

VT-Forschungsbericht 2024-02

Data-based Methods for the Screening and Design of Jet Fuels

Clemens Alexander Hall, M.Sc.

Deutsches Zentrum für Luft- und Raumfahrt
Institut für Verbrennungstechnik
Stuttgart



DLR

Deutsches Zentrum
für Luft- und Raumfahrt



Herausgeber

Deutsches Zentrum
für Luft- und Raumfahrt

**Institut für
Verbrennungstechnik**

Pfaffenwaldring 38-40
70569 Stuttgart

Telefon
Telefax

(0 7 11) 68 62 - 3 08
(0 7 11) 68 62 - 5 78

Als Manuskript gedruckt.
Abdruck oder sonstige Verwendung
nur nach Absprache mit dem Institut gestattet

D93, Stuttgart

Data-based Methods for the Screening and Design of Jet Fuels

A thesis accepted by the Faculty of
Aerospace Engineering and Geodesy
of the University of Stuttgart
in partial fulfillment of the requirements for the degree of
Doctor of Engineering Sciences (Dr.-Ing.)

by

Clemens Alexander Hall, M. Sc.

born in Villingen-Schwenningen

Main referee: Prof. Dr.-Ing. Manfred Aigner

Co-referee: Prof. Alexander Mitsos PhD

Date of defense: 23.05.2024

Institute of Combustion Technology for Aerospace Engineering
University of Stuttgart

2024

Danksagung

Die vorliegende Arbeit entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Verbrennungstechnik des Deutschen Zentrums für Luft- und Raumfahrt e.V. (DLR) in Stuttgart. Mein erster Dank gilt daher meinem ehemaligen Institutsleiter und Hauptberichter dieser Arbeit, Herrn Prof. Dr.-Ing. Manfred Aigner für die Möglichkeit zur Durchführung dieser Arbeit und die exzellenten Arbeitsbedingungen am Institut. Herrn Prof. Alexander Mitos PhD danke ich für die Übernahme des Mitberichts sowie für das große Interesse an meiner Arbeit.

Ein besonderer Dank gebührt meinen Betreuern Dr.-Ing. Bastian Rauch und Dr.-Ing. Uwe Bauder für ihre fachliche Unterstützung und die intensiven, stets produktiven Diskussionen, die maßgeblich zu den vorliegenden Ergebnissen beigetragen haben. Ebenso danke ich Dr. Patrick Le Clercq als Abteilungsleiter für seine Unterstützung und die gewährten Freiräume.

Die hervorragende Zusammenarbeit der Abteilungen Mehrphasenströmungen und Alternative Treibstoffe (MAT) sowie Chemische Kinetik und Analytik (CKA) im Rahmen meines Promotionsvorhabens möchte ich ausdrücklich hervorheben. Mein Dank gilt hier insbesondere den Abteilungsleitern Markus Köhler und erneut Dr. Patrick Le Clercq. Den Kolleginnen und Kollegen der Abteilung MAT danke ich für die großartige Arbeitsatmosphäre und die vielen gemeinsamen Erlebnisse, die über den Arbeitsalltag hinausgingen.

Allen Mitarbeiterinnen und Mitarbeitern des Instituts bin ich sehr dankbar für die intensive fachliche Unterstützung und die kollegiale Zusammenarbeit. Exemplarisch möchte ich hier Stephan Ruoff, Florian Pütz und Georg Eckel nennen.

Abschließend danke ich von ganzem Herzen meiner Familie und meiner Freundin für ihren Rückhalt und die stetige Unterstützung, insbesondere in der Schlussphase dieser Arbeit.

Stuttgart, Juni 2024

Contents

Contents.....	i
List of Figures.....	iv
List of Tables.....	ix
Nomenclature.....	xi
Abstract.....	xv
Kurzfassung.....	xvii
1 Introduction.....	1
1.1 Motivation.....	1
1.2 Sustainable Aviation Fuel development and approval.....	4
1.2.1 Jet fuel specifications and synthetic aviation fuel approval process.....	4
1.2.2 Jet fuel prescreening.....	7
1.2.3 Challenges for the prescreening concept implementation.....	9
1.3 Objectives and research questions.....	13
1.4 Chapter outline.....	14
2 Fuel Property Modeling.....	15
2.1 Principles of data-based modeling methods.....	18
2.2 Fuel property modeling methods.....	20
2.2.1 QSPR with sampling method.....	20
2.2.2 Direct Correlation method.....	24
2.2.3 Mean Quantitative Structure-Property Relationship method.....	25
2.3 Probabilistic Machine Learning correlation models.....	26
2.3.1 Working principles of Artificial Neural Networks.....	27
2.3.2 Working principle of Monte-Carlo Dropout Neural Networks.....	29
2.4 Model development and validation.....	30
2.4.1 Training and validation.....	31
2.4.2 Hyperparameter optimization.....	31
2.5 Predictive capability assessment methods for models.....	34
2.5.1 Predictive capability metrics.....	35
2.5.2 Example for the predictive capability assessment.....	38
3 Composition and Property Database.....	40

3.1	Data collection	40
3.2	Data preprocessing and outlier detection.....	43
3.3	Data characterization.....	44
3.3.1	Data characterization of fuels	44
3.3.2	Data characterization of pure compounds.....	49
4	Predictive Capability Assessment of Models and Adequacy Assessment for Fuel Screening	53
4.1	Part 1: Predictive capability assessment of models	54
4.1.1	Density.....	55
4.1.2	Surface tension.....	57
4.1.3	Net heat of combustion.....	58
4.1.4	Kinematic viscosity.....	60
4.1.5	Flash point.....	62
4.1.6	Freezing point	64
4.1.7	Cetane number.....	65
4.1.8	Distillation line	67
4.2	Part 2: Adequacy assessment of models for fuel screening.....	70
4.3	Summary and conclusion	76
5	Development of Fuel Design Tools.....	79
5.1	Molecular descriptors	81
5.2	Property correlation metrics	83
5.3	Investigation of fuel component structure-property relations	84
5.3.1	Density.....	85
5.3.2	Surface tension.....	90
5.3.3	Kinematic viscosity.....	93
5.3.4	Net heat of combustion.....	97
5.3.5	Flash point.....	100
5.3.6	Freezing point	103
5.3.7	Cetane number.....	106
5.3.8	Boiling point	109
5.3.9	Yield sooting index	112
5.4	Summary table of the structure-property correlations.....	115
5.4.1	Usage example of the structure-property correlations.....	117

5.5	Summary and conclusion	119
6	Fuel Design and Blending Analysis	121
6.1	Prescreening of the untreated jet fuel candidate.....	121
6.2	Fuel design and prescreening of fuel variants	125
6.3	Blending study of fuel variant	132
6.3.1	Case 1: blending study without uncertainty consideration	134
6.3.2	Case 2: blending study with uncertainty consideration	136
6.4	Summary and conclusion	139
7	Summary and Outlook	141
7.1	Summary.....	141
7.2	Outlook	144
	References	146
A.	Descriptions of Jet Fuel Screening Properties.....	159
B.	Approved and Pending Jet Fuel Production Routes.....	162
C.	Utilized Structural Molecular Features	166
D.	Fuel Database Schema	169
E.	Cross-validation Results of Models for Training and Testing	170
F.	Reference Models.....	184
G.	Pure Compound Descriptor Plots	189

List of Figures

Figure 1.1: Contribution of measures for reducing prognosed international aviation net CO ₂ emissions [6].	1
Figure 1.2: Restriction of possible fuel composition after ASTM D1655 for conventional crude oil-based jet fuels [29].	5
Figure 1.3: Flow diagram of the approval process of a new aviation turbine fuel after ASTM D4054 [24].	6
Figure 1.4: Schematic illustration of the screening plots for Fuel A and Fuel B as part of the jet fuel prescreening.	9
Figure 1.5: Plot GCxGC measurement of conventional Jet A-1 fuel (upper left), FT-SPK (upper right), HEFA-SPK (middle left), ATJ-SPK (middle right) IH ² fuel (lower left).	10
Figure 1.6 GCxGC measurement signal of Jet-A fuel [42].	11
Figure 2.1: Family tree of approaches for the modeling of fuels.	15
Figure 2.2: Spectrum of modeling approaches from physical models and empirical models to data-based models.	18
Figure 2.3: Quantified molecular features of 2,3-hydro-2-methyl-1h-idene, number behind SMART key shows count of molecular feature[76].	22
Figure 2.4: Schematic illustration of QSPR sampling modeling method from quantification of the molecular structure of components to the property estimation of the fuel.	22
Figure 2.5: Schematic illustration of direct correlation modeling method.	24
Figure 2.6: Schematic illustration of pseudo mean quantitative structure estimation of a jet fuel using GCxGC measurement and a mean occurrence matrix calculated from averaging structural features of possible isomers.	26
Figure 2.7: Schematic figures of artificial neural network neuron (left), connected artificial neural network (right).	27
Figure 2.8 Schematic representation of MCNN with dropout functionality during prediction. Network neurons are deactivated randomly (gray) to generate a distribution of prediction values.	30
Figure 2.9: Schematic illustration of a cross-validation.	31
Figure 2.10: Schematic workflow of the utilized hyperparameter optimization with cross-validation.	32
Figure 2.11: Schematic representation of application and validation domain of a model, adapted from [57].	35
Figure 2.12: Schematic illustration of predictive capability aspects accuracy, validity and precision for probabilistic models.	35
Figure 2.13: Schematic predicted distribution of a probabilistic model with mean prediction y and lower and upper prediction intervals y_l , y_u for a confidence level of 95 % with the associated risk and certainty.	36
Figure 2.14: Schematic illustration of unity plot for validation of probabilistic model.	38

Figure 2.15: Schematic illustration of the screening plots for Fuel A and Fuel B as part of the jet fuel prescreening.	39
Figure 3.1: Number of measurements vs temperature for density (upper left) kinematic viscosity (upper right) and surface tension (lower left).	42
Figure 3.2: Evaluated GCxGC measurement of a conventional jet fuel with representative molecules for each family.	45
Figure 3.3: Overview of current jet fuel production processes, extended from Blakey et al. 2011 [101].	47
Figure 3.4: Scatter plot of GCxGC measurements with summed hydrocarbon families. Blue: conventional fuels. Green: synthetic fuels and blends. Blue and green shaded areas indicate the observed range.	48
Figure 3.5: Scatter plot visualizing the compositional similarity of fuels and fuel components based on the dimensional reduced representation of their respective mean quantitative structure and quantitative structure representation.	51
Figure 4.1: Validation results of the density prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.	56
Figure 4.2: Validation results of the surface tension in air prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.	57
Figure 4.3: Validation results of the net heat prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.	59
Figure 4.4: Validation results of the kinematic viscosity prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.	60
Figure 4.5: Validation results of the flash point prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.	63
Figure 4.6: Validation results of the freezing point prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.	64
Figure 4.7: Validation results of the cetane number prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.	66
Figure 4.8: Validation results of the distillation with 10 vol% evaporated volume prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.	68
Figure 4.9: Validation results of the distillation with 50 vol% evaporated volume prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.	68
Figure 4.10: Validation results of the distillation with 90 vol% evaporated volume prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.	68
Figure 4.11: Composition plots of jet fuels used for Tier α prescreening: conventional oil-based fuel (upper left), SAF produced by the Fischer-Tropsch process (upper right) and SAF produced by Alcohol-to-Jet process (lower left).	71
Figure 4.12: Results of Tier α prescreening: Jet A-1 fuel (left row), FT SPK (middle) and ATJ SPK (right). DC model predictions are displayed in blue, of the M-QSPR model in green and the QSPR sampling model in purple.	72
Figure 4.13: Predicted mean values for iso-alkanes isomers with a carbon number of 12 and 16 for the flash point (left), freezing point (middle) and cetane number (right).	74

Figure 4.14: Results of Tier α prescreening for ATJ SPK with constrained isomer selection, predictions of the QSPR sampling model.	75
Figure 5.1: Molecular structure of isopentane (left) and its calculated adjacency matrix (right).	82
Figure 5.2: Density values at 15 °C of the hydrocarbon families over the carbon number nC. .	85
Figure 5.3: Density values at 15 °C of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b) and partial positive surface area PPSA (c). ...	87
Figure 5.4: Density values at 15 °C of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b) and number of ring atoms nR (c).....	87
Figure 5.5: Density values at 15 °C of mono-aromatics over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b) and partial positive surface area PPSA (c).	87
Figure 5.6: Surface tension values at 22 °C of the hydrocarbon families over the carbon number nC.	90
Figure 5.7: Surface tension values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	92
Figure 5.8: Surface tension values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	92
Figure 5.9: Surface tension values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	92
Figure 5.10: Kinematic viscosity values at 0 °C of the hydrocarbon families over the carbon number nC.	94
Figure 5.11: Kinematic viscosity values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b) and partial positive surface area PPSA (c). ...	96
Figure 5.12: Kinematic viscosity values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b) and partial positive surface area PPSA (c). ...	96
Figure 5.13: Kinematic viscosity values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b) and partial positive surface area PPSA (c).	96
Figure 5.14: Net heat of combustion values of the hydrocarbon families over the carbon number nC.	97
Figure 5.15: Net heat of combustion values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	99
Figure 5.16: Net heat of combustion values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b) and number of ring atoms nR (c).....	99
Figure 5.17: Net heat of combustion values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	99
Figure 5.18: Flash point values of the hydrocarbon families over the carbon number nC.....	100
Figure 5.19: Flash point values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	102
Figure 5.20: Flash point values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	102
Figure 5.21: Flash point values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	102

Figure 5.22: Freezing point values of the hydrocarbon families over the carbon number nC..	103
Figure 5.23: Freezing point values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	105
Figure 5.24: Freezing point values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	105
Figure 5.25: Freezing point values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	105
Figure 5.26: Cetane number values of the hydrocarbon families over the carbon number nC.	106
Figure 5.27: Cetane number values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	108
Figure 5.28: Cetane number values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	108
Figure 5.29: Cetane number values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	108
Figure 5.30: Boiling point values of the hydrocarbon families over the carbon number nC.	109
Figure 5.31: Boiling point values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	111
Figure 5.32: Boiling point values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b) and number of ring atoms nR (c).	111
Figure 5.33: Boiling point values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	111
Figure 5.34: Yield sooting index values of the hydrocarbon families over the carbon number nC.	112
Figure 5.35: Yield sooting index values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	114
Figure 5.36: Yield sooting index values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	114
Figure 5.37: Yield sooting index values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index $\eta\mathbf{B}$ (b).	115
Figure 6.1: Plot of composition of the untreated jet fuel candidate.	121
Figure 6.2: Results of the jet Tier α jet fuel prescreening for the untreated jet fuel candidate.	124
Figure 6.3: Conversion curves for hydrocracking of n-tetradecane, n-pentadecane, n-heptadecane and n-octadecane after Weitkamp [143] and Coonradt und Garwood [144].	126
Figure 6.4: Composition plots of the untreated fuel candidate (upper left), hydroisomerized variant (upper right), hydrocracked fuel variant (lower left) and hydrocracked and distilled variant (lower right).	127
Figure 6.5: Results of the jet Tier α jet fuel prescreening for the three fuel variants: hydroisomerized (left), hydrocracked (middle) hydrocracked and distilled (right).	129
Figure 6.6: Results of deterministic blending analysis, left: property plot with property ranges of fuel variant and blends, right: maximum blending fraction (top) of fuel variant, reduction of yield sooting index of blends (middle), CO ₂ reduction of blends (bottom).	135

Figure 6.7: Results of blending analysis under consideration of uncertainties, left: parallel line property plot with property ranges of fuel variant and blends, right: maximum blending fraction (top) of fuel variant, reduction of yield sooting index of blends (middle), CO ₂ reduction of blends (bottom).	137
Figure 6.8: Number of properties preventing blends for the blending study under consideration of uncertainties.....	138

List of Tables

Table 1: Critical jet fuel properties for jet fuel prescreening.....	8
Table 2: Utilized parameters for hyperparameter optimization of Monte-Carlo Neural Networks	33
Table 3: Number of unique fuels and pure compounds #FL and corresponding datapoints #DP used for the training and validation of the models.	41
Table 4: Table of considered chemical families with the corresponding formula, structural criteria and an illustrated representative.	46
Table 5: Comparison of the number of representative molecules available in the database (DB) and the number of theoretically possible molecules, calculated by MOLGEN (MG).	50
Table 6: Reproducibilities of ASTM property measurement methods [106].....	54
Table 7: Predictive capabilities of density models.	56
Table 8: Predictive capabilities of surface tension models.	58
Table 9: Predictive capabilities of net heat of combustion models.	59
Table 10: Predictive capabilities of kinematic viscosity models.	61
Table 11: Predictive capabilities of flash point models.	63
Table 12: Predictive capabilities of freezing point models.	65
Table 13: Predictive capabilities of cetane number models.....	67
Table 14: Predictive capabilities of distillation models at 10 vol%.....	69
Table 15: Predictive capabilities of distillation models at 50 vol%.....	69
Table 16: Predictive capabilities of distillation models at 90 vol%.....	70
Table 17: Summary influence of different hydrocarbon families and molecular descriptors on the density at 15 °C, * indicates an annotation for the correlation, \emptyset indicates insufficient data ...	88
Table 18: Summary influence of different hydrocarbon families and molecular descriptors on the surface tension at 22 °C, * indicates an annotation for the correlation, \emptyset indicates insufficient data.....	93
Table 19: Summary influence of different hydrocarbon families and molecular descriptors on the kinematic viscosity at 0 °C, * indicates an annotation for the correlation, \emptyset indicates insufficient data.	94
Table 20: Summary influence of different hydrocarbon families and molecular descriptors on the net heat of combustion, * indicates an annotation for the correlation, \emptyset indicates insufficient data.....	98
Table 21: Summary influence of different hydrocarbon families and molecular descriptors on the flash point, * indicates an annotation for the correlation, \emptyset indicates insufficient data.	101
Table 22: Summary influence of different hydrocarbon families and molecular descriptors on the freezing point, * indicates an annotation for the correlation, \emptyset indicates insufficient data.	104
Table 23: Summary influence of different hydrocarbon families and molecular descriptors on the cetane number, * indicates an annotation for the correlation, \emptyset indicates insufficient data....	107

Table 24: Summary influence of different hydrocarbon families and molecular descriptors on the boiling point, * indicates an annotation for the correlation, Ø indicates insufficient data.....	110
Table 25: Summary influence of different hydrocarbon families and molecular descriptors on the yield sooting index, * indicates an annotation for the correlation, Ø indicates insufficient data.	113
Table 26: Summary table for the relationship and influence of structural descriptors of fuel components on their physicochemical properties, average differences Δ are given in %.	116
Table 27: Summary table of fuel design process operations.....	126
Table 28: Specification properties for jet fuels blends after ASTM D7566.	133

Nomenclature

Acronyms

Abbreviation	Description
ANN	Artificial Neural Network
ASTM	American Society of Testing and Materials
ATJ-SKA	Alcohol-to-Jet Synthetic Kerosene with Aromatics
ATJ-SPK	Alcohol-to-Jet Synthetic Paraffinic Kerosene
CAAFI	Commercial Aviation Alternative Fuels Initiative
CHJ	Catalytic Hydrothermolysis Synthesized Kerosene
CTM	Continuous Thermodynamics Model
CRC	Coordinating Research Council
CV	Cross-validation
DC	Direct Correlation Method
DCM	Discrete Component Model
DLR	German Aerospace Center
ECLIF	Emission and Climate Impact of Alternative Fuels
EU	European Union
FAA	Federal Aviation Agency
FT	Fisher-Tropsch
FT-SPK	Fisher-Tropsch Hydroprocessed Synthesized Paraffinic Kerosene
GC	Group Contribution
GCxGC	Two-dimensional Gas Chromatography
HDO-SAK	Hydro-Deoxygenation Synthetic Kerosene
HEFA-SPK	Synthesized Paraffinic Kerosene from Hydroprocessed Esters and Fatty Acids
HFP HEFA-SK	High Freeze Point Hydroprocessed Esters and Fatty Acids Synthetic Kerosene
HHC-SPK	Hydroprocessed Hydrocarbons, Esters, and Fatty Acids Synthetic Paraffinic Kerosene
IATA	International Air Transport Association
IFPEN	French Institute of Petroleum
IH ²	Integrated Hydropyrolysis and Hydroconversion
ICAO	International Civil Aviation Organization
ILUC	Indirect Land Use and Change
JETSCREEN	Project for Jet Fuel Screening and Optimization
JSON	JavaScript Object Notation

Abbreviation	Description
MCNN	Monte-Carlo Dropout Neural Network
M-QSPR	Mean Quantitative Structure-Property Property Relationship
NJFCP	National Jet Fuel Combustion Project
NO _x	Nitrogen Oxides
OEM	Original Equipment Manufacturer
PI	Prediction Interval
QSPR	Quantitative Structure-Property Relationship
RMSE	Root Mean Squared Error
ReFGen	Representative Fuel Generator
SAF	Sustainable Aviation Fuel
SIP	Synthesized Isoparaffins from Hydroprocessed Fermented Sugars
SPK/A	Synthesized Paraffinic Kerosene with Aromatics
US	United States
UNIFAC	Unified Functional Group Activity Theory
VEM	Valence Electron Mobile Environment
XTL	Electricity- or Solar-Radiation-to-Liquid

List of Symbols

Greek symbols

Symbol	Unit	Description
α	-	Atom core count
β	-	VEM environment count
γ	-	VEM vertex count
Δ_{nC}	%	Average change of property value by the change of the number of contained carbon atoms
Δ_{η_B}	%	Average change of property value by the change of the branching index
Δ_{HC}	%	Average difference of property value of a hydrocarbon family to reference fuels
η_B	-	Topochemical atom index for branching
η^{loc}	-	Local topochemical atom index
η_N^{loc}	-	Local topochemical atom index for an unbranched molecule
η_R^{loc}	-	Local topochemical atom index of a reference molecule containing only σ -bonds
ν	mN/m	Kinematic viscosity

Latin symbols

Symbol	Unit	Description
a_{ij}	-	Molecular adjacency matrix
$cXH0$	-	Number of aromatic carbon atoms connected only to other carbon atoms
f_σ	-	Atomic vertex for non-hydrogen σ -bonds
f_π	-	Atomic vertex for non-hydrogen π -bonds
MAE	Depends on physical property	Mean Absolute Error
MAOE	Depends on physical property	Mean Absolute Error of Outliers
nC	-	Number of contained carbon atoms
nR	-	Number of contained ring atoms
NMPIW	%	Normalized Mean Prediction Interval Width

Symbol	Unit	Description
PICP	%	Prediction Interval Coverage Probability
r	-	Pearson correlation coefficient
$PPSA$	-	Partial Positive Surface Area
SA_a^+	-	Sum of positive surface area
	kg/kg,	
w	mol/mol, vol/vol	Mass, volume, or molar fraction

Abstract

To achieve climate neutrality in the aviation sector, research on new sustainable aviation fuels (SAF) is needed as the growing demand will exceed the production potential of established sustainable pathways. The focus is thereby not only on the exploration of sustainable feedstocks and the development of new production processes but also on the facilitation and acceleration of the whole fuel development process, from its conceptualization to its approval. The critical evaluation of a new production pathway guarantees the safe application and performance of a new fuel. The approval poses a major challenge for fuel producers, requiring a tremendous commitment of time, fuel volume and cost. Concepts that allow a fast-iterative, low-cost screening and design of new candidate fuels, to assess and optimize their chances for approval are thereby seen as key enablers.

Established fuel screening concepts rely on model-based prediction, which, together with state-of-the-art compositional analytics, allow the fast assessment of SAF candidates from volumes as low as 5 mL. The design of new fuels, on the other hand, requires a comprehensive understanding of the composition of a jet fuel and properties considered critical for the fuel approval. This work describes the research and development of tools for the screening and design of jet fuels. Focusing on data-based methods, the tools are built from a database composed of both jet fuels and fuel components. It is thereby investigated whether and how data-based tools are able to support the screening and design of new SAF candidates and what their limitations are.

For the jet fuel screening, three different modeling methods to predict physicochemical properties from compositional measurements are adapted and investigated: Direct correlation (DC), Mean Quantitative Structure-Property Relationship Modeling (M-QSPR) and Quantitative Structure-Property Relationship Modeling (QSPR) with sampling. All developed models are probabilistic, since the safety-relevant use case of jet fuel screening makes the consideration of uncertainties necessary. Rather than estimating one deterministic property value, probabilistic models estimate a distribution of values and with it the associated uncertainty. The predictive capabilities of the developed models are assessed using specially developed metrics and compared on the prediction of conventional and synthetic jet fuels. To put the developed models into reference, they are compared to established deterministic models from the literature. Identifying strengths and limitations of the different approaches, the models are applied to jet fuel screening to test their adequacy for the assessment of new SAF candidates.

To support the design of new SAF candidates, the relationships between the fuel composition and critical physicochemical properties are investigated. The relationships are investigated on the basis of fuel components and the influence of their chemical families as well as the structural aspects size and the branching. Trends and relations are characterized with graphs and quantitative metrics that illustrate correlation and state the average value for a change in composition.

Both the developed models and design tools are applied to the use case of screening and then optimizing a real SAF candidate to maximize its chances for successful fuel approval. The SAF candidate and three optimized fuel variants with reformulated compositions are thereby screened to assess the most suitable production route. Afterwards, a blending analysis of the SAF candidate and the variants is conducted to estimate their maximum volume fraction in the mixture with representative conventional jet fuels, considering both the safety requirements as well as the potential reduction of CO₂ and soot emissions.

As potential next steps, this work identifies the need for advancements in the analytics of the fuel composition as well as the extension of the existing fuel property databases. The former would reduce the uncertainty in the property modeling, while the latter would increase both the predictive capability of the models and the understanding of the fuel property relations.

Kurzfassung

Der wachsende zivile Luftfahrtsektor und die begrenzte Steigerungsfähigkeit etablierter Produktionspfade für nachhaltige, synthetische Treibstoffe (SAF) erfordern intensive Forschung um das gesetzte Ziel der Klimaneutralität bis 2050 zu erreichen. Neben der Erforschung neuartiger Rohstoffe und Produktionsverfahren liegt der Fokus auf einer generellen Beschleunigung des gesamten Entwicklungsprozesses, von der initialen Treibstoffformulierung bis hin zur finalen Zulassung.

Die Zulassung stellt für Treibstoffhersteller eine besondere Herausforderung dar, da sie enorme finanzielle und zeitliche Ressourcen sowie die Bereitstellung großer Treibstoffmengen erfordert. Innovative, kostengünstige Konzepte, die eine frühzeitige Bewertung und Optimierung von Treibstoffkandidaten anhand geringer Mengen ermöglichen, haben das Potenzial, den Entwicklungsprozess und die Zulassung signifikant zu beschleunigen. Diese neuen Bewertungskonzepte basieren auf einer Kombination moderner Treibstoffanalytik und modellbasierter Vorhersage kritischer Treibstoffeigenschaften und ermöglichen so die Bewertung des Kandidaten bereits ab einem Volumen von 5 mL. In der anschließenden Optimierung können die Treibstoffeigenschaften des Kandidaten durch gezielte Modifikation der Zusammensetzung verbessert werden, um die Chancen auf die eigentliche Zulassung zu erhöhen. Die Anforderungen an die Eigenschaftsmodelle und Designwerkzeuge sind hoch, da sie auch auf neuartige Treibstoffzusammensetzungen außerhalb des bisherigen Erfahrungsbereichs anwendbar sein müssen. Diese Arbeit untersucht das Potenzial und die Limitierungen datenbasierter Methoden als Werkzeuge für die beschriebene Treibstoffbewertung und das Treibstoffdesign. Unter Nutzung neuester Machine-Learning-Algorithmen und Datenbanken soll geklärt werden, ob und wie datenbasierte Methoden die frühe Phase der Treibstoffentwicklung und Zulassung unterstützen können.

Für die Bewertung der Treibstoffe werden drei verschiedene Methoden zur Modellierung von acht kritischen Treibstoffeigenschaften auf Basis der Zusammensetzung entwickelt und untersucht: Direkte Korrelation (DC), Mean Quantitative Structure-Property Relationship Modeling (M-QSPR) und Quantitative Structure-Property Relationship Modeling (QSPR) mit Sampling. Alle drei Methoden greifen dabei auf probabilistische Modelle zurück, welche nicht nur einen deterministischen Wert pro Treibstoffeigenschaft vorhersagen, sondern einen möglichen Wertebereich abschätzen und so inhärente Unsicherheiten abbilden. Die Vorhersagefähigkeiten der entwickelten Modelle werden anhand eigens entwickelter Metriken sowohl für konventionelle als auch synthetische Treibstoffe bewertet und untereinander sowie mit etablierten deterministischen Modellen aus der Literatur verglichen. Die Eignung der Modelle für die eigentliche Bewertung von neuen Treibstoffkandidaten wird anschließend in einer simulierten Treibstoffbewertung von drei Kandidaten festgestellt.

Für das Treibstoffdesign werden eigens Werkzeuge anhand von systematischen Untersuchungen der Beziehungen von Treibstoffzusammensetzung und den kritischen Eigenschaften erstellt. Der Einfluss der jeweiligen chemischen Familie, der Größe und der Topologie der Treibstoffkomponenten auf die

Eigenschaften wird anhand von Grafiken und quantitativer Metriken untersucht und in Korrelationen erfasst.

Die entwickelten Modelle und Designwerkzeuge wurden anschließend kombiniert, um einen Treibstoffkandidaten zu bewerten und zu optimieren und so dessen Chancen für die Zulassung zu maximieren. In einem ersten Schritt wurde hierbei der Treibstoffkandidat und die drei optimierten Varianten bewertet, um die Variante mit den größten Zulassungschancen zu ermitteln. Anschließend wird eine Mischungsanalyse der aussichtsreichsten Variante durchgeführt, um den maximalen Volumenanteil in Mischungen mit konventionellen Treibstoffen und die zu erwartenden CO_2 und Rußemission zu ermitteln.

Im Rahmen dieser Arbeit wurden datenbasierte Methoden erfolgreich sowohl für die Bewertung als auch das Design von Treibstoffen entwickelt, untersucht und angewendet. Limitierungen wurden hierbei vor allem aufgrund von Unsicherheiten in den Zusammensetzungsmessungen und eingeschränkter Verfügbarkeit von Daten für das Training der Modelle und die Entwicklung der Designwerkzeuge festgestellt. Nächste mögliche Schritte sind somit weitere Forschung und Verbesserung der Treibstoffanalytik, sowie die Erweiterung der verfügbaren Datenbanken durch gezielt durchgeführte Messkampagnen. Ersteres würde die Unsicherheit in der Modellierung der Eigenschaften signifikant verringern, Zweiteres die Vorhersagefähigkeit der Modelle und die Verwendbarkeit und Aussagekraft der Designwerkzeuge verbessern.

1 Introduction

1.1 Motivation

The consequences of man-made climate change make an adaptation and realignment of the aviation industry inescapable. Politically set strategies like the “European Green Deal” of the European Union (EU) and the “Sustainable Aviation Fuel Grand Challenge” of the Government of the United States of America (US) foresee a need for an emission reduction in aviation of 90% [1] and 100 % [2] respectively, to achieve the goal of climate neutrality by 2050. As a globally growing industry, the aviation sector is expected to grow approximately 4% p.a. until 2050, depending on the region [3]. Hence, a rapid adaptation of alternative technology is necessary to establish a sustainable aviation industry. Recent reports by the Intergovernmental Panel on Climate Change (IPCC) clearly state that prompt actions are required to achieve the set emission reduction goals with technology that has high technical readiness and high chances of application at large scale [4]. This has been recognized by aviation associations like the International Civil Aviation Organization (ICAO). They rank the use of sustainable aviation jet fuel (SAF) as the technology with the highest technological readiness and the highest potential emission reduction for the aviation industry [5]. Figure 1.1 shows the potential contributions of measures for net CO₂ reduction as part of the long-term high aspirational goal of ICAO from 2022 prognosed to 2050 [6]. According to ICAO's projections, sustainable aviation fuels (SAF) are expected to play a crucial role in reducing CO₂ emissions from international aviation in the future. Despite the anticipated growth of the aviation industry, particularly in developing and emerging countries, the widespread adoption of SAF has the potential to decrease CO₂ emissions below the levels seen during the 2021 COVID-19 pandemic low.

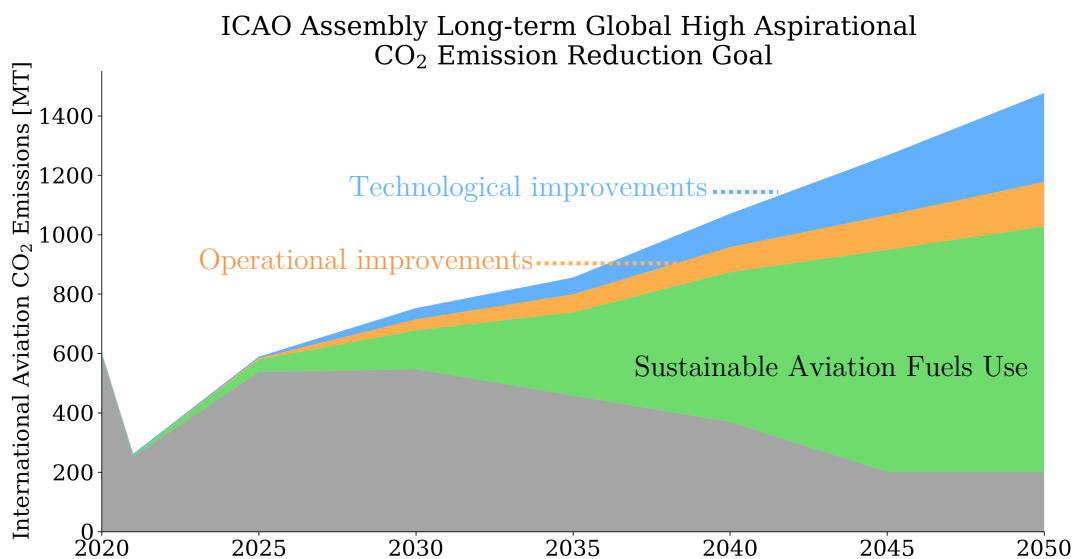


Figure 1.1: Contribution of measures for reducing prognosed international aviation net CO₂ emissions [6].

Apart from the reduction of the greenhouse effect of the CO₂, the use of SAF has also the potential to reduce parts of the so-called non-CO₂ climate effects, which result from contrails formed of emitted soot particulates. According to recent studies by Lee et al. [7] as well as Voigt et al. [8] and Faber et al. [9], the contribution of non-CO₂ effects in aviation on the climate is larger than the one of CO₂ emissions [8,9]. The emission of soot particles is strongly influenced by the jet fuel composition, with low aromatic SAF fuels showing significantly lower emissions with current technologies [8].

It is expected that the use of SAF and market-based measures like an emission trading system, will excel the potential reductions by improvements of the burner technology significantly, especially in later years [10]. The use of SAF on a large scale in the civil aviation industry is therefore a necessity to reach the emission reduction targets.

The need for the large-scale application of SAF has been recognized by major political institutions, which have released legislative proposals for SAF use and emission reduction. The “ReFuelEU” aviation proposal of the European Parliament sets the minimum share of SAF to 2 % by 2025, 5 % by 2030, 32 % by 2040 and 63 % by 2050 in the “Fit for 55” concept for climate neutrality [11]. The US Government announced a significant increase of SAF production from currently 136’000 tons in 2020 [12] to 9.08 million tons by 2030 and 106 million tons by 2050 in their “Sustainable Aviation Fuel Grand Challenge”[2].

Currently, production rates of approved SAF technologies are only able to provide a fraction of the needed sustainable fuel with 200 000 tons, which corresponds to less than 0.1 % of worldwide jet fuel demand in 2019 [13]. Drastic increases in production are planned by companies like Neste, with a planned production of 1.5 million tons by 2023 [14], Shell with 2 million tons by 2025 [13] and World Energy with 5 million tons in 2024 [15]. However, the sufficient supply of SAF volumes required to achieve the set milestones for climate neutrality is highly uncertain, with 6.4 million tons required in 2025, 18.3 million tons in 2030 and 359.2 million tons by 2050, as recently estimated by the International Air Transport Association (IATA) in 2022 [16]. This becomes especially apparent, considering that current SAF production and the production planned until 2030 consist and will consist predominantly of bio-based SAF from feedstocks like rapeseed, soy, palm oil etc. [17]. These feedstocks are however not available in sufficient quantities without interfering with other industries, e.g. the food industry [18], or negatively impacting existing natural high carbon stocks through indirect land use and change (ILUC) [19]. ILUC summarizes the potential net release of CO₂ from vegetation and soil when lands with high carbon stocks like forests and grasslands are converted to agricultural lands to compensate for the diversion of existing croplands to biofuel production. These biogenic production routes are therefore not expected to meet the rising demand of the growing aviation industry in the long term. The EU therefore increasingly supports the transition from food-based biofuels and fuels with high potential ILUC with the recast of the Renewable Energy Directive. Alternative feedstocks like lignocellulose, byproducts and wastes, as well as alternative non-biogenic production routes like

Power-to-Liquid pathways [18] are thereby especially promoted. These production routes, however, have negligible market shares compared to biogenic routes or an overall small market readiness [5].

Further research and development of new SAF production routes is therefore needed, alongside the strongly growing SAF market. The focus is thereby not only on the identification of adequate feedstocks and the development of new production processes, but also on the facilitation and acceleration of the whole development process. From a laboratory concept, the process must be developed to industrial scale under compliance of the final product with the required approval protocol after ASTM D4054 [20], which guarantees the safe application of the produced fuel in the aviation industry. Historically, the approval of a new SAF production pathway alone can last up to several years and require multiple millions of dollars as well as hundreds of tons of fuel for the extensive testing [21]. Early production capabilities of a fuel candidate are however often on a laboratory scale and uncertain chances of success prevent additional investments for upscaling as well as the willingness to fund the required test program. Considering the given timeframe for the envisaged emission reduction in aviation, the process of designing a new jet fuel and optimizing it to pass the approval process has to be reduced to a minimum in order to meet the set goals of climate neutrality.

Extensive research projects like the National Jet Fuel Combustions Program (NJFCP) funded by the US Government [22] and the project for Jet Fuel Screening and Optimization (JETSCREEN) [23] of the EU were initiated to facilitate and streamline the jet fuel approval process. Based on the findings of these research projects, Heyne and Rauch developed the concept of prescreening in 2020, which allows the assessment of new jet fuel candidates at an early stage of development with minimal cost and required fuel volume [24]. The prescreening assesses the chances of a jet fuel candidate to pass the approval process and gives fast feedback to the producer to redesign the composition and optimize the fuel accordingly. The concept thereby focuses on a few jet fuel properties that are regarded as particularly critical for the jet fuel approval by both the NJFCP and the JETSCREEN project. To reduce time, cost and required fuel volume for the measurements of these critical properties, the prescreening procedure utilizes predictive models, combined with modern analytical measurement methods. Together, these methods allow prediction of the critical properties from fuel volumes below 5 mL.

The requirements for the models are high, since the predictions are expected to be comparable to property measurements and substitute them if not yet available. To meet the requirements, the predictions have to be accurate, highly reliable and reflect potential uncertainties for their risk-informed usage. Furthermore, the models need to adequately predict desired properties not only for the known range of jet fuel compositions but also for the compositions of new SAF that might significantly deviate from the known compositional range. Apart from the models, extensive knowledge about the relationship between the fuel composition and the desired critical properties is required to design a fuel and optimize its chances of passing the approval process.

It is against this background that the following scope of this doctoral thesis is set, with the goal of developing highly accurate and reliable tools for the described use cases of screening and designing new SAF candidates and supporting their development at an early stage. In the future, these tools could be the basis for the screening and design of jet fuels under consideration of ecological aspects like the described non-CO₂ effects, saving cost, time and fuel volume in the fuel approval process.

1.2 Sustainable Aviation Fuel development and approval

The discrepancy between developing a new ecologically sustainable SAF production route under optimal economic conditions and ensuring its safe use is a serious challenge for fuel producers, the aviation industry and certification associations. Since 2008, seven unique production paths and with them seven SAF types have been developed by the fuel industry and certified by the American Society for Testing and Materials (ASTM) [25]. The approval by the ASTM is necessary for every newly developed production path and the corresponding fuel type. It guarantees the safe application of the fuel in the existing infrastructure of the aviation industry, from production, transport, storage, and handling to the operability in the aircraft.

