, H., and Xu, P. (2009). Enhanced 2,3-butanediol production by *Klebsiella pneumoniae* SDM. *Applied Microbiology and Biotechnology*, 82(1):49–57.

Magnusson, R. and Nilsson, C. (2011). The influence of oxygenated fuels on emissions of aldehydes and ketones from a two-stroke spark ignition engine. *Fuel*, 90(3):1145–1154.

Mamman, A. S., Lee, J.-M., Kim, Y.-C., Hwang, I. T., Park, N.-J., Hwang, Y. K., Chang, J.-S., and Hwang, J.-S. (2008). Furfural: Hemicellulose/xylose-derived biochemical. *Biofuels, Bioproducts and Biorefining*, 2(5):438–454.

Marquardt, W., Dahmen, M., Hechinger, M., Recker, S., Ulonska, K., Viell, J., and Victoria Villeda, J. J. (2013). Bioeconomy science - A chemical engineering perspective. In *9th World Congress of Chemical Engineering (WCCE 9), 18.-23.08.2013, Seoul, South Korea*.

Marquardt, W., Harwardt, A., Hechinger, M., Kraemer, K., Viell, J., and Voll, A. (2010). The biorenewables opportunity - Toward next generation process and product systems. *AIChE Journal*, 56(9):2228–2235.

Marrero, J. and Gani, R. (2001). Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria*, 183–184:183–208.

Marshall, A.-L. and Alaimo, P. J. (2010). Useful products from complex starting materials: Common chemicals from biomass feedstocks. *Chemistry -A European Journal*, 16(17):4970–4980.

Marsili-Libelli, S., Guerrizio, S., and Checchi, N. (2003). Confidence regions of estimated parameters for ecological systems. *Ecological Modelling*, 165(2):127–146.

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.

Masuch, K., Fatemi, A., Murrenhoff, H., and Leonhard, K. (2011). A COSMO-RS based QSPR model for the lubricity of biodiesel and petrodiesel components. *Lubrication Science*, 23(6):249–262.

Matijošius, J. and Sokolovskij, E. (2009). Research into the quality of fuels and their biocomponents. *Transport*, 24(3):212–217.

McCormick, R. L., Ratcliff, M. A., Christensen, E., Fouts, L., Luecke, J., Chupka, G. M., Yanowitz, J., Tian, M., and Boot, M. (2015). Properties of oxygenates found in upgraded biomass pyrolysis oil as components of spark and compression ignition engine fuels. *Energy & Fuels*, 29(4):2453–2461.

McDonald, C. M. and Floudas, C. A. (1995). Global optimization for the phase stability problem. *AIChE Journal*, 41(7):1798–1814.

McEwen, J. T. and Atsumi, S. (2012). Alternative biofuel production in non-natural hosts. *Current Opinion in Biotechnology*, 23(5):744–750.

McLean, K. A. and McAuley, K. B. (2012). Mathematical modelling of chemical processes - Obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *The Canadian Journal of Chemical Engineering*, 90(2):351–366.

Meusinger, R. and Moros, R. (1999). Determination of quantitative structure–octane rating relationships of hydrocarbons by genetic algorithms. *Chemometrics and Intelligent Laboratory Systems*, 46(1):67–78.

Michalik, C., Stuckert, M., and Marquardt, W. (2009). Optimal experimental design for discriminating numerous model candidates: The AWDC criterion. *Industrial & Engineering Chemistry Research*, 49(2):913–919.

Miller, D. and Thodos, G. (1963). On the reduced Frost-Kalkwarf vapor pressure equation. *Industrial & Engineering Chemistry Fundamentals*, 2(1):78–80.

Milpied, J., Jeuland, N., Plassat, G., Guichaous, S., Dioc, N., Marchal, A., and Schmelzle, P. (2009). Impact of fuel properties on the performances and knock behaviour of a downsized turbocharged DI SI engine - Focus on octane numbers and latent heat of vaporization. *SAE International Journal of Fuels and Lubricants*, 2(1):118–126.

Mitsos, A. and Barton, P. I. (2007). A dual extremum principle in thermodynamics. *AIChE Journal*, 53(8):2131–2147.

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.

Mohs, A., Jakob, A., and Gmehling, J. (2009). Analysis of a concept for predicting missing group interaction parameters of the UNIFAC model using connectivity indices. *AIChE Journal*, 55(6):1614–1625.

Moity, L., Molinier, V., Benazzouz, A., Barone, R., Marion, P., and Aubry, J.-M. (2014). In silico design of bio-based commodity chemicals: Application to itaconic acid based solvents. *Green Chemistry*, 16(1):146–160.

Moity, L., Molinier, V., Benazzouz, A., Joossen, B., Gerbaud, V., and Aubry, J.-M. (2016). A top-down in silico approach for designing ad hoc bio-based solvents: Application to glycerol-derived solvents of nitrocellulose. *Green Chemistry*, 18(11):3239–3249.

Morgan, H. (1965). The generation of a unique machine description for chemical structures - A technique developed at Chemical Abstracts Service. *Journal of Chemical Documentation*, 5(2):107–113.

Mueller, C. J., Cannella, W. J., Bruno, T. J., Bunting, B., Dettman, H. D., Franz, J. A., Huber, M. L., Natarajan, M., Pitz, W. J., Ratcliff, M. A., and Wright, K. (2012). Methodology for formulating diesel surrogate fuels with accurate compositional, ignition-quality, and volatility characteristics. *Energy & Fuels*, 26(6):3284–3303.

Multer, A., McGraw, N., Hohn, K., and Vadlani, P. (2013). Production of methyl ethyl ketone from biomass using a hybrid biochemical/catalytic approach. *Industrial & Engineering Chemistry Research*, 52(1):56–60.

Muradov, N. Z. and Veziroğlu, T. N. (2008). Green path from fossil-based to hydrogen economy: An overview of carbon-neutral technologies. *International Journal of Hydrogen Energy*, 33(23):6804–6839.

Murphy, M. J., Taylor, J. D., and McCormick, R. L. (2004). Compendium of experimental cetane number data. Technical report. National Renewable Energy Laboratory, NREL/SR-540-36805, Golden, Colorado, USA.