1.2.1 Jet fuel specifications and synthetic aviation fuel approval process

The ASTM (American Society for Testing and Materials) oversees three crucial specifications to which sustainable fuels or their conventional blending counterparts must adhere: D1655 [26], D7566 [25] and D4054. The standard practice ASTM D4054 and the specification for jet fuels containing synthesized hydrocarbons ASTM D7566 are relevant for SAF. ASTM D4054 describes the process for the approval of a new aviation turbine fuel, while ASTM D7566 holds the standards for aviation turbine fuel blends containing synthetic hydrocarbons. Each approved production path has an annex in ASTM D7566, which states the specifications for the respective fuel type, its production path and feedstock, their maximum blending fraction and specifications for the fuel blend itself. Blends that comply with the set specifications in the annexes of ASTM D7566 and the requirements for jet fuel blends are considered “drop-in fuels” that can directly be utilized in existing infrastructure and aircrafts. At the time of writing, ASTM D7566 states a maximum fraction of up to 10 % and 50 vol% for SAF blends, depending on the SAF type. ASTM D1655 holds the two major specifications for conventional jet fuel types civil aviation: Jet A, defined by the ASTM itself, and Jet A-1, defined by the Defense Standard 91-91 of the Ministry of Defense [27]. Besides Jet A and Jet A-1, specifications exist for further civil fuel types that are country-specific and play a minor role in the commercial aviation sector: TS-1 for Russia and the Commonwealth of Independent States and RP fuels for the Republic of China [28].

ASTM D1655 and ASTM D7566

Both ASTM D1655 and D7566 are performance specifications and do not explicitly define an allowed jet fuel composition or compositional range. They rather specify a combination of minimum and maximum requirements for physicochemical and performance properties and allow fractions of certain chemical families as well as trace compounds, e.g., antioxidants. In combination with the approved production routes, these specification requirements implicitly constrain the range of possible jet fuel compositions. To illustrate the compositional restriction as a result of the requirements of property and composition, Figure 1.2 shows a schematic ternary diagram after de Klerk for the resulting compositional range of conventional crude-oil based Jet A-1 fuel after ASTM D1655 [29]. The possible jet fuel composition and property constrained are thereby presented in a simplified schematic figure. The possible compositional range of Jet A-1, indicated in gray, is graphically restricted by the minimum and maximum requirements of the specification, e.g. the minimum aromatic content and the maximum freezing point.

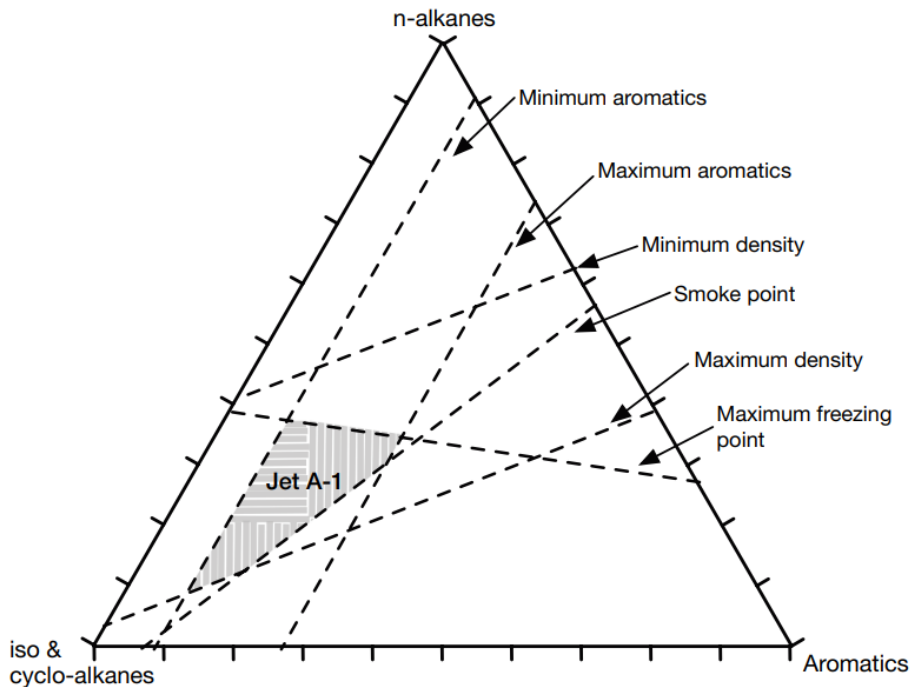


Figure 1.2: Restriction of possible fuel composition after ASTM D1655 for conventional crude oil-based jet fuels [29].

At the time of writing, ASTM D1655 and D7566 hold specification requirements that are classified into the following categories: composition, volatility, fluidity, combustion, corrosion, thermal stability, contaminants and additives. All properties and compositions of a fuel have to be measured with approved analytical methods that are also stated in the respective specification.

ASTM D4054

ASTM D4054 describes the process for approving a new SAF production path and creating a respective specification that can be included as an annex in ASTM D7566. ASTM D4054 is based on the experiences from the approval processes by the British Ministry of Defense for the first synthetic jet fuel by the company Sasol in 2009 [30]. It was developed as a guide by original equipment manufacturers (OEM) of the aviation industry with the support of ASTM members and includes property and composition targets that are known to impact the performance of the turbine engines and fuel system [31]. The approval process consists of three parts: 1) Initial screening, 2) Follow-on testing and 3) Balloting and approval. The parts have to be successfully completed in sequential order to advance to the next. Figure 1.3 shows a schematic flow diagram of the approval process, with the two testing phases, the balloting and approval.

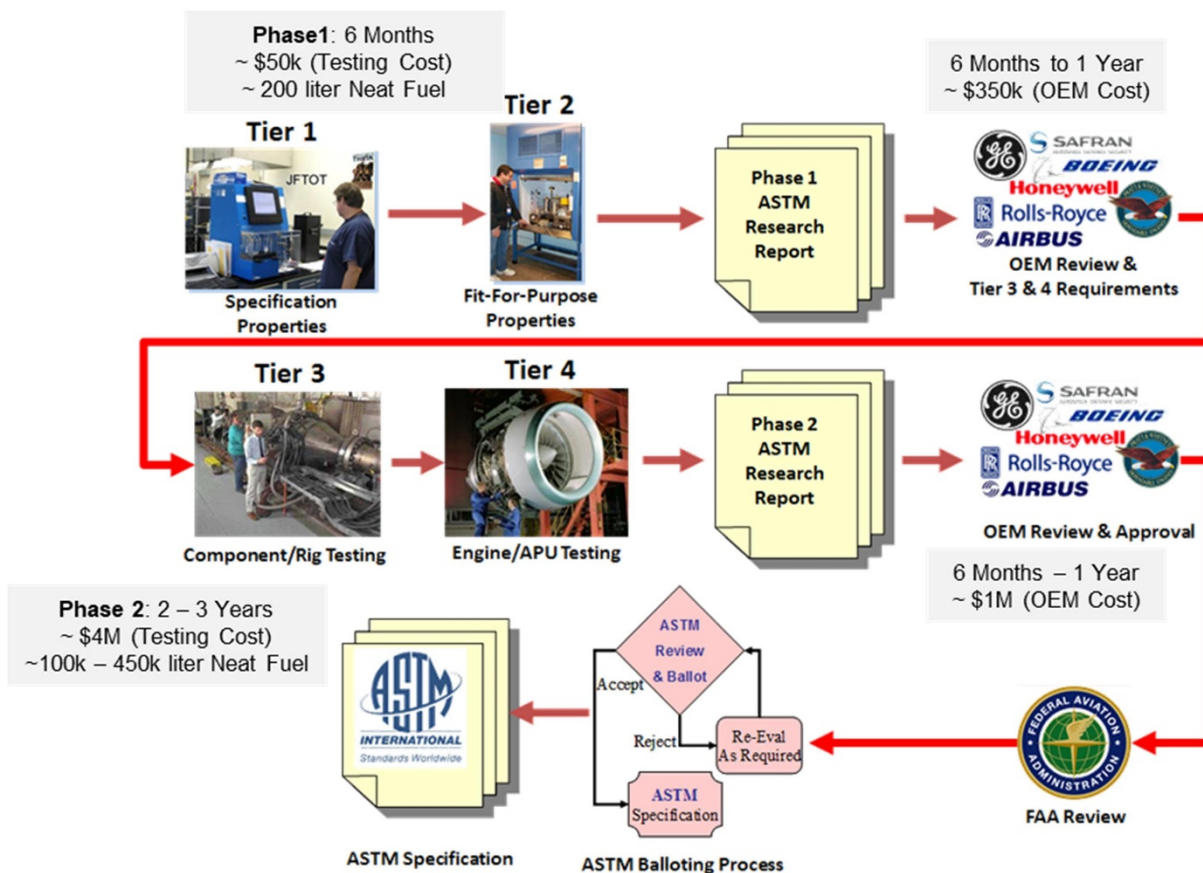


Figure 1.3: Flow diagram of the approval process of a new aviation turbine fuel after ASTM D4054 [24].

The test programs of phases 1 and 2 are comprised of four tiers that have to be completed successfully, also in sequential order. A fuel is tested for its specification properties in Tier 1, followed by fit-for-purpose properties in Tier 2, component and rig tests in Tier 3 and finally engine tests in Tier 4. If a later test tier fails, there is a risk that the entire sequential testing process will have to start all over again. All required tests or compositional analyses require a substantial amount of volume from the fuel candidate. Tier 1 and Tier 2, in which predominantly physicochemical properties and the chemical composition are measured, demand 200 liters and

around 50 000 US dollars testing cost. Tier 3 and 4 can demand between 100 000 to 450 000 liters and around 4 million US dollars in testing costs. Tier 3 and 4 investigate inter alia the spray characteristics, the ignition behavior, the cold start and lean blow-out, as well as the operability and performance of the fuel candidate [24,31]. The extent of the tests in Tier 3 are determined by engine OEMs based on the results of Tier 1 and Tier 2. Similarities in the chemical composition or the measured properties are thereby considered and influence the extent of testing in Tier 3 and Tier 4 and therefore the required fuel amounts [31]. After successfully passing the extensive test of phases, research reports are prepared and passed to the OEMs for their internal review. The report thereby states the results of the tests. The report of phase 2 furthermore has to give a detailed description of the way the production process will be controlled to ensure the same quality of the tested fuel and the fuel that will be produced in commercial quantities. In the review, OEMs, the Federal Aviation Agency (FAA) of the US Government and the ASTM decide if the new fuel candidate fits an existing annex in ASTM D7566 or if a new one has to be created [31].

The specification changes in phase 3 are the final gate of the approval process, in which the research report is balloted for comment and approval and the creation of a new annex in ASTM D7566. The balloting process allows diverse groups of stakeholders from other areas of the fuel and petroleum community to review the report and note concerns that might require additional measurements to be added in the new specification annex or stop the approval process entirely [31].

The complete ASTM D4054 approval process can take up to several years, requires a sustained commitment, millions of US dollars and up to hundreds of thousands of liters of fuel for testing [24,31]. Since fuel producers that seek approval can often not provide the necessary fuel amounts for testing, a “fast-track” approval process was added as annex 4 in ASTM D4054 in September of 2020. It reduces the approval to Tier 1 testing and selected tests from higher Tiers, a fast track research report reviewed by OEMs and FAA and the balloting and specification change [21,32]. New production paths approved after the fast-track process are however limited to a maximum blending fraction of 10 vol% [32].

1.2.2 Jet fuel prescreening

The need for an even faster, less fuel- and cost-intensive processes for the assessment and approval of a new SAF candidate was constituted by OEMs and ASTM in CAAFI 2014 [33]. Based on findings from the subsequent research projects NJFCP [22] and JETSCREEN [23], Heyne and Rauch developed a concept for an accelerated assessment process called jet fuel prescreening [24]. This concept makes it possible to assess the chances of a fuel passing the actual approval process at an early stage of development with minimal costs and fuel volume using model-based property predictions. [24]. Based on the results of the screening, a fuel producer can redesign the fuel composition to optimize it accordingly. The concept focuses on the assessment of the fuel composition and the evaluation of critical fuel properties, especially properties that influence

operability and safety issues, which may not be directly exhibited until Tier 3 and Tier 4 of the approval process. Eight properties were identified as a short list by the research projects that have a critical impact on aircraft and engine and ground handling [24]. The eight properties are summarized in Table 1.

Property	Unit	Dependency	Min	Max
Density	kg/m ³	15 °C	775	840
Kinematic viscosity	mm ² /s	-20 °C		8
		-40 °C		12
Surface	mN/s	20 °C		
Net heat of combustion	MJ/kg		42.8	
Flash point	°C		38	68
Freezing point	°C			-40
Derived cetane number	-		30	
Distillation line	°C	10 vol%	150	205
		50 vol%	165	229
		90 vol%	190	262
		100 vol%		300
		$T_{50} - T_{10}$	10	
		$T_{90} - T_{10}$	40	

Table 1: Critical jet fuel properties for jet fuel prescreening.

With the exception of the cetane number, tests for all listed critical properties are part of the Tier 1 and Tier 2 test programs of part 1 of the ASTM D4054 approval process. Detailed descriptions of the properties and their importance for aircraft and engine and ground handling are given in individual paragraphs in the Supplementary Material A.

To test composition and the outlined critical properties, the prescreening process provides two test tiers, Tier α and Tier β . Tier α is a screening based on the analyzed fuel composition with model-based property prediction and Tier β consists of experimental property measurements verifying predictions with particularly high uncertainties [24]. For Tier α , a fuel sample of just 5 mL is required. From this sample, the fuel composition is characterized using the analytical GCxGC method [24]. The model-based predictive models subsequently predict the outlined critical properties based on this compositional measurement. Tier β requires 150-500 mL of fuel, depending on the conducted tests. Based on the findings of Tier α and Tier β figures of merit for the performance in spray and engine operations, relevant for Tier 3 and Tier 4, can be estimated [24].

For a screened jet fuel candidate to have high chances of passing the actual ASTM D4054 test program, the estimated and measured properties of Tier α and Tier β should lie inside set specification limits of ASTM D4054 and ASTM D7566. If properties lie outside the specification

limits, the jet fuel composition should be redesigned to meet the specification in the next iterations. Figure 1.4 shows an exemplary case for the screening a property for the two fuels A and B. While the property value of Fuel A lies inside the allowed value range, the property value of Fuel B lies below the lower specification limits, as indicated in red. The composition of Fuel B has therefore to be adjusted and screened again in another iteration. Ideally, the compositional redesign of Fuel B can be conducted virtually using simulative tools to further save time and fuel volume, resolving the need for cost- and time-intensive iterations of the process parameters. This requires comprehensive knowledge about the relationships between fuel composition and all relevant properties, as well as appropriate simulative tools of the production process.

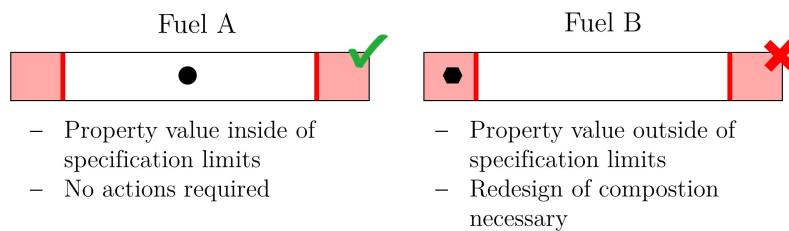


Figure 1.4: Schematic illustration of the screening plots for Fuel A and Fuel B as part of the jet fuel prescreening.

1.2.3 Challenges for the prescreening concept implementation

Parts of the prescreening concept were implemented and tested in the scope of the JETSCREEN project to assess the availability and adequacy of predictive property models for the Tier α testing [23]. The assessment identified limitations and challenges in both the availability and adequacy of the models, which consequently limited the application of the prescreening concept [34]. The different limitations and challenges are explored in more detail in the following.

Large variety of possible jet fuel compositions

As outlined in Section 1.2.1, jet fuel specifications do not directly specify an allowed composition range but rather the limits for the possible value range of jet fuel properties. As a result, the compositions of jet fuel candidates that enter the screening process can differ drastically from the compositions of fuels from known and approved production routes. To illustrate the variation of the possible composition range, Figure 1.5 shows plots of the composition of four representative synthetic fuels from the DLR Jet Fuel Database: FT-SPK, HEFA-SPK, ATJ-SPK, IH² and one conventional fuel, Jet A-1, as a reference. With the exception of the IH² fuel, all fuels are already approved by ASTM. Detailed descriptions of the different approved and pending fuel types and their corresponding production paths are given in Supplementary Material B.

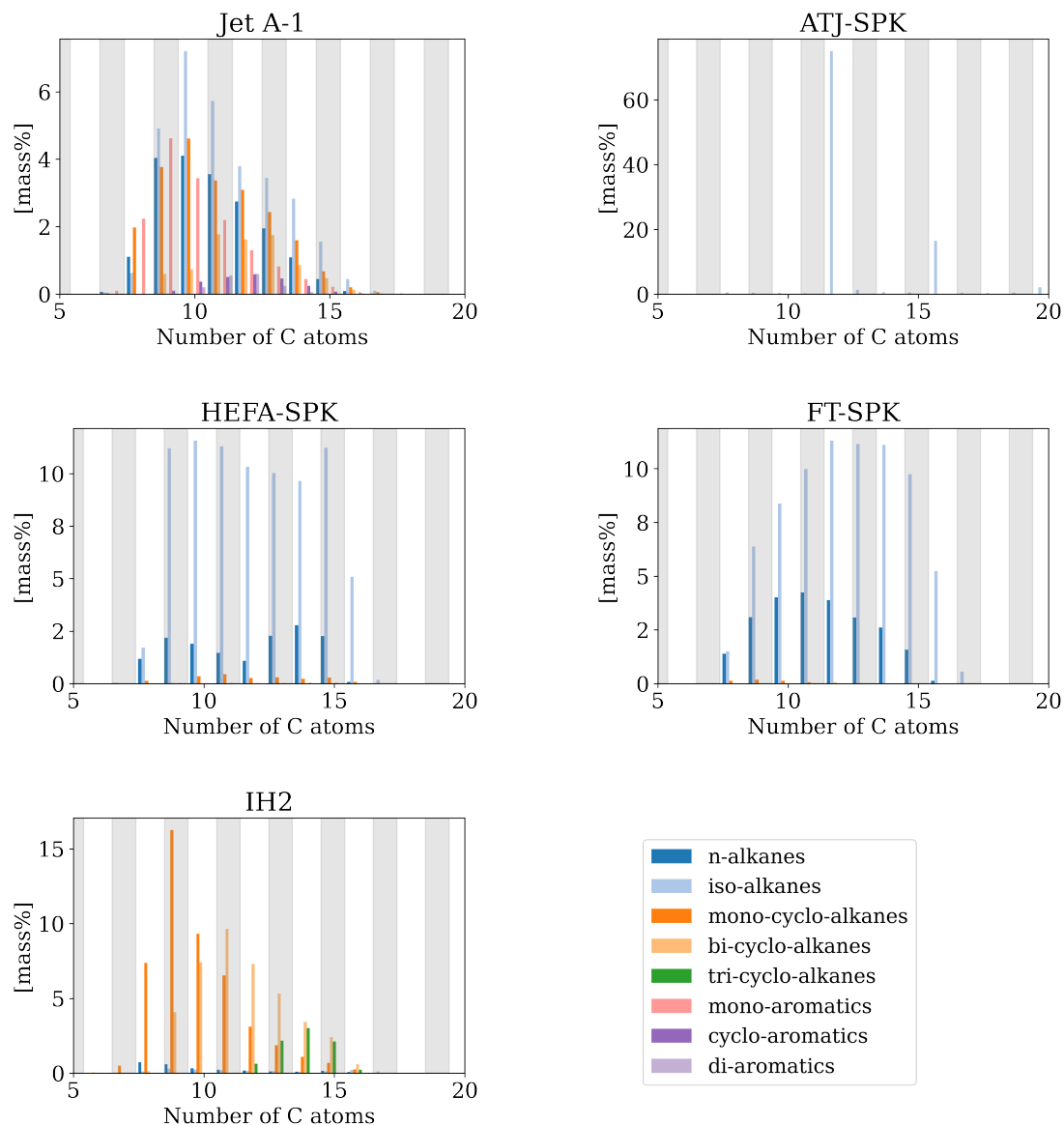


Figure 1.5: Plot GCxGC measurement of conventional Jet A-1 fuel (upper left), FT-SPK (upper right), HEFA-SPK (middle left), ATJ-SPK (middle right) IH² fuel (lower left).

The comparison of the plots visualizes the drastic compositional differences between the fuel types and the known and established conventional Jet A-1 fuel on the upper left. While the Jet A-1 fuel shows a broad Gaussian-like distribution for all considered families over the number of contained carbon atoms, the compositions of the SAF fuels are dominated by one or two families with distinct distributions. For FT-SPK and HEFA-SPK the compositions are dominated by n-alkanes and iso-alkanes and for the IH² by mono- and bi-cyclo-alkanes. The composition of the ATJ-SPK is made up almost entirely of two iso-alkanes with 12 and 16 carbon atoms. The wide variety of fuel compositions and the constant formulation of new candidates can pose a challenge to predictive models, as they may be confronted with fuels for which they have not been developed and validated.

Limitations in compositional analytics of the fuel composition

Both the compositional comparison of a new jet fuel candidate with already approved fuels and the property modeling of fuels require a compositional characterization of the candidate. There exist various analytical measurement methods that can be applied for the characterization of a jet fuel composition. ASTM D4054 lists both Mass Spectroscopy after ASTM D2425 [35] and High-Pressure Liquid Chromatography ASTM D6379 [36]. However, these methods only yield information about the cumulative fraction of compounds from the different hydrocarbon families. The accurate modeling of fuels requires more detailed information beyond the hydrocarbon family. Most modern laboratories use Two-dimensional Gas Chromatography (GCxGC) for the compositional analysis of jet fuels [24,37,38]. This measurement method uses two sequential gas chromatography columns for the separation of the fuel constituents with a subsequent mass spectroscopy or a flame ionization detector [39]. The two gas chromatography columns allow for a more precise identification of the fuel components, both by their chemical family and their number of carbon atoms they contain. However, the identification of the exact chemical component/isomer is currently not always possible for jet fuels due to the overlay of measurement signals [40,41]. Figure 1.6 illustrates this for an exemplary GCxGC measurement of conventional Jet A fuel. The colors in Figure 1.6 indicate the strength of the signal and the detected fraction of a fuel component, going from no signal (blue) to medium signal (green) to high signal (red). Figure 1.6 shows that signals lie in part very close to each other and can overlap. The clear classification of a signal and the identification of every individual component is therefore not possible.

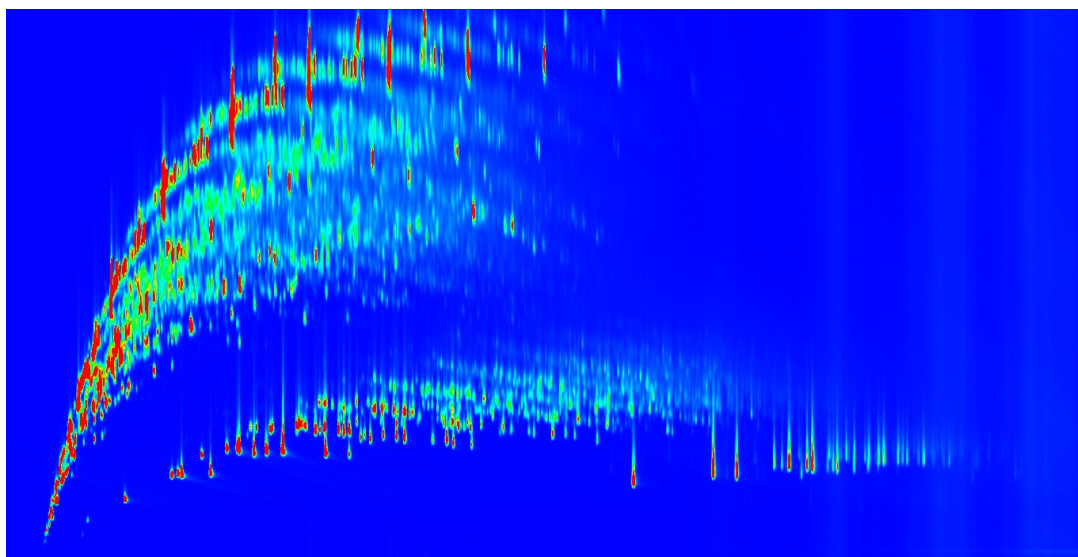


Figure 1.6 GCxGC measurement signal of Jet-A fuel [42].

The unidentified isomers can thereby have drastically different property values, which consequently affects the uncertainty in the property value of the fuel; e.g. for the freezing point, value differences of $-110.15\text{ }^{\circ}\text{C}$ to $12.6\text{ }^{\circ}\text{C}$ are recorded in the created database for iso-alkanes containing 10 carbon atoms. These differences become increasingly significant if a fuel

composition is dominated by one or two families or even distinct components, e.g. like for the ATJ-SPK. For the fuel prescreening, the models are required to reflect this uncertainty as part of their prediction. Furthermore, the predicted uncertainties have to be put into context with the limits of the fuel specification to illustrate the potential risk of accepting a prediction.

Availability and adequacy of state-of-the-art fuel property models

To apply predictive models for the Tier α prescreening, the models need to be 1) available 2) able to model the large possible composition space and reflect existing uncertainty, as well as 3) be able to predict adequate results. Models able to predict fuel properties on the basis of composition measurements have been investigated and developed since the 1950s [43]. In a recent publication, Vozka and Kilaz reviewed published fuel property models and compared them based on accuracy metrics provided by the respective authors [43]. The review lists possible models for six of the eight required properties able to predict on the basis of GCxGC measurements. All recommended models, e.g. by Shi et al. [40] and Vozka et al. [43] are deterministic data-based correlation models. These modeling methods directly correlated the GCxGC measurement, or averaged values of representative species, with the property returning one value. This means that uncertainties, e.g. due to unidentified isomers or other sources, cannot be reflected by the models and that the outlined prescreening requirements can therefore not be fulfilled. The provided accuracy metrics, in most cases averaged prediction errors, do furthermore not allow an estimation of the adequacy of the models for the application of prescreening. This is because the composition range of the fuels used for the publication and the calculation of the accuracy metric might not cover the composition range relevant for the screening. The ability of the models to predict adequate results is therefore highly uncertain. To assess the adequacy of the models, they must actually be tested on a representative selection of fuels relevant to prescreening, as shown in Figure 1.5.

To actually assess the adequacy of available models, three different state-of-the-art property models were investigated in the scope of the JETSCREEN project for their ability to adequately predict for a selection of conventional and synthetic fuels [34]. The project compared the Representative Fuel Generator (ReFGen) model and the Quantitative Structure Property Relationship (QSPR) model of the French Institute of Petroleum (IFPEN) as well as the Continuous Thermodynamics Model (CTM) the German Aerospace Center (DLR) by Le Clercq [44]. Models for three of the eight properties were thereby available for the comparison. Likewise to the models of Vozka and Shi, these models are deterministic and approximate the fuel either by representative species like the ReFGen or QSPR model, or by fitted distributions for the hydrocarbon families like the CTM model. Inherent uncertainties of the GCxGC measurements could therefore not be reflected and the outlined requirements were therefore also not met. The assessment of the models showed in part significant deviations of up to 47 % from measurement data for the CTM model, especially for synthetic fuels [34]. The models were therefore rated inadequate for prescreening purposes. The deviations were explained by the simplified fuel

representation of the models, which approximate the composition by only a few isomers and or distributions for the families, as well as their original fields of application, with the CTM model being mainly optimized for conventional fuels.

This review illustrates the need for new property models, developed and tested specifically for the use case of jet fuel screening and design. Existing models do not fulfill the identified prescreening requirements and were found to be inadequate or tested with metrics that do not guarantee their predictive capability for the intended application.

Following the increased development of data-based models, new modeling methods should be explored and created using newly developed Machine Learning algorithms. The new models should thereby be tested using predictive capability metrics that allow the assessment of their adequacy for the intended application of jet fuel prescreening.

1.3 Objectives and research questions

As outlined over the course of the last chapter, the aviation industry and fuel producers are in need of linking concepts that accelerate and streamline the development and approval of new sustainable aviation fuels. Recent scientific work developed those concepts, like the prescreening concept introduced by Heyne and Rauch [24]. The prescreening concept itself however relies on predictive models and a comprehensive understanding of the relations between fuel composition and properties to optimize the fuel composition.

This thesis investigates the question whether and how new data-based models are able to provide the tools for the outlined prescreening process. The research aims to develop both property models for predicting critical fuel properties and design tools to optimize fuel candidates for approval. The main objectives of this study are:

- Development of models for the prediction of the critical jet fuel properties from GCxGC composition measurements under consideration of uncertainties
- Development of an adequate database for the development and testing of the models
- Development of predictive capability metrics to assess the adequacy of the models for the application of jet fuel prescreening
- Development of tools for the jet fuel design based on the investigation of the relationships between fuel composition and the critical properties
- Finally, the application of the developed tools for the screening and design of new jet fuel candidates to optimize their chances for approval

To fulfill these objectives, this work focuses on the use of data-based methods both for the development of the models and the investigation of fuel composition and property relations. For the development of the necessary database, already existing data from different sources and databases is utilized. This reduces the need for own extensive measurement campaigns, and

allows the focus on different data-based modeling methods and correlation algorithms from the field of Machine Learning. Critical gaps in the utilized data as well as limitations that hinder the development of accurate and reliable tools are furthermore outlined, along with recommendations for future research.

1.4 Chapter outline

To answer the research question and address the individual objectives, this work is structured into the following dedicated chapters.

In Chapter 2 the theory and inner workings of the developed modeling methods are presented. The chapter furthermore holds sections for the model training and validation, as well as the developed predictive capability metrics and assessment process.

The database for the development of the predictive models and the fuel design is described in Chapter 3. This chapter describes and characterizes the data and illustrates the utilized preprocessing and outlier detection.

Chapter 4 holds the results of validation and adequacy assessment of the predictive models. The models are thereby compared with each other as well as with established models from the literature to relate their predictive capability with known modeling approaches. The adequacy of the models for the jet fuel screening is subsequently assessed based on a simulated prescreening of three fuels, which were excluded from the training and validation.

In Chapter 5 the influence of structural aspects of fuel components on the considered properties is investigated. The influence of the chemical family, size and branching of the fuel components on the different properties is thereby summarized in tools, as basis for the subsequent fuel design.

Chapter 6 applies all developed tools for a combined workflow of fuel screening and design for a real jet fuel candidate. The original jet fuel candidate and reformulated fuel variants, created as part of the fuel design, are thereby screened to assess their chances as potential applicants for the approval process. In a subsequent blending study the variant with the highest chances for approval is blended with a representative selection of conventional fuels to estimate their maximum blending fraction and potential as a synthetic blending component.

Chapter 7 relates the results of this work to the set research question and objectives. Based on this discussion, possible next steps and recommendations for further research are suggested.

2 Fuel Property Modeling

The modeling of physicochemical properties has always been of great interest both for the scientific fuel community and the fuel industry. The ability to predict properties, e.g., the net heat of combustion of a fuel solely based on its composition reduces the need for respective measurements, thereby saving time and cost and allowing the assessment of fuels [24] or subsequent simulations of processes like evaporation [45,46].

Fundamentally, there exist two approaches of modeling a fuel: 1) modeling a fuel as a mixture of constituents and 2) modeling a fuel as an entity. The first approach describes a fuel as a mixture of more fundamental and underlying constituents that either exist in the fuel as components or are sufficiently representative for the fuel composition. The bulk property of the fuel is thereby calculated from the property values of the individual constituents using an adequate mixing rule. The second approach directly correlates a compositional measurement, e.g. a GCxGC measurement, or a chemometric measurement signal of the fuel with the physicochemical properties using a regression algorithm. In contrast to the first approach, this one does not rely on mixing rules. Over the years, research has produced several modeling methods for both of the two approaches. The development was thereby strongly coupled to the available compositional analytics and the availability of data for the development of the models and their desired applications. Figure 2.1 shows a schematic illustration of the family tree of the two approaches and their respective modeling methods, which will be explained in the following.

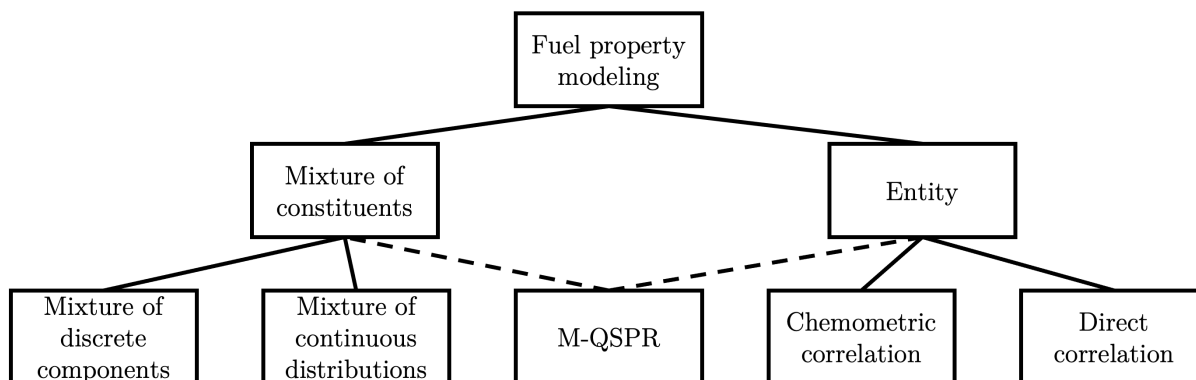


Figure 2.1: Family tree of approaches for the modeling of fuels.

For the mixture of constituents two major modeling methods emerged over the years. The method of modeling a fuel as a mixture of discrete components and the method of modeling a fuel as mixture of continuous distributions. The first method describes a fuel as a mixture of pure components, that have either been identified as fuel component, or are assumed to exist in the fuel and are sufficiently representative. The second method does not require the identification or assumption of individual fuel components. The fuel is rather modelled as a mixture of those continuous family distributions, where the distribution parameters are calculated from the

compositional measurement, e.g. the GCxGC measurement. Both methods estimate the bulk property of the fuel from the property values of the individual constituents.

For models using the method of discrete components, the property values of the fuel components can be provided from predictions, like in the Discrete Component Model (DCM) of Le Clercq of the German Aerospace Center (DLR) [45,47] or from measurements as in the model of Yang et al. [48]. Mixture of continuous distribution models on the other hand, like the DLR Continuous Thermodynamics Model (CTM) of Le Clercq [44], predict the property for each family by an underlying correlation that relates the family distribution with the property [30]. The utilization of models from the two methods is often strongly restricted to the use case models were designed for (e.g. fluid dynamic simulation) as well as the available compositional analytics and validation data. For the use case of simulating complex physical phenomena e.g. evaporation, computational limitations often constrain the number of possible fuel constituents since each constituent requires its own set of equations for the mass balance. For discrete component models, this limits the number of fuel components to one representative compound per family and carbon number, as for the DCM model of Le Clercq, or even fewer if the fuel is approximated using a surrogate, e.g. the model of Bell [49]. CTM models are especially suitable for the study of complex physical phenomena like evaporation [46].

The approach of modeling a fuel as an entity stands in strong contrast to the presented methods of modeling a fuel as mixture of constituents. Both the chemometric and the direct correlation methods directly correlate the compositional measurement of a fuel as a whole, or the chemometrical measurement signal with the physicochemical properties. Information and assumptions about potential fuel components are not necessary. The first models of the entity approach were developed by Cookson et al. [50–54] in the 1980s. They followed the direct correlation method and correlated the mass or volume fraction of the identified hydrocarbon families using a multilinear regression algorithm. The fractions of the hydrocarbon families were determined using GC, nuclear magnetic resonance spectroscopy, and high-pressure liquid chromatography. Chemometric models for the application of jet fuels were firstly developed by Morris et al. for the application of the prediction of critical fuel properties from near-infrared absorption spectra for the US Navy [55,56]. Direct correlation methods have the distinct advantage of using evaluated measurements in a standardized format. The standardized format allows the utilization of measurements from multiple different laboratories for the training of the direct correlation models. For chemometric models, the standardization of measurement signals is very challenging, which often limits the usable data to one reference laboratory or one particular measurement apparatus. With the increasing use of GCxGC for the compositional analysis of fuels, direct correlation methods with GCxGC measurements as input were developed, inter alia by Shi et al. [40] and Vozka et al. [43].

The modeling methods described up to this point are all deterministic, meaning they predict one property value for a given fuel composition. This is inherent to the described modeling

approaches, where the DCM model of LeClercq uses just one species to represent a family with a certain carbon number and the outlined direct correlation methods solely use deterministic correlation algorithms. Research conducted on jet fuel modeling and screening by the University of Dayton [48] and the JETSCREEN project [47] revealed, that deterministic modeling methods are not sufficient to adequately predict desired properties. The return of just one property value and the neglect of uncertainties which inherently exist e.g. due to unidentified isomers, proved to be insufficient. This is especially problematic for synthetic jet fuels, where differences in the properties are significant for isomers of a family at a certain carbon number and need to be reflected in the modeling. The necessity of the consideration of uncertainties induced by unidentified isomers in the GCxGC measurement was therefore directly adopted and implemented in the prescreening process by Heyne and Rauch [24].

To create a modeling method tailored for fuel screening, Yang et al. [48] developed a probabilistic discrete component model. This modeling method considers multiple possible isomers by sampling property values of isomers assumed to be present in the fuels from a measurement database using Monte-Carlo sampling [48]. This improved the predictive capability of the model by allowing the estimation of a possible value range of the fuel property, reflecting the inherent uncertainty of the modeling problem. However, the model of Yang et al. proved to strongly rely on the availability of property measurements of multiple isomers, which are often not available in current property databases. If the number of available measurements is too low or the set of available isomers is not representative, deviations and invalid uncertainty estimations can occur [48].

This thesis extends selected modeling methods of the previously outlined work, with the explicit goal to tailor these models to meet the unique needs of jet fuel screening and design processes. The developed models should be able to accurately and reliably predict properties and uncertainties for the prescreening process. Identified limitations of the outlined deterministic modeling methods and the dependence of the probabilistic method of Yang et al. on measurement data should thereby be overcome. For this, probabilistic models from both modeling approaches are developed to model jet fuels both as entities and as mixtures of constituents. From the two approaches, three different modeling methods are derived: 1) Monte-Carlo sampling of predicted fuel component properties, 2) direct probabilistic correlation and 3) Mean Quantitative Structure-Property Relationship (M-QSPR) modeling. The M-QSPR is a specifically developed hybrid method that has characteristics of both the 1) and the 2) modeling approach. It models a fuel as an entity, however it requires a selection of representative components. The method therefore sits between both approaches in the family tree of Figure 2.1.

The three methods differ fundamentally from each other. This allows the comparison of their individual advantages, disadvantages and limitations. Furthermore, potential benefits of using multiple modeling methods simultaneously for the use cases of jet fuel screening and design can be investigated. As part of this work, the developed models are also compared with existing

models from literature to outline the benefits of probabilistic modeling, both with respect to accuracy and the additional value of the estimated uncertainty.

2.1 Principles of data-based modeling methods

Data-based models have long been used in the field of fuel property modeling and most of the methods presented in the introduction of this chapter are data-based. This section outlines the principles of data-based modeling in comparison to physical modeling and elaborates the rationale for the use of data-based methods in the field of fuel property modelling.

The differentiation of the physical and data-based modeling approach requires a comparison of the underlying modeling philosophies and procedures. On a fundamental level, all modeling approaches for physical applications have the same intention: the replication of an objective reality in order to simulate possible events as bases for present decisions, for which experience or measurements are missing [57]. However, the physical and data-based modeling approach differ significantly in the way they replicate objective reality. Physical modeling approaches rely on a combination of known physical theory and observations derived from measurements. The theory itself must be derived from the measurements themselves or be already available from previous evaluations. Based on both theory and measurement, a conceptual model is prepared by human analysis, often in the form of a mathematical formula. This mathematical formula can then be implemented as a computational model, validated and if the validation is successful, utilized to simulate the desired events. A schematic flow diagram of the physical modeling process is given on the left of Figure 2.2 [57].