Murtaugh, P. A. (2009). Performance of several variable-selection methods applied to real ecological data. *Ecology Letters*, 12(10):1061–1068.

Mustaffa, A. A., Kontogeorgis, G. M., and Gani, R. (2011). Analysis and application of GC^+ models for property prediction of organic chemical systems. *Fluid Phase Equilibria*, 302(1):274–283.

Naik, S., Goud, V. V., Rout, P. K., and Dalai, A. K. (2010). Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 14(2):578–597.

Nakamura, C. E. and Whited, G. M. (2003). Metabolic engineering for the microbial production of 1,3-propanediol. *Current Opinion in Biotechnology*, 14(5):454–459.

Naser, S. and Fournier, R. (1991). A system for the design of an optimum liquid-liquid extractant molecule. *Computers & Chemical Engineering*, 15(6):397–414.

Netzeva, T. I., Worth, A. P., Aldenberg, T., Benigni, R., Cronin, M. T., Gramatica, P., Jaworska, J. S., Kahn, S., Klopman, G., Marchant, C. A., Myatt, G., Nikolova-Juliazkova, N., Patlewicz, G. Y., Perkins, R., Roberts, D. W., Schultz, T. W., Stanton, D. T., van de Sandt, J. J., Tong, W., Veith, G., and Yang, C. (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *ATLA*, 33:1–19.

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.

Nielsen, T. L., Abildskov, J., Harper, P. M., Papaeconomou, I., and Gani, R. (2001). The CAPEC database. *Journal of Chemical & Engineering Data*, 46(5):1041–1044.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, New York, New York, USA.

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(33):1–14.

O'Connor, C., Forrester, R., and Scurrell, M. (1992). Cetane number determination of synthetic diesel fuels. *Fuel*, 71(11):1323–1327.

Oehlschlaeger, M. A., Shen, H.-P. S., Steinberg, J., and Vanderover, J. (2009). The ignition of C7-C16 normal and branched alkanes at elevated pressures. In *Proceedings of the 6th US National Meeting on Combustion, 17.-20.05.2009, Ann Arbor, Michigan, USA*.

Oinuma, H., Miyake, K., Yamanaka, M., Nomoto, K., Katoh, H., Sawada, K., Shino, M., and Hamano, S. (1990). Neural networks applied to structure-activity relationships. *Journal of Medicinal Chemistry*, 33(3):905–908.

Olden, J. D. and Jackson, D. A. (2000). Torturing data for the sake of generality: How valid are our regression models? *Ecoscience*, 7(4):501–510.

Osecky, E. (2013). *The Ignition Quality Tester: An Alternative for Characterizing the Combustion Kinetics of Low Volatility Fuels*. PhD thesis, Colorado School of Mines, Golden, Colorado, USA.

Park, S. H., Suh, H. K., and Lee, C. S. (2009). Effect of bioethanol- biodiesel blending ratio on fuel spray behavior and atomization characteristics. *Energy & Fuels*, 23(8):4092–4098.

Park, S. H., Suh, H. K., and Lee, C. S. (2010). Nozzle flow and atomization characteristics of ethanol blended biodiesel fuel. *Renewable Energy*, 35(1):144–150.

Pepiot-Desjardins, P., Pitsch, H., Malhotra, R., Kirby, S., and Boehman, A. (2008). Structural group analysis for soot reduction tendency of oxygenated fuels. *Combustion and Flame*, 154(1):191–205.

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.

Perez, P. L. and Boehman, A. L. (2012). Experimental investigation of the autoignition behavior of surrogate gasoline fuels in a constant-volume combustion bomb apparatus and its relevance to HCCI combustion. *Energy & Fuels*, 26(10):6106–6117.

Petersen, B., Gernaey, K., and Vanrolleghem, P. A. (2001). Practical identifiability of model parameters by combined respirometric-titrimetric measurements. *Water Science and Technology*, 43(7):347–355.

Petrus, L. and Noordermeer, M. A. (2006). Biomass to biofuels, a chemical perspective. *Green Chemistry*, 8(10):861–867.

Phoon, L. Y., Hashim, H., Mat, R., and Mustaffa, A. A. (2015). Tailor-made green diesel blends design using a decomposition-based computer-aided approach. In *Proceedings of*

the 12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering, 31.05.-04.06.2015, Copenhagen, Denmark, pages 1085–1090.

Pickett, L. M. and Siebers, D. L. (2004). Non-sooting, low flame temperature mixing-controlled DI diesel combustion. Technical report. SAE Technical Paper 2004-01-1399.

Pitz, W. J. and Mueller, C. J. (2011). Recent progress in the development of diesel surrogate fuels. *Progress in Energy and Combustion Science*, 37(3):330–350.

Poling, B. E., Prausnitz, J. M., and O'Connell, J. P. (2001). *The Properties of Gases and Liquids - Fifth Edition*. McGraw-Hill, New York, New York, USA.

Przyjazny, A. and Kokosa, J. M. (2002). Analytical characteristics of the determination of benzene, toluene, ethylbenzene and xylenes in water by headspace solvent microextraction. *Journal of Chromatography A*, 977(2):143–153.

Ptasinski, K. J., Prins, M. J., and Pierik, A. (2007). Exergetic evaluation of biomass gasification. *Energy*, 32(4):568–574.

Qian, Y., Zhu, L., Wang, Y., and Lu, X. (2015). Recent progress in the development of biofuel 2,5-dimethylfuran. *Renewable and Sustainable Energy Reviews*, 41:633–646.

Quaiser, T. and Mönnigmann, M. (2009). Systematic identifiability testing for unambiguous mechanistic modeling - Application to JAK-STAT, MAP kinase, and NF-κb signaling pathway models. *BMC Systems Biology*, 3(50):1–21.

Rabinovitch-Deere, C. A., Oliver, J. W., Rodriguez, G. M., and Atsumi, S. (2013). Synthetic biology and metabolic engineering approaches to produce biofuels. *Chemical Reviews*, 113(7):4611–4632.

Rackemann, D. W. and Doherty, W. O. (2011). The conversion of lignocellulosics to levulinic acid. *Biofuels, Bioproducts and Biorefining*, 5(2):198–214.