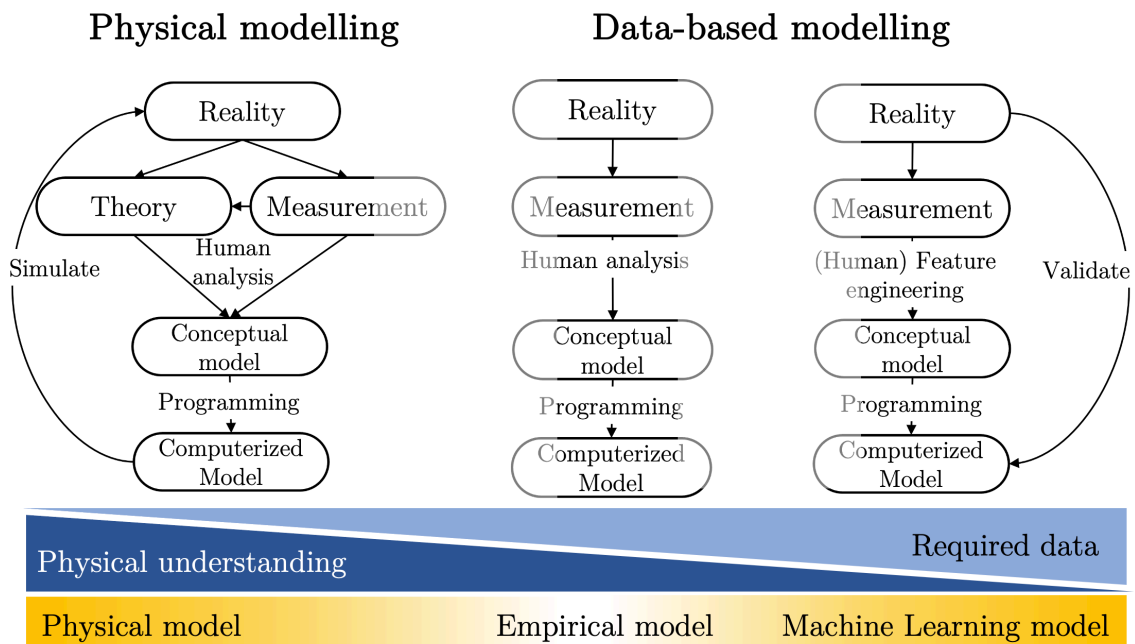


Figure 2.2: Spectrum of modeling approaches from physical models and empirical models to data-based models.

Many scientists consider this modeling philosophy and procedure the true way of scientific modeling. However, physical approaches rely on three necessary preconditions that make the modeling procedure possible in the first place: 1) A problem with human comprehensible complexity and a physical theory that either already exists or can be derived from the available measurement data or knowledge. 2) Measurements that allow the derivation of all important and influential features; and 3) a modeling problem that can be simulated with existing computational resources [57]. The three preconditions of cause influence each other and are themselves interdependent, e.g. measurement methods often depend on previous knowledge about the features of interest, which can only be derived if the problem itself is comprehensible for current human understanding. Also, if the necessary theory does not fully exist yet, the available measurement methods are not able to identify all influential features, or the problem exceeds the current human understanding. If this is the case, reality can be approximated using data-based approaches.

Data-based approaches approximate, meaning they do not exactly simulate the underlying mechanisms of reality, they imitate them based on previously made observations using correlation algorithms [58]. They therefore do not rely on a full existing theory and measurements with all influential features, but rather try to approximate the problem with available formulas, data and computational resources. The conceptual data-based model is not primarily derived from human understanding and identified physical formulas but directly from the observed data [57]. The approach thereby assumes that the available measurement methods are able to capture data, which intrinsically provides enough variance and influential features that a sufficient conceptual and mathematical model can be derived from it. Data is thereby generally needed in greater amounts. Data-based models can be differentiated into empirical and Machine Learning models [57]. For empirical models, the conceptual model is derived from human analysis such as investigating the data using statistical analysis and fitting a selected, often low dimensional mathematical model with foreseeable course. The influence and general correlations of the features are thereby often known and built implicitly in the computerized model, see middle of Figure 2.2. For Machine Learning models, the correlations of features and target value are in most cases unknown and the conceptual and computerized model is more flexible to allow an optimal correlation. Human analysis is thereby often limited to the selection and engineering of appropriate features. In many cases the conceptual model can directly constitute the computerized model, and all feature target relations are purely derived from the data, see Figure 2.2 [57]. However even for Machine Learning models meta information and domain knowledge can be directly built into model, e.g. via hybrid modelling [59–62]. For the case of physical applications, constraints, relations or trends can be built into or constrained in the model development.

The level of known theory incorporated into a model and the extent of human analysis can vary across different use cases. The boundaries between physical, empirical, and Machine Learning-based modeling approaches often become blurred, making it more appropriate to view them as a spectrum rather than distinct categories. The bottom part of Figure 2.2 shows a schematic illustration of this spectrum from, physical models to data-based models. Generally physical models require more physical understanding and less data for model development and validation, while data-based models generally require more data and less understanding. Examples of physical models are Newtonian mechanics [63] and kinetic gas theory [64]. Empirical models are often found in engineering e.g. for the correlation of the Reynolds and Prandtl number to calculate the heat transfer coefficient of a heat exchanger [65]. Examples of Machine Learning are found in the fields of natural language processing [66] and computer vision [66], research fields where the underlying investigated phenomena are often not completely understood and have to be approximated. This is also the case for the field of chemoinformatics, particularly the modeling of physicochemical properties of pure compounds, mixtures and fuels and the reason for the use of data-based models in this work. Similar to the field of computer vision, the complex underlying laws in this case the relations between intra- and inter-molecular relations on physicochemical properties required the use Machine Learning algorithms for a long time. Famous Machine Learning based modeling methods in cheminformatics are the Group Contribution method of Joback [67] and the Unified Functional Group Activity Theory (UNIFAC) [68]. Beyond the complexity of the modeling problem, limitations in the compositional analytics of fuels, as the unidentified isomers explained in Section 1.2.3 also prevent exact knowledge of the fuel composition, which makes an approximation of the fuel necessary.

2.2 Fuel property modeling methods

2.2.1 QSPR with sampling method

Sampling methods model a fuel as a mixture of components. Since GCxGC can presently not identify all isomers contained in the fuel, the composition has to be approximated by representative components for each chemical family and carbon number in the GCxGC matrix. These components are thereby assumed to be part of the fuel. Up to this point, a proven list of possible representative molecules that describe the variability of unidentified isomers relevant for jet fuels does not exist. Such a list might furthermore strongly depend on the respective fuel production process. Yang et al. therefore approximated the fuel composition by using all components tabulated in physicochemical property databases [48]. The method of Yang et al. samples measured property values from databases like the NIST Standard Reference Database 103a [69] or the DIPPR 801 database [70]. The method has great extensibility, since new hydrocarbon families, can easily be added if property measurements of the respective hydrocarbon

family are available. Yang et al. illustrated the predictive capability of this measurement sampling method for the modeling of 20 SAF using measurements of up to 1223 molecules. However, the method is strongly limited by the availability of property data. If no measurements are available for a certain chemical family at a certain carbon number, measurements of similar components with a similar number of carbon atoms have to be chosen. Furthermore, outliers and erroneous measurements have to be filtered before the sampling. This can induce a modeling error. For this reason, this work utilizes property models that predict the physicochemical properties of all components assumed to be in the fuel. The properties of the fuel components are predicted from structural information of the molecules and are therefore not limited to the availability of measurement data, assuming accurate model predictions.

There exist various approaches for the modeling of physicochemical properties based on the structural information of molecules. Saldana et al. demonstrated the applicability of Quantitative Structure-Property Relationship (QSPR) models for the modeling of pure compounds, which could be present in fuels as components. Models were developed for most of the desired properties, considering compounds from all relevant hydrocarbon families [71–73]. Following the conclusions of Saldana et al., the QSPR method is chosen for the property modeling of fuel components in this work.

The QSPR modeling method postulates a relationship between the structure and physicochemical properties and assumes similar property values for components with similar structures and substructures. In this approach, quantitative structures like the number of contained carbon atoms or the number of methyl groups are quantified for every molecule. This approximates the molecular structure by a quantitative description, that can be correlated with the property. The selection of the molecular features depends on the respective modeling task. From conclusions drawn in previous work by Saldana et al. [71–73] solely functional group count descriptors are considered. These descriptors quantify relevant functional chemical substructures that characterize the molecules of the considered chemical families. The substructures are quantified based on the simplified molecular input line entry system (SMILES) of the respective molecule, using SMILES arbitrary target specifications (SMARTS) [74]. Figure 2.3 shows a visual example of the quantification of the functional count descriptors for 2,3-hydro-2-methyl-1h-idene. The colored substructures and the count of the occurrence of the quantitative structure SMARTS key are displayed in the legend: e.g. for the substructure [CX4H3], the count of 1 indicates that the molecule contains one methyl-group, while [R] 9 shows that the molecule has 9 ring atoms etc. A full table of the utilized 49 substructures, the corresponding SMART key and a description is listed in the Supplementary Material C. The SMART keys are quantified using the RDKit Python package [75].

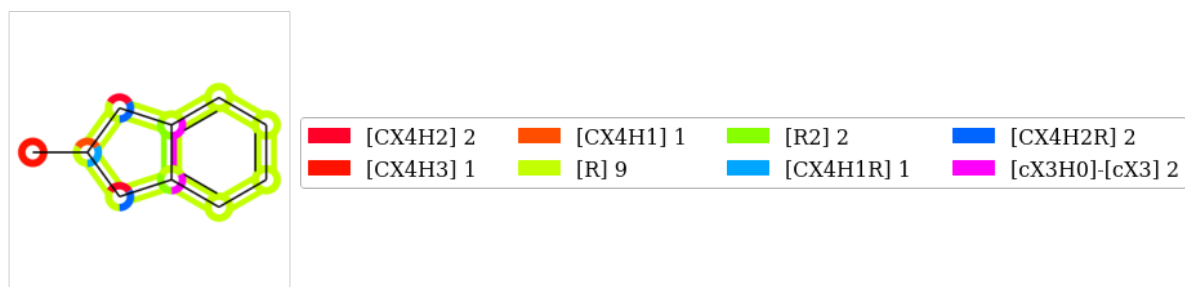


Figure 2.3: Quantified molecular features of 2,3-hydro-2-methyl-1h-idene, number behind SMART key shows count of molecular feature[76].

A schematic representation of the whole process, from the quantification of the molecular structure of the components to the prediction of the property and the estimation of the bulk property with random sampling is shown in Figure 2.4. The process is shown for three exemplary isomers of the iso-alkanes with 10 C-atoms and four exemplary substructures.

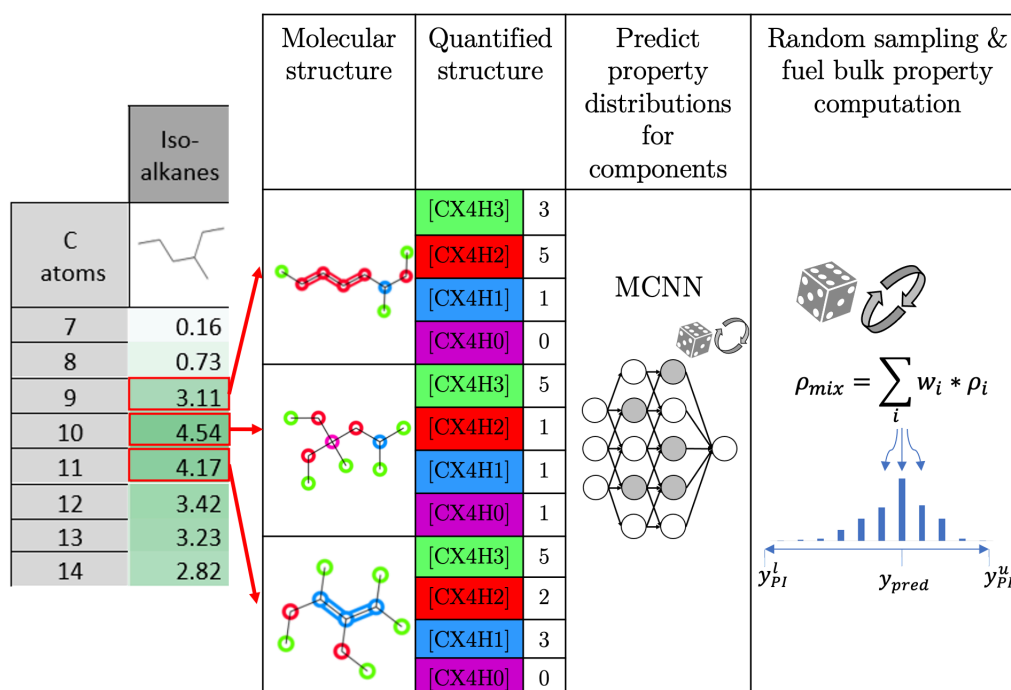


Figure 2.4: Schematic illustration of QSPR sampling modeling method from quantification of the molecular structure of components to the property estimation of the fuel.

In the first step, compounds that are assumed to be present in the fuel as fuel components based on the GCxGC measurement, are selected and converted into the quantified structure representation by the functional count descriptors. Up to 1870 compounds are considered, as described in Chapter 3, and assumed to be representative and potentially present in the fuel as components. In the second step, property values of all components are predicted with a trained probabilistic QSPR property model, computing a distribution of property values for each component. The pure compound property models are trained on the measurement data of pure compounds from the utilized database described in Chapter 3. Unity plots of the cross-validation

results of the training and testing for fuels and pure compounds of all models can be found in the Supplementary Material E, Figure E 1 to Figure E 39. In a last step, random samples are drawn from the predicted value distributions of randomized isomers, as indicated by the rolling dice on the far right in Figure 2.4. The property samples are multiplied with the fraction given by the GCxGC measurements and summed up using a suitable mixing rule to compute the bulk property of the fuel. By randomly sampling values from the predicted value distributions of random isomers, both uncertainties from the predicted property values of the individual component and uncertainty due to unidentifiable isomers are considered. The sampling process is repeated 100 times, which proved sufficient, to create a value distribution of the bulk property of the fuel. This distribution is described by the median value y_{pred} and the lower and upper prediction intervals y_{PI}^l and y_{PI}^u with 95 % confidence intervals.

The modeling of different fuel properties requires suitable mixing rules for each property, which are presented in the following. All mixing rules were chosen based on experience and literature recommendations [77]. For the modeling of the density, surface tension, net heat of combustion, flash point, freezing point and cetane number, a linear mixing rule after Equation (2.1) is used, where the bulk property value y_{mix} is calculated by summing over the predicted property value y_i of the individual components weighted by their fraction in the fuel w_i . All isomers are assumed to be present with the same probability and are not weighted individually. For the density, surface tension and net heat of combustion the mass fractions of the components are used in Equation (2.1). For the cetane number, the mole fractions and for flash point and freezing point the volume fractions of the components yield the best results. The necessary densities for the calculation of the volume fractions are predicted with the QSPR density property model for each component.

$$y_{mix} = \sum_i w_i y_i \quad (2.1)$$

For the kinematic viscosity the Grunberg-Nissan mixing rule without binary interaction coefficient is utilized, see Equation (2.1) [78].

$$\ln(v_{mix}) = \sum_i w_i \ln(v_i) \quad (2.2)$$

To predict the viscosity at low temperatures (< -10 °C) a fitting formula proposed by the ASTM D341 [79] see Equation (2.3) and Equation (2.4) is utilized. The predicted viscosity values at -10, 0, 20, and 60 °C are thereby utilized as base points using the QSPR viscosity model. Afterwards, a linear Bayesian regression model is used to fit the linearized formula of Equation (2.3) and draw random samples from the fitted distribution of the regression model, to estimate the property value distribution.

$$\log_{10}(\log_{10}(Z)) = A - B \log(T) \quad (2.3)$$

$$Z = v + 0.7 + \exp(-1.47 - 1.84v - 0.51v^2) \quad (2.4)$$

For the calculation of the distillation line, the simulated distillation approach after ASTM D2887 [80] is chosen. The distillation of the fuel is thereby estimated from the predicted boiling points of the individual components. In an iterative process, the temperature of the simulation is thereby increased and components are assumed to evaporate immediately when the temperature reaches the predicted boiling point. The predicted distillation line is then corrected using the correction formula defined by ASTM D2887. To calculate the evaporated volume, the corresponding densities of the components are calculated at 15 °C.

2.2.2 Direct Correlation method

In the Direct Correlation method (DC), the two-dimensional matrix of the GCxGC composition measurement is reshaped into a one-dimensional vector and directly used for the correlation with the measured properties. Assumptions about representative components, data about pure compounds and appropriate mixing rules, as for the QSPR sampling model, are not necessary. The property model is directly trained and validated solely on the fuel data. The probability distribution of property values is calculated directly by the probabilistic regressor, which predicts both a mean value and uncertainties. The uncertainties are thereby derived purely from the measurements. The utilized probabilistic regression algorithm will be introduced in Section 2.3. To describe the distribution, again the median value y_{pred} and the lower and upper prediction intervals y_{PI}^l and y_{PI}^u with 95 % confidence are used, likewise to the previous method. A schematic representation of the reshaping of a GCxGC fuel measurement and the property prediction is shown in Figure 2.5.

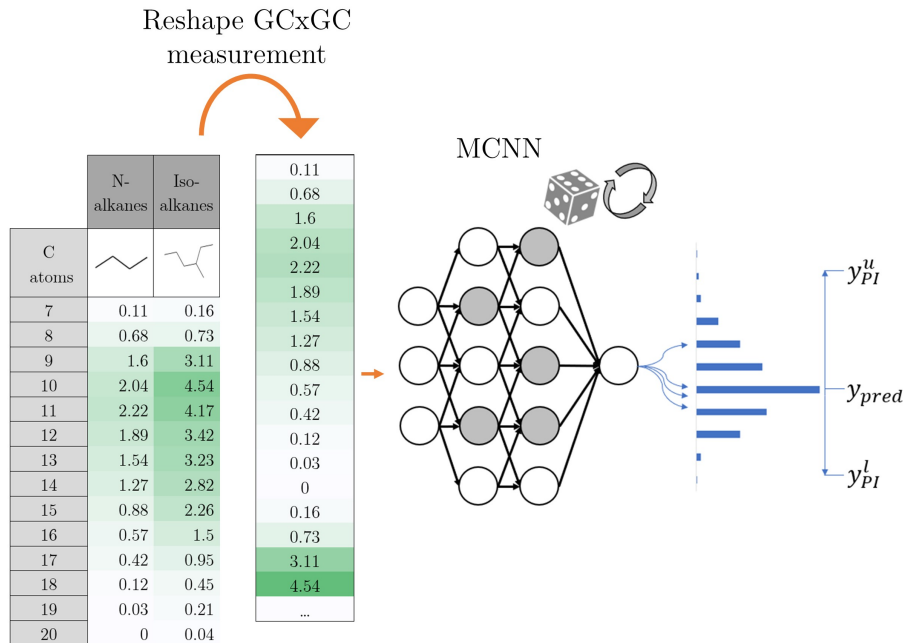


Figure 2.5: Schematic illustration of direct correlation modeling method.

The limitation of the DC method to only train on fuel data restricts its flexibility and extensibility. Up to this point the number of available GCxGC fuel measurements and corresponding property measurements is low (in the few hundreds until now) compared to the data available for the pure compounds. The potential composition and value range for training the property model is therefore very limited. Combined with the high number of input features (8 families * 25 carbon atoms = 200) this can lead to potential overfitting of the model and the limitation of a high predictive capability of the property model to the training and validation domain only. As a result, predictions outside of the training and validation domain can consequently deviate significantly from the true values. Since the direct correlation method estimates uncertainty based on the intrinsic noise of the training data, uncertainties due to isomers and measurements are only estimated, if they exist in the training data. Hence, multiple GCxGC measurements with different isomers and property measurements from different laboratories have to be available to estimate these uncertainties as part of the predicted uncertainty of the model. If this is not the case, the model only estimates the intrinsic uncertainty of the training data, e.g. the distribution and distance of the individual datapoints.

2.2.3 Mean Quantitative Structure-Property Relationship method

The Mean Quantitative Structure-Property Relationship method (M-QSPR) is developed to overcome limitations and requirements that inherently exist for the two other modeling methods: The availability of data and the requirement of mixing rules to describe the interactions of the fuel components. Based on the method of Ajmani et al. [81] of mole averaged quantitative structure descriptions for mixtures, the M-QSPR approach extends the approximation of the structure of molecules by a quantified structure representation to fuels. The fuel is thereby not modelled as a mixture of components but as a mole averaged quantitative pseudo structure. This pseudo structure representation is described by the same 49 structural features used for the QSPR sampling method, see Supplementary Material C. This pseudo structure representation can be directly used for the correlation with the physicochemical property together with the data of pure compounds. A M-QSPR model can therefore be trained on both pure compound data like a QSPR model and fuel data like a direct correlation model. This increases the size of available training data and allows for a direct learning of the mixing behavior, resolving the need for an appropriate mixing rule.

To calculate this M-QSPR representation, the molar fractions of the GCxGC measurement are multiplied with the average occurrence of the 49 structural features of all possible isomers of a family at a certain carbon number and then summed up over all cells in the GCxGC matrix. The mean occurrence of the quantitative structures is precalculated for all possible cells in the GCxGC, creating a mean occurrences matrix for all structural features, which speeds up the calculation of the M-QSPR representation. Likewise to the QSPR sampling method, all isomers are assumed to be representative and potentially present in the fuel, both for conventional and synthetic fuels. However, if further information about the possible isomers of a certain fuel or a

certain production pathway exist, an individual mean occurrence matrix can be calculated for the composition measurement or the production pathway. A schematic illustration of the calculation of the quantitative pseudo structure representation is given in Figure 2.6 for the example of iso-alkanes with 8 carbon atoms and the methyl group descriptor [CX4H3]. The occurrence of the [CX4H3] descriptor in all iso-alkanes 8 isomers N is thereby averaged, multiplied with the respective molar fraction of the GCxGC measurement and summed up for all GCxGC cells.

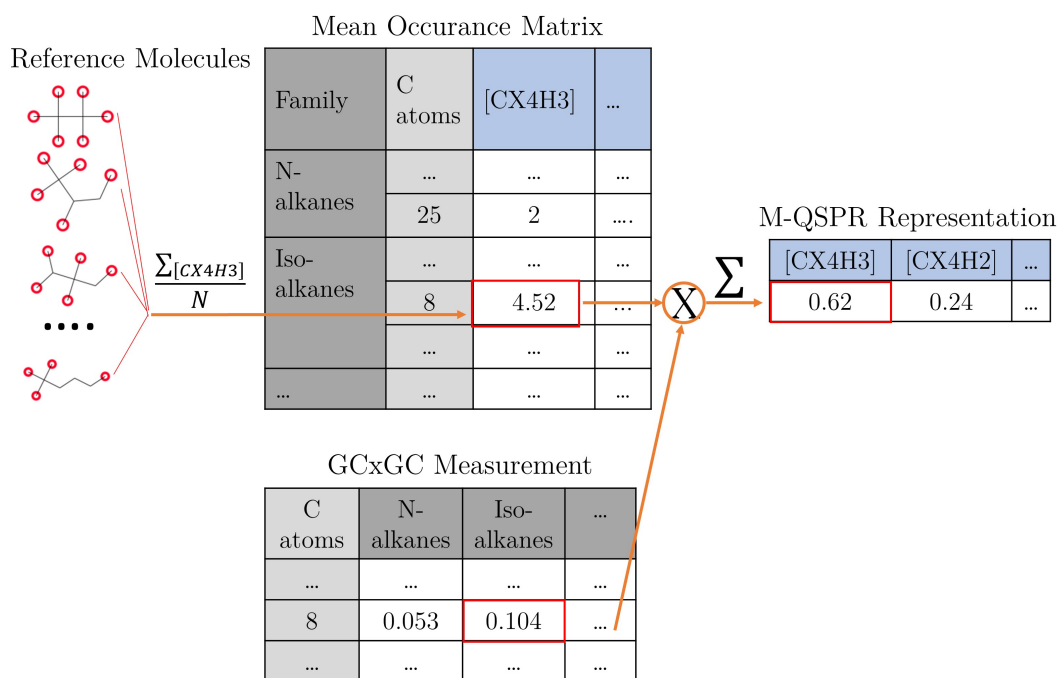


Figure 2.6: Schematic illustration of pseudo mean quantitative structure estimation of a jet fuel using GCxGC measurement and a mean occurrence matrix calculated from averaging structural features of possible isomers.

The M-QSPR representation can be directly used for the training and validation of the model together with pure compound data in their QSPR representation, because they share same feature space. The distribution of possible property values can then be directly predicted by a probabilistic regressor, similar to the DC method. The distribution is again described by the median value y_{pred} and the lower and upper prediction intervals y_{PI}^l and y_{PI}^u with 95 % confidence.

2.3 Probabilistic Machine Learning correlation models

Each of the presented modeling methods relies on a correlation model to correlate the representation of the fuel compositions (reshaped GCxGC for the DC method, mean quantitative pseudo structure for the M-QSPR method and quantitative structure of components for the QSPR sampling method) with the property values of pure components and fuels. For all methods in this work, the deep neural network algorithm with the Monte-Carlo dropout technique (MCNN) is used. Apart from the MCNN, other regression algorithms, like the Gaussian Process

Regressor, Bootstrap Neural Networks, Probabilistic Support Vector Machine Regressor were investigated. However, the MCNN showed the best and most consistent performance over all considered modeling methods.

The MCNN algorithm was developed by Gal and Ghahramani [82] and uses an artificial neural network (ANN) for the correlation of the input features with the desired output. In the following, the working principles of the ANN and the MCNN are presented, followed by descriptions of the validation and hyperparameter optimization methods for Machine Learning models.

2.3.1 Working principles of Artificial Neural Networks

ANNs mimic biological neural systems. They are composed of connected layers of individual neurons that propagate an input signal if a threshold of an activation function is exceeded. The connections are weighted to regulate the importance of inputs [83]. To illustrate the inner workings of an ANN, the left part of Figure 2.7 shows a schematic representation of a single neuron calculating the output y from multiple inputs x_i . In this neuron, the inputs x_i are multiplied with an individual weight w_i summed up under the addition of a bias b and fed through an activation function σ to calculate the desired output y , see Equation (2.5) [83]. In the case of the direct correlation method, y represents the properties and would x_i represents the individual composition features (e.g. mass fractions of the GCxGC or structural features for the QSPR and M-QSPR models). If the activation function σ would be unity function, multiplying the term in the brackets of Equation (2.5) just by one, this single neuron model would be equivalent to a multi-linear regression.



Figure 2.7: Schematic figures of artificial neural network neuron (left), connected artificial neural network (right)

$$y(x) = \sigma \left(\sum_i^n w_i x_i + b \right) \quad (2.5)$$

In general, other activation functions are used for σ , that only propagate the signal if a threshold value is exceeded. This allows the correlations of complex non-linear relations between the input features x_i and the output variable y , which can excel the capabilities of multilinear regression models. For this work, four different activation functions (Sigmoid, ReLU, ELU und CELU) are utilized. They are described in the respective equations Equation (2.6) to Equation (2.9). z thereby abbreviates the terms in the brackets of Equation (2.5).

$$\sigma = \frac{1}{1 + e^{-z}} \quad (2.6)$$

$$\sigma = \max(0, z) \quad (2.7)$$

$$\sigma = \begin{cases} z & ; z > 0 \\ e^z - 1 & ; z \leq 0 \end{cases} \quad (2.8)$$

$$\sigma = \max(0, z) + \min(0, e^z - 1) \quad (2.9)$$

If the single neuron model is extended both by multiple neurons in a layer j , and multiple stack layers k , a multilayer ANN, also called deep neural network (DNN) is generated, as shown on the right side of Figure 2.8. The individual output of the previous layer is thereby the input of the next layer. Equation (2.10) and Equation (2.11) show the formulas for the calculation of the output y^k for a layer of an ANN and the calculation of the final output y for the ANN for a two layered deep neural network.

$$y^k(x) = \sigma^k \left(\sum_j^m \sum_i^n (w_{ij}^k x_i^{k-1} + b_j^k) \right) \quad (2.10)$$

$$y(x) = y^k(y^{k-1}(x)) \quad (2.11)$$

Before accurate predictions with ANN are possible, the respective model parameters have to be fitted / trained based on the input and output data of a dataset [83]. For neural networks, the parameters adjusted in the training are the weights of the individual neuron connections w_{ij}^k and the biases of the neuron b_j^k . The optimization is conducted by minimizing a loss function over all data points M . In the scope of this work, the mean squared error loss function is used, see Equation (2.12), which computes the squared deviation of the prediction for all data points $y(X_m)$ to the measurement values \hat{y}_m .

$$L(X, y) = \frac{1}{2M} \sum_m^M (y(X_m) - \hat{y}_m)^2 \quad (2.12)$$

The parameters are adjusted or optimized over a set number of training intervals using the method of backpropagation. In the backpropagation, the model predicts a property value for each data point in a forward pass, calculates the loss using the defined loss function and then adjusts the individual weights and biases in a backward pass accordingly [83]. The individual changes Δw_{ij}^k and Δb_j^k are calculated from the derivatives of loss function to the respective parameters $\frac{\partial L}{\partial w_{ij}^k}$ and $\frac{\partial L}{\partial b_j^k}$, see Equation (2.13) and Equation (2.14). The learning rate ε weights the calculated change of an iteration. The respective derivatives for the individual weights and biases are calculated using the chain rule, see Equation (2.15) to Equation (2.20). The local gradient $\frac{\partial L}{\partial y_j^k}$, which states the individual change of the loss function for a change of the predicted

values, is calculated by applying the chain rule to the nested function in Equation (2.11) for each respective neuron [83,84].

$$\Delta w_{ij}^k := \varepsilon \frac{\partial L}{\partial w_{ij}^k} \quad (2.13)$$

$$\Delta b_j^k := \varepsilon \frac{\partial L}{\partial b_j^k} \quad (2.14)$$

$$\frac{\partial L}{\partial w_{ij}^k} = \frac{\partial L}{\partial y_j^k} \frac{\partial y_j^k}{\partial w_{ij}^k} \quad (2.15)$$

$$\frac{\partial y_j^k}{\partial w_{ij}^k} = x_i^{k-1} \quad (2.16)$$

$$\frac{\partial L}{\partial w_{ij}^k} = \frac{\partial L}{\partial y_j^k} x_i^{k-1} \quad (2.17)$$

$$\frac{\partial L}{\partial b_j^k} = \frac{\partial L}{\partial y_j^k} \frac{\partial y_j^k}{\partial b_j^k} \quad (2.18)$$

$$\frac{\partial y_j^k}{\partial b_j^k} = 1 \quad (2.19)$$

$$\frac{\partial L}{\partial b_j^k} = \frac{\partial L}{\partial y_j^k} \quad (2.20)$$

The minimization of the loss function and adjustment of the weights and biases are carried out by an optimizer. In the scope of this work the adaptive moment estimation (ADAM) is utilized, a proven and widely used optimizer for regression tasks [85].

2.3.2 Working principle of Monte-Carlo Dropout Neural Networks

ANN regression models are deterministic models by default, meaning that they only predict one value. To extend an ANN to a probabilistic model, that predicts a distribution of possible values, the concept of Monte-Carlo Dropout Neural Networks by Gal and Ghahramani is utilized [82]. This concept applies the technique of randomly deactivating neurons of the ANN, a technique frequently utilized to prevent the overfitting of the model during the training, also during the prediction. The general equation of an ANN with the dropout technique is represented with a Bernoulli distribution for each neuron $P(w_{ij}^k)$, that returns 0 and 1 based on a set dropout probability, see Equation (2.21) [82]. The dropout probability is thereby a hyperparameter that demands tuning to optimally align with the specific requirements of the use case.

$$y^k(x) = \sigma^k \left(\sum_j^m \sum_i^n P(w_{ij}^k) (w_{ij}^k x_i^{k-1} + b_j^k) \right) \quad (2.21)$$

A prediction of the ANN is thereby repeated multiple times in a Monte-Carlo sampling, each time deactivating random neurons, which produces varying values that resemble a probability distribution. A schematic illustration of the prediction process of the MCNN is given in Figure 2.8.

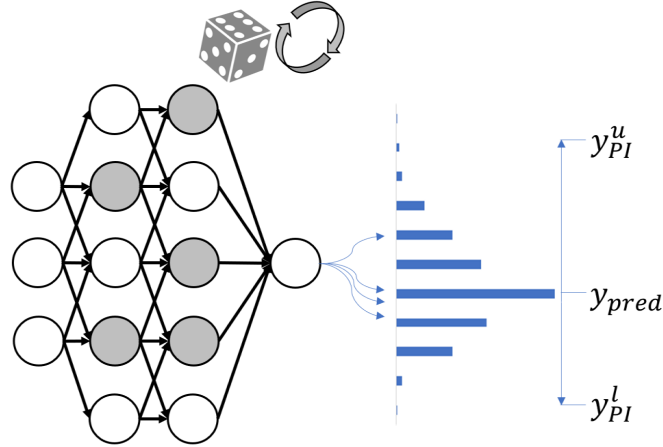


Figure 2.8 Schematic representation of MCNN with dropout functionality during prediction. Network neurons are deactivated randomly (gray) to generate a distribution of prediction values.

The predicted distribution is interpreted by Gal and Ghahramani as a Bayesian approximation, that describes the prediction uncertainty due to noise in the data (conflicting measurements) as well as dissimilarity between the training and test data. The predicted distribution therefore reflects the certainty of the prediction, with a narrow distribution indicating a certain prediction and a wider distribution indicating an uncertain one [82]. The distribution itself is described by the mean value y_{pred} and the lower and upper prediction intervals y_{PI}^l and y_{PI}^u . The validity of the predicted uncertainty intervals as prediction intervals (PI), is reviewed as part of the predictive capability assessment of the models. The estimated PI therefore have to enclose a set percentage of values for a set percent of certainty, e.g. 95 % of measurements for 95 % of certainty. The MCNN is written in Python using the PyTorch library [86].

2.4 Model development and validation

The development of a Machine Learning model consists of three main steps: the selection and configuration of the model, the training and validation of the model and finally its application [66]. In the first step, the utilized model and the configuration of its respective hyperparameters have to be selected. Hyperparameters are parameters that define a model but are themselves not adjusted during the training process, e.g. for ANN the number of layers and number of layers per neuron. In the second step the model is trained by adjusting the trainable parameters based on the provided training data and afterwards validated on test data not utilized in training. The

following section gives further insights into the utilized methods for the training and testing of the models as well as the prefixed hyperparameter optimization. Training and testing will thereby be explained first, since the utilized hyperparameter optimization method requires both processes.

2.4.1 Training and validation

As Machine Learning models infer the correlation of input and output variables solely based on the provided data, training and testing are of utmost importance for the applicability of the model. Both processes have to be conducted carefully, to validate that the model correctly captures the underlying problem and does not overfit to the provided training data. In the field of Machine Learning cross-validation (CV) is established and widely used concept for the validation of models on a set amount of data [66,87]. The CV process simulates the procedure of training and testing several times, using different training and test data derived from the same data set. The different training and test datasets are created by randomly shuffling the data and splitting it into set fractions for training and testing, e.g. 85 % for training 15 % for testing. The process is repeated over several folds. In each fold the model is trained on training data and subsequently tested on testing data. Figure 2.9 schematically illustrates the splitting of the data for a 4-fold CV.

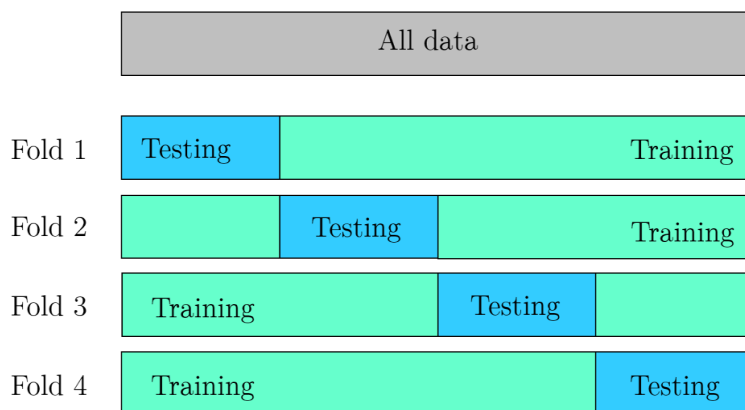


Figure 2.9: Schematic illustration of a cross-validation.

By comparing the predictions from several folds, potential overfitting of the model to the training data can be investigated and prevented. After a successful cross-validation, the model is trained on the complete dataset.

2.4.2 Hyperparameter optimization

As laid out in the introduction of this section, the first step of the model development is the selection of the model and its configuration. Since in this work the MCNN proved to be the best predictive model overall, only the configuration with the hyperparameters, e.g. the number of layers of the MCNN, has to be selected. The selection of the model configuration has to be carried out before the actual cross-validation on the full dataset, typically on a subset of the data.

Hyperparameters can be chosen based on prior or expert knowledge, a random grid search, or by the use of an optimizer that automatically adjusts the hyperparameters based on a defined loss function. In the scope of this work the hyperparameter selection by optimization is chosen, since no prior knowledge for the modeling of physicochemical properties with the utilized MCNN algorithm existed. Compared to a random grid search of different hyperparameter configurations, the hyperparameter optimization by optimization was found to yield better results in less time.

As optimizer, a Bayesian optimization with a Gaussian Process and Matern kernel is used. The underlying principles of the Gaussian Process were published [47] and will not be explained further as part of this work. In this optimization, the hyperparameters are adjusted over 30 iterations to find an optimal configuration of the MCNN for each modelled property. In every iteration a 4-fold-cross-validation is conducted using 800 samples with set fraction of 15% for the testing and 85 % for the training. The set numbers for the optimization iterations, training samples and fractions for the cross-validation proved to be sufficient and were not investigated further. Each iteration, the average over the prediction results of all cross validations is used to compute a loss according to a loss function, which will be explained in the following. The model configuration with the lowest loss over the set number of iterations is then utilized for the cross-validation on the full dataset. Figure 2.10 shows a schematic illustration of the hyperparameter optimization.

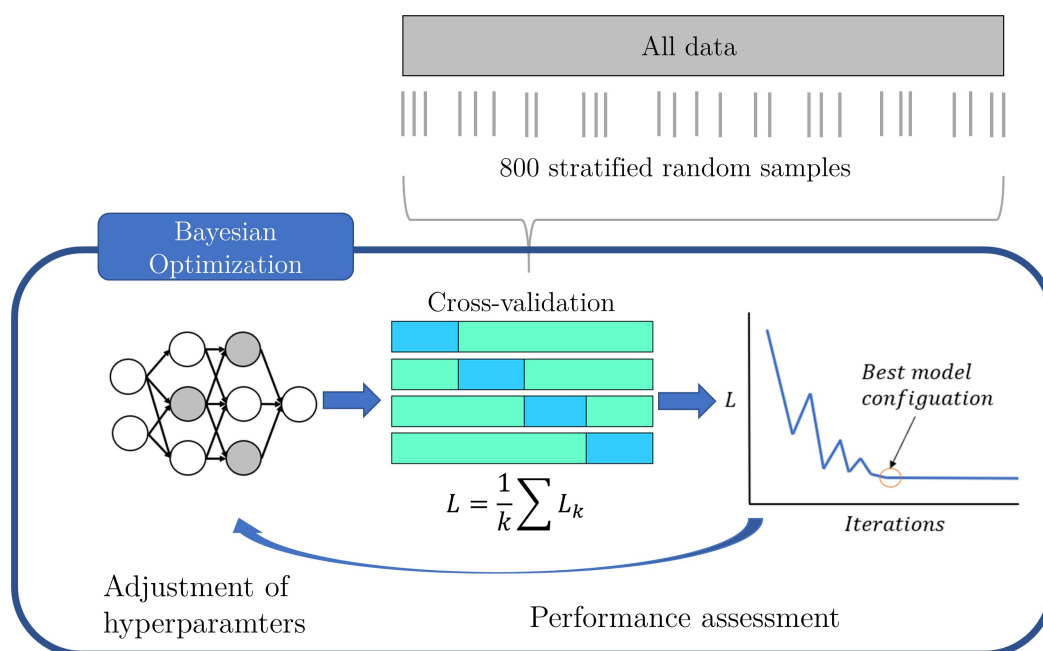


Figure 2.10: Schematic workflow of the utilized hyperparameter optimization with cross-validation.

As hyperparameters, the topology of the network, e.g. the number of neurons per layer, as well as the training conditions, e.g. the learning rate are optimized. Additionally, the choice of an adequate scaler for the scaling of the input and output data is considered as a hyperparameter. As scalers the *MinMaxScaler* and *RobustScaler* are utilized, see Equation (2.22) and Equation

(2.23). The complete list of hyperparameters as well as their considered options for the optimization are summarized in Table 2. The variable num_{feat} denotes the number of input features for the model, e.g. for the QSPR model for the flash point 49, while the count of num_{feat} in the brackets indicates the number of layers.