Rangarajan, S., Bhan, A., and Daoutidis, P. (2010). Rule-based generation of thermochemical routes to biomass conversion. *Industrial & Engineering Chemistry Research*, 49(21):10459–10470.

Rangarajan, S., Bhan, A., and Daoutidis, P. (2012). Language-oriented rule-based reaction network generation and analysis: Description of RING. *Computers & Chemical Engineering*, 45:114–123.

Rangarajan, S., Bhan, A., and Daoutidis, P. (2014a). Identification and analysis of synthesis routes in complex catalytic reaction networks for biomass upgrading. *Applied Catalysis B: Environmental*, 145:149–160.

Rangarajan, S., Brydon, R. R., Bhan, A., and Daoutidis, P. (2014b). Automated identification of energetically feasible mechanisms of complex reaction networks in heterogeneous catalysis: Application to glycerol conversion on transition metals. *Green Chemistry*, 16(2):813–823.

Rangarajan, S., Kaminski, T., Van Wyk, E., Bhan, A., and Daoutidis, P. (2014c). Language-oriented rule-based reaction network generation and analysis: Algorithms of RING. *Computers & Chemical Engineering*, 64:124–137.

Rasmussen, B. P. (2003). *Prediction Interval Estimation Techniques for Empirical Modeling Strategies and their Applications to Signal Validation Tasks*. PhD thesis, The University of Tennessee, Knoxville, Tennessee, USA.

Reddemann, M., Mathieu, F., Martin, D., and Kneer, R. (2010). The influence of fuel properties on spray propagation, atomization and evaporation. In *Proceedings of the 23rd Annual Conference on Liquid Atomization and Spray Systems, September 2010, Brno, Czech Republic*. ILASS Europe 2010.

Reddemann, M. A., Mathieu, F., Martin, D., and Kneer, R. (2011). Impact of physical properties on primary breakup for a diesel nozzle configuration. *Atomization and Sprays*, 21(3):221–235.

Reynolds, C. H., Druker, R., and Pfahler, L. B. (1998). Lead discovery using stochastic cluster analysis (SCA): A new method for clustering structurally similar compounds. *Journal of Chemical Information and Computer Sciences*, 38(2):305–312.

Rogers, D. and Hopfinger, A. J. (1994). Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *Journal of Chemical Information and Computer Sciences*, 34(4):854–866.

Román-Leshkov, Y., Chheda, J. N., and Dumesic, J. A. (2006). Phase modifiers promote efficient production of hydroxymethylfurfural from fructose. *Science*, 312:1933–1937.

Rosell, M., Lacorte, S., and Barceló, D. (2006). Analysis, occurrence and fate of MTBE in the aquatic environment over the past decade. *TrAC Trends in Analytical Chemistry*, 25(10):1016–1029.

Rothamer, D. A. and Donohue, T. J. (2013). Chemistry and combustion of fit-for-purpose biofuels. *Current Opinion in Chemical Biology*, 17(3):522–528.

Rothamer, D. A. and Jennings, J. H. (2012). Study of the knocking propensity of 2, 5-dimethylfuran-gasoline and ethanol-gasoline blends. *Fuel*, 98:203–212.

Roy, K. (2007). On some aspects of validation of predictive quantitative structure-activity relationship models. *Expert Opinion on Drug Discovery*, 2(12):1567–1577.

Ryan, T. W. and Matheaus, A. C. (2003). Fuel requirements for HCCI engine operation. Technical report. SAE Technical Paper 2003-01-1813.

Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V., and Todeschini, R. (2012). Comparison of different approaches to define the applicability domain of QSAR models. *Molecules*, 17(5):4791–4810.

Saldana, D. A., Creton, B., Mougin, P., Jeuland, N., Rousseau, B., and Starck, L. (2013). Rational formulation of alternative fuels using QSPR methods: Application to jet fuels. *Oil & Gas Science and Technology–Revue dIFP Energies nouvelles*, 68(4):651–662.

Saldana, D. A., Starck, L., Mougin, P., Rousseau, B., Pidol, L., Jeuland, N., and Creton, B. (2011). Flash point and cetane number predictions for fuel compounds using quantitative structure property relationship (QSPR) methods. *Energy & Fuels*, 25(9):3900–3908.

Samudra, A. P. and Sahinidis, N. V. (2013). Optimization-based framework for computer-aided molecular design. *AIChE Journal*, 59(10):3686–3701.

Santana, R. C., Do, P. T., Santikunaporn, M., Alvarez, W. E., Taylor, J. D., Sughrue, E. L., and Resasco, D. E. (2006). Evaluation of different reaction strategies for the improvement of cetane number in diesel fuels. *Fuel*, 85(5):643–656.

Sarathy, S., Westbrook, C., Mehl, M., Pitz, W., Togbe, C., Dagaut, P., Wang, H., Oehlschlaeger, M., Niemann, U., Seshadri, K., Veloo, P., Egolfopoulos, F., and Lu, T. (2011). Comprehensive chemical kinetic modeling of the oxidation of 2-methylalkanes from C7 to C20. *Combustion and Flame*, 158(12):2338–2357.

Sarathy, S. M., Oßwald, P., Hansen, N., and Kohse-Höinghaus, K. (2014). Alcohol combustion chemistry. *Progress in Energy and Combustion Science*, 44:40–102.

Sarathy, S. M., Vranckx, S., Yasunaga, K., Mehl, M., Oßwald, P., Metcalfe, W. K., Westbrook, C. K., Pitz, W. J., Kohse-Höinghaus, K., Fernandes, R. X., and Curran, H. (2012). A comprehensive chemical kinetic combustion model for the four butanol isomers. *Combustion and Flame*, 159(6):2028–2055.

Sauerbrei, W., Royston, P., and Binder, H. (2007). Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Statistics in Medicine*, 26(30):5512–5528.

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.

Scior, T., Medina-Franco, J., Do, Q.-T., Martínez-Mayorga, K., Yunes Rojas, J., and Bernard, P. (2009). How to recognize and workaround pitfalls in QSAR studies: A critical review. *Current Medicinal Chemistry*, 16(32):4297–4313.

Seber, G. and Wild, C. (2003). *Nonlinear Regression*. Wiley, New York, New York, USA.

Segura, H., Polishuk, I., and Wisniak, J. (2000). Phase stability analysis in binary systems. *Physics and Chemistry of Liquids*, 38(3):277–331.