Parameter	Optimization range
(Neurons per Layer)	$(num_{feat}), (num_{feat}, num_{feat}), (num_{feat}, num_{feat}, num_{feat}),$ $(num_{feat}, num_{feat}, num_{feat}, num_{feat}), (2*num_{feat}), (2*num_{feat},$ $2*num_{feat}), (2*num_{feat}, 2*num_{feat}, 2*num_{feat}), (2*num_{feat},$ $2*num_{feat}, 2*num_{feat}, 2*num_{feat}),$
Droprate	$0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$
Learning rate	$1e-4, 1e-3, 1e-2$
Learning weight decay	$0, 1e-8, 1e-6, 1e-4, 1e-2$
Number of training epochs	$500, 1000, 1500, 2000$
Activation function	$ReLU, ELU, Sigmoid, CELU$
X feature scaler	$None, MinMaxScaler, RobustScaler$
y feature scaler	$None, MinMaxScaler, RobustScaler$

Table 2: Utilized parameters for hyperparameter optimization of Monte-Carlo Neural Networks

Utilized Scalers are:

$$MinMaxScaler = \frac{x_i - \bar{x}_i}{x_{i,max} - x_{i,min}} \quad (2.21)$$

$$RobustScaler = \frac{x_i - Q_{50}(x_i)}{Q_{75}(x_i) - Q_{25}(x_i)} \quad (2.22)$$

As outlined, the hyperparameters are adjusted according to a loss function. For this work, a custom function is implemented, see Equation (2.25). The loss function is based on the Root Mean Squared Error (RMSE), see Equation (2.24), and furthermore includes the two predictive capability metrics PICP and NMPIW that will be explained in the next Section 2.5.1. For the PICP a target confidence level of 95 % is set for the confidence level of the model. The NMPIW relates the precision of the prediction to the reproducibility limits of the measurements. To constrain the maximum influence of the precision of the predictions on the hyperparameter optimization, a maximum of 2 is set in Equation (2.25). Without this constraint the hyperparameter optimization could tend to dominantly focus on the precision via the NMPIW and disregard the validity of the PI via the PICP. The optimization aims at returning hyperparameters of an optimal model that is as accurate and precise as possible. The most optimal model will however probably not comply with all set thresholds of accuracy, validity and precision since the loss function describes a trade-off between the metrics. The introduction of

additional weights for the different metrics in the loss function could force the compliance of one metric to a critical threshold, if required. The Bayesian optimization is carried out using the Python library `scikit-optimize` [88].

$$RMSE = \sqrt{\frac{1}{n_{Test}} \sum_{i=1}^{n_{Test}} (y_{pred,i} - y_{test,i})^2} \quad (2.24)$$

$$L_{opt} = RMSE \left(1 + \frac{\max(0, 95 - PICP)}{100} + \min\left(2, \frac{\max(0, NMPIW - 100)}{100}\right) \right) \quad (2.25)$$

2.5 Predictive capability assessment methods for models

The screening and design of jet fuels solely on the basis of model predictions makes a comprehensive assessment of the predictive capability of the models necessary. It has to be determined if the predictions of a model comply with the requirements of the modeling problem. Oberkampff and Roy define the predictive capability as “extrapolation or interpolation of the model to specific conditions defined by the intended use of the model” [57]. The predictive capability of a model and the decision of its adequacy are therefore always relative to the actual use case itself. The use case defines the application domain of a model, which thus defines the requirements to determine the adequacy of the model [57]. The predictive capability of the models has to be quantified by suitable metrics and compared to requirements that are derived from the intended use case.

The predictive capability assessment proves to be an everlasting problem since data for the model assessment of the whole application domain is most of the time not available. Figure 2.11 visualizes this problem on the example of two schematic domains, the application domain A and the validation domain V. V thereby often only encloses a subfraction of the intended application domain, that a model can be trained and validated on. Especially for the use case of jet fuel screening, this is a present problem since the composition of the new fuel candidates often lies outside the known domain. The adequacy of the developed models is therefore conducted as a two-step process in this work. In the first step, the models are trained and validated on property measurements, determining the predictive capability by metrics. In the second step, the intended use case of fuel prescreening is simulated by predicting properties for the screening of fuels excluded from training and validation data. This will be demonstrated in the predictive capability assessment of the models in Section 4.2.

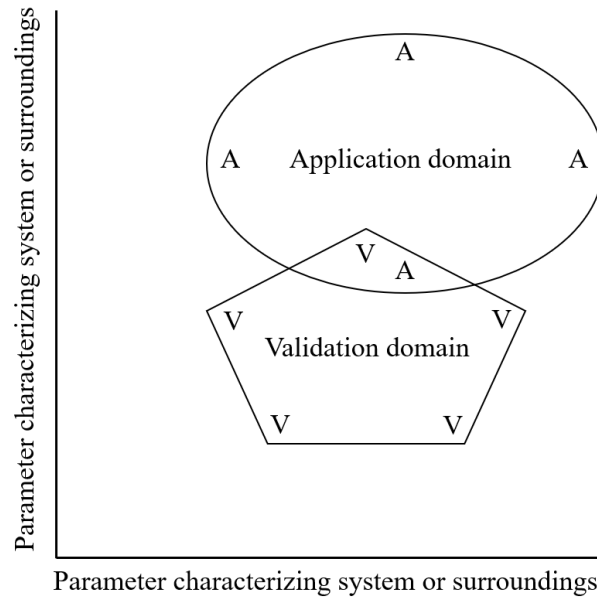


Figure 2.11: Schematic representation of application and validation domain of a model, adapted from [57].

2.5.1 Predictive capability metrics

Rather than predicting just a single value as deterministic models, probabilistic models predict a distribution of possible values. This characteristic requires special metrics that quantify the different aspects of probabilistic model predictions and allow an assessment and comparison of the models [76,89,90]. In the scope of this work, three aspects are considered: accuracy, validity and precision. Accuracy measures the average proximity of the predicted from the true values, precision the size of the predicted distribution and validity describes if the true values are on average enclosed by the predicted distribution. To illustrate the three aspects, Figure 2.12 gives visual examples on the basis of shots on target signs. The target sign illustrates the distance to the true test value at the center, also indicated as an orange line down below.

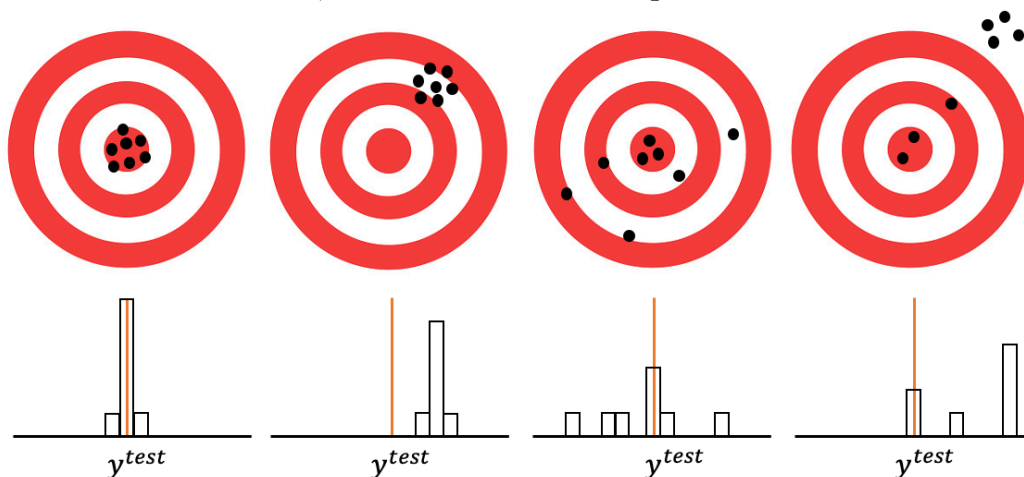


Figure 2.12: Schematic illustration of predictive capability aspects accuracy, validity and precision for probabilistic models.

A prediction should in the best case be highly accurate, valid and precise. This is shown by the first target sign on the outer left, where all predictions lie on the target sign close to the center. However, the three aspects are not necessarily linked together as the other cases and their target signs visualize. For the second case, all predictions are precise because the predictions are in close distance to each other, however they lie far from the center of the target sign and therefore have low accuracy. In the third case from the left, the accuracy is higher, with the majority of the predictions in close proximity to the center. However, the distribution is wide and therefore the predictions are unprecise. In the fourth case from the left, the majority of the predictions lie completely off the target sign, the distribution is therefore not valid.

In Figure 2.12, the predicted distributions are represented by a collection of only a few samples. In reality, Monte-Carlo sampling returns an almost continuous distribution of samples. The distribution can thereby be characterized by the value of the highest likelihood, i.e. the mean value \bar{y} , and a lower and upper prediction interval (PI). The PI mark the limits of the distribution for a set confidence level. Figure 2.13 shows a schematic distribution with an indicated mean value \bar{y} , lower and upper limit y^l and y^u of the PI, for a set confidence level of 95 %. The PI enclose the set certainty range, while the distribution range outside the PI resemble the existing associated risk of 5% for a confidence level of 95 %.

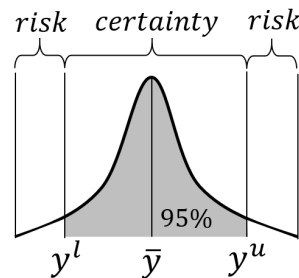


Figure 2.13: Schematic predicted distribution of a probabilistic model with mean prediction \bar{y} and lower and upper prediction intervals y^l , y^u for a confidence level of 95 % with the associated risk and certainty.

The three predictive capability aspects, validity and precision outlined above can be inferred from the distribution given in Figure 2.13. The distance of the mean value of the distribution to the measurement states the accuracy. The range of the distribution between the lower and upper limit of the PI reflects the certainty of the prediction; a narrow distribution indicates a certain prediction while a wider predicted distribution indicates an uncertain one [82]. The validity of the predicted distribution is estimated based on the accordance of the set level of confidence and the percentage of measurement values lying inside the PI of their corresponding prediction. E.g. for a predicted distribution with a set confidence level of 95 %, the PI that mark the lower and upper PI have to enclose 95 % of the measurements. If this is not the case, the distribution is invalid and the PI not meaningful.

To quantify the three predictive capability aspects, a concept is used that was developed and published for the assessment of probabilistic models in the safety-relevant area of jet fuel screening [76,89]. As metric for accuracy, the mean absolute error (MAE) of the mean of the predicted distribution $y_{pred,i}$ and the measurement value $y_{test,i}$ is calculated, see Equation (2.26). Compared to other accuracy metrics like the mean squared error or the root mean squared error it measures the average deviation, in the same unit of the prediction, weighting all validation values equally.

$$MAE = \frac{1}{n_{Test}} \sum_{i=1}^{n_{Test}} y_{pred,i} - y_{test,i} \quad (2.26)$$

As a metric for the validity of the predicted distribution, described by the PI, the Prediction Interval Coverage Probability (PICP) is calculated, see Equation (2.27). The PICP states the average probability, that a measured value lies inside the predicted upper and lower PI y_{PI}^l and y_{PI}^u [90]. In (2.27) c_i is a boolean value; it is 1 if $y_{pred,i}$ lies inside the interval and therefore $y_{PI}^l < y_{test} < y_{PI}^u$ and 0 otherwise. If the PICP and the set confidence level of PI of the prediction are comparable, or if the PICP is even greater than the set PI, predictions do on average lie inside the PI and the PI can be considered valid. If the PI are valid both in training and testing, the PI is considered reliable.

$$PICP = \frac{1}{n_{Test}} \sum_{i=1}^{n_{Test}} c_i * 100 \% \quad (2.27)$$

As metric for the precision of the distribution, the Normalized Mean Prediction Interval Width (NMPIW) is chosen. The NMPIW calculates the mean width of the PI relative to a reference Δ_{ref} , see Equation (2.28) [91]. Δ_{ref} always has to be set individually for each use case. For the scope of this work the Δ_{ref} is set to the individual uncertainty of the property measurement methods, the reproducibility, to put the predicted uncertainty in reference to the uncertainty of the measurements. Δ_{ref} could also be set to a range of experience, as demonstrated in a recent paper, to determine the adequacy of models for a use case based on the estimated uncertainty [89]. The use of the metrics for the assessment of the predictive capability of probabilistic models will be explained in the following Section 2.5.2.

$$NMPIW = \frac{1}{n_{Test}} \sum_{i=1}^{n_{Test}} \frac{y_{PI,i}^u - y_{PI,i}^l}{\Delta_{ref}} * 100 \% \quad (2.28)$$

To handle cases where the predicted PI are not valid since the calculated PICP differs significantly from the set confidence level, an additional correction factor is utilized. The Mean

Absolute Outlier Error (MAOE) quantifies the mean deviation of outliers to the next upper or lower PI, see Equation (2.29). The size of invalid PI can therefore be increased by the MAOE to achieve valid PI on average at the price of an increased NMPIW. This metric is necessary if a probabilistic prediction may not be used safely because the set PICP differs significantly from the set confidence.

$$MAOE = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} c_i * \min(|y_{PI,i}^u - y_{test,i}|, |y_{test,i} - y_{PI,i}^l|) \quad (2.29)$$

2.5.2 Example for the predictive capability assessment

To illustrate the introduced concept of a predictive capability assessment as well as the influence of the utilization of the predicted PI for the application or probabilistic models, two examples are given in the following; 1) A general validation of a probabilistic model, 2) Evaluation and application of probabilistic models for the prediction of a critical property.

The first example illustrates the validation of the probabilistic models both visually and based on the introduced predictive capability metrics. Figure 2.14 shows the schematic graph of a unity plot, where the predicted values of a probabilistic models y_{pred} are plotted against the measurement values y_{test} . The unity line is plotted in solid black, whereas the reproducibility of the property measurement is indicated by grey dashed lines. For a probabilistic prediction to be valid, estimated prediction intervals have to cross the grey dash lines of the uncertainty region of the measurement. Predictions that comply with this constraint are considered valid and are colored blue predictions for which this is not the case are considered invalid and are colored red. To estimate the validity based on the introduced PICP metric, the percentage of valid predictions is related to the total number of predictions. If the calculated PICP percentage is equal to or higher than the set confidence of the probabilistic model e.g. 95%, the model is considered valid.

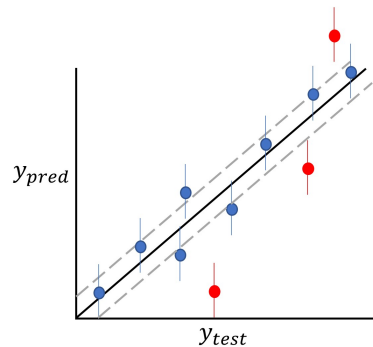


Figure 2.14: Schematic illustration of unity plot for validation of probabilistic model.

The second example in Figure 2.15 shows the application of probabilistic models for the use case of jet fuel screening for two example fuels: Fuel A and Fuel B. Similar to the example of jet fuel screening based on measurements in Figure 1.4 the predicted value has to lie inside the set

specification limits of ASTM D4054 and D7566, indicated in red. In contrast to the previous example, not only the predicted mean value, indicated as a solid dot, but also the estimated prediction intervals, indicated as error bars, have to lie inside the limits. In the given example, this is only true for Fuel A, whereas for Fuel B, the lower prediction interval crosses the lower specification limit, while the mean value still lies inside the specification limit. Therefore, for Fuel B there is a risk, which can be communicated to the stakeholders, that the prediction does not comply with the specification limit. If this risk is not acceptable, an additional property measurement is required.

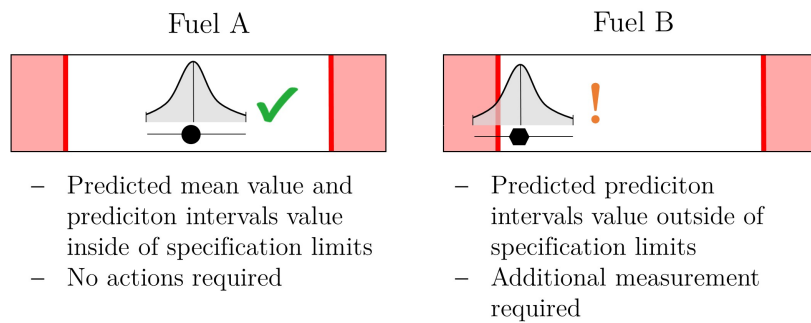


Figure 2.15: Schematic illustration of the screening plots for Fuel A and Fuel B as part of the jet fuel prescreening.

3 Composition and Property Database

Data is the cornerstone of this work. Both the development of the data-based models as well as the intended fuel design tools require respective datasets. For the development of the property models, both composition and property data of fuels and fuel components are necessary. The DC method requires data of fuels, the QSPR sampling method needs data of fuel components, and the M-QSPR method, able to train on both data from fuels and pure compounds, requires data from both. The investigation of the fuel composition and fuel properties for the fuel design tools are carried out on the basis of pure component data. The purpose of this work to develop tools that can be applied to safety-relevant use cases furthermore requires high reliability and uniformity of the data. In summary, the utilized data should be highly reliable, measured with identical or comparable measurement methods, and provided in a uniform format for both fuels and pure components.

Since at the time of writing no available database complied with the set requirements, providing data for both fuels and pure components for all eight critical fuel properties, a custom database is set up as part of this work. Data from multiple commercial, public and internal databases are collected, unified and stored in a central database using a standardized schema. The following chapter outlines the development of the database, the collection of the data, as well as the preprocessing and removal of potential outliers. In the last section, the database is characterized by estimating and visualizing the extent of the data to illustrate the database domain and compare it to the intended application domain of the use cases. Since the database is in part comprised of data from commercial databases, it is not published as part of this work.

3.1 Data collection

For jet fuels, the data is taken from the DLR Jet Fuel Database. This fuel database holds GCxGC composition and property data of 75 conventional crude oil-based jet fuels and 56 synthetic jet fuels at the time of writing. Data for conventional fuels mainly originates from the commercially available world fuel survey of 2006 by the Coordinating Research Council (CRC) [92]. The data of the synthetic fuels was systematically gathered from research projects like JETSCREEN, NJFCP and Emission and Climate Impact of Alternative Fuels (ECLIF) [8]. For the fuel components, the data is taken from the established commercially available pure component databases like the NIST Standard Reference Database 103a [69], DIPPIR 801 database [70] as well as the public databases Pubchem [93] and Chemspider [94]. Data from Pubchem and Chemspider are especially used for the properties flash point and freezing point. Data for the cetane number and the yield sooting index is taken from publications of the National Renewable Energy Laboratory [95,96] and Li et al. [97]. All in all, data of 1870 pure compounds are used for the training and validation of the QSPR and M-QSPR models. Table 3 gives a detailed summary

of the number of data points utilized for the training and validation of the models for each property. Column #FL denotes the number of unique fuels, and column #DP lists the total number of individual measurements. The summary is given for fuel components (Comp.), conventional fuels (Conv.), and non-conventional synthetic fuels and blends (Syn.) as well as the sum of all.

Properties	Comp.		Conv.		Syn.		All	
	#FL	#DP	#FL	#DP	#FL	#DP	#FL	#DP
Density	1865	27347	75	425	56	260	1996	28032
Surface tension	294	4455	62	157	22	79	377	4691
Kinematic viscosity	1795	20046	74	197	60	166	1929	20409
Net heat of combustion	1503	1875	74	85	45	62	1622	2022
Flash point	268	640	75	80	62	54	397	791
Freezing point	402	1053	74	80	50	58	526	1201
Cetane number	159	211	7	22	29	48	195	281
Distillation / boiling point	1825	4333	75	772	63	524	1963	5629
Yield sooting index	98	98	0	0	0	0	98	98

Table 3: Number of unique fuels and pure compounds #FL and corresponding datapoints #DP used for the training and validation of the models.

Table 3 illustrates the differences in the number of data points available for the properties and individual compounds. For properties such as density and net heat of combustion, more data points are available for more individual compounds compared to flash point or cetane number. The impact on the predictive ability of the models will be investigated in Section 4.1.

For the jet fuel screening, property values are needed at 15 °C for the density, 22 °C for the surface tension, and -20 °C and -40 °C for the kinematic viscosity at 1 atm. Flash point, freezing point, net heat of combustion, distillation line and cetane number are measured under laboratory standard conditions. In the scope of this work, measurements for the density, surface tension and kinematic viscosity are used from an extended temperature range of -40 to 140 °C, due to the availability of data. The kinematic viscosity values are largely computed from dynamic viscosity

measurements divided by density measurements at the same temperature. This is done because most of the viscosity measurements of pure compounds are only available as dynamic viscosity. For dynamic viscosity measurements without a corresponding density measurement, the density is predicted using the trained QSPR model. Due to the high accuracy with of the developed QSPR density model, with an average error of less than 6 kg/m³ as illustrated in Figure E 5, the induces error is deemed negligible. To visualize the number of available measurements for the temperature dependent properties density, kinematic viscosity and surface tension Figure 3.1 shows histograms with the cumulative numbers over the temperature. The number of data points for compounds are shown in orange, for conventional fuels in blue and for synthetic fuels in green.

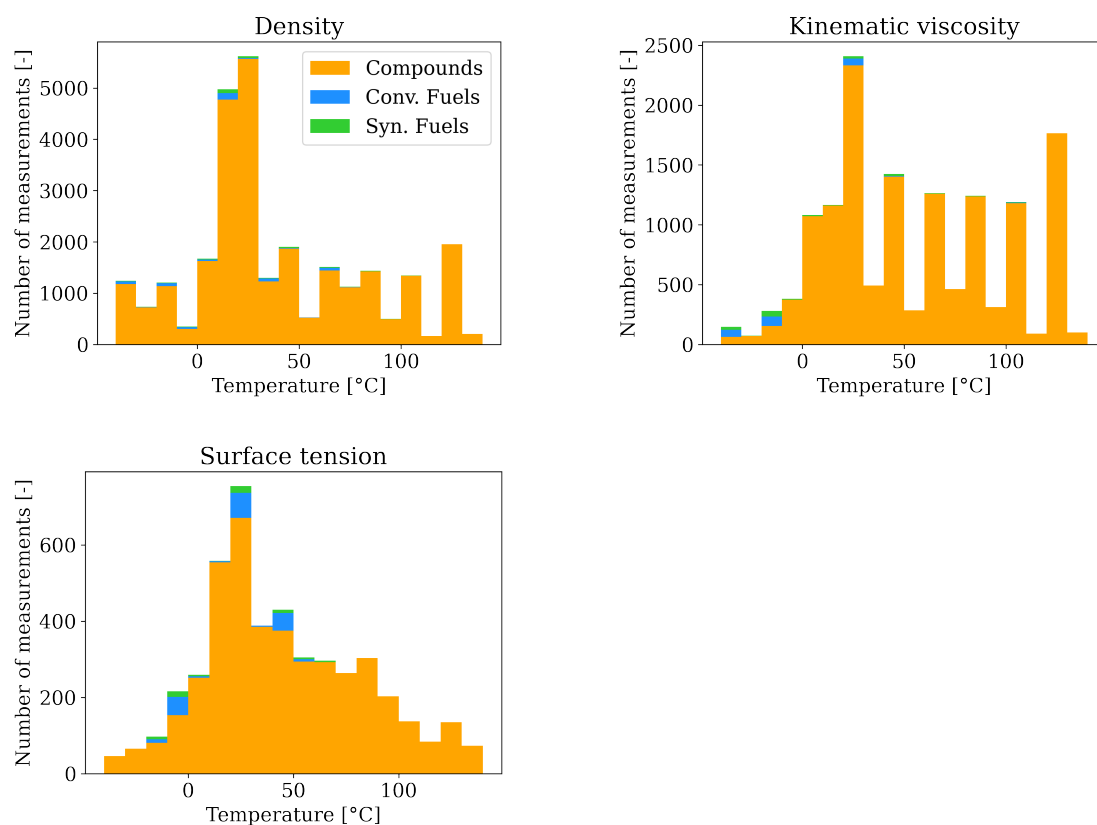


Figure 3.1: Number of measurements vs temperature for density (upper left) kinematic viscosity (upper right) and surface tension (lower left).

Figure 3.1 highlights the abundance of datasets for individual compounds compared to fuels, owing to extensive past studies on pure compounds summarized in databases. However, a notable gap exists for measurements below 0°C, especially kinematic viscosity of pure compounds. This scarcity arises from solidification of many pure compounds at low temperatures and historically limited research focus on low-temperature component properties. The influence of the availability of the data points at different temperatures will also be investigated as part of the model validation in Section 4.1.

To store the data uniformly in the database, the fuel property schema developed by Blakey, Rauch, Oldani and Lee is utilized [98]. The schema has originally been developed for the storage

and exchange of fuel data, but is flexible and extensible enough to also allow the storage of pure compound data. The schema makes use of the JavaScript Object Notation (JSON) format, an unstructured storage method that can be integrated in databases like MongoDB. An example of the schema for the storage of density values is given in Supplementary Material D.

3.2 Data preprocessing and outlier detection

As mentioned in the introduction of this chapter, the reliability of the different composition and property measurements is of utmost importance for the development of data-based tools. Ideally only measurements from the same or comparable standardized measurement protocols are utilized, under consideration of the reproducibility of the respective measurement method. The reproducibility of a measurement method thereby states the expected deviation of the measurements if a measurement is repeated by a different operator at a different site [99]. Measurements that deviate significantly from the other measurements and lie outside of the stated reproducibility, should be identified as outliers and removed in the preprocessing of the data. In practice and also in this work, this procedure can often not be carried out to its full extent, for three reasons:

1. For GCxGC composition measurements of jet fuels a standardized measurement method does not exist at the time of writing. The identification of outliers of the composition data is therefore not possible. The reliability of GCxGC composition measurements depends solely on the trustworthiness of the analytical laboratory.
2. For some measurements, the respective measurement method is not provided in the database. This is the case especially for data from public databases like Pubchem and Chemspider. Due to the availability of properties like flash point or freezing point, which are not listed in commercial databases like the NIST Standard Reference Database, the use of these measurements is unavoidable. Therefore, it is assumed that the property measurement methods of fuels and pure components are comparable, an assumption that can only be verified in the later property modeling.
3. The detection of measurement outliers requires a minimum number of reliable measurements for the calculation of an average, from which outliers can be identified and subsequently removed. For this work, the minimum number is set to 3. However, for some components and their respective physicochemical properties, this number is not achieved with the available data. The calculation of a respective measurement norm and the detection of outliers are therefore not possible. For these cases, all measurements are assumed to be reliable and used for the development of the data-based tools.

Due to the three outlined reasons, the preprocessing of the data, and with it the removal of potential outliers, can only be carried out for property measurements of fuels and pure

compounds. For the outlier detection, two detection methods are utilized: one for the detection of significant outliers from the whole dataset of a property and one for the detection of outliers for measurements of each individual component / fuel. For the detection of outliers from the whole dataset, measurements with a deviation of more than 3.5 standard deviations σ are removed see Equation (3.1).

$$|y_i - \bar{y}| > 3.5 * \sigma \quad (3.1)$$

Outliers of each individual pure compound / fuel on the other hand are removed using the modified Z-score by Iglewicz and Hoaglin, an outlier detection method for small sample sizes as recommended by NIST, see Equation (3.2) [100]. Data points with a score greater than 3.5 are removed from the dataset.

$$\frac{0.675 * (|y_i - \bar{y}|)}{\text{median}(|y_i - \bar{y}|)} > 3.5 \quad (3.2)$$

For the temperature-dependent properties, e.g. the density, the outlier detection is carried out for measurements at different temperatures with a temperature window of ± 1 °C, both for the whole dataset of a property and the measurements of an individual compound / fuel.

3.3 Data characterization

To illustrate the kind and extent of data utilized for this work, the following section describes and visualizes the data for both jet fuels and fuel components. The characterization furthermore serves to compare the available data to the possible application domain of the property modeling of jet fuels. This section thereby focuses on the compositional data, to investigate the extent of the compositional data available for training and validation of the models. Detailed illustrations and discussions for the value range of the property measurements are also given in the later chapter 5.

3.3.1 Data characterization of fuels

Description of fuel composition data

The compositions of fuels in this work are characterized by GCxGC measurements. As explained in Section 1.2.3, GCxGC measurements do often not allow the exact identification of the fuel components/isomers. Identified species are therefore classified with respect to their chemical family and the number of carbon atoms they contain. Unidentified isomers are consequently lumped together, yielding a two-dimensional matrix that lists the mass fractions of the detected species. Figure 4.1 shows a representative GCxGC measurement of a conventional oil-based jet fuel for 7 to 20 carbon atoms and eight hydrocarbon families.







	N-alkanes	Iso-alkanes	Mono-cyclo-alkanes	Bi-cyclo-alkanes	Mono-aromatics	Cyclo-aromatics	Di-aromatics
C atoms							
7	0.11	0.16	0.39	0	0.18	0	0
8	0.68	0.73	1.38	0.04	0.99	0	0
9	1.6	3.11	4.13	0.82	2.41	0.16	0
10	2.04	4.54	5.13	1.68	2.23	1.09	0.2
11	2.22	4.17	5.78	1.84	1.78	1.61	0.54
12	1.89	3.42	4.26	1.69	1.4	1.23	0.74
13	1.54	3.23	3.22	1.62	0.95	1.17	0.47
14	1.27	2.82	2.23	1.14	0.7	0.65	0.4
15	0.88	2.26	1.52	0.38	0.5	0.54	0
16	0.57	1.5	0.75	0.05	0.39	0	0
17	0.42	0.95	0.33	0.02	0	0	0
18	0.12	0.45	0.1	0	0	0	0
19	0.03	0.21	0.03	0	0	0	0
20	0	0.04	0	0	0	0	0

Figure 3.2: Evaluated GCxGC measurement of a conventional jet fuel with representative molecules for each family.

The eight hydrocarbon families in Figure 4.1 are the most relevant for jet fuels and the standard chemical families considered in scope of this work: n-alkanes, iso-alkanes, mono-cyclo-alkanes, bi-cyclo-alkanes, tri-cyclo-alkanes, cyclo-aromatics, mono-aromatics and di-aromatics. Components with up to 25 carbon atoms are thereby considered. Table 4 shows representative molecules for all eight hydrocarbon families, as well as their general molecular formula and the structural criteria for their classification in the respective families.

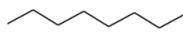
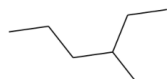
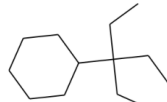
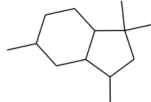

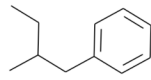
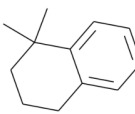
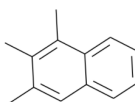
Family	Formula	Structural criteria	Example
N-alkanes	C_nH_{2n+2}	- No atoms in rings - No branched carbon bonds	
Iso-alkanes	C_nH_{2n+2}	- No atoms in rings - At least one branched carbon bond	
Mono-cyclo-alkanes	C_nH_{2n}	- One aliphatic ring - No aromatic rings	
Bi-cyclo-alkanes	C_nH_{2n-2}	- Two aliphatic rings - No aromatic rings	
Tri-cyclo-alkanes	C_nH_{2n-4}	- Three aliphatic rings - No aromatic rings	
Mono-aromatics	C_nH_{2n-6}	- No aliphatic rings - One aromatic ring	
Cyclo-aromatics	C_nH_{2n-8}	- One aliphatic ring - One aromatic ring	
Di-aromatics	C_nH_{2n-12}	- No aliphatic rings - Two aromatic rings	

Table 4: Table of considered chemical families with the corresponding formula, structural criteria and an illustrated representative.

Characterization of variability and extent of the fuel data

Apart from conventional oil-based jet fuel production, there currently exist seven approved synthetic production routes with a registered annex in ASTM D7566. Two co-production routes, for production sights optimized for e.g. for bio-diesel, are furthermore annexed to ASTM D1655. Four new production pathways actively seek certification. Of those seven approved and four pending production routes, the fuel dataset holds fuels of six different fuel types as well as blends with conventional fuels: Alcohol-To-Jet process (ATJ-SPK), Fischer-Tropsch process (FT-SPK), Synthesized kerosene with aromatics (SPK/A), Hydroprocessed Esters and Fatty Acids process (HEFA-SPK), Catalytic Hydrothermal Conversion process (CHJ) and Integrated Hydrolysis process of Shell (IH²). For conventional fuels, the datasets hold representatives from all major fuel types both for commercial and military applications: Jet A, Jet A-1, JP-5, JP-8 and TS-1.

The different production paths differ substantially from each other, using different chemical process operations and conditions. The production paths can however be classified based on the respective utilized resources for the feedstock: Fossil resources, biomass and XTL that utilize either electricity or solar radiation to produce fuel from carbon dioxide and water. Figure 3.5 gives a schematic overview of the production paths, with their resources on the left, the principal processing operation in the middle and the corresponding fuel type on the right. Approved SAF that can be flown as drop-in fuel as blends is thereby shown in blue boxes with solid borders. SAF that is still in the approval process is shown in blue boxes with dashed borders. The variety of feedstocks from the three resources and processing technology results in a large variety in jet fuel composition, which is only insinuated by the number of fuel types on the right.

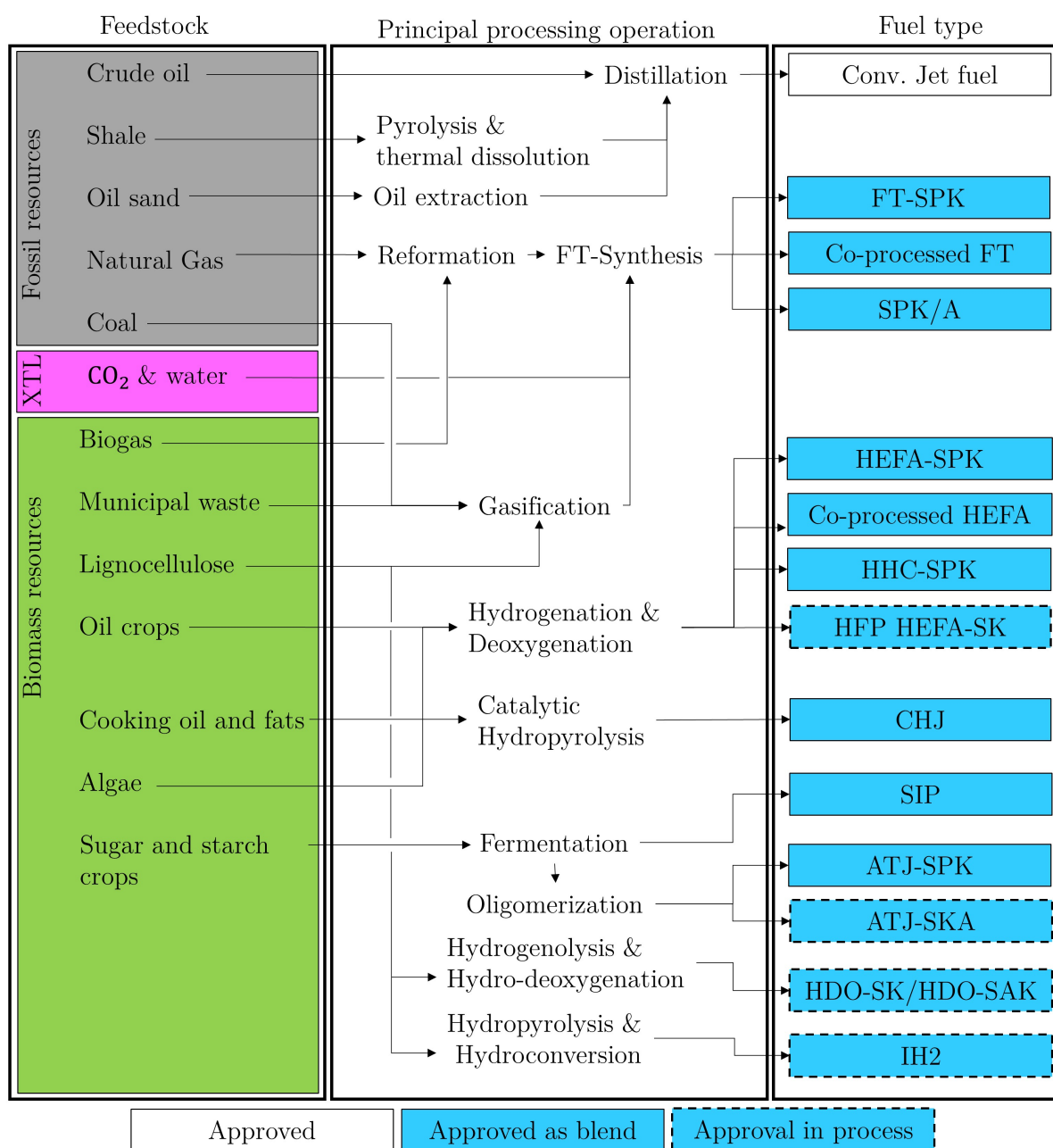


Figure 3.3: Overview of current jet fuel production processes, extended from Blakey et al. 2011 [101].

To visualize the extent and variety of the jet fuel compositions, Figure 3.4 shows a scatter plot for the mass fractions from the GCxGC measurements, summed up for each hydrocarbon family. Conventional fuels (Conv.) are thereby displayed in blue, synthetic fuels (Syn.) in green.

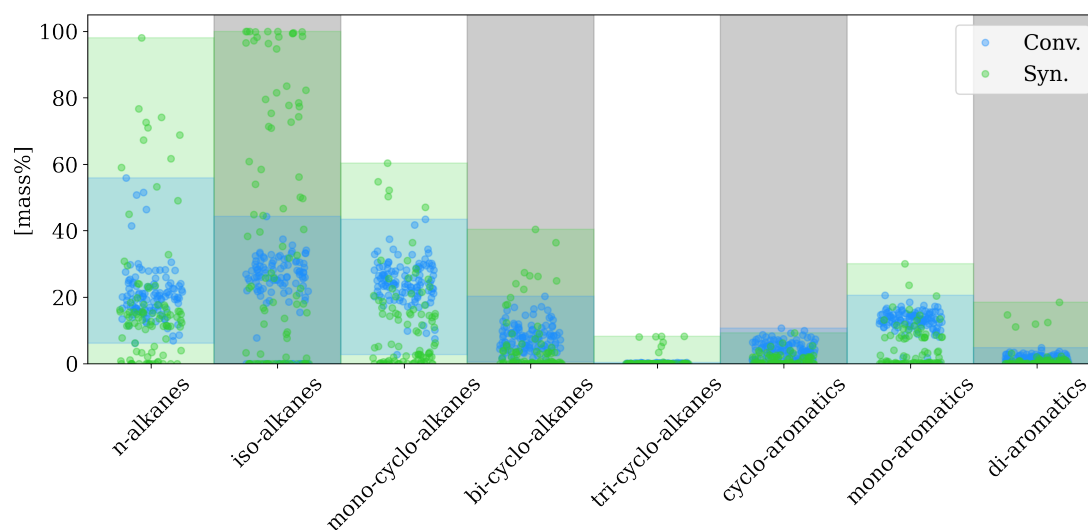


Figure 3.4: Scatter plot of GCxGC measurements with summed hydrocarbon families. Blue: conventional fuels. Green: synthetic fuels and blends. Blue and green shaded areas indicate the observed range.

The plot illustrates on the one hand the extent and distribution of available fuel composition measurements and on the other hand the variety of both conventional and synthetic jet fuels. For all families except the cyclo-aromatics synthetic fuels cover a broader and more evenly distributed composition range. Differences of up to 90 mass% in a family are observed for synthetic fuels. For conventional fuels, the differences in their composition range are less extensive; however depending on the family, differences of 20 mass% are possible. This shows the variability of conventional jet fuels and illustrates, that one cannot speak of one typical conventional fuel composition. The broader composition range overall of the synthetic fuels can be explained by the large variety of different crude oils and other feedstocks outlined in Figure 3.3. The extent of this composition space illustrates the challenge for the modeling of fuel properties, especially considering that the information about the distribution along the number of carbon atoms of the individual fuels is not displayed. The variability of the fuel composition makes a continuous adaptation and extension of the fuel dataset beyond this work necessary, both for the development and validation of the models.

3.3.2 Data characterization of pure compounds

Description of pure compound data

As mentioned in Section 2.2.1 pure compounds are represented in this work using the simplified molecular input line entry system (SMILES) [74]. Using the canonical SMILES representation of the RDKit Python package [75] allows a unique string representation of each pure compound. From the SMILES representation, a quantitative structure representation of each compound is derived using a selection of distinct molecular substructures, see Section 2.2.1, which can now be utilized for the characterization and visualization of the dataset.