Sen, S. M., Henao, C. A., Braden, D. J., Dumesic, J. A., and Maravelias, C. T. (2012). Catalytic conversion of lignocellulosic biomass to fuels: Process development and technoeconomic evaluation. *Chemical Engineering Science*, 67(1):57–67.

Sennott, T., Gotianun, C., Serres, R., Mack, J., and Dibble, R. (2013a). Cetane number prediction from molecular structure using artificial neural networks. In *Proceedings of the 8th US National Combustion Meeting 2013, Park City, Utah, USA*. Combustion Institute - Western States Section, Paper No. 070CO-0401.

Sennott, T., Gotianun, C., Serres, R., Ziabasharhagh, M., Mack, J., and Dibble, R. (2013b). Artificial neural network for predicting cetane number of biofuel candidates based on molecular structure. In *Proceedings of the ASME 2013 Internal Combustion Engine Division Fall Technical Conference, 13.-16.10.2013, Dearborn, Michigan, USA*. American Society of Mechanical Engineers, Paper No. ICEF2013-19185.

Serrano-Ruiz, J. C. and Dumesic, J. A. (2011). Catalytic routes for the conversion of biomass into liquid hydrocarbon transportation fuels. *Energy & Environmental Science*, 4(1):83–99.

Serrano-Ruiz, J. C. and Dumesic, J. A. (2012). *Catalysis for Alternative Energy Generation*, chapter Catalytic Production of Liquid Hydrocarbon Transportation Fuels, pages 29–56. Springer, New York, New York, USA.

Serrano-Ruiz, J. C., Luque, R., and Sepúlveda-Escribano, A. (2011). Transformations of biomass-derived platform molecules: From high added-value chemicals to fuels via aqueous-phase processing. *Chemical Society Reviews*, 40(11):5266–5281.

Shao, Z., Li, C., Di, X., Xiao, Z., and Liang, C. (2014). Aqueous-phase hydrogenation of succinic acid to γ-butyrolactone and tetrahydrofuran over Pd/C, Re/C, and Pd–Re/C catalysts. *Industrial & Engineering Chemistry Research*, 53(23):9638–9645.

Sheldon, R. A. (2014). Green and sustainable manufacture of chemicals from biomass: State of the art. *Green Chemistry*, 16(3):950–963.

Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., Cho, K. M., and Liao, J. C. (2011). Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Applied and Environmental Microbiology*, 77(9):2905–2915.

Shimizu, K.-i., Uozumi, R., and Satsuma, A. (2009). Enhanced production of hydroxymethylfurfural from fructose with solid acid catalysts by simple water removal methods. *Catalysis Communications*, 10(14):1849–1853.

Shinnar, R. and Citro, F. (2006). A road map to US decarbonization. *Science*, 313:1243–1244.

Siebers, D. L. (1999). Scaling liquid-phase fuel penetration in diesel sprays based on mixing-limited vaporization. Technical report. SAE Technical Paper 1999-01-0528.

Sirignano, M., Conturso, M., and DAnna, A. (2015). Effect of furans on particle formation in diffusion flames: An experimental and modeling study. *Proceedings of the Combustion Institute*, 35(1):525–532.

Smagala, T. G., Christensen, E., Christison, K. M., Mohler, R. E., Gjersing, E., and McCormick, R. L. (2013). Hydrocarbon renewable and synthetic diesel fuel blendstocks: Composition and properties. *Energy & Fuels*, 27(1):237–246.

Smith, B. L. and Bruno, T. J. (2007). Improvements in the measurement of distillation curves. 3. Application to gasoline and gasoline+methanol mixtures. *Industrial & Engineering Chemistry Research*, 46(1):297–309.

Smolenskii, E., Bavykin, V., Ryzhov, A., Slovokhotova, O., Chuvaeva, I., and Lapidus, A. (2008). Cetane numbers of hydrocarbons: Calculations using optimal topological indices. *Russian Chemical Bulletin*, 57(3):461–467.

Song, J. (2004). *Building Robust Chemical Reaction Mechanisms: Next Generation of Automatic Model Construction Software*. PhD thesis, Department of Chemical Engineering, Massachusetts Institute of Technology, Boston, Massachusetts, USA.

Sorenson, S. C. (2001). Dimethyl ether in diesel engines: Progress and perspectives. *Journal of Engineering for Gas Turbines and Power*, 123(3):652–658.

Starck, L., Lecointe, B., Forti, L., and Jeuland, N. (2010). Impact of fuel characteristics on HCCI combustion: Performances and emissions. *Fuel*, 89(10):3069–3077.

Stichlmair, J. and Fair, J. R. (1998). *Distillation: Principles and Practices*. Wiley, New York, New York, USA.

Stickney, M. J., Jones Jr, E. M., and Chandrasekharaiah, M. S. (2005). Fuel blend for an internal combustion engine. US Patent 6,923,839.

Storch, M., Hinrichsen, F., Wensing, M., Will, S., and Zigan, L. (2015). The effect of ethanol blending on mixture formation, combustion and soot emission studied in an optical DISI engine. *Applied Energy*, 156:783–792.

Straathof, A. J. (2014). Transformation of biomass into commodity chemicals using enzymes or cells. *Chemical Reviews*, 114(3):1871–1908.

Sudholt, A., Cai, L., Heyne, J., Haas, F. M., Pitsch, H., and Dryer, F. L. (2015). Ignition characteristics of a bio-derived class of saturated and unsaturated furans for engine applications. *Proceedings of the Combustion Institute*, 35(3):2957–2965.

Sutter, J. M., Dixon, S. L., and Jurs, P. C. (1995). Automated descriptor selection for quantitative structure-activity relationships using generalized simulated annealing. *Journal of Chemical Information and Computer Sciences*, 35(1):77–84.

Tanaka, S., Ayala, F., Keck, J. C., and Heywood, J. B. (2003). Two-stage ignition in HCCI combustion and HCCI control by fuels and additives. *Combustion and Flame*, 132(1):219–239.

Taylor, J., McCormick, R., and Clark, W. (2004). Report on the relationship between molecular structure and compression ignition fuels, both conventional and HCCI. Technical report. National Renewable Energy Laboratory, NREL/MP-540-36726, Golden, Colorado, USA.

The MathWorks Inc. (2016). Matlab R2016b. The MathWorks Inc., Natick, Massachusetts, USA. `www.mathworks.com`.

Thewes, M., Muether, M., Pischinger, S., Budde, M., Brunn, A., Sehr, A., Adomeit, P., and Klankermayer, J. (2011a). 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., Muther, M., Brassat, A., Pischinger, S., and Sehr, A. (2011b). 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.

Todeschini, R. and Consonni, V. (2008). *Handbook of Molecular Descriptors*. Wiley-VCH, Weinheim, Germany.

Todeschini, R., Consonni, V., Mauri, A., and Pavan, M. (2009). dragonX software version 1.4. Talete srl, Milan, Italy. `www.talete.mi.it`.

Topliss, J. G. and Costello, R. J. (1972). Chance correlations in structure-activity studies using multiple regression analysis. *Journal of Medicinal Chemistry*, 15(10):1066–1068.

Topliss, J. G. and Edwards, R. P. (1979). Chance factors in studies of quantitative structure-activity relationships. *Journal of Medicinal Chemistry*, 22(10):1238–1244.

Török, B., Bucsi, I., Beregszászi, T., Kapocsi, I., and Molnár, Á. (1996). Transformation of diols in the presence of heteropoly acids under homogeneous and heterogeneous conditions. *Journal of Molecular Catalysis A: Chemical*, 107(1):305–311.

Török, B., London, G., and Bartók, M. (2000). Reduction of carbonyl compounds to hydrocarbons by catalytic hydrogenation: A novel one-pot method using Pt/K-10 montmorillonite catalyst. *Synlett*, 2000(5):631–632.

Tran, L. S., Sirjean, B., Glaude, P.-A., Fournet, R., and Battin-Leclerc, F. (2012). Progress in detailed kinetic modeling of the combustion of oxygenated components of biofuels. *Energy*, 43(1):4–18.

Tree, D. R. and Svensson, K. I. (2007). Soot processes in compression ignition engines. *Progress in Energy and Combustion Science*, 33(3):272–309.

Tripathi, R., Lee, C., Fernandes, R. X., Olivier, H., Curran, H. J., Sarathy, S. M., and Pitsch, H. (2017). Ignition characteristics of 2-methyltetrahydrofuran: An experimental and kinetic study. *Proceedings of the Combustion Institute*, 36(1):587 – 595.

Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics*, 29(6-7):476–488.

Tropsha, A. and Golbraikh, A. (2007). Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Current Pharmaceutical Design*, 13(34):3494–3504.

Tropsha, A. and Golbraikh, A. (2010). *Handbook of Chemoinformatics Algorithms*, chapter Predictive Quantitative Structure-Activity Modeling. Chapman and Hall/CRC Press, Boca Raton, Florida, USA.

Tropsha, A., Gramatica, P., and Gombar, V. K. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, 22(1):69–77.

Trost, B. M. (1991). The atom economy – A search for synthetic efficiency. *Science*, 254:1471–1477.

Ugray, Z., Lasdon, L., Plummer, J., Glover, F., Kelly, J., and Martí, R. (2007). Scatter search and local NLP solvers: A multistart framework for global optimization. *INFORMS Journal on Computing*, 19(3):328–340.

Ulonska, K., Skiborowski, M., Mitsos, A., and Viell, J. (2016a). Early-stage evaluation of biorefinery processing pathways using process network flux analysis. *AIChE Journal*, 62(9):3096–3108.

Ulonska, K., Voll, A., and Marquardt, W. (2016b). Screening pathways for the production of next generation biofuels. *Energy & Fuels*, 30(1):445–456.

Upadhye, A. A., Qi, W., and Huber, G. W. (2011). Conceptual process design: A systematic method to evaluate and develop renewable energy technologies. *AIChE Journal*, 57(9):2292–2301.

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., and van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics*, 7(1):142.

Van Dien, S. (2013). From the first drop to the first truckload: Commercialization of microbial processes for renewable chemicals. *Current Opinion in Biotechnology*, 24(6):1061–1068.

Van Dongen, D. B., Doherty, M. F., and Haight, J. R. (1983). Material stability of multicomponent mixtures and the multiplicity of solutions to phase-equilibrium equations. 1. Nonreacting mixtures. *Industrial & Engineering Chemistry Fundamentals*, 22(4):472–485.

Van Dyk, B. and Nieuwoudt, I. (2000). Design of solvents for extractive distillation. *Industrial & Engineering Chemistry Research*, 39(5):1423–1429.

van Leeuwen, B. N., van der Wulp, A. M., Duijnstee, I., van Maris, A. J., and Straathof, A. J. (2012). Fermentative production of isobutene. *Applied Microbiology and Biotechnology*, 93(4):1377–1387.

van Putten, R.-J., van der Waal, J. C., de Jong, E., Rasrendra, C. B., Heeres, H. J., and de Vries, J. G. (2013). Hydroxymethylfurfural, a versatile platform chemical made from renewable resources. *Chemical Reviews*, 113(3):1499–1597.

Vancoillie, J., Sileghem, L., and Verhelst, S. (2014). Development and validation of a quasi-dimensional model for methanol and ethanol fueled SI engines. *Applied Energy*, 132:412–425.

Vancoillie, J., Verhelst, S., and Demuynck, J. (2011). Laminar burning velocity correlations for methanol-air and ethanol-air mixtures valid at SI engine conditions. Technical report. SAE Technical Paper 2011-01-0846.

Vandersickel, A., Hartmann, M., Vogel, K., Wright, Y., Fikri, M., Starke, R., Schulz, C., and Boulouchos, K. (2012). The autoignition of practical fuels at HCCI conditions: High-pressure shock tube experiments and phenomenological modeling. *Fuel*, 93:492–501.

Vemuri, G., Eiteman, M., and Altman, E. (2002). Succinate production in dual-phase *Escherichia coli* fermentations depends on the time of transition from aerobic to anaerobic conditions. *Journal of Industrial Microbiology and Biotechnology*, 28(6):325–332.

Verhelst, S. and Sheppard, C. (2009). Multi-zone thermodynamic modelling of spark-ignition engine combustion - An overview. *Energy Conversion and Management*, 50(5):1326–1335.