Characterization of variability and extent of the fuel data

Before the visual characterization of the pure compound dataset, the extent of available compound data and the number of compounds that could theoretically exist as possible fuel components are illustrated. As explained in Section 1.2.3, the exact molecular composition can presently often not be identified with state-of-the-art GCxGC measurement methods. Since no widely accepted selection of pure compounds as potential fuel components exists, this work assumes that all compounds of the eight considered hydrocarbon families can potentially be present as components in both conventional and synthetic fuels. This approximation is undoubtedly a very cautious one, as the range of possible components that could realistically exist in conventional and synthetic fuels is likely to be significantly narrower. The pure compound databases used for this work list 1870 relevant compounds. However, these are only a small fraction of all theoretically possible molecules. Table 5 shows a comparison of the number of theoretically possible compounds and the number of compounds listed in the databases and considered in this work. The theoretically possible compounds were systematically generated using the molecule generator MOLGEN version. 5 [102]. Their number is listed in the column MG, the number of molecules listed in databases and used for this work is shown in the column DB. The compounds are classified into the eight considered hydrocarbon families using characteristic molecular substructures outlined in Section 3.3.1. Due to the exponential increase of possible isomers and computational limitations for the classification, only calculations up to molecules with 12 carbon atoms are executed. For iso-alkanes the numbers above C 12 were taken from theoretical calculations [103].

C number	n-alkanes		iso-alkanes		mono-cyclo-alkanes		bi-cyclo-alkanes		cyclo-aromatics		mono-aromatics		di-aromatics	
	DB	MG	DB	MG	DB	MG	DB	MG	DB	MG	DB	MG	DB	MG
1	1	1												
2	1	1												
3	1	1			1	1								
4	1	1	1	1	2	2	0	1						
5	1	1	2	2	5	5	2	4						
6	1	1	4	4	12	12	5	14	0	4	1	1		
7	1	1	8	8	27	29	5	46	0	9	1	1		
8	1	1	17	17	43	73	8	150	0	31	4	4		
9	1	1	32	34	52	185	18	477	2	75	8	8		
10	1	1	49	74	58	475	18	1503	8	218	22	22	0	5
11	1	1	37	158	33	1231	12	4680	16	588	42	51	2	21
12	1	1	45	354	29	3232	18	14461	27	1657	60	136	9	103
13	1	1	34	801	14	>3232	15	>14461	32	>1657	51	>136	24	>103
14	1	1	29	1857	17	>3232	16	>14461	31	>1657	60	>136	38	>103
15	1	1	27	4345	18	>3232	14	>14461	29	>1657	34	>136	17	>103
16	1	1	32	10358	8	>3232	15	>14461	16	>1657	22	>136	9	>103
17	1	1	11	24893	6	>3232	8	>14461	7	>1657	15	>136	3	>103
18	1	1	12	60 522	8	>3232	17	>14461	6	>1657	22	>136	13	>103
19	1	1	14	147 283	13	>3232	6	>14461	12	>1657	11	>136	4	>103
20	1	1	20	366 318	4	>3232	10	>14461	9	>1657	13	>136	9	>103
21	1	1	9	>366 317	6	>3232	2	>14461	0	>1657	3	>136	1	>103
22	1	1	14	>366 317	5	>3232	4	>14461	6	>1657	14	>136	4	>103
23	1	1	5	>366 317	3	>3232	0	>14461	0	>1657	4	>136	0	>103
24	1	1	17	>366 317	4	>3232	3	>14461	1	>1657	9	>136	0	>103
25	1	1	2	36 797 587	4	>3232	5	>14461	1	>1657	7	>136	1	>103

Table 5: Comparison of the number of representative molecules available in the database (DB) and the number of theoretically possible molecules, calculated by MOLGEN (MG).

Table 5 illustrates the large extent of potential fuel components based on the assumption that all isomers of a hydrocarbon group (chemical family, carbon number) can possibly exist in a fuel. The number of isomers increases exponentially with the carbon number for all families. For isomers containing over nine carbon atoms, the pure compound database utilized for this work holds only a small fraction of the potential fuel components. These differences even increase for higher carbon numbers, where the number of available measurements decreases, while the number of possible isomers increases. Since jet fuels are typically composed of components containing 7 to 16 carbon atoms, uncertainty might actually be underestimated, because the set assumption states, that all isomers of the hydrocarbon groups can be present in a fuel. The results of the validation of the developed property models will show the validity of the set assumptions and illustrate if further compositional information and a constrained selection of the isomers are necessary.

To visualize the extent of the pure compound data, the molecular structures of the fuel components are converted into the introduced quantitative structure representation, as explained in Section 2.2.1. Each pure compound is represented by a vector with 49 dimensions as a quantitative structure representation. Additionally, the composition of the jet fuel dataset is visualized as stars, in blue for conventional fuels and green for synthetic fuels. Similar to the fuel components, the GCxGC compositions of the fuels are converted to their mean quantitative

structure representation, introduced in Section 2.2.3. The 200 dimensions of the GCxGC matrix are converted to the same 49-dimensional feature space and then reduced to two dimensions using a t-distributed stochastic neighbor embedding algorithm. The dimension reduction algorithm allows the transformation of the 49 dimensional space into a two-dimensional one, keeping the relative distance between the data points as similar as possible [104]. The results of the dimension reduction are visualized in Figure 3.5, fuel components are thereby displayed as dots in a color corresponding to the hydrocarbon family.

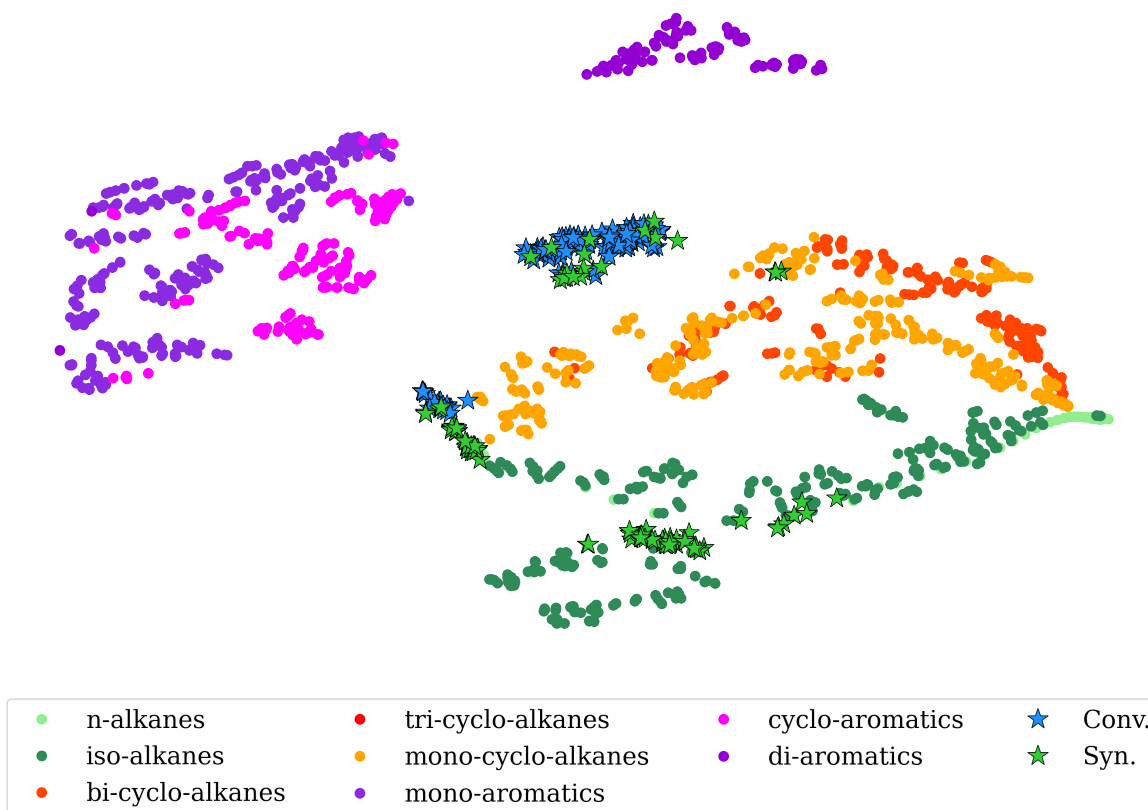


Figure 3.5: Scatter plot visualizing the compositional similarity of fuels and fuel components based on the dimensional reduced representation of their respective mean quantitative structure and quantitative structure representation.

The scatter points in Figure 3.5 form a clear pattern and the position and distance of the respective scatter points visualize the affiliation of the fuel components to their respective hydrocarbon families and the similarity to other components / fuels. Different clusters are visible for different hydrocarbon families and fuels. N- and iso-alkanes form a cluster, as do cyclo-alkanes, mono- and cyclo-aromatics and the di-aromatics. The cluster of the n- and iso-alkanes, cyclo-alkanes and aromatics lie thereby close together, illustrating their structural similarity and the respective absence / presence of cyclic and or aromatic bonds. For the fuels, four clusters are visible. One big cluster composed of both conventional and synthetic fuels is located in the middle of the figure, with two clusters close to the n- and iso-alkanes and one cluster close to the cyclo-alkanes. The central fuel cluster contains fuels that are composed of components from all

hydrocarbon families, with no particularly dominating family. As a result, fuels from this cluster cannot be affiliated with one of the clusters of the hydrocarbon families and form their own cluster, separate from the others. The fuels in this cluster are predominantly conventional fuels and blends, with broad compositions of all other hydrocarbon families. The three residual fuel clusters all overlay one of the hydrocarbon families. The fuels of these clusters are thereby themselves mostly composed of the respective family, ATJ-SPK, FT-SPK and HEFA-SPK form the clusters overlaying the iso-alkanes and IH^2 fuels form the cluster overlaying the cyclo-alkanes. On the one hand, the figure therefore visualizes the compositional similarity of both fuel components and fuels. On the other hand, it illustrates the compositional space and extent of the data utilized for the training and testing of the models, both for pure compounds and fuels.

4 Predictive Capability Assessment of Models and Adequacy Assessment for Fuel Screening

The predictive capability assessment of the models is presented in two parts. In the first part, the results of the cross-validation and testing of the property prediction for both conventional (Conv.) and synthetic (Syn.) fuels are presented. This part investigates the general ability of the different methods to model the respective properties and reviews their predictive capability on the basis of the metrics introduced in Section 2.5.1. Models are thereby compared with respect to each other and deterministic reference models from the literature. The comparison with existing literature models will investigate the accuracy of the developed models relative to available state-of-the-art models and demonstrate the use of the additional estimated prediction uncertainties. In the second part, the developed models are applied for a simulated jet fuel prescreening after the Tier α protocol of Heyne and Rauch [24] to review the model adequacy for the desired use case.

In the first part, only predictions from the testing, for fuels not seen by the models in training, are used to determine predictive capability metrics. Since the DC model and the M-QSPR model are directly trained on fuel data, the predictions are taken from the testing of the 4-fold cross-validation of the models described in 2.4.1, which is performed with a 15 % test fraction. To investigate the influence/importance of synthetic fuels for the predictive capability of the DC and M-QSPR models, results are also provided for models trained solely on conventional fuels (no-Syn.).

For the comparison with state-of-the-art models from the literature, models like the DLR Discrete Component Model (DCM) developed by Le Clercq [45,47] or respective multilinear regression models (MLR) by Liu et al. [105], are utilized. At the time of writing, these models were the only ones available in full code for comparison, able to predict properties from the GCxGC format outlined in Section 3.3.1. The DCM models a fuel as a mixture of constituents, similar to the QSPR sampling method, as described in the introduction of Chapter 2. The MLR models of Liu et al. are direct correlation models that predict the fuel properties based on the summed-up fractions of the respective fuel families, e.g. iso-alkanes, mono-cyclo-alkanes etc. Short descriptions of both models can be found in the Supplementary Material F.

For the adequacy assessment of the models, hold-out fuels are screened that have not been part of the previous training and testing process. The second part critically cross-checks the results of the first part and evaluates the adequacy of the models for the actual practical use case of jet fuel screening. The predictions are thus evaluated not only in relation to the measurements, but also in relation to the specification limits as adequacy requirements, to examine the impact of limited predictive capability on decision making in prescreening.

4.1 Part 1: Predictive capability assessment of models

Before the validation results of the individual properties are presented, the schema for the presentation of the validation results is outlined. This schema will be used in the following for the results of all properties. First, the general ability of the method to model the property without a systematic error on the basis of the GCxGC measurement is investigated. The prediction results are thereby presented in unity plots, where the predicted values are plotted against the corresponding measurements, similar to the predictive capability example in 2.5.2. Markers for conventional fuels (Conv.) are shown in blue and synthetic fuels (Syn.) in green. If the prediction and the measurement are in perfect agreement, the markers lie on the unity line, which is displayed in black. If strong deviations and even systematic deviations of the predictions from the unity line are visible, the model is unable to model the property on the basis of the GCxGC data.

For the developed models, the prediction intervals (PI) are indicated as error bars for a 95 % confidence level. As a reference for the accuracy and precision of the predictions the reproducibilities of the measurements, taken from the CRC Report No. AV-23-15/17 [106], are indicated as grey dashed lines. For a prediction to be valid, the corresponding PI have to lie inside or cross the indicated grey reproducibility lines. The reproducibilities of the corresponding measurement methods are listed in Table 6.

Property	ASTM Method	Reproducibility
Density [kg/m ³]	D4052	0.52
Flash point [°C]	IP 170	3.2
Freezing point [°C]	D5972	0.8
Net heat of combustion [MJ/kg]	D4809	0.324
Surface Tension, [mN/m]	D971	0.1*X
Viscosity, [mm ² /s]	D445	0.019*X
Cetane number [-]	D6890	0.0385*(X+18)
Distillation 10 vol% [°C]	D86	3.15
Distillation 50 vol% [°C]	D86	3.46
Distillation 90 vol% [°C]	D86	3.83

Table 6: Reproducibilities of ASTM property measurement methods [106]

In the second step, the predictive capability of the models is assessed based on the introduced metrics. The calculated metrics are provided in separate tables. The accuracy is calculated using MAE from Equation (1.26). In the assessment, the accuracy of the developed models is compared with respect to each other and the reference models from the literature. The validity of the predicted PI is quantified by the PICP, see Equation (1.27). For a model to be considered valid, the calculated PICP must be on average equal to or greater than the set confidence level of 95 %. For the evaluation of the precision, the NMPIW from Equation (1.28) is calculated with the measurement reproducibility to compare the uncertainties of the models with those of the measurements. Based on these comparisons, the models with the highest predictive capability and therefore the highest potential for fuel screening are outlined.

In the upcoming figures, the testing results of the developed models (DC, M-QSPR and QSPR sampling model (a-c)) trained on all datapoints are shown in the first row of the unity plots. The second row shows the DC model (d) and M-QSPR model (e) trained solely on conventional fuels as well as the validation results of the reference model (f) from the literature. The respective models are additionally labeled with an alphabetical suffix for clarity (a-f).

4.1.1 Density

For the density, almost all predictions of the considered models lie close to the unity line, as visible in Figure 4.1. Exceptions are visible for the predictions of Syn. fuels for the DC model, trained only on conventional data (DC -no Syn.) Figure 4.1 (d), and the reference DCM model Figure 4.1 (f). All developed models (DC, M-QSPR and QSPR sampling (a-c)) are therefore generally able to model the density of fuels from the GCxGC data without a systematic error; however the DC method requires synthetic fuels to be present in the training data. The deviations of the predictions of DC -no Syn.(d) and the DCM model (f) correspond to synthetic fuels with a low fraction of aromatics. Synthetic fuels with a high fraction of aromatics are still predicted correctly. For the DC -no Syn. model (d), this indicates a dependence of the DC method on similar fuels in training and testing. For the reference DCM model (f), the systematic deviation might be due to the wrong selection of the representative isomer or the consideration of just one isomer per chemical family and carbon number. The deviations are not observed for the QSPR sampling model (c), which likewise to the DCM model (f), models the fuel as a mixture of components. This illustrates the benefit of considering multiple isomers in a probabilistic modeling approach compared to the deterministic approach of the DCM model.

The striking outlier at 726 kg/m³ is visible for almost all models. Since this is the only consistent outlier observed for different models, an erroneous GCxGC or density measurement might be the reason. The measurement was not made in house a verification of this observation is therefore not possible.

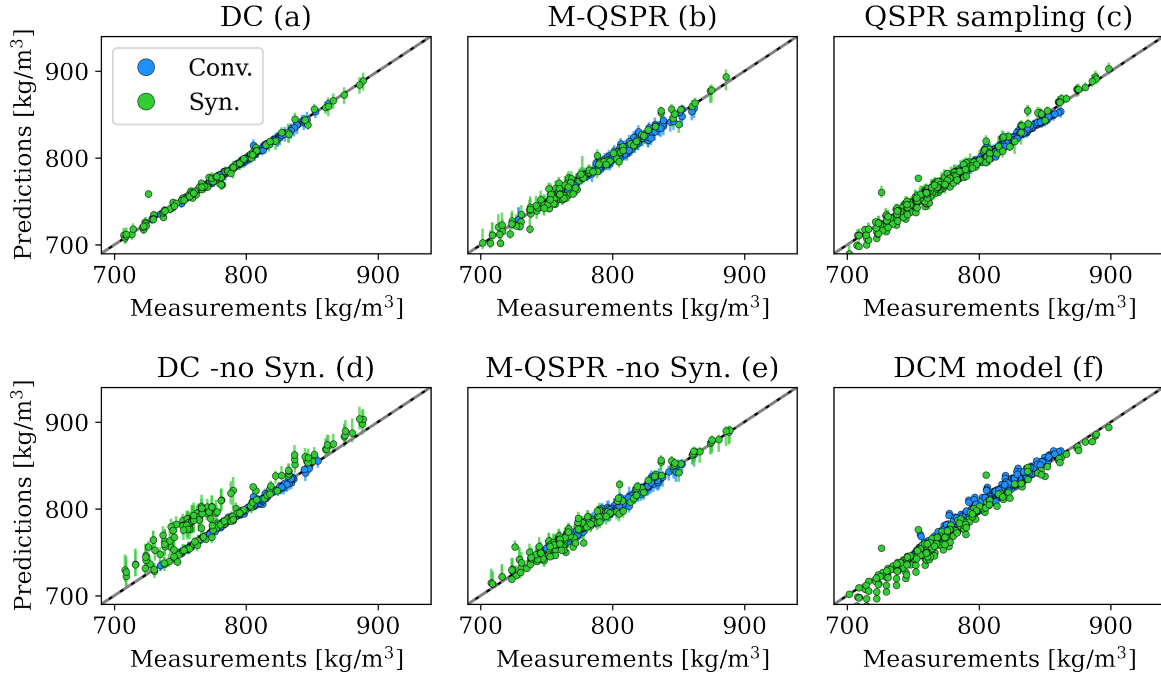


Figure 4.1: Validation results of the density prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.

	Conv. Fuels				Syn. Fuels			
	MAE [kg/m ³]	PICP [%]	NMPIW [%]	MAOE [kg/m ³]	MAE [kg/m ³]	PICP [%]	NMPIW [%]	MAOE [kg/m ³]
DC (a)	0.96	99.4	1667.9	0.42	2.39	81.8	1847.6	2.86
M-QSPR (b)	2.59	98.1	2927.5	0.52	4.59	88.2	3057.4	3.37
QSPR sampling (c)	3.6	87.7	1874.9	0.93	4.67	73.4	2257.1	4.17
DC -no Syn. (d)	1.23	96.6	1280.3	1.11	14.4	44.0	3496.0	11.97
M-QSPR -no Syn. (e)	2.16	98.8	2889.7	1.52	5.55	74.7	2943.3	4.14
DCM model (f)	4.0				7.6			

Table 7: Predictive capabilities of density models.

The previous observations are reflected in the metrics in Table 7. High accuracies are achieved for the DC (a), M-QSPR (b), QSPR sampling (c) models with a MAE of 0.96 kg/m³ to 3.6 kg/m³ for conventional and 2.39 kg/m³ to 4.67 kg/m³ for synthetic fuels, which corresponds to less than 0.5 % relative error. The developed models are highlighted in dark grey in Table 7. The accuracy of the deterministic reference DCM model (f) with 7.6 kg/m³ for synthetic fuels is thereby excelled by all developed probabilistic models (DC, M-QSPR and QSPR sampling (a-c)). For the validity, PICP values close or even above the desired 95 % threshold are achieved for the conventional fuels. This means that on average 95 % of the measurements are enclosed by the PI. For the synthetic fuels the 95 % threshold is not reached by any model. Therefore, the use of the estimated Mean Absolute Outlier Error (MAOE), see Section 2.5.1, is recommended to additionally extend the estimated PI for the prediction of synthetic fuels, to statistically enclose

the measurements. The NMPIW exceeds the reproducibilities of the measurements multiple times. The uncertainty of a model prediction is therefore significantly greater than the one of a measurement. This is comprehensible, since the density measurement method is highly accurate, with a reproducibility 0.52 kg/m^3 . In the later prescreening the predicted PI will be set in reference to the specification limits to see if they are adequate for the actual use case.

4.1.2 Surface tension

Similar to the predictions of the density, the predictions for the surface tension closely follow the unity line and most of the predictions are located inside the reproducibility region of the measurement method, see Figure 4.2. Exceptions are again visible for predictions of synthetic fuels of the DC -no Syn. Model Figure 4.2 (d). For the DC -no Syn. model, this again shows the dependency on the presence of synthetic fuels in the training data. Striking is the systematic deviation of predictions from the QSPR sampling model (c) and the reference DCM model (f) for a few synthetic fuels with values above 30 mN/m , see Figure 4.2 (c) and Figure 4.2 (f). These predictions correspond to research fuels from the JETSCREEN project with an unusually high fraction of aromatics (up to 30 mass%). This deviation could be due to a false prediction of the property values for the components in the QSPR model or a false selection of isomers for QSPR sampling and DCM model. In general, all models are able to model the surface tension without a significant systematic error outside of the reproducibility region on the basis of the GCxGC data.

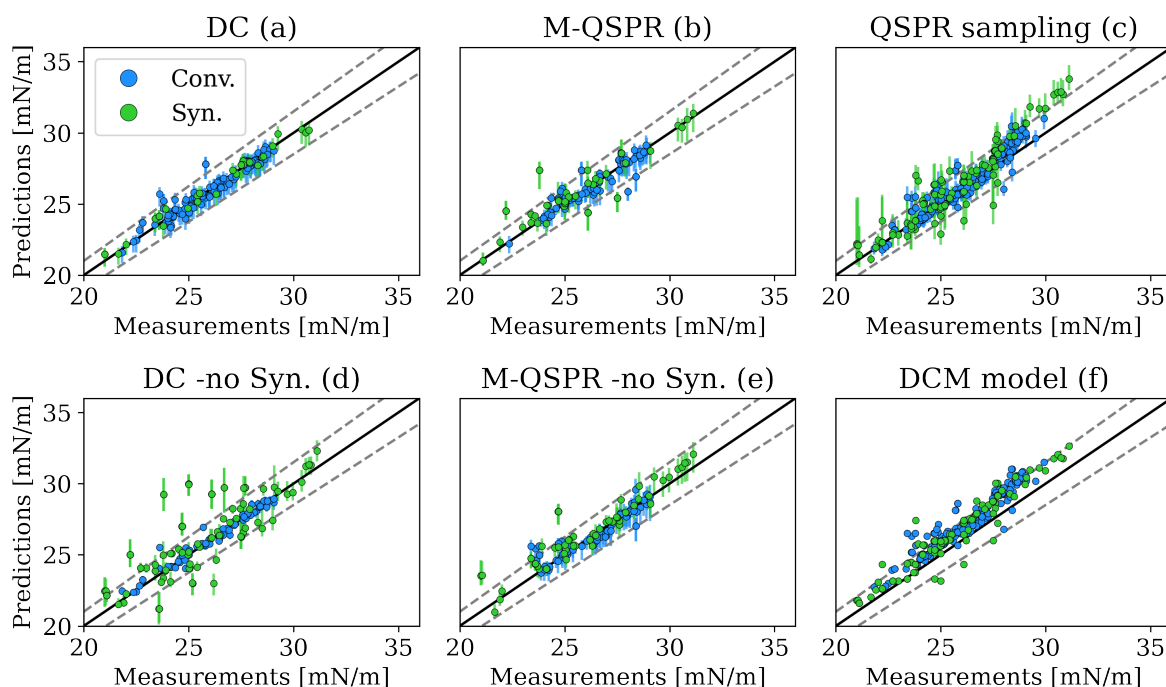


Figure 4.2: Validation results of the surface tension in air prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.

The metrics in Table 8 show a similar predictive capability for all models, with a PICP over 95 % and a NMPIW that is comparable to the reproducibility of the measurements. With respect to the accuracy, MAE of 0.36 mN/m to 0.54 mN/m are achieved for conventional fuels and 0.35 mN/m to 0.95 mN/m for synthetic fuels. The surface tension can therefore be accurately predicted with all methods. The accuracy of the DCM model (f) of 1.04 mN/m is slightly exceeded by all developed models (DC, M-QSPR and QSPR sampling (a-c)), however the main advantage of the probabilistic models are the PI that correctly estimate uncertainty.

	Conv. Fuels				Syn. Fuels			
	MAE [mN/m]	PICP [%]	NMPIW [%]	MAOE [mN/m]	MAE [mN/m]	PICP [%]	NMPIW [%]	MAOE [mN/m]
DC (a)	0.36	100.0	51.4	0.37	0.35	100.0	54.3	0.08
M-QSPR (b)	0.38	100.0	46.3	0.55	0.78	100.0	57.6	1.19
QSPR sampling (c)	0.54	100.0	44.3	0.51	0.96	98.9	78.3	0.79
DC -no Syn. (d)	0.22	100.0	17.4	0.26	1.49	93.5	51.1	1.2
M-QSPR -no Syn. (e)	0.42	100.0	49.1	0.53	0.82	98.3	48.2	0.8
DCM model (f)	1.15				1.04			

Table 8: Predictive capabilities of surface tension models.

4.1.3 Net heat of combustion

For the net heat of combustion, stronger differences between the accuracies of the different models are visible in Figure 4.3. For both conventional and synthetic fuels, only the predictions of the developed DC and M-QSPR models lie inside the reproducibility regions, see Figure 4.3 (a) and (b). For the QSPR sampling, the DC -no Syn. and the M-QSPR -no Syn. models in Figure 4.3 (c-e), deviations and outliers are visible, especially for synthetic fuels. For the DC and M-QSPR method, this illustrates again the need for synthetic fuels in the training data for this method. For the reference MLR model by Liu et al. Figure 4.3 (e) only some predictions for conventional fuels lie inside the reproducibility region, while most of the predictions lie almost parallel to the measurement axis, indicating a systematic error. The low accuracy and the systematic error of the reference MLR model can probably be explained by the missing information about the average carbon numbers of the hydrocarbon families [105]. The model of Liu et al. was referenced by Wang et al. [107] in a recent review and the only deterministic model fully available in literature, able to predict the net heat of combustion on the basis of the utilized GCxGC format.

For the QSPR sampling model (c), a systematic offset is visible for the predictions of conventional fuels with the QSPR sampling model. The offset could be explained by the inapplicability of the mass fraction based mixing rule, or differences in the measurement methods of pure compounds and fuels, which could not be verified. For the M-QSPR model (b), this offset is not observed the model is able to adapt the offset with the presence of fuels in the dataset. Since the offset is smaller than the reported reproducibility, the offset is not significant, but further investigations

might be necessary to understand the observation. Another striking systematic error is observed for the M-QSPR -no Syn. model (e) and the QSPR sampling model (c) for a group of synthetic fuels in the value range of 43.8 to 44.3 MJ/kg. The models predict very similar values for different fuels with different measurement values. These systematic errors were also observed by Yang et al. [48]. The corresponding synthetic fuels all contain a large fraction of iso-alkanes. Yang et al. explained the systematic errors with errors in the net heat of combustion measurements of the fuels and the high uncertainties due to the low reproducibility of the measurement method. Since most of the predictions with this systematic error still lie inside the reproducibility region, this is not significant.

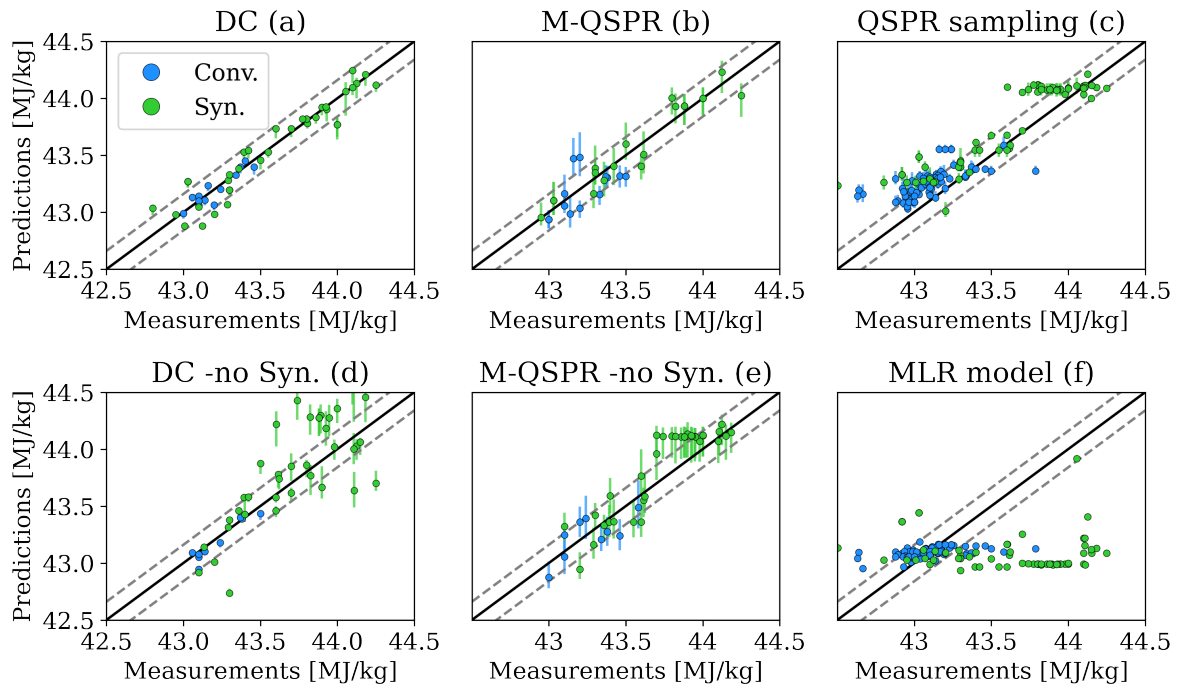


Figure 4.3: Validation results of the net heat prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.

	Conv. Fuels				Syn. Fuels			
	MAE [MJ/kg]	PICP [%]	NMPIW [%]	MAOE [MJ/kg]	MAE [MJ/kg]	PICP [%]	NMPIW [%]	MAOE [MJ/kg]
DC (a)	0.05	100.0	17.7	0.03	0.09	100.0	28.6	0.1
M-QSPR (b)	0.14	100.0	67.9	0.1	0.09	100.0	87.0	0.09
QSPR sampling (c)	0.18	92.9	29.6	0.16	0.19	86.4	27.6	0.18
DC -no Syn. (d)	0.05	100.0	15.3	0.04	0.26	87.2	58.6	0.19
M-QSPR -no Syn. (e)	0.13	100.0	92.3	0.03	0.17	100.0	89.9	0.07
MLR model (f)	0.11				0.61			

Table 9: Predictive capabilities of net heat of combustion models.

Considering the reproducibility of 0.324 MJ/kg of the measurement method and the explanation for the systematic outliers of the synthetic fuels, the DC, M-QSPR and QSPR sampling models (a-c) are able to accurately predict the net heat of combustion on the basis of the GCxGC measurement.

In the predictive capability metrics, the high accuracies are visible in the MAE, ranging from 0.05 MJ/kg to 0.18 MJ/kg for conventional fuels from 0.09 MJ/kg to 0.19 MJ/kg for synthetic fuels, see Table 9. The developed probabilistic models (DC, M-QSPR and QSPR sampling (a-c)), thereby clearly outperform the MLR model (f) of Liu et al, with comparable accuracies for conventional fuels with a MAE of 0.11 MJ/kg and significantly higher accuracies for synthetic fuels with a MAE of 0.61 MJ/kg. The calculated PICP close to or even above of 95 % and NMPIW values below 30 % illustrate, that all predictions are valid and highly precise.

4.1.4 Kinematic viscosity

For the kinematic viscosity, the predictions of the DC, M-QSPR and QSPR sampling model, closely follow the unity line for measurements up to 12 mm²/s in Figure 31 (a-c). Above this range, significant deviations for a group of synthetic fuels are visible, especially for the QSPR sampling model and the M-QSPR -no Syn model, see Figure 31 (c) and Figure 31 (e). The erroneous predictions belong to research fuels from the JETSCREEN project, with an unusually high fraction of aromatics, particularly di-aromatics with mass fractions up to 18 mass% and temperatures below -30 °C.

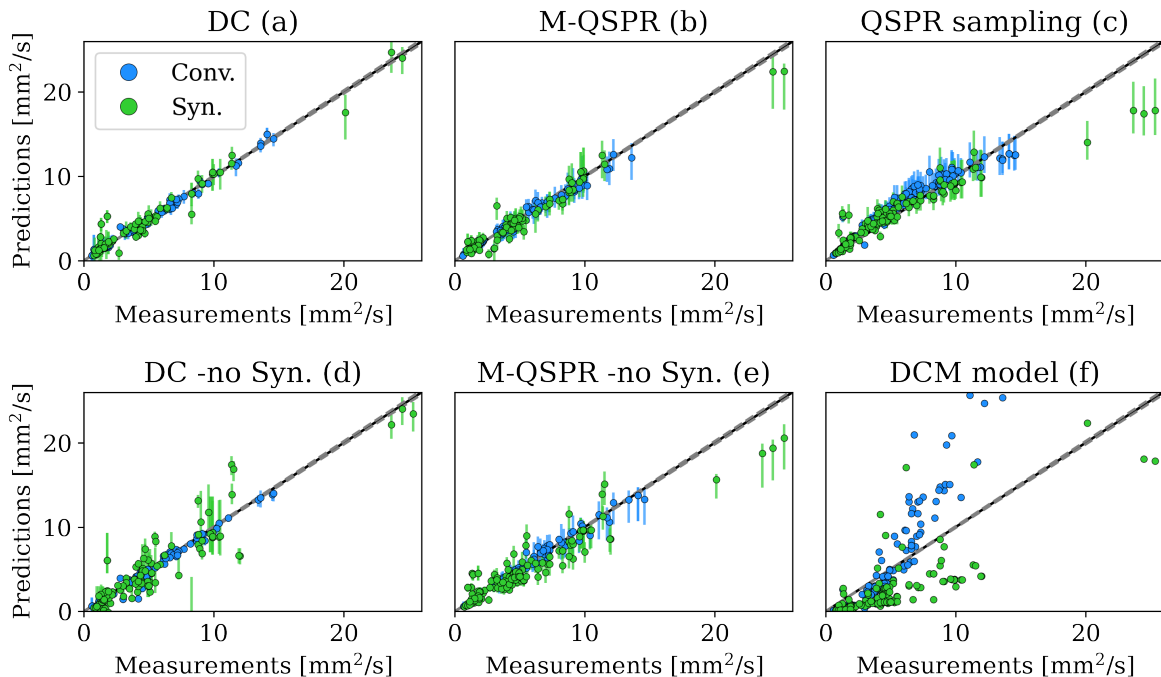


Figure 4.4: Validation results of the kinematic viscosity prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.

For the reference DCM model shown in Figure 31 (f), significant deviations are visible both for conventional and synthetic fuels. While the predictions for the conventional fuels start to

significantly deviate at around 8 mm²/s, corresponding to temperatures below -20 °C, most of the synthetic fuels are underpredicted over the whole value range. Generally, the DC, M-QSPR and QSPR sampling (a-c) are therefore able to model the kinematic viscosity from the GCxGC measurements without systematic errors, while the reference model shows significant systematic deviations for both conventional and synthetic fuels.

	Conv. Fuels				Syn. Fuels			
	MAE [mm ² /s]	PICP [%]	NMPIW [%]	MAOE [mm ² /s]	MAE [mm ² /s]	PICP [%]	NMPIW [%]	MAOE [mm ² /s]
DC (a)	0.24	92.5	2001.8	0.24	0.61	75.9	2775.1	0.86
M-QSPR (b)	0.33	97.1	2699.5	0.46	0.58	84.9	2807.1	0.58
QSPR sampling (c)	0.89	31.3	1207.7	0.46	0.81	54.5	1437.2	0.66
DC -no Syn. (d)	0.26	86.4	1650.8	0.45	1.44	59.5	11003.9	1.07
M-QSPR -no Syn. (e)	0.32	97.8	2572.1	0.21	0.94	63.5	2147.4	0.93
DCM model (f)	2.65				3.03			

Table 10: Predictive capabilities of kinematic viscosity models.

The visual observations are reflected in the accuracy metrics in Table 10. The highest accuracies are observed for the DC model (a) and the M-QSPR model (b) with a MAE of 0.24 mm²/s and 0.33 mm²/s for conventional fuels and 0.61 mm²/s and 0.58 mm²/s for synthetic fuels. The accuracies of the reference DCM model (f) are significantly lower, with MAE of 2.65 mm²/s for conventional and 3.03 mm²/s for synthetic fuels. The developed models (DC, M-QSPR, QSPR sampling (a-c)) excel the accuracy of the reference DCM model therefore 10 times for conventional fuels and 5 times for synthetic fuels. The DC (a) and M-QSPR (b) model, which directly learn from fuel data, show thereby the highest accuracies. The scattering and deviations of the DC -no Syn. model and the M-QSPR -no Syn. model (e) illustrate the need for synthetic fuel data as part of the training data for those models. The deviations of the predictions from the QSPR sampling model (c) above 12 mm²/s can probably be explained by the low number of available measurements of single components at low temperature ranges, below -10 °C. For the pure compounds, only a small number of low temperature viscosity measurements are available, as illustrated in Figure 3.1. As a result, the underlying QSPR model might predict the kinematic viscosity value of the sampled aromatic components to low, which subsequently leads to an underestimation of the kinematic viscosity of the fuel. Intermolecular interactions of the fuel components at the lower temperature ranges of -20 °C to -40 °C not covered by the utilized mixing rule also play a role. Since the reference DCM model (f) also relies on the selection of isomers, but only one per chemical family and carbon number, the deviations can probably be attributed to the wrong selection of the respective isomers. Furthermore, the underlying viscosity model of the DCM model that predicts the viscosity over the temperature range could also be a reason for the strong deviations. For this work, the property model of Mehrotra [108] is utilized,

which differs from the temperature model recommended in ASTM D341 utilized by the QSPR sampling model.

With respect to the validity the M-QSPR model (b) showed the highest PICP with 97 % for conventional and 85 % synthetic fuels. The PICP of the other models differ significantly from the set confidence level of 95 %. Generally, the use of the estimated MAOE is recommended to increase the predicted PI, especially for synthetic fuels. The elevated NMPIW for the models is understandable, as the measurements come from various laboratories and the measurement technique employed is highly precise. A model prediction has therefore higher uncertainty than a measurement.

All in all, the DC model (a) and the M-QSPR model (b) show the highest predictive capability. For the QSPR sampling model (c) significant deviations can occur for fuels with a high fraction of aromatic components, at low temperatures below $-20\text{ }^{\circ}\text{C}$ and high viscosity ranges above $12\text{ mm}^2/\text{s}$.

4.1.5 Flash point

The results of the flash point are presented in Figure 4.5. For the DC, M-QSPR and QSPR model (a-c) most predictions follow the unity line however, stronger scattering and significant deviations are observed in part, especially for synthetic fuels. Striking is the erroneous prediction at $49\text{ }^{\circ}\text{C}$, which is observed for all models. It corresponds to a synthetic fuel produced by the Alcohol-To-Jet process, which is solely composed of iso-alkanes. Other outliers also correspond to synthetic fuels composed of only one or two hydrocarbon families with GCxGC measurements dominated by components with distinct carbon numbers. For the reference MLR model shown in Figure 4.5 (f), the predictions for both conventional and synthetic fuels lie almost parallel to the measurement axis, indicating a systematic error. Likewise to the net heat of combustion, the model of Liu et al. was the only one fully available and able to model the flash point on the basis of the given GCxGC measurements. The reference model is therefore not able to model the flash point from the GCxGC data. A comparison with the predictions of the DC model Figure 4.5 (a), which also uses the method of direct correlation, shows that information about the number of carbon atoms is probably needed to estimate the flash point accurately. For the DC no-Syn. model in Figure 4.5 (d), strong deviations are visible for synthetic fuels. The modeling of the flash point with the DC method therefore strongly relies on the presence of synthetic fuels in the training data. In sum, all developed models (DC, M-QSPR and QSPR sampling (a-c)), are generally able to model the flash point of fuels from the GCxGC measurements without a systematic error.