Vickers, C. E., Klein-Marcuschamer, D., and Krömer, J. O. (2012). Examining the feasibility of bulk commodity production in *Escherichia coli*. *Biotechnology Letters*, 34(4):585–596.

Victoria Villeda, J. J. (2017). *Reaction Network Generation and Evaluation for the Design of Biofuel Value Chains*. PhD thesis, RWTH Aachen University, AVT - Process Systems Engineering, Fortschritt-Berichte VDI, Reihe 3 (Verfahrenstechnik), Nr. 950, VDI-Verlag, Duesseldorf, Germany.

Victoria Villeda, J. J., Dahmen, M., Hechinger, M., Voll, A., and Marquardt, W. (2012a). Towards model-based design of biofuel value chains. *Current Opinion in Chemical Engineering*, 1(4):465–471.

Victoria Villeda, J. J., Dahmen, M., Hechinger, M., Voll, A., and Marquardt, W. (2012b). Towards model-based identification of next generation biofuels - Aspects of product and process design. In *244th ACS National Meeting & Exposition, 19.-23.08.2012, Philadelphia, Pennsylvania, United States*.

Victoria Villeda, J. J., Dahmen, M., Hechinger, M., Voll, A., and Marquardt, W. (2015). Towards model-based design of tailor-made fuels from biomass. In Klaas, M., Pischinger, S., and Schröder, W., editors, *Fuels From Biomass: An Interdisciplinary Approach*, pages 193–211. Springer, Heidelberg, Germany.

Victoria Villeda, J. J., Kremer, F., Dahmen, M., Hechinger, M., Pischinger, S., and Marquardt, W. (2012c). Fuel requirement definition and model-based biofuel identification within the fuel design process of TMFB. In *Book of Extended Abstracts of the 5th TMFB International Workshop of the Cluster of Excellence "Tailor-Made Fuels from Biomass", 13.-14.06.2012, Aachen, Germany*.

Villalba, H. E. G. (2009). *Development of Group Contribution Plus Models for Properties of Organic Chemical Systems*. PhD thesis, Computer Aided Process-Product Engineering Center, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Lyngby, Denmark.

Visco, D. P., Pophale, R. S., Rintoul, M. D., and Faulon, J.-L. (2002). Developing a methodology for an inverse quantitative structure-activity relationship using the signature molecular descriptor. *Journal of Molecular Graphics and Modelling*, 20(6):429–438.

Voll, A. (2014). *Model-based Screening of Reaction Pathways for Processing of Biorenewables*. PhD thesis, Aachener Verfahrenstechnik - Process Systems Engineering, RWTH Aachen University, Shaker Verlag GmbH, Aachen, Germany.

Voll, A. and Marquardt, W. (2012a). Benchmarking of next-generation biofuels from a process perspective. *Biofuels, Bioproducts and Biorefining*, 6(3):292–301.

Voll, A. and Marquardt, W. (2012b). Reaction network flux analysis: Optimization-based evaluation of reaction pathways for biorenewables processing. *AIChE Journal*, 58(6):1788–1801.

Wackett, L. P. (2011). Engineering microbes to produce biofuels. *Current Opinion in Biotechnology*, 22(3):388–393.

Walker, J. D., Jaworska, J., Comber, M. H., Schultz, T. W., and Dearden, J. C. (2003). Guidelines for developing and using quantitative structure-activity relationships. *Environmental Toxicology and Chemistry*, 22(8):1653–1665.

Walter, E. and Pronzato, L. (1997). *Identification of Parametric Models*. Springer, New York, New York, USA.

Wang, C., Xu, H., Daniel, R., Ghafourian, A., Herreros, J. M., Shuai, S., and Ma, X. (2013). Combustion characteristics and emissions of 2-methylfuran compared to 2,5-dimethylfuran, gasoline and ethanol in a DISI engine. *Fuel*, 103:200–211.

Wang, H., Zheng, Z., Yao, M., and Reitz, R. D. (2014). An experimental and numerical study on the effects of fuel properties on the combustion and emissions of low-temperature combustion diesel engines. *Combustion Science and Technology*, 186(12):1795–1815.

Wang, X., Huang, Z., Kuti, O. A., Zhang, W., and Nishida, K. (2011). An experimental investigation on spray, ignition and combustion characteristics of biodiesels. *Proceedings of the Combustion Institute*, 33(2):2071–2077.

Wasylkiewicz, S. K., Sridhar, L. N., Doherty, M. F., and Malone, M. F. (1996). Global stability analysis and calculation of liquid-liquid equilibrium in multicomponent mixtures. *Industrial & Engineering Chemistry Research*, 35(4):1395–1408.

Watkins, D. S. (2002). *Fundamentals of Matrix Computations - Second Edition*. John Wiley and Sons, New York, New York, USA.

Weaver, S. and Gleeson, M. P. (2008). The importance of the domain of applicability in QSAR modeling. *Journal of Molecular Graphics and Modelling*, 26(8):1315–1326.

Weidlich, U. and Gmehling, J. (1987). A modified UNIFAC model. I: Prediction of VLE, h^E, and γ^∞. *Industrial & Engineering Chemistry Research*, 26(7):1372–1381.

Weinebeck, A., Hoppe, F., Dahmen, M., Hechinger, M., Mottweiler, J., Kremer, F., Murrenhoff, H., Marquardt, W., Bolm, C., and Pischinger, S. (2014). Biomass-based ketones as successful example for the fuel design process. In *Book of Extended Abstracts of the 2nd TMFB International Conference, 16.-18.06.2014, Aachen, Germany*.

Weinebeck, A. and Murrenhoff, H. (2013). Lubricity of new tailor-made fuels from biomass. In *Proceedings of the 13th Scandinavian International Conference on Fluid Power SICFP2013, 03.-05.06.2013, Linkoeping, Sweden*, pages 345–354. Linkoeping University Electronic Press.

Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36.

Weininger, D., Weininger, A., and Weininger, J. L. (1989). SMILES. 2. Algorithm for generation of unique SMILES notation. *Journal of Chemical Information and Computer Sciences*, 29(2):97–101.

Werpy, T., Petersen, G., Aden, A., Bozell, J., Holladay, J., White, J., Manheim, A., Eliot, D., Lasure, L., Jones, S., Gerber, M., Ibsen, K., Lumberg, L., and Kelley, S. (2004). Top value added chemicals from biomass - Volume 1 - Results of screening for potential candidates from sugars and synthesis gas. Technical report. Pacific Northwest National Laboratory (PNNL), National Renewable Energy Laboratory (NREL), Office of Biomass Program (EERE), U.S. Department of Energy.

Westbrook, C., Pitz, W., Mehl, M., and Curran, H. (2011). Detailed chemical kinetic reaction mechanisms for primary reference fuels for diesel cetane number and spark-ignition octane number. *Proceedings of the Combustion Institute*, 33(1):185–192.

Westbrook, C. K. (2000). Chemical kinetics of hydrocarbon ignition in practical combustion systems. *Proceedings of the Combustion Institute*, 28(2):1563–1577.

Westbrook, C. K. (2013). Biofuels combustion. *Annual Review of Physical Chemistry*, 64:201–219.

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.

Whittingham, M. J., Stephens, P. A., Bradbury, R. B., and Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, 75(5):1182–1189.

Wikel, J. H. and Dow, E. R. (1993). The use of neural networks for variable selection in QSAR. *Bioorganic & Medicinal Chemistry Letters*, 3(4):645–651.

Wikström, C., Albano, C., Eriksson, L., Fridén, H., Johansson, E., Nordahl, Å., Rännar, S., Sandberg, M., Kettaneh-Wold, N., and Wold, S. (1998). Multivariate process and quality monitoring applied to an electrolysis process: Part I. Process supervision with multivariate control charts. *Chemometrics and Intelligent Laboratory Systems*, 42(1):221–231.

Wojcik, B. (1948). Catalytic hydrogenation of furan compounds. *Industrial & Engineering Chemistry*, 40(2):210–216.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1):37–52.

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130.

Won, S. H., Dooley, S., Veloo, P. S., Wang, H., Oehlschlaeger, M. A., Dryer, F. L., and Ju, Y. (2014). The combustion properties of 2,6,10-trimethyl dodecane and a chemical functional group analysis. *Combustion and Flame*, 161(3):826–834.

Wright, M. E. (2015). Process for the dehydration of aqueous bio-derived terminal alcohols to terminal alkenes. US Patent Application Publication No. US 2015/0025288 A1.

Wright, W. R. and Palkovits, R. (2012). Development of heterogeneous catalysts for the conversion of levulinic acid to γ-valerolactone. *ChemSusChem*, 5(9):1657–1667.

Wyman, C. E., Dale, B. E., Elander, R. T., Holtzapple, M., Ladisch, M. R., and Lee, Y. (2005). Coordinated development of leading biomass pretreatment technologies. *Bioresource Technology*, 96(18):1959–1966.

Yamada, T. and Gunn, R. D. (1973). Saturated liquid molar volumes. Rackett equation. *Journal of Chemical and Engineering Data*, 18(2):234–236.

Yang, F., Liu, Q., Yue, M., Bai, X., and Du, Y. (2011). Tantalum compounds as heterogeneous catalysts for saccharide dehydration to 5-hydroxymethylfurfural. *Chemical Communications*, 47(15):4469–4471.

Yang, H., Fairbridge, C., and Ring, Z. (2001). Neural network prediction of cetane numbers for isoparaffins and diesel fuel. *Petroleum Science and Technology*, 19(5-6):573–586.

Yang, H., Ring, Z., Briker, Y., McLean, N., Friesen, W., and Fairbridge, C. (2002). Neural network prediction of cetane number and density of diesel fuel from its chemical composition determined by LC and GC–MS. *Fuel*, 81(1):65–74.

Yang, Y., Boehman, A. L., and Santoro, R. J. (2007). A study of jet fuel sooting tendency using the threshold sooting index (TSI) model. *Combustion and Flame*, 149(1):191–205.

Yang, Y. and Dec, J. E. (2013). Bio-ketones: Autoignition characteristics and their potential as fuels for HCCI engines. *SAE International Journal of Fuels and Lubricants*, 6(3):713–728.

Yang, Y., Dec, J. E., Dronniou, N., and Simmons, B. (2010). Characteristics of isopentanol as a fuel for HCCI engines. *SAE International Journal of Fuels and Lubricants*, 3(2):725–741.

Yanowitz, J., Christensen, E., and McCormick, R. L. (2011). Utilization of renewable oxygenates as gasoline blending components. Technical report. National Renewable Energy Laboratory, Technical Report, NREL/TP-5400-50791, Golden, Colorado, USA.

Yanowitz, J., Ratcliff, M., McCormick, R., Taylor, J., and Murphy, M. (2014). Compendium of experimental cetane numbers. Technical report. National Renewable Energy Laboratory, NREL/TP-5400-61693, Golden, Colorado, USA.

Yao, M., Wang, H., Zheng, Z., and Yue, Y. (2010). Experimental study of n-butanol additive and multi-injection on HD diesel engine performance and emissions. *Fuel*, 89(9):2191–2201.

Yao, M., Zheng, Z., and Liu, H. (2009). Progress and recent trends in homogeneous charge compression ignition (HCCI) engines. *Progress in Energy and Combustion Science*, 35(5):398–437.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H., Andrae, S., Yang, T., Lee, S. Y., Burk, M., and Van Dien, S. (2011). Metabolic engineering of *Escherichia coli* for direct production of 1, 4-butanediol. *Nature Chemical Biology*, 7(7):445–452.

Yoneda, H., Tantillo, D. J., and Atsumi, S. (2014). Biological production of 2-butanone in *Escherichia coli*. *ChemSusChem*, 7(1):92–95.

You, C., Zhang, C., Chen, L., and Qi, Z. (2015). Highly dispersed palladium nanoclusters incorporated in amino-functionalized silica spheres for the selective hydrogenation of succinic acid to γ-butyrolactone. *Applied Organometallic Chemistry*, 29(10):653–660.

Yousefinejad, S. and Hemmateenejad, B. (2015). Chemometrics tools in QSAR/QSPR studies: A historical perspective. *Chemometrics and Intelligent Laboratory Systems*, 149:177–204.