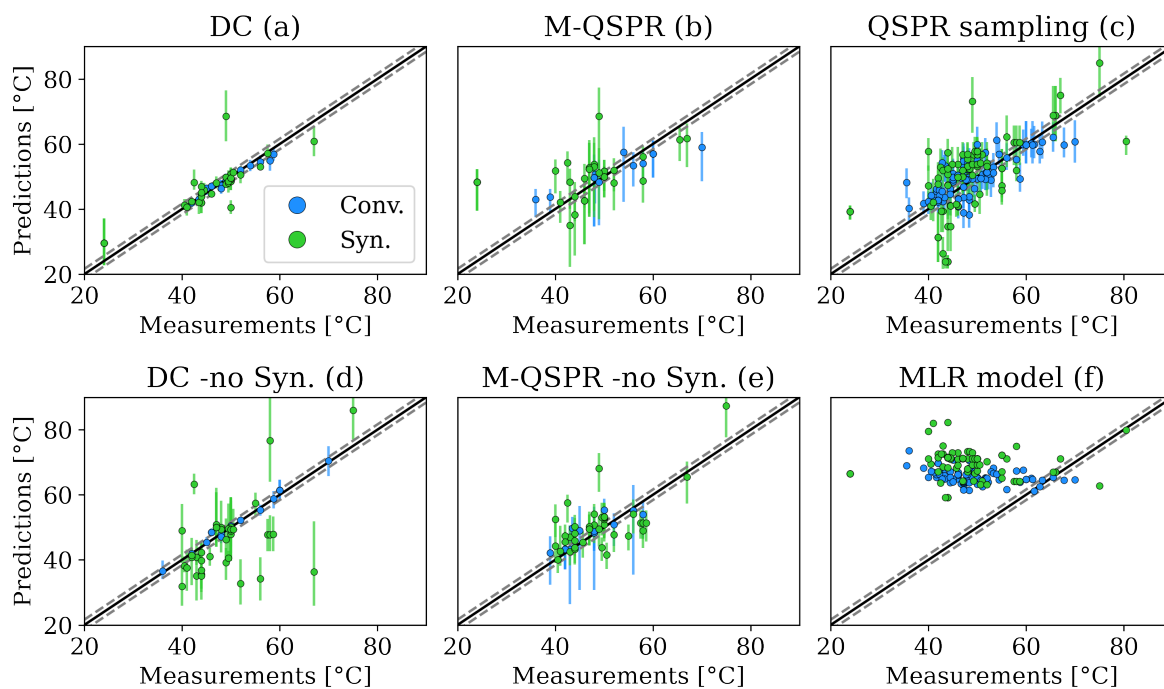


Figure 4.5: Validation results of the flash point prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.

	Conv. Fuels				Syn. Fuels			
	MAE [°C]	PICP [%]	NMPIW [%]	MAOE [°C]	MAE [°C]	PICP [%]	NMPIW [%]	MAOE [°C]
DC (a)	0.99	100.0	98.3	0.08	2.4	92.1	158.2	2.39
M-QSPR (b)	3.77	91.7	451.0	2.78	6.8	80.0	468.1	6.65
QSPR sampling (c)	3.63	89.3	258.2	2.43	7.36	77.2	322.1	4.96
DC -no Syn. (d)	0.81	100.0	116.8	1.59	6.37	73.8	405.8	6.29
M-QSPR -no Syn. (e)	2.13	100.0	444.4	0.24	4.62	83.3	295.8	4.12
MLR model (f)	16.9				21.39			

Table 11: Predictive capabilities of flash point models.

The observed deviations of the outlined synthetic fuels can probably be explained by the smaller datasets of fuels and pure compounds and larger variance of the property values for the flash point of isomers. Only data for 397 unique pure compounds and fuels is available for the flash point, compared to 1622 unique pure compounds and fuels for the net heat of combustion. Simultaneously, the variance of the property values increases for different isomers. E.g. for isoalkanes with 10 carbon atoms the flash points recorded in the database range between 15.85 °C and 51.7 °C. This will be shown and discussed in more detail in Section 5.3.5. This has a significant influence on the QSPR sampling model shown in Figure 4.5 (c), where the stronger deviations and larger uncertainties for synthetic fuels can be explained by the differences in the values of the fuel components. For conventional fuels, the influence of individual isomers is

smaller due to more evenly distributed compositions. Therefore, the QSPR sampling method requires more detailed compositional information, down to the individual fuel component or a restriction of the possible isomers, especially for synthetic fuels, to reduce the uncertainties in the prediction. For the DC model displayed in Figure 4.5 (a), the comparably small PI can be explained by the utilized data, composed solely of fuels without the strong variances due to possible isomers.

Comparing the computed metrics, the highest accuracies are observed for the DC model (a), with a MAE of 0.99 °C for conventional and 2.4 °C for synthetic fuels. The PICP of the DC model are comparable to 95 %, while the NMPIW of 100 to 500 %, which illustrates that the prediction uncertainty exceeds the measurement reproducibility multiple times.

4.1.6 Freezing point

For the freezing point strong, differences in the accuracy of the predictions for synthetic and conventional fuels are visible in Figure 4.6. While the predictions of the developed DC, M-QSPR and QSPR sampling models for conventional fuels are aligned with the unity line, the predictions for synthetic fuels strongly deviate in part, see Figure 4.6 (a-c).

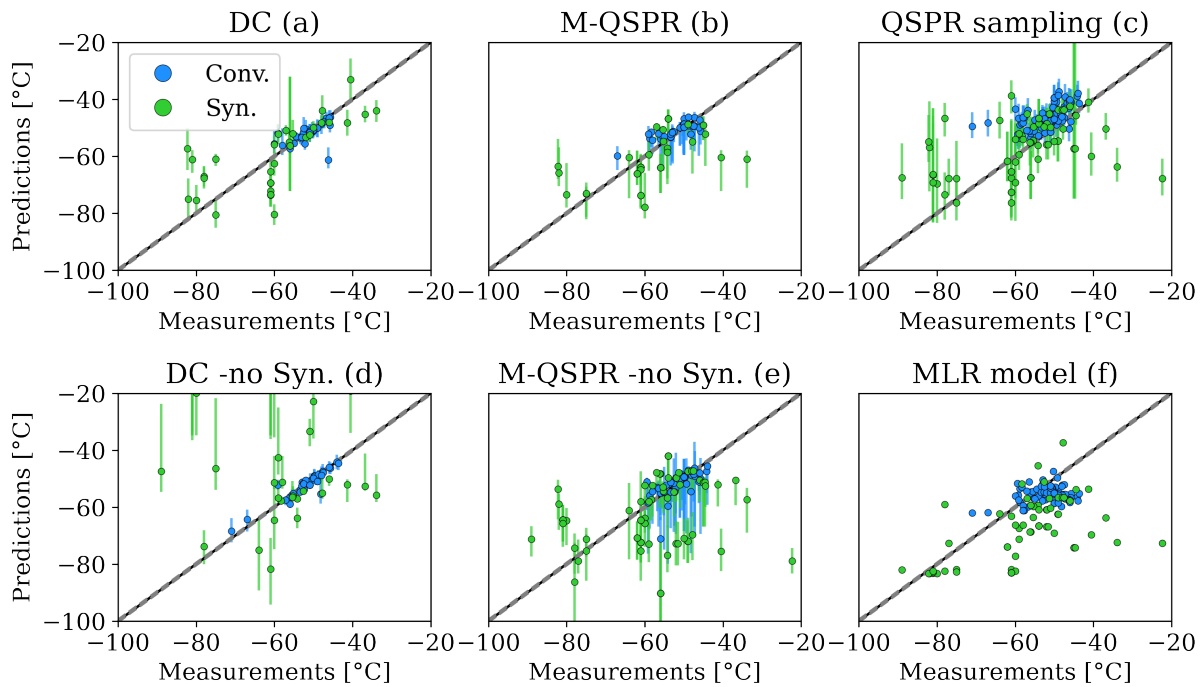


Figure 4.6: Validation results of the freezing point prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.

For the DC model and M-QSPR in Figure 4.6 (a) and (b) strong deviations are observed, particularly for synthetic fuels with freezing points below -60 °C. For the QSPR sampling model in Figure 4.6 (c) the large uncertainties are striking for the predictions of most of the synthetic fuels. The developed models trained solely on conventional fuels, DC no-Syn. and M-QSPR no-Syn. model, over- and underestimate the freezing points of the synthetic fuels significantly, see

Figure 4.6 (d) and (e). For the reference MLR model in Figure 4.6 (f), the predictions for synthetic fuels also deviate and some form a parallel line to the measurement axis, indicating a systematic error. Likewise, to the flash point, the large uncertainties and significant deviations of the M-QSPR model (b) and the QSPR sampling model (c) can be explained by the smaller dataset and the strong influence of the different isomers on the freezing point. For iso-alkanes with 10 carbon atoms, value differences of $-110.15\text{ }^{\circ}\text{C}$ to $-12\text{ }^{\circ}\text{C}$ are listed in the pure compound databases. Examples for other families will be given in later in Section 5.3.6. This explains the strong deviations and the uncertainties for the QSPR sampling model (c) and shows the need for more detailed compositional information down to the individual fuel component or a restriction of the sampled isomers. Generally, the influence of the isomers on the modeling of the freezing point is observed to be significant. Based on the unity plots, only the DC model (a) and the M-QSPR (b) show the ability to model the freezing point without a systematic error.

	Conv. Fuels				Syn. Fuels			
	MAE [°C]	PICP [%]	NMPIW [%]	MAOE [°C]	MAE [°C]	PICP [%]	NMPIW [%]	MAOE [°C]
DC (a)	1.51	89.7	633.3	54.07	7.31	47.2	1303.5	62.78
M-QSPR (b)	2.24	78.9	1148.0	52.42	7.3	50.0	1557.7	57.45
QSPR sampling (c)	5.75	41.6	949.2	49.16	9.16	63.8	2397.5	55.67
DC -no Syn. (d)	1.12	92.7	464.3	52.52	30.6	21.1	3088.0	39.72
M-QSPR -no Syn. (e)	2.33	88.6	2023.0	50.7	11.62	42.0	1780.4	62.87
MLR model (f)	5.02				12.24			

Table 12: Predictive capabilities of freezing point models.

The predictive capability metrics in Table 12 also show the highest accuracies for the DC and the M-QSPR models with MAE of $1.51\text{ }^{\circ}\text{C}$ and $2.24\text{ }^{\circ}\text{C}$ for conventional fuels and $7.14\text{ }^{\circ}\text{C}$ and $7.3\text{ }^{\circ}\text{C}$ for synthetic fuels, see Table 12. The strong deviations observed for fuels with freezing points below $-60\text{ }^{\circ}\text{C}$ would however limit the application domain to a value range above $-60\text{ }^{\circ}\text{C}$. Furthermore, the use of the calculated MAOE is recommended since the calculated PICP are significantly lower compared to the confidence level of 95 %, with 89.7 % and 79 % for conventional fuels and 48.65 % and 50 % for synthetic fuels for the DC (a) and M-QSPR (b) model respectively. The precision of the predictions expressed by the NMPIW exceeds the reproducibility of the measurements multiple times, illustrating the high level of uncertainty associated with the predictions.

4.1.7 Cetane number

Since the number of available fuel data points for the modeling of the cetane number is strongly limited with only 36 unique fuels, the input format of the DC model is modified to prevent overfitting. Therefore, the mass fractions of carbon atoms in the ranges 1-12 and 12-25 are summed up, creating the features n-alkanes 1-12, n-alkanes 12-25, iso-alkanes 1-12 etc. The DC

input features are thereby reduced from 200 to 16. The results for this modified DC model and all other models are shown in Figure 4.7. According to Vozka and Kilaz [43], there is currently no deterministic reference model that predicts the property value of GC or GCxGC measurements of fuels for the cetane number. Therefore, no predictions of a reference model are provided in the following.

The unity plots for the different models are displayed in Figure 4.7. For both conventional and synthetic fuels, only the predictions of the DC model and the M-QSPR model in Figure 4.7 (a) and (b) follow the unity line. For the QSPR sampling model and the models trained solely on conventional fuels in Figure 4.7 (c-e), the predictions deviate from the measurements, especially for values higher or lower than the range of conventional fuels. For the QSPR sampling model (c), large uncertainties are thereby observed.

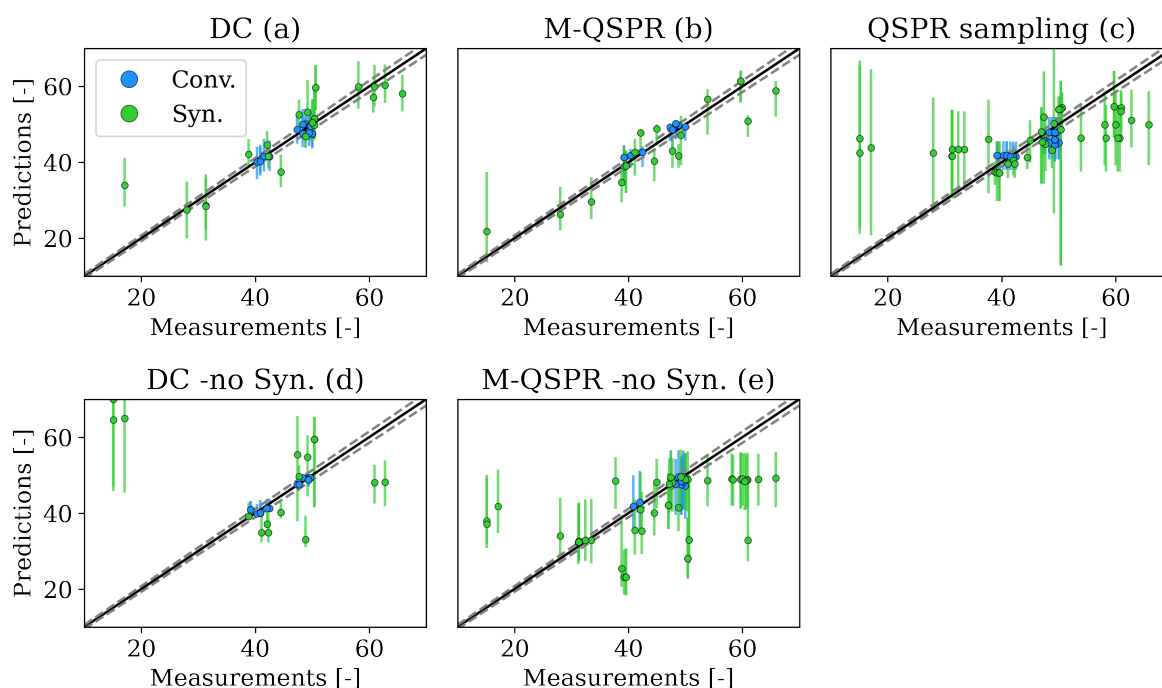


Figure 4.7: Validation results of the cetane number prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.

The cetane number is strongly influenced by the branching of a component. Low-branched components similar to cetane (hexadecane) have a high cetane number, while highly branched isomers show lower cetane numbers. For iso-alkanes with 8 carbon atoms differences between 4.9 to 47 are recorded in the data. More detailed examples that illustrate the difference in the cetane number for different isomers will be given in Section 5.3.7. Since the QSPR sampling method assumes all isomers of the considered hydrocarbon families to be equally present, significant deviations and large uncertainties are possible for synthetic fuels composed of only a few specific isomers in Figure 4.7 (c). For the QSPR sampling model, more detailed composition information down to the component level of the fuel or a restriction of possible isomers is necessary to improve

the accuracy for those synthetic fuels. Based on the unity plots, only the DC (a) and M-QSPR (b) models are able to model the cetane number without a systematic error.

The predictive capability metrics in Table 13 list the accuracies of the DC (a) and M-QSPR (b) models with a MAE of 1.02 and 1.4 for the conventional fuels and 3.74 and 3.7 for the synthetic fuels. For conventional fuels, the PI are valid with a PICP of 100 %, for synthetic fuels a PICP of 82-84 % would however recommend the use of the MROE. The NMPIW is over 5 and up to 10 times higher than the reproducibility of the measurement. Predictions are therefore associated with significantly higher uncertainties than measurements.

	Conv. Fuels				Syn. Fuels			
	MAE [-]	PICP [%]	NMPIW [%]	MAOE [-]	MAE [-]	PICP [%]	NMPIW [%]	MAOE [-]
DC (a)	1.02	100.0	321.8	0.0	3.74	82.6	444.8	3.96
M-QSPR (b)	1.4	100.0	236.9	0.0	3.7	84.2	448.3	3.23
QSPR sampling (c)	2.39	100.0	323.5	0.89	7.51	82.2	789.4	3.59
DC -no Syn. (d)	0.71	100.0	103.0	1.06	15.79	62.5	555.6	13.39
M-QSPR -no Syn. (e)	1.21	100.0	576.2	0.0	9.56	57.8	710.6	7.2

Table 13: Predictive capabilities of cetane number models.

4.1.8 Distillation line

ASTM D4054 lists limits for the distillation of jet fuels at 10, 50 and 90 vol% recovery. Likewise to the cetane number, the current literature provides no alternative deterministic model, that predicts the distillation line from GCxGC measurement data, according to Vozka and Kilaz [43]. The results of the distillation line prediction are therefore presented without the comparison to a reference model.

Figure 4.8, Figure 4.9 and Figure 4.10 show the results for 10, 50 and 90 vol% evaporated respectively, for the DC model (a) and (c) and the QSPR sampling model (b). Modeling the distillation line with the M-QSPR method would be equal to the DC method since only fuel data can be used for the correlation. For the distillation, strong differences are visible between the accuracy of the predictions for conventional and synthetic fuels. For conventional fuels, the predictions of both the DC model and the QSPR model lie close to the unity line. For synthetic fuels, this is only the case for the DC model trained on both conventional and synthetic fuels, see Figure 4.8, Figure 4.9 and Figure 4.10 (a). The predictions of the QSPR sampling model show deviations for some synthetic fuels, especially at 10 vol%, see Figure 4.8 (b). For the DC -no Syn almost all predictions for the synthetic fuels strongly deviate from the unity line, see Figure 4.8, Figure 4.9 and Figure 4.10 (c). For the DC model, this again shows the need for synthetic fuels as part of the training data.

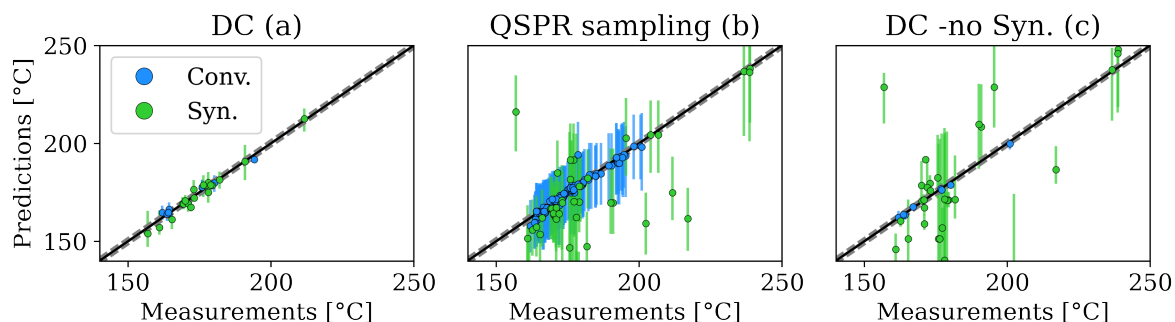


Figure 4.8: Validation results of the distillation with 10 vol% evaporated volume prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.

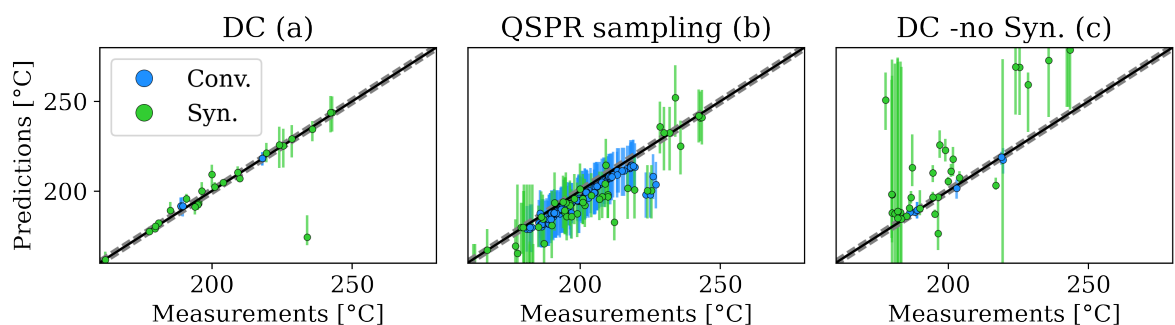


Figure 4.9: Validation results of the distillation with 50 vol% evaporated volume prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.

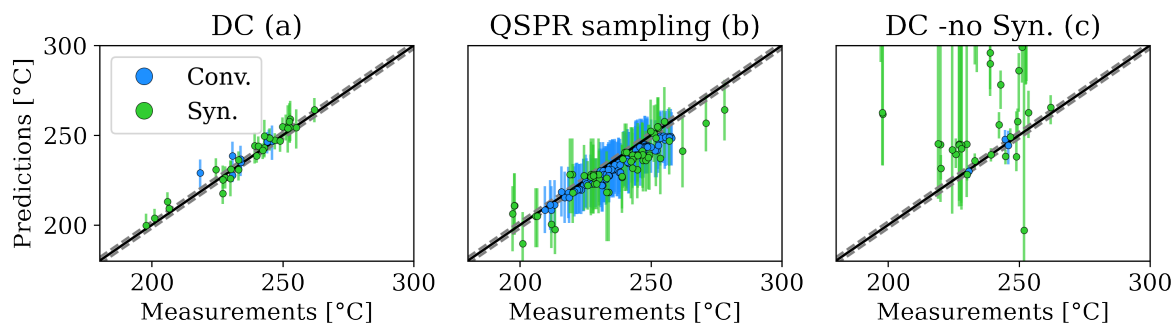


Figure 4.10: Validation results of the distillation with 90 vol% evaporated volume prediction. Results for conventional fuels are displayed in blue, results of synthetic fuels in green.

The deviating predictions of the QSPR sampling model (b) correspond to fuels with distinct compositions composed solely of iso-alkanes with certain carbon numbers like fuels produced by the ATJ process and fuel surrogates. These fuels are composed of only a small number of distinct components, which results in an unsteady distillation line. It was observed that the utilized correlation of the ASTM D2887, which is used to convert the simulated distillation line calculated on the basis of the boiling points to ASTM D86 values [109], returns erroneous and unphysical values for these synthetic fuels: Increasing temperatures are partially calculated for decreasing evaporated volumes. In those cases, the temperature at the higher evaporated value is utilized.

An update of the ASTM D2887 correction for synthetic fuels like ATJ-SPK might therefore be necessary in the future.

Compared to the predictions of the DC model (a), the predictions of the QSPR sampling model in Figure 4.8, Figure 4.9 and Figure 4.10 (b) show relatively large uncertainties both for conventional and synthetic fuels. The large uncertainties are mainly due to the uncertainties of the ASTM D2887 correction of 10.73 % for 10 vol%, 6.96 % for 50 vol% and 8.85 % for 90 vol% that are propagated in the QSPR sampling model. For conventional fuels, the reproducibilities clearly overstate the true uncertainty of the measurements, as indicated also by PICP values close to 100 % in Table 14 to Table 16. For synthetic fuels, the estimated uncertainties are on average sufficient with PICP close to 95 %. This underlines the necessity of more detailed composition information for synthetic fuels composed of only a few molecules. More detailed information about the fuel composition would probably not only increase the accuracy of the predictions but also allow a reduction of the reproducibilities of the ASTM D2887 correction model.

	Conv. Fuels				Syn. Fuels			
	MAE [°C]	PICP [%]	NMPIW [%]	MAOE [°C]	MAE [°C]	PICP [%]	NMPIW [%]	MAOE [°C]
DC (a)	1.27	100.0	206.3	0.78	2.2	100.0	321.0	1.77
QSPR sampling (b)	1.26	100.0	985.9	0.0	18.15	78.2	1292.1	25.49
DC -no Syn. (c)	0.65	100.0	157.1	0.0	38.5	62.2	1457.5	44.22

Table 14: Predictive capabilities of distillation models at 10 vol%.

	Conv. Fuels				Syn. Fuels			
	MAE [°C]	PICP [%]	NMPIW [%]	MAOE [°C]	MAE [°C]	PICP [%]	NMPIW [%]	MAOE [°C]
DC (a)	1.6	100.0	266.5	0.0	4.38	95.8	291.1	13.06
QSPR sampling (b)	4.74	96.7	748.8	11.08	12.2	86.4	926.4	16.76
DC -no Syn. (c)	0.92	100.0	244.1	0.0	30.13	52.6	1543.1	24.46

Table 15: Predictive capabilities of distillation models at 50 vol%.

	Conv. Fuels				Syn. Fuels			
	MAE [°C]	PICP [%]	NMPIW [%]	MAOE [°C]	MAE [°C]	PICP [%]	NMPIW [%]	MAOE [°C]
DC (a)	3.67	100.0	331.0	2.35	3.22	100.0	354.5	1.71
QSPR sampling (b)	4.49	100.0	785.1	0.0	6.89	100.0	917.6	1.22
DC -no Syn. (c)	1.54	100.0	287.4	0.0	35.62	52.8	1687.9	32.75

Table 16: Predictive capabilities of distillation models at 90 vol%.

The observations from the unity plots are reflected in the respective predictive capability metrics in Table 14 to Table 16. The accuracies for the predictions of conventional fuels are high, with a MAE around below 5 °C for both models. For synthetic fuels, the observed erroneous predictions reduce the accuracy with a MAE of 6.89 °C to 18.15 °C QSPR sampling model (b), while the accuracy of the DC model (a) remains high at 2.2 °C to 4.38 °C. For the DC model, the PI are valid with PICP values over 95 %. For the QSPR sampling model, this is only true for the conventional fuels. For both models, the NMPIW is several times greater than the reproducibility of the measurements, illustrating the significantly higher uncertainty of the predictions.

4.2 Part 2: Adequacy assessment of models for fuel screening

In the second part of the predictive capacity assessment, the adequacy of the developed models for the fuel prescreening is assessed based on a simulated fuel prescreening of hold-out fuels, neither seen by the model in the training nor in the testing. A model is thereby rated adequate, if the fuel screening based on the predictions yields the same results and leads to the same conclusions as the measurement. In the prescreening the predicted properties are compared with the specification limits of ASTM D4054, considering both the mean prediction and the prediction intervals (PI). If the predicted mean value and PI completely lie either inside or outside the specification limits, the prediction of the model is completely accepted; if the PI cross the specification line, a measurement is recommended, similar to the example in Section 2.5.2. If a model correctly predicts all properties inside or outside the specification limits with or without a measurement recommendation and the measurement values are enclosed by the PI, the model is considered adequate for fuel screening. In addition to the adequacy of each model, the potential benefits of considering multiple models and comparing their results to assess a jet fuel are also examined.

Three fuels are screened: one conventional oil-based fuel and two SAF fuels, one produced by the Alcohol-To-Jet process (ATJ-SPK), one produced by a Fischer-Tropsch process (FT-SPK). The conventional fuel and the ATJ-SPK fuels were composed and measured as part of the JETSCREEN project, whereas the FT-SPK fuel with the POSF 5018 was investigated in the NJFCP project. Figure 4.11 shows the plots of the GCxGC compositions of the fuels.

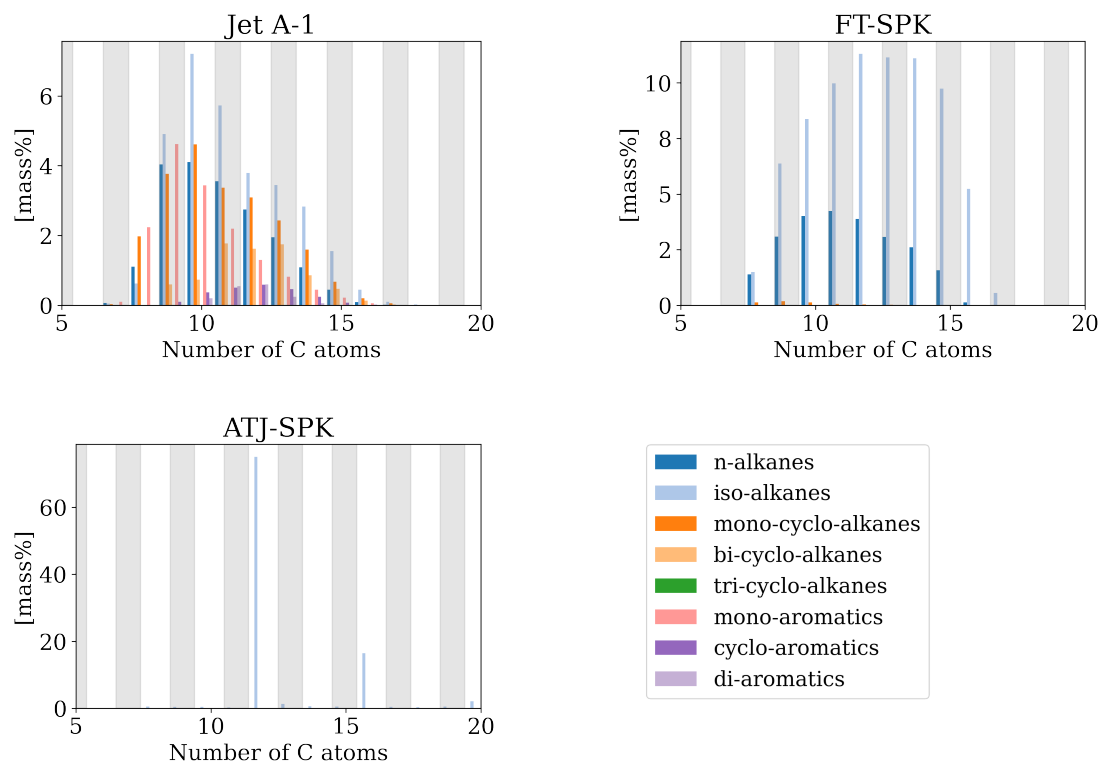


Figure 4.11: Composition plots of jet fuels used for Tier α prescreening: conventional oil-based fuel (upper left), SAF produced by the Fischer-Tropsch process (upper right) and SAF produced by Alcohol-to-Jet process (lower left).

The compositions of the three fuels differ significantly. The fuels are chosen intentionally to investigate the suitability of the models for a wide range of possible fuel compositions: a conventional fuel consisting of all hydrocarbon families with a broad distribution of components with different carbon numbers, a FT-SPK fuel with a composition dominated by two hydrocarbon families but with a broad distribution of components with different carbon numbers, and an ATJ-SPK fuel with a composition dominated by isomers with two specific carbon numbers from one hydrocarbon family.

To review the results from the first part of the predictive capability assessment, all developed models are used for the fuel screening. In the following plots, predictions of the DC model are shown in blue, the predictions of the M-QSPR model in green and the predictions of the QSPR model in purple. The property measurements of the fuels are displayed in black, with error bars indicating the reproducibility of the measurement method. The specification limits of the respective properties are shown as red lines. The range outside of these specifications are shaded in red. To additionally indicate if the property is predicted inside the specification, the frame of the individual plots is colored: green if both the mean prediction and the predicted PI of all models lie inside the specifications; orange if the PI cross the specification limits; and red if the mean prediction of one model lies outside the specification limit. The results of the prescreening

are displayed in Figure 4.12 with individual columns for each fuel and rows for each property value.

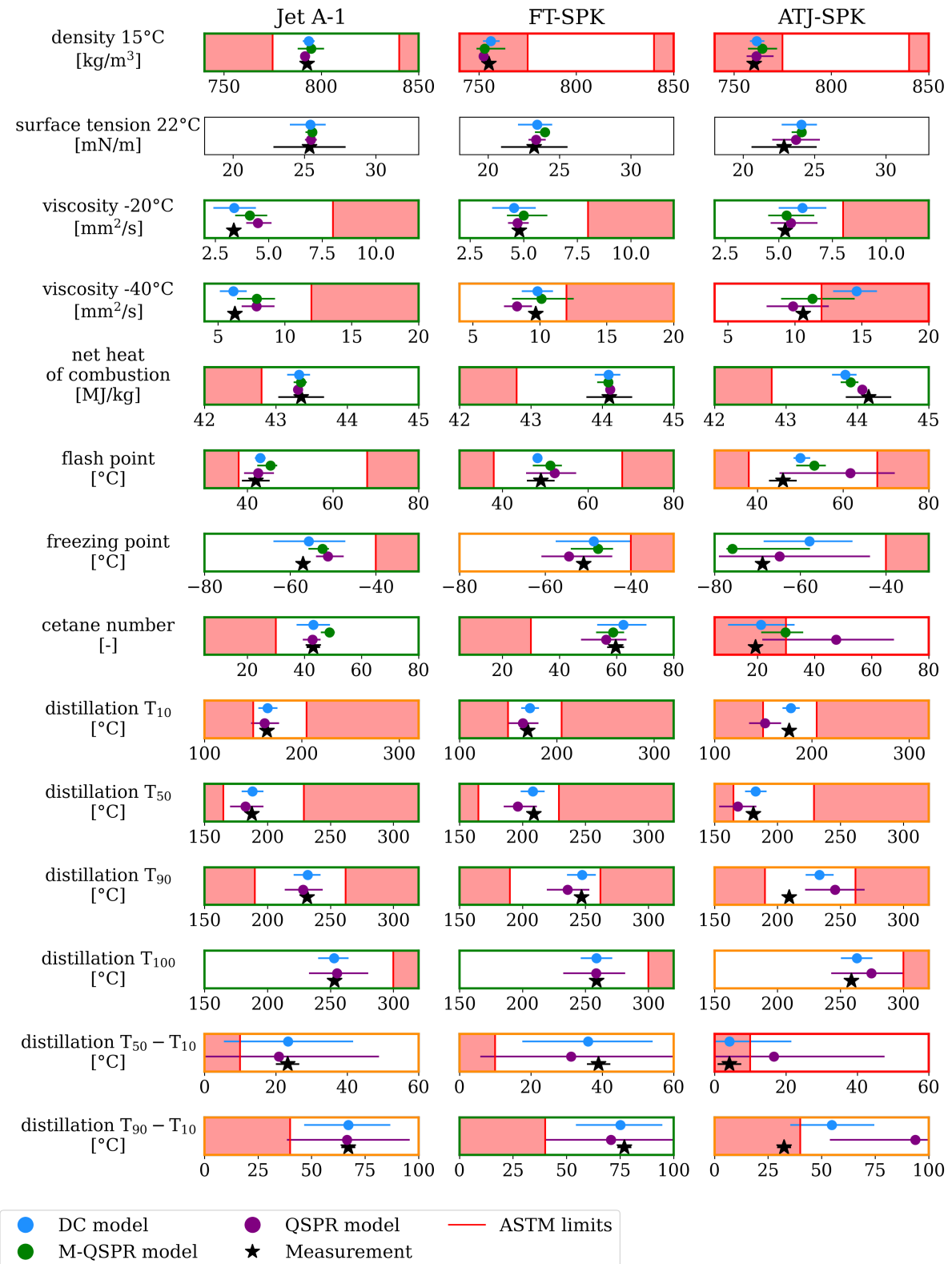


Figure 4.12: Results of Tier α prescreening: Jet A-1 fuel (left row), FT SPK (middle) and ATJ SPK (right). DC model predictions are displayed in blue, of the M-QSPR model in green and the QSPR sampling model in purple.

For the Jet A-1 fuel and the FT-SPK fuel, in the first and second columns of Figure 4.12, the predictions of all models are in high agreement with the measurements. The mean values are predicted correctly either inside or outside of the specification limits. With the exception of distillation $T_{50} - T_{10}$ and $T_{90} - T_{10}$, as well as the kinematic viscosity at -40 °C and the freezing point for the FT-SPK this is also true for the PI. For these properties, measurements would be recommended. As explained in the distillation testing Section 4.1.8, the high level of uncertainty for the distillation is mainly due to the conversion correlation of ASTM D2887. Uncertainties could therefore be reduced if the conversion correlation would be optimized. For the kinematic viscosity at -40 °C and the freezing point the PI of all models are comparable and the specifications are exceeded due to the proximity of the values to the limits. For the Jet A-1 and the FT-SPK, the high accuracy and precision of all models can probably be attributed to the composition of the two fuels. Both fuels have broad distributions of components for multiple families. For the M-QSPR model and the QSPR sampling the broad distribution over the families corresponds to the underlying assumption, that all isomers are present with the same likelihood. For the DC model the high accuracy and precision can be explained by the high similarity of the two hold-out fuels, to fuels in training data, as observed in the previous Section 4.1.

As result of the screening, the Jet A-1 would correctly be assessed as fuel with high chances for approval with the recommendation for a measurement of the distillation line. For the FT-SPK the density would be correctly assessed as too low, while the other properties indicate high chances for approval. The low density could be adjusted in a blending operation. Due to the exceeding PI of the viscosity at -40 °C, the freezing point and the distillation line, measurements would however be recommended for these properties. For these two fuels, all three developed models with their underlying assumptions showed high predictive capability and are rated adequate for the fuel prescreening. The high agreement of the model predictions with overlapping PI and close mean predictions was thereby observed as a trust indicator that correctly illustrated that the true value was close to the predictions.

For the ATJ-SPK fuel differences between the predictions of the models increase and also greater deviations from the measurements are visible. Accurate predictions and high agreement between the models are only observed for the density, surface tension and kinematic viscosity at -20 °C. For the other properties, the model predictions differ in part from each other and even contradict in their prediction of the properties as being inside or outside of the specification limits. Large uncertainties are thereby observed, particularly for the QSPR prediction of the flash point, freezing point and cetane number. For the kinematic viscosity at -40 °C the M-QSPR and the QSPR sampling model predict the viscosity inside the specification limits with a recommendation for a measurement. The DC model however, falsely predicts the viscosity value completely outside the specification limit with the mean value and PI. For the cetane number and the distillation $T_{90}-T_{10}$ and $T_{50}-T_{10}$ the mean values are falsely predicted as being inside the specification limits by both the QSPR model and for $T_{90}-T_{10}$ and the DC model. For these predictions, measurements

are recommended as the PI cross the specification line. But if the measurements are not conducted, the results of the screening would be erroneous and would not correspond to the measurements. Only the M-QSPR model correctly predicted all properties as inside and outside of the specification, pointing out both the low density and cetane number. However, the inability for modeling the distillation line prevents a screening based on this model alone. Considering the predictions of all models simultaneously does not help in this situation. Kinematic viscosity and cetane number are predicted correctly by the majority of the models, however the T_{90} - T_{10} is predicted wrong both with the DC and the QSPR sampling model. Therefore, none of the models is adequate for the screening of ATJ-SPK.

The lower accuracy and precision of the model for the ATJ fuels can be explained considering the composition of the ATJ-SPK, and the set modeling assumptions. The ATJ-SPK consists to 91 mass% of certain iso-alkanes isomers with 12 and 16 carbon atoms. For the DC model no fuel similar enough e.g. with the same isomers is present in the training data to explain the deviation for the distillation at T_{90} . For the QSPR sampling and M-QSPR models the assumption, that all isomers are present with the same likelihood in the composition is no longer viable. For the QSPR sampling model this results in the large observed PI and inaccurate predictions. To illustrate the influence of different isomers on the prediction of the ATJ-SPK with the QSPR sampling model, the predicted values for the flash point, freezing point and cetane number of all considered iso-alkanes isomers with 12 and 16 carbon atoms are shown in Figure 4.13.

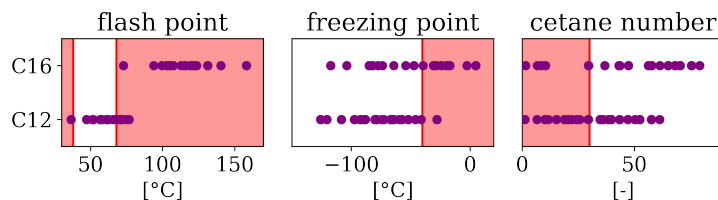


Figure 4.13: Predicted mean values for iso-alkanes isomers with a carbon number of 12 and 16 for the flash point (left), freezing point (middle) and cetane number (right).