Yunus, N. A., Gernaey, K. V., Manan, Z. A., Woodley, J. M., and Gani, R. (2011). Design of tailor-made chemical blend using a decomposition-based computer-aided approach. In *Proceedings of the Fourth International Conference on Modeling, Simulation and Applied Optimization (ICMSAO), 19.-21.04.2011, Kuala Lumpur, Malaysia*. IEEE.

Yunus, N. A., Gernaey, K. V., Woodley, J. M., and Gani, R. (2012). An integrated methodology for design of tailor-made blended products. In *Proceedings of the 22nd European Symposium on Computer Aided Process Engineering 17.-20.06.2012, London, United Kingdom*, pages 752–756.

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.

Yunus, N. A. B. (2014). *Systematic Methodology for Design of Tailor-Made Blended Products: Fuels and Other Blended Products*. PhD thesis, Computer Aided Process-Product Engineering Center, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Lyngby, Denmark.

Yushan, Z. and Zhihong, X. (1999). A reliable method for liquid-liquid phase equilibrium calculation and global stability analysis. *Chemical Engineering Communications*, 176(1):133–160.

Zalfen, K. (2014). Computer-Aided Design of Fuel Blends Derived from the Selective Transformation of Biomass Monomers. *Master's thesis*. Aachener Verfahrenstechnik - Process Systems Engineering (Systemverfahrenstechnik), RWTH Aachen University, Aachen, Germany.

Zelle, R. M., de Hulster, E., van Winden, W. A., de Waard, P., Dijkema, C., Winkler, A. A., Geertman, J.-M. A., van Dijken, J. P., Pronk, J. T., and van Maris, A. J. (2008). Malic acid production by *Saccharomyces cerevisiae*: Engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. *Applied and Environmental Microbiology*, 74(9):2766–2777.

Zhang, G.-S., Zhu, M.-M., Zhang, Q., Liu, Y.-M., He, H.-Y., and Cao, Y. (2016). Towards quantitative and scalable transformation of furfural to cyclopentanone with supported gold catalysts. *Green Chemistry*, 18(7):2155–2164.

Zhang, L., Cignitti, S., and Gani, R. (2015a). Generic mathematical programming formulation and solution for computer-aided molecular design. *Computers & Chemical Engineering*, 78:79–84.

Zhang, M., Chen, L., Jiang, Z., and Ma, J. (2015b). Effects of dehydration rate on the yield of ethyl lactate in a pervaporation-assisted esterification process. *Industrial & Engineering Chemistry Research*, 54(26):6669–6676.

Zhang, W., Yu, D., Ji, X., and Huang, H. (2012). Efficient dehydration of bio-based 2,3-butanediol to butanone over boric acid modified HZSM-5 zeolites. *Green Chemistry*, 14(12):3441–3450.

Zhao, X., Cheng, K., and Liu, D. (2009). Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis. *Applied Microbiology and Biotechnology*, 82(5):815–827.

Zhong, S., Daniel, R., Xu, H., Zhang, J., Turner, D., Wyszynski, M. L., and Richards, P. (2010). Combustion and emissions of 2,5-dimethylfuran in a direct-injection spark-ignition engine. *Energy & Fuels*, 24(5):2891–2899.

Zhu, S., Zhu, Y., Hao, S., Zheng, H., Mo, T., and Li, Y. (2012). One-step hydrogenolysis of glycerol to biopropanols over Pt–H4SiW12O40/ZrO2 catalysts. *Green Chemistry*, 14(9):2607–2616.

Online-Buchshop für Ingenieure

VDI nachrichten **BUCHSHOP**

Online-Shops

Fachliteratur und mehr - jetzt bequem online recherchieren & bestellen unter: www.vdi-nachrichten.com/Der-Shop-im-Ueberblick

Täglich aktualisiert: Neuerscheinungen VDI-Schriftenreihen

Im Buchshop von vdi-nachrichten.com finden Ingenieure und Techniker ein speziell auf sie zugeschnittenes, umfassendes Literaturangebot.

Mit der komfortablen Schnellsuche werden Sie in den VDI-Schriftenreihen und im Verzeichnis lieferbarer Bücher unter 1.000.000 Titeln garantiert fündig.

Im Buchshop stehen für Sie bereit:

VDI-Berichte und die Reihe **Kunststofftechnik:**

Berichte nationaler und internationaler technischer Fachtagungen der VDI-Fachgliederungen

Fortschritt-Berichte VDI:

Dissertationen, Habilitationen und Forschungsberichte aus sämtlichen ingenieurwissenschaftlichen Fachrichtungen

Newsletter „Neuerscheinungen":

Kostenfreie Infos zu aktuellen Titeln der VDI-Schriftenreihen bequem per E-Mail

Autoren-Service:

Umfassende Betreuung bei der Veröffentlichung Ihrer Arbeit in der Reihe Fortschritt-Berichte VDI

Buch- und Medien-Service:

Beschaffung aller am Markt verfügbaren Zeitschriften, Zeitungen, Fortsetzungsreihen, Handbücher, Technische Regelwerke, elektronische Medien und vieles mehr – einzeln oder im Abo und mit weltweitem Lieferservice

Die Reihen der Fortschritt-Berichte VDI:

1 Konstruktionstechnik/Maschinenelemente
2 Fertigungstechnik
3 Verfahrenstechnik
4 Bauingenieurwesen
5 Grund- und Werkstoffe/Kunststoffe
6 Energietechnik
7 Strömungstechnik
8 Mess-, Steuerungs- und Regelungstechnik
9 Elektronik/Mikro- und Nanotechnik
10 Informatik/Kommunikation
11 Schwingungstechnik
12 Verkehrstechnik/Fahrzeugtechnik
13 Fördertechnik/Logistik
14 Landtechnik/Lebensmitteltechnik
15 Umwelttechnik
16 Technik und Wirtschaft
17 Biotechnik/Medizintechnik
18 Mechanik/Bruchmechanik
19 Wärmetechnik/Kältetechnik
20 Rechnerunterstützte Verfahren (CAD, CAM, CAE CAQ, CIM ...)
21 Elektrotechnik
22 Mensch-Maschine-Systeme
23 Technische Gebäudeausrüstung

ISBN 978-3-18-395403-2