Figure 4.13 clearly illustrates the large range of the property values, spanning multiple magnitudes and crossing the specification limits. To therefore reduce uncertainty and increase the accuracy of the QSPR sampling model, the selection of possible isomers must be constrained.

Prediction with constrained isomer selection for ATJ-SPK

To investigate if a constrained selection of the isomers changes the predictions and overall adequacy of the QSPR sampling model for the ATJ-SPK fuel, the prescreening with this model is repeated. For the investigated ATJ-SPK a special situation applies, since the fuel components that make up over 90 mass% of the composition are actually known: 2,2,4,6,6-pentamethyl heptane for iso-alkanes C 12 and 2,2,4,4,6,8,8-heptamethyl nonane for iso-alkanes C 16. Since the QSPR sampling is a flexible modeling approach, that allows the direct selection of identified components, the properties can again be predicted using these two isomers and randomly

sampling the isomers for the other carbon numbers. The results of the repeated screening are shown in Figure 4.14. The predictions with the constrained isomer selection are shown in a lighter purple, the previous predictions of the QSPR sampling model considering all isomers are also indicated as reference.

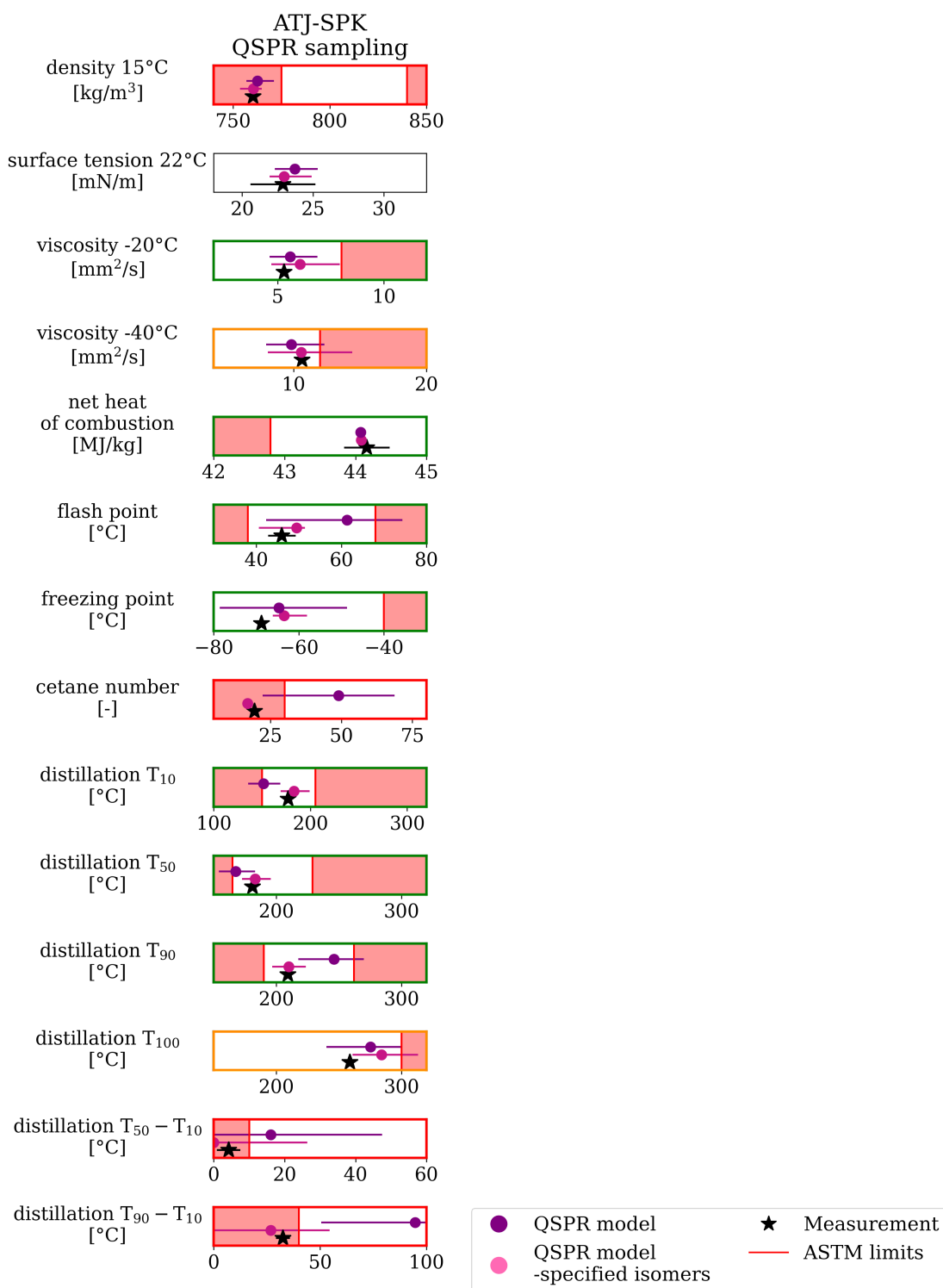


Figure 4.14: Results of Tier α prescreening for ATJ SPK with constrained isomer selection, predictions of the QSPR sampling model.

For the flash point, the freezing point and the cetane number, significantly lower PIs are visible for the restricted isomer selection compared to the reference predictions taking all isomers into account. Increases in the accuracies are observed for the flash point and cetane number where the property values are correctly predicted inside and outside the specification limits respectively. Increased accuracy is also observed for the distillation line for T_{10} , T_{50} and T_{90} . The increased uncertainty for the viscosity prediction can probably be attributed to the uncertainty of the predictions for the individual pure compounds at these low temperatures. The individual predicted value ranges of the components have a greater impact when fewer molecules are considered, resulting in more uncertainty in the prediction of the bulk property.

The predictions with the constrained isomer selection change the adequacy of the prediction results for the three outlined properties. Flash point, freezing point and cetane number are now predicted correctly inside and outside of the specification limits, yielding the correct screening result. With the constrained isomer selection, the QSPR sampling model can therefore be rated adequate. Measurements would still be recommended for the viscosity at -40 °C and the distillation line.

4.3 Summary and conclusion

Probabilistic modeling methods play a central role in the prescreening concept of Heyne and Rauch and enable the assessment of jet fuel candidates from low volume composition measurements. Their ability to predict critical physicochemical properties and corresponding uncertainties accurately and reliably, allows for risk-informed decision making in the early development stage of fuel candidates.

In this chapter, probabilistic property models from three different methods were investigated and compared: Direct correlation (DC), Mean Quantitative Structure-Property Relationship Modeling (M-QSPR) and Quantitative Structure-Property Relationship Modeling (QSPR) with sampling. The models are compared on their predictive capability for the modeling of physicochemical properties in general and investigated for their adequacy to support the jet fuel prescreening. To put the predictive capability of the developed models in context, they are compared with state-of-the-art deterministic models from the literature, to compare differences in accuracy and illustrate the benefit of predicted uncertainties. As reference models the DLR Discreate Component Model (DCM) by Le Clercq and multilinear regression models (MLR) by Liu et al. were selected.

Using a Monte-Carlo Dropout Neural Network as probabilistic regression algorithm, the models were trained and tested for their predictive capability to model the eight critical physicochemical properties relevant for the jet fuel prescreening. The assessments were conducted on 75 conventional oil-based jet fuels and 56 synthetic fuels, as well as fuel blends.

The predictive capability assessment proved that all developed models are in general able to model the eight properties on the basis of the GCxGC measurements, without a systematic error. The highest predictive capability was on average observed for the DC method. With the exception of the freezing point, highly accurate and precise predictions were achieved for fuels of all considered fuel types and production paths. However, a comparison with the results of DC models trained solely on conventional fuels revealed the need of this method for fuels similar to the predicted fuels as part of the training data. If the DC model is trained solely on conventional fuels, accuracy and precision for the prediction of synthetic fuels are significantly reduced.

The M-QSPR model showed less dependence on the synthetic fuel data, and the second-best performance for most of the properties. The possibility to learn from both data of pure compounds and fuels proved to be beneficial, especially for cases with a low number of available fuel measurements as training data, like for the modeling of the cetane number.

For the QSPR sampling model, the predictive capability was highly dependent on the fuel composition and the predicted property. For fuels where the underlying assumption, that all considered isomers are present with the same likelihood, was valid, a high predictive capability was observed for all properties. For synthetic fuels, where this assumption was invalid, large uncertainties and deviations were visible, especially for properties where the values of the different isomers differs significantly, e.g. freezing point and cetane number. Since the QSPR sampling model is the most flexible modeling method, that allows the definition of the fuel down to the component level, higher accuracies could probably be achieved, if more detailed fuel composition is known, or if the range of possible isomers is constrained.

In comparison with the deterministic state-of-the-art reference models, all developed probabilistic models showed higher accuracy and outperformed the predictive capability of the reference models for all eight critical properties. This was especially visible for the prediction of the kinematic viscosity, where the strong deviations of the DCM predictions can be attributed to the use of just one isomer as representative component for a family and carbon number. The additionally estimated prediction intervals proved to be particularly useful to illustrate uncertainty associated with each prediction, an important information deterministic models cannot provide.

To review the adequacy of the models for the actual fuel prescreening, the prescreening was simulated for three different hold-out fuels that were not seen by any model, neither in training nor in validation. One conventional Jet A-1 fuel and two synthetic fuels produced by the Fischer-Tropsch process (FT-SPK) and the Alcohol-to-Jet process (ATJ-SPK). For both the Jet A-1 and the FT-SPK all three models proved to be adequate for the screening correctly predicting all properties as either inside or outside the specification limit. The high accuracy and precision of the model predictions were explained by the compositions of the fuels. Jet A-1 and the FT-SPK are composed of components with a broad range of different carbon numbers, for the Jet A-1

from different families. The underlying assumption of the models, that all considered isomers are present in the fuels with the same likelihood, was therefore found valid for these fuels.

For the ATJ-SPK stronger differences in the property predictions of the different models were visible. Only the predictions of the M-QSPR models correctly predicted all properties inside and outside the specification, pointing out the low density and cetane number of the fuel. For the DC and QSPR sampling models, erroneous predictions were observed, which would have yielded a false screening result for the fuel. For the QSPR models enlarged PI were visible that indicated high uncertainty in the property predictions. For both models this was again attributed to the composition of the ATJ-SPK, which in contrast to the previous models is dominated by certain iso-alkanes isomers with 12 and 16 carbon atoms. For the DC models no fuel similar enough was available in the training data, while for the QSPR sampling models the assumption that all isomers occurred with the same likelihood was no longer valid. To investigate if a constrained isomer selection would change the adequacy of the QSPR sampling model, the prescreening was repeated. Known isomers were thereby set for the composition input of the QSPR sampling model. The constrained selection led to significantly improved results with the QSPR sampling model. The uncertainties of the predictions for the flash point, freezing point and cetane number were reduced and the cetane number correctly predicted as too low. The QSPR sampling model with the constrained isomer selection was thereby rated adequate.

5 Development of Fuel Design Tools

The design of jet fuels is the second intended application area of the data-based tools developed in this work. Fuel design methods shall support the formulation of jet fuel candidates to optimize the chances for passing the fuel approval process after ASTM D4054. In contrast to the fuel prescreening, the fuel design not only explains the reasons for the non-compliance but also proposes reformulation measures for the fuel to optimize the properties accordingly.

To support the fuel design, tools are required that illustrate the relationship between fuel composition and properties. In contrast to previously described predictive models, these tools should provide insights into how modifications in fuel composition can affect its properties. The influence of various aspects of fuel composition on properties, such as the proportion of hydrocarbon families or the size of components, must be understood in order to derive tools that allow estimation of changes in properties with corresponding changes in composition. This chapter investigates the correlations between fuel composition and properties of interest. From these investigations, tools are created that serve as basis for the later fuel design in Chapter 6.

Fuels are mixtures of hundreds of compounds, which are present in the fuel as components with a set mass fraction. The properties of fuels are consequently determined by the contributions of the individual components as well as by their interactions. The design of jet fuels therefore requires a comprehensive understanding of the components and their properties relevant for the fuel approval. This is becoming increasingly important as synthetic jet fuels often consist of a smaller number of families and thus a smaller number of components compared to conventional crude oil-based jet fuels. As shown in Section 4.2, the bulk properties of fuels like the ATJ-SPK are primarily determined by a few distinct components that make up a large portion of the fuel composition.

The underlying processes of the production of synthetic fuels allow a more and more precise tailoring of the fuel composition, close to the individual component. Chemical family, size, and topology of components are key design parameters that can be adjusted to formulate a promising fuel candidate [73,110,111].

To support the fuel design and understand the relations between fuel compositions and properties, the present research investigates the influence of structural aspects of the fuel components relevant for jet fuels based on their chemical family, size and topology. Braun-Unkhoff et. al. investigated the relation between the size of typical fuel components and their physicochemical properties on the basis of measurements from the DIPPR 801 Database [70]. They especially pointed out the potential influence of the branching as a molecular topology parameter on properties like the freezing point [111]. Wang et. al. summarized the work of different publications for a selection of five critical jet fuel properties (e.g., density, net heat of combustion, and flash point) [107] and identified the hydrogen content and molar mass as

important component features for correlations with properties. The search for possible single compound jet fuels illustrated the importance of the chemical family by pointing out certain cyclo-alkanes as most promising candidates due to the similarity of their physical properties to the average properties of conventional jet fuels [112–114]. Work on the fuel design of diesel and gasoline fuels investigated the structural influence of compounds on the ignition behavior characterized by cetane and octane numbers, pointing out the influence of the fraction of certain chemical families, e.g., with high fractions of n-alkanes leading to increased cetane number, but also the number and size of side chains [115,116].

The influence of quantitative structural features on the modeling of properties like the density, flash point, freezing point, and cetane number was described *inter alia* by Saldana et al. [71,72,117], Creton et. al. [118], and Cai et al. [119] as part of their development of QSPR predictive property models for fuel components. By sensitivity analysis or investigation of importance of the utilized QSPR features, they outlined the influence of chain length and hydrocarbon family, but also structural aspects like the number and position of methyl groups on the listed properties. Apart from statistical analysis and analysis of the trained models, direct molecular design is an emerging research topic [120,121]. This technique couples predictive property models with a molecular generator in an optimization loop to iteratively find / optimize the molecules with properties as components the intended fuel application. Up to now, recent research focusses predominantly on fuel components and low compound mixtures for gasoline fuels, with the goal to optimize properties that characterize the combustion behavior like the octane number, and other relevant properties for the field, e.g., vapor pressure, distillation behavior, density, and viscosity [122–125]. The molecular optimization is thereby carried out in a closed loop, where the molecular design is fully conducted by the interplay of property models and optimizer [125–127]. The property models estimate the desired characteristics of candidate molecules, while the optimizer generates new molecular structures based on these predictions, iteratively refining the search towards molecules with optimal properties. Information about structural aspects is not directly investigated but could be inferred by sensitivity analysis of the utilized models or comparison and analysis of the molecular structures over the optimization iterations.

The listed publications investigated the relationships between the fuel components and their properties either primarily qualitatively, with the focus on investigating trends, low compound mixtures and in part other application areas. No publication has yet provided a comprehensive investigation for all eight critical properties, relevant for the jet fuel screening with the focus on identifying and quantifying the relations between compound structures and properties. This work extends the outlined publications, providing investigations for all critical properties and all relevant hydrocarbon families. Using a significantly larger database compared to previous publications, with 1870 possible fuel components, the influence of the chemical family, size and branching on the properties of the component is investigated. Investigations are conducted both visually and with the use of uniform quantitative metrics. The investigations solely rely on

measurement data, since predicted property values, e.g. with a QSPR model, would assume the correct property prediction for each individual component. This can however not be guaranteed, especially for properties with a smaller number of measurements for the training of the model, see e.g. flash point in Section 3.3.2. Erroneous predictions for individual components could lead to the misidentification of correlations, in contrast to fuel modeling, where the high number of components reduces the individual influence for most of the fuels, especially for conventional ones. A comprehensive investigation of the relations between composition and properties based on the available measurement data is therefore the safest approach to derive reliable fuel design tools, while simultaneously pointing out missing data and the need for further research.

The component-property relations are quantified for the three aspects: hydrocarbon family, size and branching of the components. Eight relevant physicochemical properties are considered: density, surface tension, kinematic viscosity, net heat of combustion, flash point, freezing point, cetane number and the distillation line. Since non-CO₂ effects of aircraft emissions, especially via soot emissions and contrail formation, are of growing concern [8], the sooting tendency of the components is considered additionally. As a quantitative metric for the sooting tendency, the yield sooting index of the components is utilized [128,129].

5.1 Molecular descriptors

To investigate the influence of the topology of the fuel components, five molecular descriptors are introduced that quantify different aspects of the topology of the fuel components. To quantify the size of the components, the number of contained carbon atoms or carbon number (nC), is utilized.

To quantify the branching, the Topochemical Atom Index for branching (η_B) is utilized [130]. This index quantifies the branching based on the Valence Electron Mobile environment (VEM). The VEM considers the number of connections, the kind of bond, as well as the position of the atoms. Calculating the VEM involves two steps. In a first step, the vertex count γ_i is calculated for every atom i , by relating the atomic number of an atom α_i to the VEM count β_i [131], see Equation (5.1).

$$\gamma_i = \frac{\alpha_i}{\beta_i} \quad (5.1)$$

$$\beta_i = \sum_{\sigma(i)} f_{\sigma} + \sum_{\pi(i)} f_{\pi} \quad (5.2)$$

For a carbon atom the α_i in Equation (5.1) equals to 0.5. The VEM count β_i is calculated from two local vertex invariants f_{σ} and f_{π} , accounting for non-hydrogen σ bonds and π bonds, see Equation (5.2). The VEM count β_i therefore depends on the bonds of the individual atom to all atoms in the fuel component it is connected to. For a carbon atom with non-hydrogen σ -bonds

f_σ equals to 0.5, for π -bonds f_π equals to 1 [132]. In the second step, the VEM vertex counts γ_i of all atoms A are multiplied with their neighboring γ_j to calculate the local Topochemical Atom Index η^{loc} , under consideration of the adjacency matrix a_{ij} , see Equation (5.3). The adjacency matrix is a two-dimensional matrix that indicates which atoms of a molecule are bonded. Figure 5.1 gives an example of the calculation of the adjacency matrix for iso-pentane. The adjacency matrix has rows and columns that correspond to the vertices of the molecular structure. Connections between atoms in Figure 5.1 are indicated by 1 while 0 indicates no connection.

$$\eta^{loc} = \sum_i^{A-1} \sum_{j=i+1}^A a_{ij} * (\gamma_i * \gamma_j)^{0.5} \quad (5.3)$$

$$\eta_B = \eta_N^{loc} - \eta^{loc} + 0.086 * nR \quad (5.4)$$

The branching index η_B is finally calculated from the difference between the η_N^{loc} of the unbranched isomer of the component (e.g. pentane for iso-pentane) and the η^{loc} of the component of interest (e.g. iso-pentane). To account for ring structures Equation (5.4) contains an additional term that takes the number of rings (nR) in the molecule into account (e.g. 1 for toluene, 2 for naphthalene)

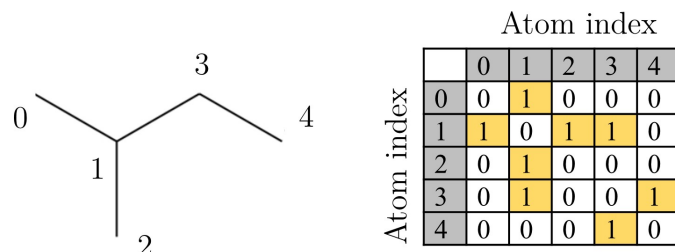


Figure 5.1: Molecular structure of isopentane (left) and its calculated adjacency matrix (right).

As additional structural descriptors for the topology, the partial positive surface area ($PPSA$) and the number of ring atoms (nR) as well as the number of aromatic carbon atoms connected only to other carbon atoms are ($cXHO$) considered. All three descriptors were identified as particularly important by Saldana et al., as features for the QSPR modeling of fuel component properties [71,72]. Likewise to the nC , nR states the number of atoms in a circular structure. nR is considered to investigate the influence of circular structures on the properties. $cXHO$ similarly counts the number of carbon atoms in an aromatic ring, connected only to other carbon atoms.

The $PPSA$ quantifies the sum of the surface area SA_a^+ of a molecule, see Equation (5.5) [131]. The individual surface areas of the molecules SA_a^+ are thereby calculated with an underlying rolling ball algorithm. In contrast to the other descriptors, the $PPSA$ is a 3D descriptor, meaning it calculates the $PPSA$ for a 3D representation of the molecule.

$$PPSA = \sum_{a+} SA_a^+ \quad (5.5)$$

All descriptors are calculated using the PaDEL software [133] and the corresponding Python wrapper PaDELPy [134], that calculates the descriptors on the basis of the SMILES representation of the component.

5.2 Property correlation metrics

To quantify the correlations of the properties with the structural descriptors nC and η_B , as well as the hydrocarbon family of the components, four respective property metrics are introduced. The influence of the nR , $cXH0$ and the $PPSA$ are only investigated based on visual property correlations, since their relevance is observed only for some hydrocarbon families and properties.

The first metric Δ_{HC} investigates the influence of the hydrocarbon family on the respective properties, keeping the size of the components and their branching constant. Therefore, the average property value of a hydrocarbon family ($\bar{y}_{C_7-C_{16}}^{HC}$) is calculated considering all components of a family with 7 to 16 carbon atoms. $\bar{y}_{C_7-C_{16}}^{HC}$ is calculated using components with 7 to 16 carbon atoms, since this is the typical carbon range for jet fuels. The conventional reference fuels hold on average 99 % of their mass fraction in the interval of 7 to 16 carbon atoms. The average difference Δ_{HC} ($\bar{y}_{C_7-C_{16}}^{HC}$) is then calculated by relating $\bar{y}_{C_7-C_{16}}^{HC}$ to the average property value of conventional crude oil-based jet fuels (\bar{y}^{fuels}), see Equation (5.6). As data for the conventional reference fuels, property measurements from the CRC world fuel survey from 2006 are used [92]. This metric provides an estimate of the difference in property values as the fraction of the hydrocarbon family in the fuel is increased or decreased.

$$\Delta_{HC} (\bar{y}_{C_7-C_{16}}^{HC}) = \frac{\bar{y}_{C_7-C_{16}}^{HC} - \bar{y}^{fuels}}{\bar{y}^{fuels}} * 100\% \quad (5.6)$$

The second metric quantifies the overall correlation of property values with the structural descriptors on the basis of the Pearson Correlation Coefficient (r). The Pearson Correlation Coefficient is a common metric for the identification of correlations between two variables x and y , see Equation (5.7). The value range for r lies between -1 and 1, where -1 indicates a strong negative and 1 indicates a strong positive correlation. A r value below -0.5 or above 0.5 is normally used as threshold for an identified correlation [57]. For reasons of clarity, the r is scaled between -10 and 10 in this work. The thresholds for the identification of a correlation are moved to -5 and 5 respectively. For the correlation of property values with nC , all components with 7 to 16 carbon atoms are considered. For the calculation of the correlation with η_B however, only the components of one set carbon number and hydrocarbon family are considered, to exclude the influence for changes in nC , e.g. iso-alkanes with 10 carbon atoms. The correlation of η_B is thereby conducted always for the components with the greatest number of available measurements.

$$r = \frac{\text{cov}(y, x)}{\sigma_x \sigma_y} \quad (5.7)$$

The third metric quantifies the average differences in the property values of fuel components with different carbon numbers from one hydrocarbon family. This metric Δ_{nC} thus provides an estimate of the difference in property values when the carbon number of the fuel components is increased or decreased by one, keeping the hydrocarbon family and average branching constant. To compute Δ_{nC} , the absolute differences of the average value of components with a set carbon number (\bar{y}_{C_i}) to the average values of components with all other carbon numbers ($\bar{y}_{C_{j \neq i}}$) are calculated, e.g. average density of iso-alkanes with 8 carbon atoms to iso-alkanes with 9 carbon atoms etc. The difference of \bar{y}_{C_i} and $\bar{y}_{C_{j \neq i}}$ are divided by the difference of the carbon numbers (C_i, C_j), see Equation (5.8). Δ_{nC} is calculated for every hydrocarbon family individually.

$$\Delta_{nC} = \frac{1}{N + M} \sum_{C_i \geq 7}^{N=C_i \leq 16} \sum_{C_{j \neq i} \geq 7}^{M=C_j \leq 16} \frac{1}{C_i - C_j} \left| \frac{\bar{y}_{C_i} - \bar{y}_{C_j}}{\bar{y}_{C_j}} \right| * 100 \% \quad (5.8)$$

To quantify the influence of the branching, the absolute average difference of isomers Δ_{η_B} at a set carbon number is estimated using Equation (5.9).

$$\Delta_{\eta_B} = \frac{1}{N} \sum_i^N \frac{|y_i - \bar{y}|}{\bar{y}} * 100 \% \quad (5.9)$$

This fourth metric estimates the average difference in a property values that can be observed for two different isomers. Due to the limited availability of the data and the clarity of the presentation, Δ_{η_B} is only calculated once for every hydrocarbon family, always for isomers with the highest number of available measurements, e.g. iso-alkanes with 10 carbon atoms. The made observations are consequently assumed to be representative for all isomers in a hydrocarbon family.

5.3 Investigation of fuel component structure-property relations

The component structure-property relations are investigated both visually by plots and quantitatively by the introduced metrics. While the plots serve the visually identification of correlations for one structural descriptor, the metrics allow the simultaneous comparison of the correlations for all properties. Detailed explanations of the individual plots and tables for the metrics are given in the upcoming Section 5.3.1 Density, but applied for all other properties.

For the temperature dependent properties density, surface tension and kinematic viscosity, ASTM D1655 and D4054 specify a measurement temperature of 15 °C for the density, 22 °C for the surface tension and -20 °C and -40 °C for the viscosity. Since most of the component property

measurements in the available databases, do not contain measurements at these exact temperatures, the values for the density and surface tension are interpolated using linear interpolation. Since all relevant fuel components do not undergo a phase change at these temperatures, the linear interpolation is very reliable. For the low temperatures of $-20\text{ }^{\circ}\text{C}$ and $-40\text{ }^{\circ}\text{C}$ however, only very few property measurements are recorded. The property values would therefore have to be extrapolated for most of the components. Since correct extrapolation can neither be guaranteed nor verified, values are interpolated at $0\text{ }^{\circ}\text{C}$ for the investigation, using the ASTM D341 equation presented in Section 2.2.1. It is assumed that the observations made at $0\text{ }^{\circ}\text{C}$ also apply to viscosity at -20 and $-40\text{ }^{\circ}\text{C}$, an assumption that cannot be verified due to the low availability of measurements at those temperatures. More low temperature measurements are therefore needed in the future to verify this assumption in later work.

5.3.1 Density

As a fundamental physicochemical property of a fuel, the density impacts the loadable weight of an aircraft, the energy content of the fuel together with the net heat of combustion, and it is an important parameter for metering and balancing of the aircraft, e.g. for fuel gauging.

To investigate how the density of fuel components and consequently the density of fuels is affected by changes in the values of the structural descriptors, correlations are first investigated visually. First, the differences of the density values of the different hydrocarbon families and the change of the density values over the carbon number of the components is investigated. Therefore, Figure 5.2 displays the density values of components for all hydrocarbon families over nC . Values are plotted for components with carbon numbers in a range of 5 to 20.

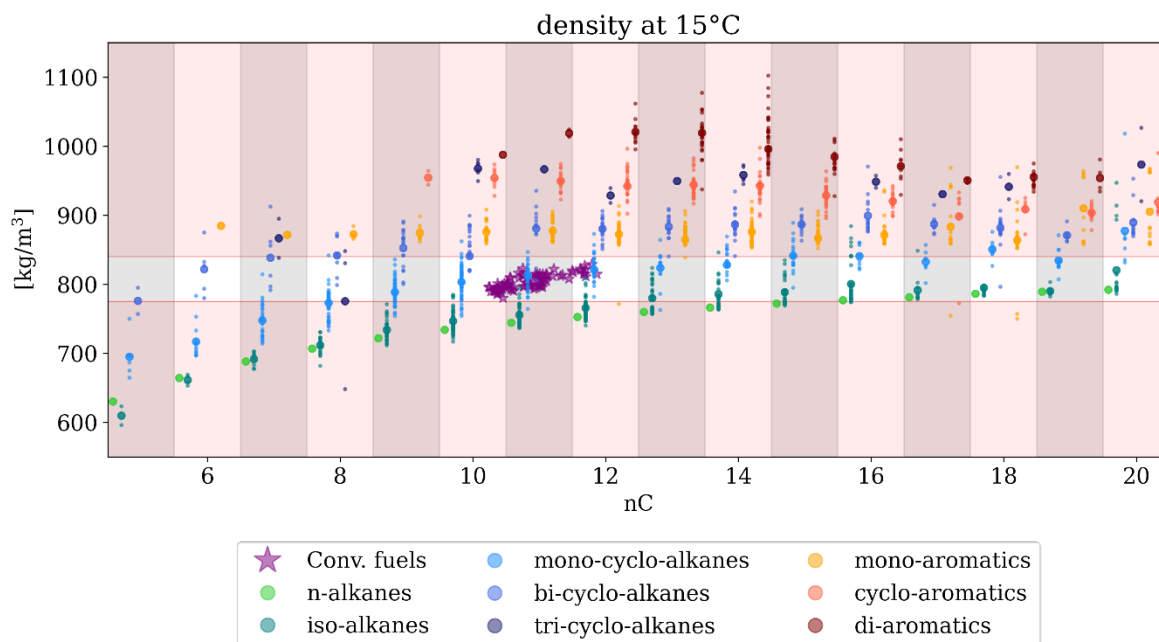


Figure 5.2: Density values at $15\text{ }^{\circ}\text{C}$ of the hydrocarbon families over the carbon number nC .

The mean density values for each nC of a hydrocarbon family are indicated by a large dot in the middle of smaller scatter dots, that display the single values of the respective isomers. As an additional reference, values of conventional jet fuels (Conv. fuels) from the utilized database are displayed as purple stars. The specification limits of the respective properties are shown as red lines. The respective descriptor values for the fuels are calculated from their mean quantitative structure representation, introduced in Section 2.2.3.

Figure 5.2 shows increasing density values over the carbon number nC for all alkanes. For mono-cyclo-aromatics no clear correlation is visible, while the density values of cyclo- and di-aromatics decrease with increasing nC . Saldana et al. [72], and Braun-Unkhoff et al. [111] also observed those trends. The correlations indicate, that increasing the carbon number of the fuel components would increase the density of a fuel for alkanes, and decrease the density for cyclo- and di-aromatics. The density range of n-alkanes and iso-alkanes lies below the value range of conventional reference fuels (Conv. fuels), while the value range for mono-cyclo-alkanes is comparable. The value ranges of the other families exceed the value range of the reference fuels. For the design of fuels, this consequently means, that a high fraction of n- and iso-alkanes with carbon numbers below C 13 can result in fuel densities that lie below the specification range. This is e.g. the case for FT-SPK and HEFA fuels as shown by the density measurements in Figure 4.12. For synthetic fuels with high fractions of cyclo-alkanes on the other hand, densities similar to conventional fuels can be expected.

Next, the influence of the branching and potentially other descriptors on the density values of the components is investigated. The correlations are thereby investigated separately for each family and descriptor in individual plots in Figure 5.3 to Figure 5.5. For reasons of clarity and to limit the number of figures in this chapter, the following figures only show the plots for n- & iso-alkanes, mono-cyclo-alkanes and mono-aromatics. The plots for the other families are listed in Supplementary Material G. The values of the n- & iso-alkanes as well as bi- & tri-cyclo-alkanes are shown together for illustration purposes and similarity of the property values. Plots with the subscript (a) on the left show the property values of the fuel components over nC , plots with subscript (b) in the middle over the branching index η_B and (c) on the right plots over additional descriptors, for which correlations are identified. To additionally illustrate the general trend over the respective first order regression lines are indicated, if more than three measurements are available. In plots with the subscripts (a) and (b), the indicated molecular structures also illustrate the context of the branching structures for selected components.

density at 15°C for n-alkanes & iso-alkanes

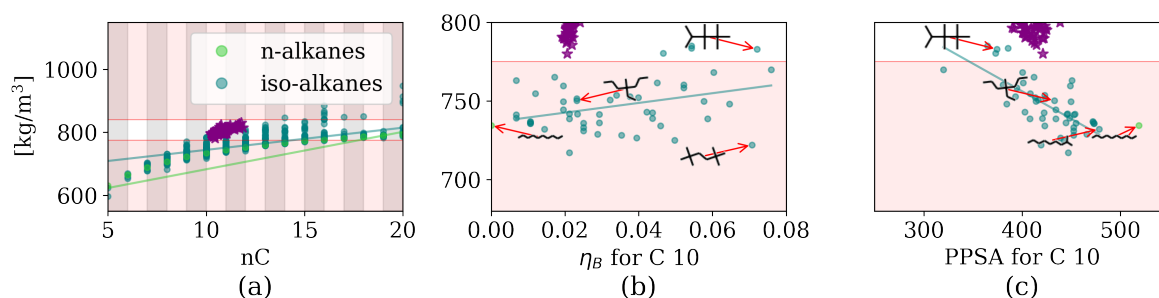


Figure 5.3: Density values at 15 °C of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b) and partial positive surface area PPSA (c).

density at 15°C for mono-cyclo-alkanes

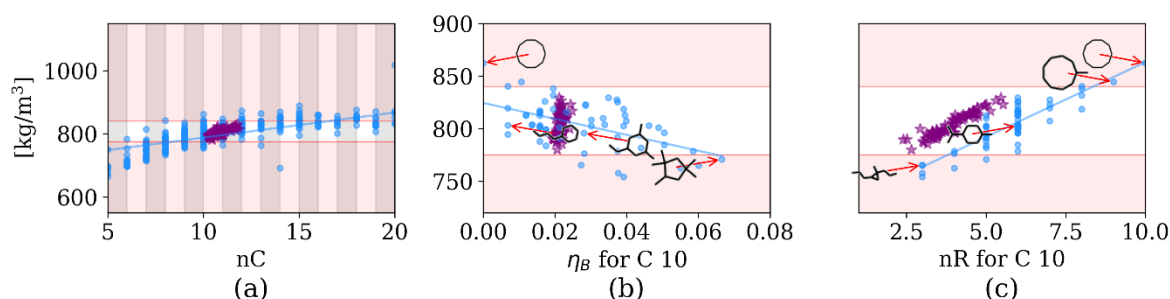


Figure 5.4: Density values at 15 °C of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b) and number of ring atoms nR (c).

density at 15°C for mono-aromatics

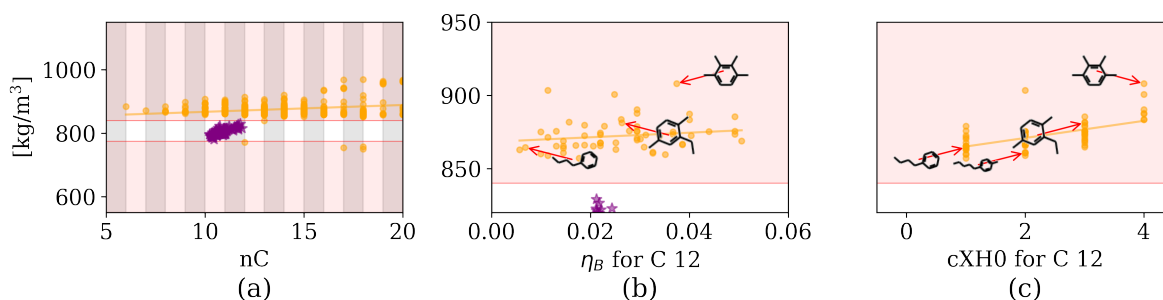


Figure 5.5: Density values at 15 °C of mono-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b) and partial positive surface area PPSA (c).

The plots with the subscript (a) again show the density of the components at different carbon numbers. The indicated trendlines confirm the correlations of the density with nC observed in Figure 5.2. For the branching index (η_B) a correlation with the density is only observed for mono-cyclo-alkanes, with decreasing density values for an increasing branching index in Figure 5.4 (b). As explained in Section 5.2, plots for the branching display the property values of isomers at one

set carbon number, e.g. iso-alkanes with 10 carbon atoms in Figure 5.3 (b). Due to the limited number of available measurements, the investigations are conducted only for isomers with the highest number of available measurements. The identified correlations are assumed to be valid for all other isomers of the family. If measurements of different isomers are available also for other carbon numbers, this assumption is verified by respective investigations. For the density, the observations are verified e.g. for the iso-alkanes by investigating isomers at 9, 11 and 12 carbon atoms.

Apart from the branching index, correlations with other descriptors could be identified. For mono-cyclo-alkanes, bi-cyclo-alkanes and cyclo-aromatics correlations with the number of ring atoms nR are observed, see Figure 5.4 (c), Figure G 1 (c) and Figure G 2 (c). For the iso-alkanes and the di-aromatics, further correlations with the partial positive surface area descriptor ($PPSA$) are visible in Figure 5.3 (c) and Figure G 3 (c). For Mono-aromatics and di-aromatics correlations with the $cXHO$ descriptor are visible in Figure 5.5 (c) and Figure G 3 (c). These findings are in accordance with Saldana et al., who list these descriptors as important features for their QSPR model [72].

To compare the correlations of all families and later all properties simultaneously, as well as to estimate the change of the property value for a change in the descriptor value, the correlations are quantified using the described metrics in Section 5.2. The metrics gives a rough estimate of the impact of the descriptors on the property and indicate how and to what extent the density would be affected when a descriptor is increased or decreased. The calculated metrics with the average differences Δ and correlation coefficients r are summarized in Table 17.

	Δ_{HC} [%]	nC		η_B		Annotation
		r	Δ_{nC} [%]	r	Δ_{η_B} [%]	
n-alkanes	-7.6	8.0	1.3			
iso-alkanes	-5.9	7.0	1.5	4.0*	1.7*	Correlation with partial positive surface area
mono-cyclo-alkanes	0.6	6.0	1.2	-5.0*	2.1*	Correlation with partial positive surface area and number of ring atoms
bi-cyclo-alkanes	8.1	4.0	0.9	-2.0*	1.4*	Correlation with partial positive surface area and number of ring atoms
tri-cyclo-alkanes	14.5	2.0	2.5	\emptyset	\emptyset	
mono-aromatics	8.6	2.0	0.2	2.0*	0.8*	Correlation with aromatic carbon atoms, connected only to other carbon atoms
cyclo-aromatics	17.2	-5.0	0.5	-1.0*	1.1*	Correlation with partial positive surface area and number of ring atoms
di-aromatics	24.4	-4.0	1.1	4.0*	2.4*	Correlation with aromatic carbon atoms, connected only to other carbon atoms

Table 17: Summary influence of different hydrocarbon families and molecular descriptors on the density at 15 °C, * indicates an annotation for the correlation, \emptyset indicates insufficient data

The metric in the first column Δ_{HC} of Table 17 quantifies the average differences of the densities of the hydrocarbon families with respect to the conventional reference fuels. Δ_{HC} thereby allows a comparison of the average densities of the families with respect to each other. The average differences Δ_{HC} state that n-alkanes have densities which are on average 7.7 % lower compared to conventional fuels. Di-aromatics on the other hand, are found to have the highest densities, 24.4 % higher than conventional fuels on average. The order of the Δ_{HC} values corresponds to the order observed visually in Figure 5.2. To additionally illustrate the order of the average property values of the families, the background of the Δ_{HC} values in Table 17 is colored. The color scheme ranges from blue, indicating lower values than the reference fuels, to orange, indicating higher values than the reference fuels. White corresponds to the value range similar as reference fuels.

The next two columns hold metrics for the influence of the carbon number nC and the branching η_B . First the strength of the correlations is indicated by the correlation coefficient r . For nC the r values correspond to the observations made in Figure 5.2, with strong positive correlations for alkanes and negative correlations for aromatics. For η_B only a correlation for the mono-cyclo-alkanes with a value of -5 is identified. Observations and correlations with other descriptors like the nR and the $PPSA$ are indicated by a “*” suffix in Table 17. For cases where the low number of measurements prevents the calculation of metrics, like for the tri-cyclo-alkanes, Table 17 holds a \emptyset symbol.

Apart from the general correlations, the average differences: Δ_{nC} and Δ_{η_B} are calculated, which state the average change of the property value for a change in the descriptor value, likewise to Δ_{HC} . Despite the large correlation coefficients, the density changes only slightly for changes in the carbon number and the branching. The calculated average differences show values below 2.4 % for Δ_{nC} and 0.8 % for Δ_{η_B} , which are significantly lower compared to the Δ_{HC} values. However, significant differences of up to 50 kg/m³ are possible between different isomers, even though small or no correlations are observed, e.g. mono-aromatics see Figure 5.5 (b).

In summary, the density is predominantly affected by the hydrocarbon family. If the density of a fuel should therefore be increased or decreased in the fuel design, the strongest effects are achieved by changing the fraction of the families accordingly. Changes in the density by increasing or decreasing the average carbon number and therefore the size of the components are only relevant for smaller components with nC below 10 to 12, see Figure 5.2 to Figure 5.5 subscript (a). For larger components with a nC above 13, the average values differ only slightly. The branching was found to have no systematic impact on the density, however the consideration of property differences of different isomers can be relevant for low component fuels with densities close to the specification limits.

5.3.2 Surface tension

The surface tension influences the atomization of a fuel, together with the density and viscosity, it is a fundamentally physicochemical property relevant for the ignition and reignition of fuels. Surface tensions that are too high can consequently lead to problems in atomization and ignition.

As for the density, first the influence of the carbon number and the hydrocarbon family on the surface tension are investigated. The order of the hydrocarbon families observed in Figure 5.6 is thereby similar to the order observed for the density in the previous Section 5.3.1. Iso- and n-alkanes show the lowest surface tension, followed by mono-cyclo-alkanes and mono-aromatics, as well as bi-cyclo-alkanes and cyclo-aromatics. For tri-cyclo-aromatics and di-aromatics the highest surface tensions are observed. This order is observed for a range of 10 to 13 carbon atoms. For higher and lower carbon numbers, changes in the relative order can be observed due to the limited availability of measurement data. For e.g. nC 16 the average surface tension of cyclo-aromatics is higher than the one of the di-aromatics.

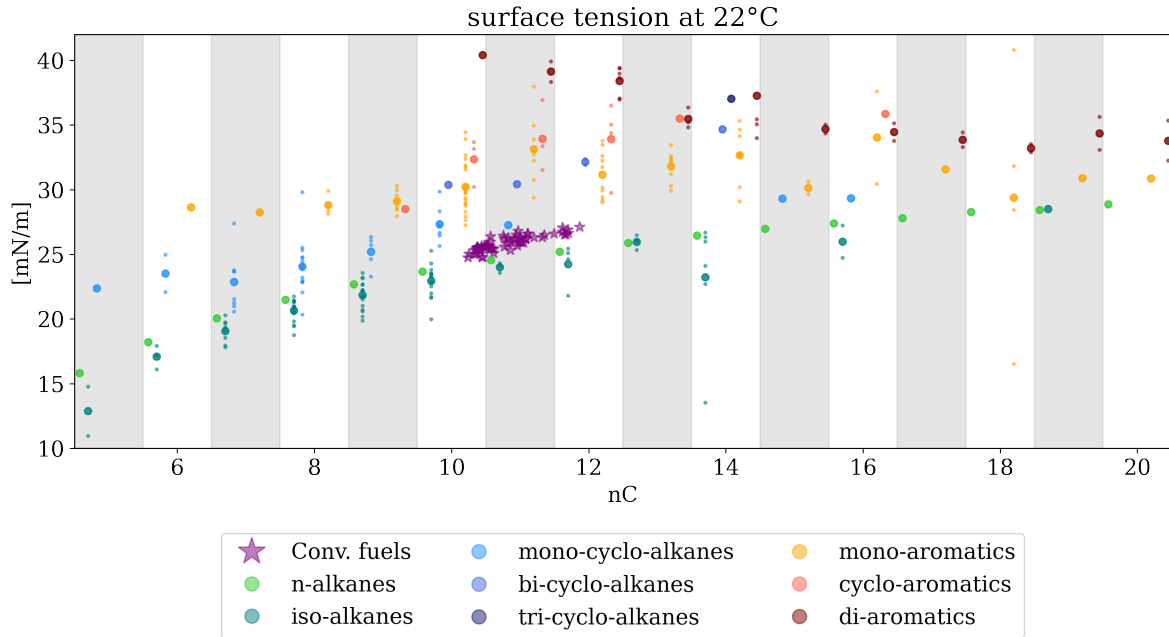


Figure 5.6: Surface tension values at 22 °C of the hydrocarbon families over the carbon number nC .

The relative order of the average surface tensions of the different hydrocarbon families is reflected in Table 18 by the average differences Δ_{HC} and the corresponding background colors. Similar to the density, the value range of n- and iso-alkanes lies below the one of the reference fuels with a relative difference Δ_{HC} of -5.6 % and -10.8 % respectively. The average surface tension of mono-cyclo-alkanes is comparable with the values of the reference fuels, while the property values of the other families exceed the range, e.g. 43.4 % for di-aromatics. Over the carbon number, similar correlations of the average surface tensions of the families are visible as for the density in Figure 5.6. Increasing surface tensions with nC are visible for all alkanes. Mono-cyclo-aromatics show

no clear correlation, while the surface tension of cyclo- and di-aromatics decreases with increasing nC .

To lower the surface tension, e.g. to improve the atomization of the fuel, the fraction of n- and iso-alkanes with low carbon numbers should be increased. Increased fractions of mono-cyclo-alkanes would lead to a surface tension similar to conventional fuels, while increased fractions of other families would increase the surface tension of a fuel on average. A change in the average carbon number would only be relevant for the surface tension of components with smaller carbon numbers, below 10.

The correlations of the property values with nC is reflected in Table 18, with high positive r values of 8 and 10 for the alkanes, negative values of -7 for di-aromatics. For η_B a negative correlation is observed for iso- and mono-cyclo-alkanes in Figure 5.7 (b) and Figure 5.8 (b). For the mono-aromatics, the positioning and number of branches has a notable effect on the surface tension see Figure 5.9 (b). While the surface tension decreases with increased branching for components with one side chain, it increases for components with multiple side chains. The trend over the $cXHO$ descriptor is able to resolve this relationship further and illustrates a correlation of the surface tension of mono-aromatics with the number of carbon atoms connected only to other carbon atoms in Figure 5.9 (c). The other families do not allow an investigation due to the limit number of available measurements. Compared to the density, the average value differences nC and Δ_{η_B} are higher. For a change in nC , Δ_{nC} values between 2.7 % to 4 % are calculated. For changes in the branching, the differences Δ_{η_B} are between 4.0 and 5.2 %.

Similar to the density, the surface tension is mostly affected by the hydrocarbon family. The surface tension of fuels can be lowered by increasing the fraction of n- and iso-alkanes and increased by increasing the fraction of mono-, bi- and tri-cyclo-alkanes as well as aromatics. Size and the branching of the components have a slightly stronger influence on the surface compared to the density, but the effects are still of minor significance.

surface tension at 22°C for n-alkanes & iso-alkanes

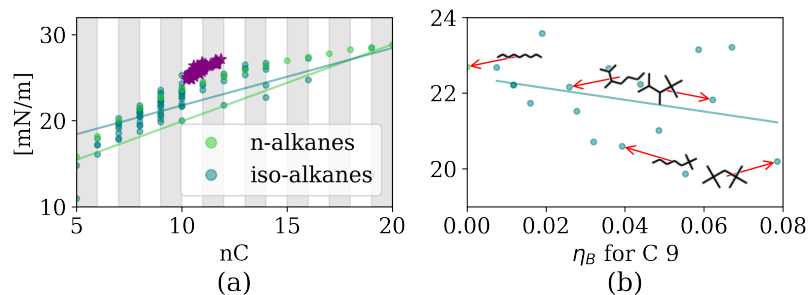


Figure 5.7: Surface tension values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

surface tension at 22°C for mono-cyclo-alkanes

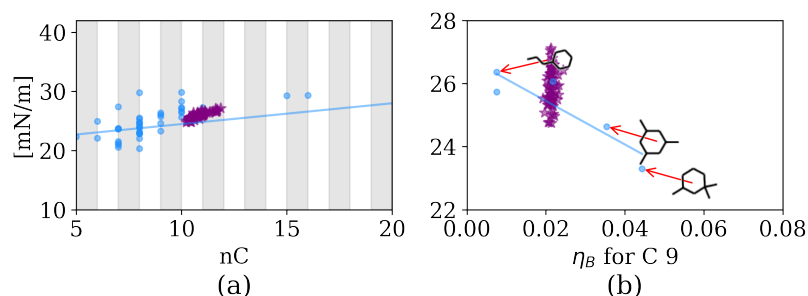


Figure 5.8: Surface tension values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

surface tension at 22°C for mono-aromatics

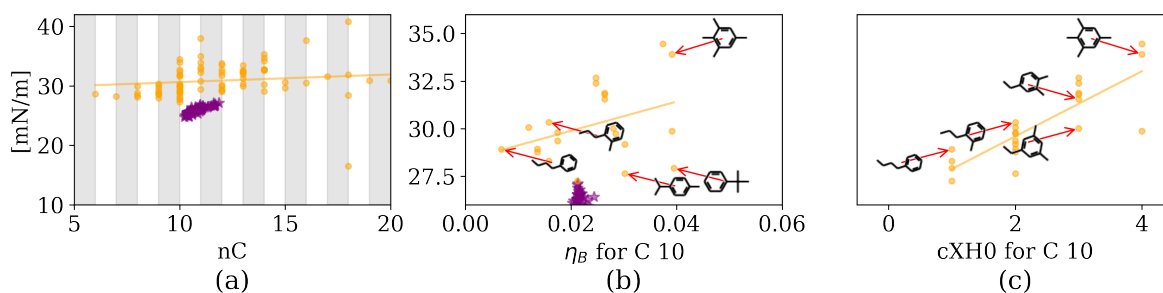


Figure 5.9: Surface tension values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

	Δ_{HC} [%]	nC		η_B		Annotation
		r	Δ_{nC} [%]	r	Δ_{η_B} [%]	
n-alkanes	-5.6	8.0	3.3			
iso-alkanes	-10.8	6.0	4.0	-3.0	4.0	
mono-cyclo-alkanes	2.3	4.0	3.3	-9.0	4.0	
bi-cyclo-alkanes	23.2	10.0	3.3	∅	∅	
tri-cyclo-alkanes	43.1	∅	∅	∅	∅	
mono-aromatics	19.5	2.0	2.7	4.0*	5.2*	Correlation with aromatic carbon atoms, connected only to other carbon atoms
cyclo-aromatics	28.8	∅	∅	∅	∅	
di-aromatics	43.4	-7.0	3.2	∅	∅	

Table 18: Summary influence of different hydrocarbon families and molecular descriptors on the surface tension at 22 °C, * indicates an annotation for the correlation, ∅ indicates insufficient data.

5.3.3 Kinematic viscosity

The kinematic viscosity is another fundamental property that is of great importance for the operability of pumps, nozzles and the atomization process, which are critical especially at low temperatures. Similar to the surface tension too high viscosity values can lead to problems in atomization and ignition.

The comparison of the different hydrocarbon families in Figure 5.10 shows an exponential increase of the viscosity, with increasing values for increasing nC values for all families. While the values of all components are very similar at small nC up to 8-10, the differences increase rapidly at an nC of 10-12. This exponential increase was also observed by Saldana et al. for n- and iso-alkanes at 15 °C [72] and by Wang et al. for alkanes and aromatics in general [107]. The observations are reflected in the correlation coefficients r for nC in Table 19 with strong positive values of up to 9 for the different families and average differences Δ_{nC} over 30 %. Apart from the exponential increase, strong scattering of the viscosity values is visible for nC values above 9 meaning, that strong differences in the property values of the different isomers exist at higher carbon numbers. The scattering also explains the comparably low correlation coefficients r of 5 and 6 for the bi- and tri-cyclo-alkanes as well as the cyclo-aromatics for nC , even though a clear correlation with the carbon number is visible in the plots.

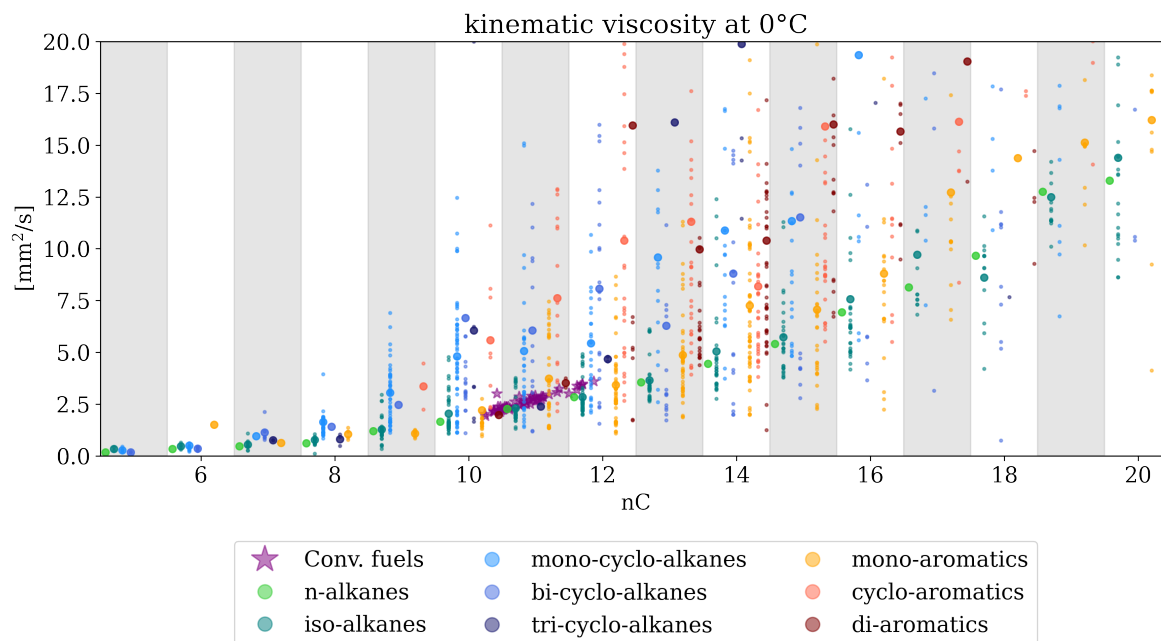


Figure 5.10: Kinematic viscosity values at 0 °C of the hydrocarbon families over the carbon number nC .

	Δ_{HC} [%]	nC		η_B		Annotation
		r	Δ_{nC} [%]	r	Δ_{η_B} [%]	
n-alkanes	-18.6	9.0	39.7			
iso-alkanes	-11.8	8.0	37.1	5.0*	36.0*	Correlation with partial positive surface area
mono-cyclo-alkanes	97.6	7.0	42.3	-4.0*	45.9*	Correlation with partial positive surface area
bi-cyclo-alkanes	95.0	5.0	43.5	4.0*	50.4*	Correlation with partial positive surface area
tri-cyclo-alkanes	176.2	5.0	104.9	∅	∅	
mono-aromatics	15.8	7.0	41.1	3.0	39.1	
cyclo-aromatics	195.3	6.0	30.9	-4.0	46.3	
di-aromatics	219.1	7.0	54.6	-3.0	34.0	

Table 19: Summary influence of different hydrocarbon families and molecular descriptors on the kinematic viscosity at 0 °C, * indicates an annotation for the correlation, ∅ indicates insufficient data.

The strong scattering of the viscosity values makes a visual identification of the relative order of the different property values difficult. The order is therefore inferred based on the Δ_{HC} values in Table 19. N- and iso-alkanes are ranked as the least viscous, followed by mono-aromatics, mono-bi-, and tri-cyclo-alkanes and cyclo- and di-aromatics. Wang et al. confirm the observed order

and list n- and iso-alkanes as the least viscous, followed by cyclo-alkanes aromatics, without differentiating between mono-, cyclo- and di-aromatics [107]. High viscosities can therefore be expected for fuels with a high average carbon number or high fraction of bi-cyclo-alkanes, cyclo- and di-aromatics.

As indicated by the scattering in Figure 5.10, the viscosity values of different isomers of the hydrocarbon families differ significantly, compared to the previous temperature dependent properties. The individual plots in Figure 5.11 to Figure 5.13 subscript (a) illustrate this more clearly. The differences in the viscosity values are reflected by high Δ_{η_B} values in Table 19 with values above 34 %. The different viscosities can partially be explained by the branching with slight positive correlations visible for iso-alkanes, bi-cyclo-alkanes and mono-aromatics in Figure 5.11, Figure G 7 and Figure 5.13 subscript (b). However, stronger correlations are visible with the PPSA for iso-alkanes, mono-, bi- and tri-cyclo-alkanes, see Figure 5.11, Figure 5.12 and Figure G 7 subscript (c). For the other hydrocarbon families slight trends are also visible. Similar to the density, the correlations of the kinematic viscosity with the *PPSA* indicate a strong influence of Van-der-Waals interactions for these families. Saldana et al. also mention descriptors related to the Van-der-Waals interactions as the most influential features for their QSPR viscosity model [72]. For the mono-aromatics the viscosity values are furthermore strongly affected by the number and position of side chains, similar to the surface tension. The indicated components in Figure 5.13 (b) show lower viscosities for components with side chains in ortho¹ and higher viscosities for components in meta² position. The influence of the branching for aromatic components and the importance of the position of the branches was also described by Wang et al. [107] and Cai et al. [119] using their QSPR model.

In summary, the kinematic viscosity of fuel components is affected by all aspects: the hydrocarbon family, the carbon number and the topology of the individual isomers. For jet fuels, the viscosity can therefore be influenced by increasing the fraction of the respective hydrocarbon families, the average size of the components, as well as the content of specific isomers. The viscosity of fuels can e.g. be decreased by increasing the fraction of n- and iso-alkanes and lowering the average carbon number of the components, or by increasing the fraction of alkanes with a high PPSA. The influence of the carbon number should be especially considered for components with a carbon number above 10, where rapid increases in the viscosity values can be expected.

¹ opposing branches in an aromatic ring structure

² opposing branches in 1,3 position in an aromatic ring structure

kinematic viscosity at 0°C for n-alkanes & iso-alkanes

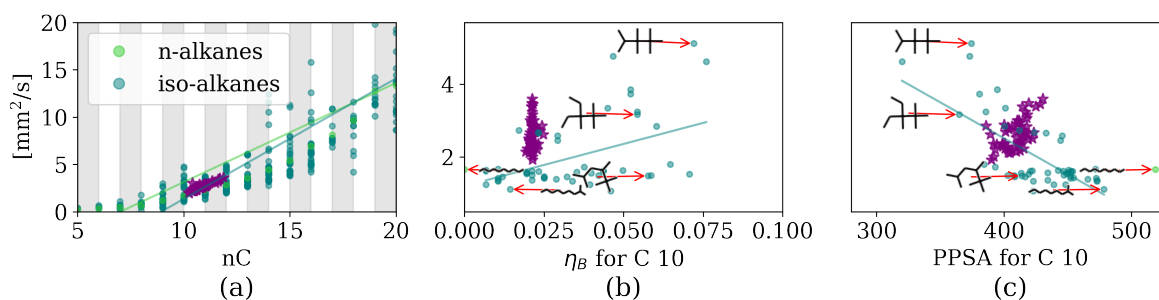


Figure 5.11: Kinematic viscosity values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b) and partial positive surface area PPSA (c).

kinematic viscosity at 0°C for mono-cyclo-alkanes

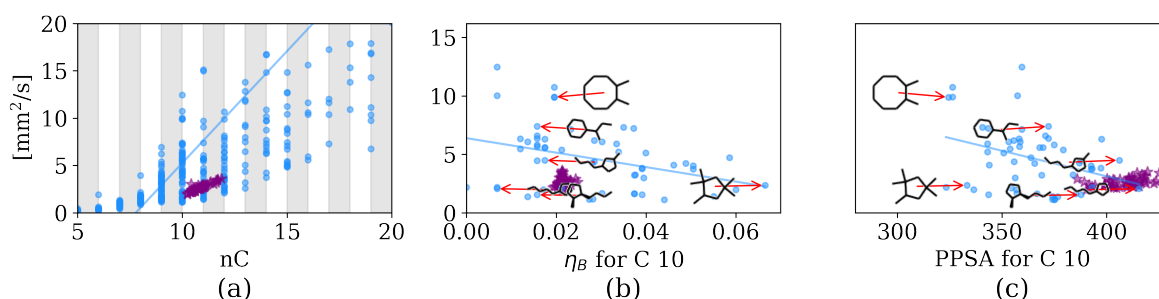


Figure 5.12: Kinematic viscosity values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b) and partial positive surface area PPSA (c).

kinematic viscosity at 0°C for mono-aromatics

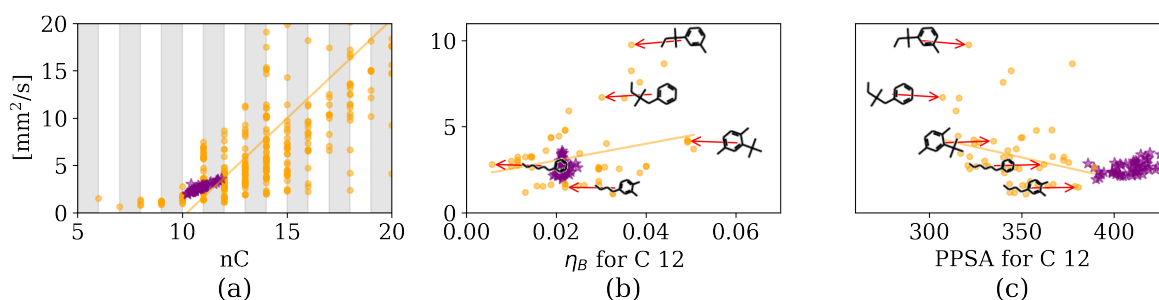


Figure 5.13: Kinematic viscosity values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b) and partial positive surface area PPSA (c).

5.3.4 Net heat of combustion

The energy released by the burning of the fuels is measured by the net heat of combustion, it therefore is the key indicator for the performance of the fuel.

Figure 5.14 shows the values of the different hydrocarbon families over nC . Comparing the average property values of the hydrocarbon families, the n- and iso-alkanes show the highest values, followed by mono-cyclo-alkanes, bi- and tri-cyclo-alkanes and the aromatics, with di-aromatics having the lowest net heat of combustion. High net heat of combustion can therefore be expected for fuels with a high fraction of n- and iso-alkanes. The order is again reflected by the Δ_{HC} values in Table 20 and indicated by the color scheme. Figure 5.14 furthermore shows, that the net heat of combustion values of n-, iso- and mono-cyclo-alkanes decrease over nC , while for bi- and tri-cyclo-alkanes no clear correlation is visible. The aromatic families show the reversed trend, with values that increase with an increasing nC .

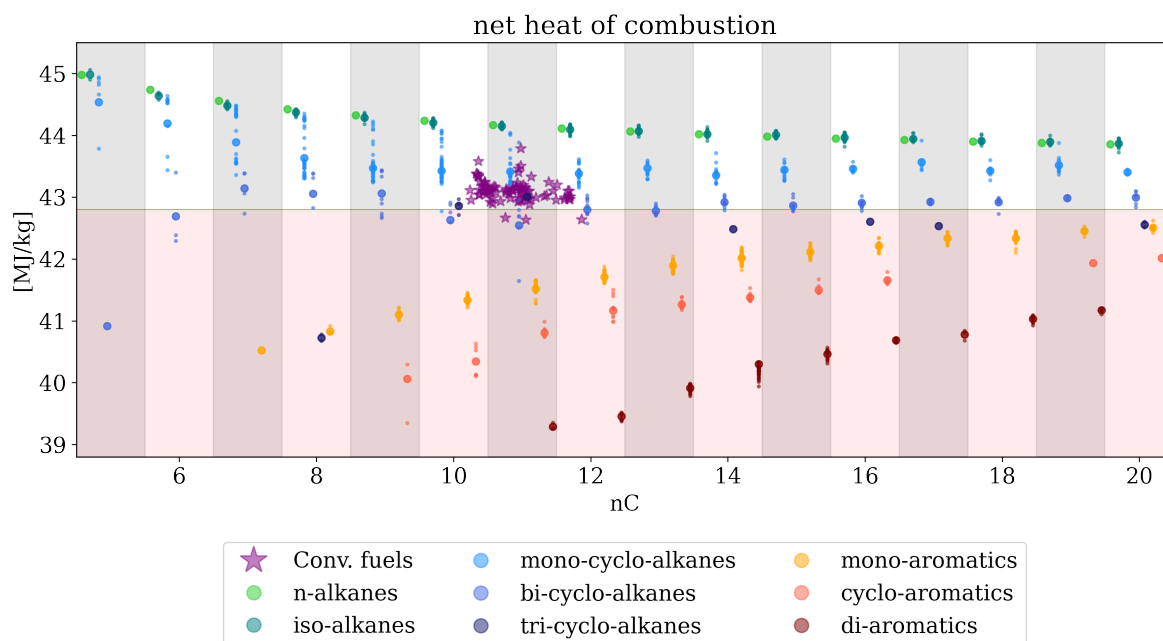


Figure 5.14: Net heat of combustion values of the hydrocarbon families over the carbon number nC .

Likewise to the density and the surface tension, the differences in the property values are most significant for components with a lower nC . For higher nC , above 10 to 12, the net heat of combustion for most families changes only slightly, especially for alkanes. The calculated correlation coefficients r in Table 20 reflect those observations, with negative values for n-, iso- and mono-cyclo-alkanes and positive ones for the aromatics. Wang et al. also observed the described correlations for n- and iso-alkanes as well as for mono-cyclo-alkanes and mono-aromatics [107]. For the branching, slight negative correlations are visible for the iso- and bi-cyclo-alkanes as well as mono-aromatics and di-aromatics in the Figure 5.15 (b), Figure G 7 (b), Figure 5.17 (b), and Figure G 9 (b). For mono-cyclo-alkanes the number of ring atoms nR greatly

affects the property value, as visible in Figure 5.16 (b). Components with less than five and more than six ring atoms show significantly higher net heat of combustion values. For mono- and di-aromatics also positive correlations with the $cXH0$ descriptor are visible in Figure 5.17 (c) and Figure G 9 (c). The net heat of combustion of the cyclo-aromatics does not show a clear correlation. For the tri-cyclo-alkanes, the low number of measurements again prevents an investigation.

Comparing the average differences in Table 20, the greatest differences are observed between the hydrocarbon families, with values of up to -7.1 % for Δ_{HC} . For the carbon number and the branching, the values below 0.7 % for Δ_{nC} and 0.4 % Δ_{η_B} show that those descriptors are of minor significance for the net heat of combustion values.

	Δ_{HC} [%]	nC		η_B		Annotation
		r	Δ_{nC} [%]	r	Δ_{η_B} [%]	
n-alkanes	2.5	-8.0	0.1			
iso-alkanes	2.5	-9.0	0.1	-4.0	0.1	
mono-cyclo-alkanes	1.0	-4.0	0.1	1.0*	0.4*	Dependence on number of ring atoms
bi-cyclo-alkanes	-0.4	2.0	0.2	-5.0	0.1	
tri-cyclo-alkanes	-1.8	5.0	0.7	∅	∅	
mono-aromatics	-3.7	9.0	0.4	-6.0*	0.1*	Correlation with aromatic carbon atoms, connected only to other carbon atoms
cyclo-aromatics	-4.9	9.0	0.6	-2.0	0.1	
di-aromatics	-7.2	10.0	0.7	-6.0*	0.2*	Correlation with aromatic carbon atoms, connected only to other carbon atoms

Table 20: Summary influence of different hydrocarbon families and molecular descriptors on the net heat of combustion, * indicates an annotation for the correlation, ∅ indicates insufficient data.

The net heat of combustion is therefore primarily affected by the hydrocarbon family of the fuel components. The carbon number is only relevant for smaller components, while the branching is found to be generally of minor significance. Thus, the net heat of combustion of fuels can mainly be adjusted by changing the respective fractions of the hydrocarbon families. An increase of n-alkanes and iso-alkanes would therefore generally increase the net heat of combustion of a fuel, while an increase of aromatic components would lower it.

net heat of combustion for n-alkanes & iso-alkanes

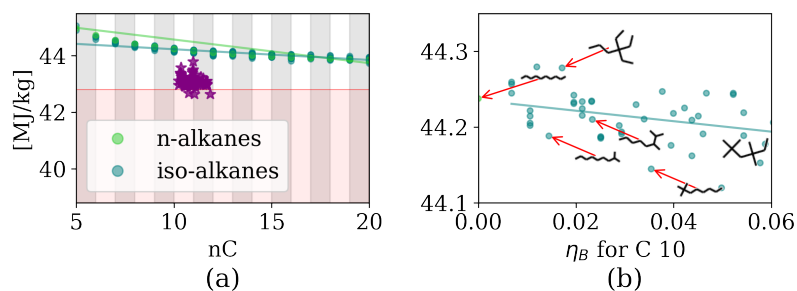


Figure 5.15: Net heat of combustion values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

net heat of combustion for mono-cyclo-alkanes

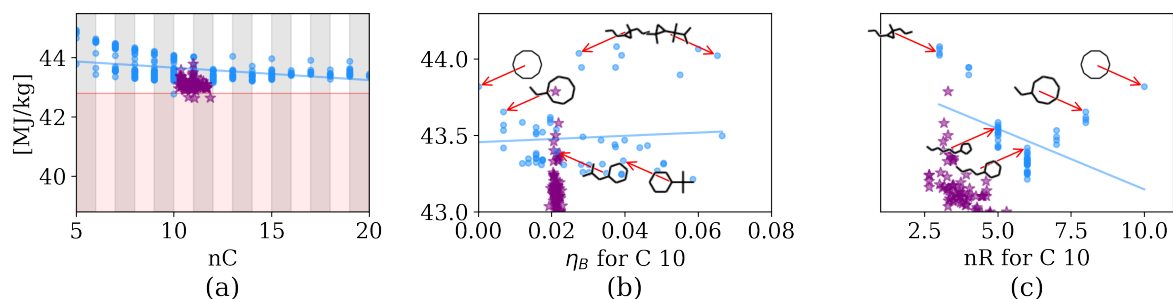


Figure 5.16: Net heat of combustion values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b) and number of ring atoms nR (c).

net heat of combustion for mono-aromatics

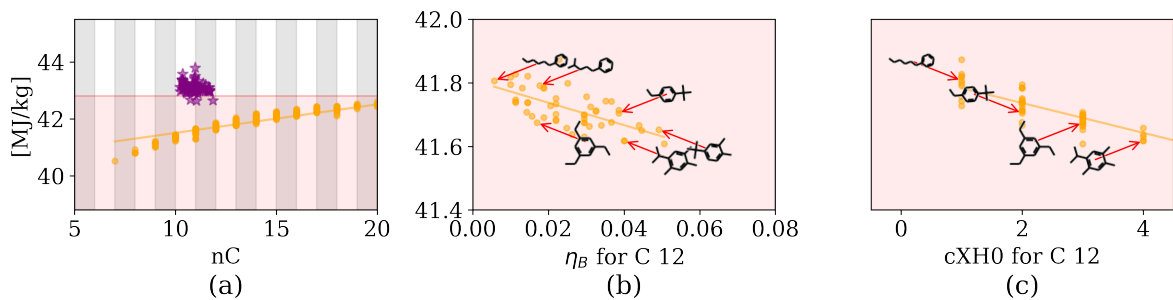


Figure 5.17: Net heat of combustion values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

5.3.5 Flash point

The flash point characterizes the flammability of a fuel and is a key property for the fire safety and handling of a fuel. Low flash points could lead to an ignition of the fuel at low temperatures, the ASTM D4054 therefore sets a minimum at 38 °C, meaning that the fuel is not allowed to ignite an external flame at temperatures below this limit.

Compared to other properties like the density and net heat of combustion, the flash point values of all families in Figure 5.18 lie significantly closer together, for carbon numbers up to 14-16. For higher carbon numbers, the spread between the different flash points of the families increases. All families furthermore show a strong positive correlation with the carbon number, with increasing flash points as nC increases.

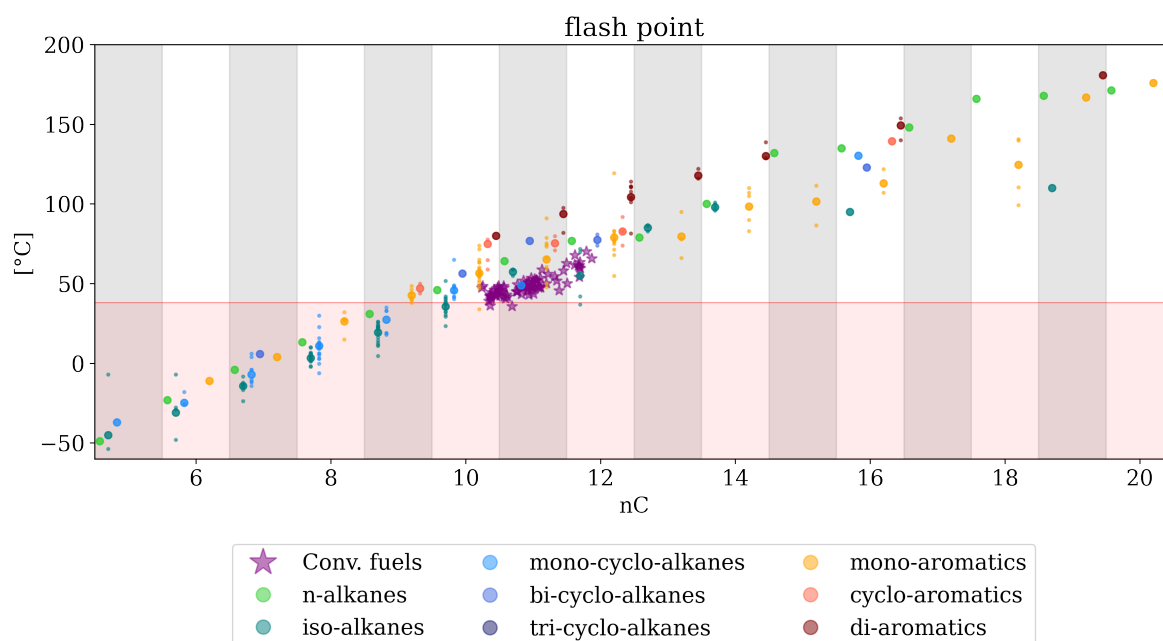


Figure 5.18: Flash point values of the hydrocarbon families over the carbon number nC .

These observations are reflected in the metrics of Table 21. The Δ_{HC} values of the different families are similar, with the exception of the cyclo- and di-aromatics. For the carbon number, the correlation coefficients r in Table 21 indicate strong positive correlations with values of 8 to 10. The comparison of the Δ_{HC} values of hydrocarbon families ranks mono-cyclo-alkanes and iso-alkanes as components with the lowest flash points, followed by n-alkanes, mono-aromatics, bi- and tri-cyclo-alkanes and cyclo-aromatics. Di-aromatics show the highest flash point values on average. However, the low ranking of the mono-cyclo-alkanes is due to missing measurement data at carbon numbers between 11 to 16, see left plot in Figure 5.20 (a). In reality, the average flash point for the mono-cyclo-alkanes is therefore expected to be higher. Wang et al. confirm this presumption and list the average flash point of cyclo-alkanes higher than the ones of n- and iso-

alkanes [107]. The distortion of the order is indicated by an additional “*” annotation for Δ_{HC} of mono-cyclo-alkanes in Table 21.

With respect to the branching, decreasing flash point values are visible over the branching index for all hydrocarbon families in Figure 5.19 to Figure 5.21 subscript (b).

The comparison of the average differences for the carbon number Δ_{nC} and the branching Δ_{η_B} in Table 21 shows high values for both aspects. Both the carbon number and the branching therefore strongly influence the flash point, with values up to 4.8 % for Δ_{nC} and up to 14 % for Δ_{η_B} . In many cases, the difference of the branching is higher than the influence of the carbon number. This is particularly visible for the iso-alkanes in Figure 5.19 (b), where depending on the isomer, the ASTM specification limit is either met or not. This is the reason for the large predicted uncertainty intervals for the flash point predictions of synthetic fuels in Section 4.2.

	Δ_{HC} [%]	nC		η_B		Annotation
		r	Δ_{nC} [%]	r	Δ_{η_B} [%]	
n-alkanes	5.7	10.0	4.6			
iso-alkanes	-0.2	10.0	4.6	-8.0	14.0	
mono-cyclo-alkanes	-1.9*	10.0	4.8	-5.0	9.7	Low ranking due to missing data in carbon range 11 to 14
bi-cyclo-alkanes	5.9	10.0	3.8	∅	∅	
tri-cyclo-alkanes	5.9	∅	∅	∅	∅	
mono-aromatics	5.5	10.0	3.5	-10.0	5.1	
cyclo-aromatics	10.4	10.0	3.5	-9.0	10.5	
di-aromatics	19.7	10.0	3.2	8.0	2.3	

Table 21: Summary influence of different hydrocarbon families and molecular descriptors on the flash point, * indicates an annotation for the correlation, ∅ indicates insufficient data.

In conclusion, the branching of the isomers and the carbon number have the greatest effect on the flash point value, while most of the average values of the different families are quite close. For the design of jet fuels, the branching of components is therefore an important aspect to consider. It can determine whether a fuel's flash point is within or outside specification, especially

for fuels with low constituents such as ATJ-SPK. Low flash points of fuels can be increased by increasing the average carbon number of the fuel components and reducing their branching.

flash point for n-alkanes & iso-alkanes

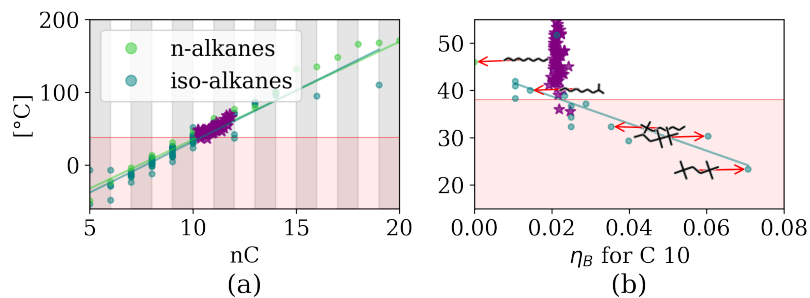


Figure 5.19: Flash point values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

flash point for mono-cyclo-alkanes

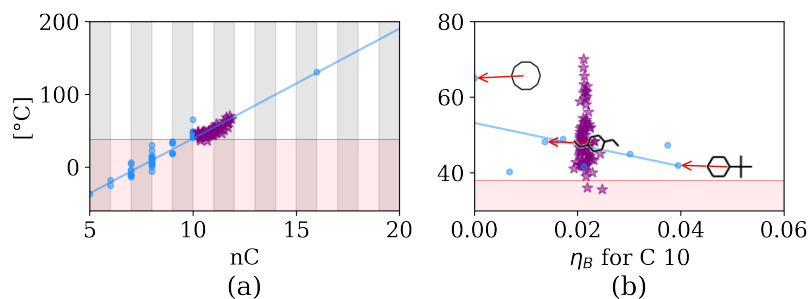


Figure 5.20: Flash point values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

flash point for mono-aromatics

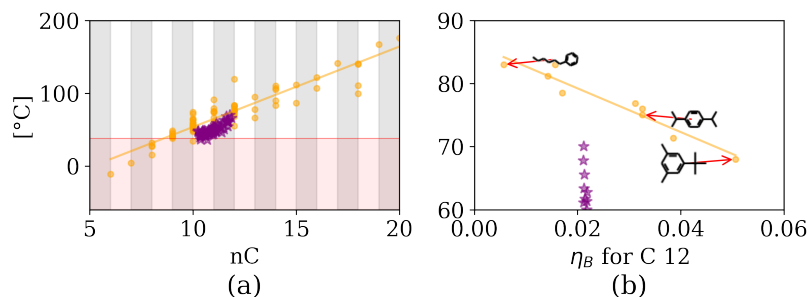


Figure 5.21: Flash point values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

5.3.6 Freezing point

The freezing point characterizes the low temperature behavior of a fuel and states the temperature at which fuel starts to solidify. Together with the viscosity, it is a key property that guarantees the safe operation of the fuel at low temperatures.

Figure 5.22 shows a positive correlation of the freezing point values of the n-, iso- and mono-cyclo-alkanes as well as mono-aromatics with an increasing nC . This is visible in Figure 5.22 and even clearer in the individual plots in Figure 5.23 to Figure 5.25 subscript (a). The values of the bi-cyclo-alkanes and di-aromatics on the other hand, show no clear correlation, while for the tri-cyclo-alkanes and cyclo-aromatics the low number of available measurements prevent an investigation.

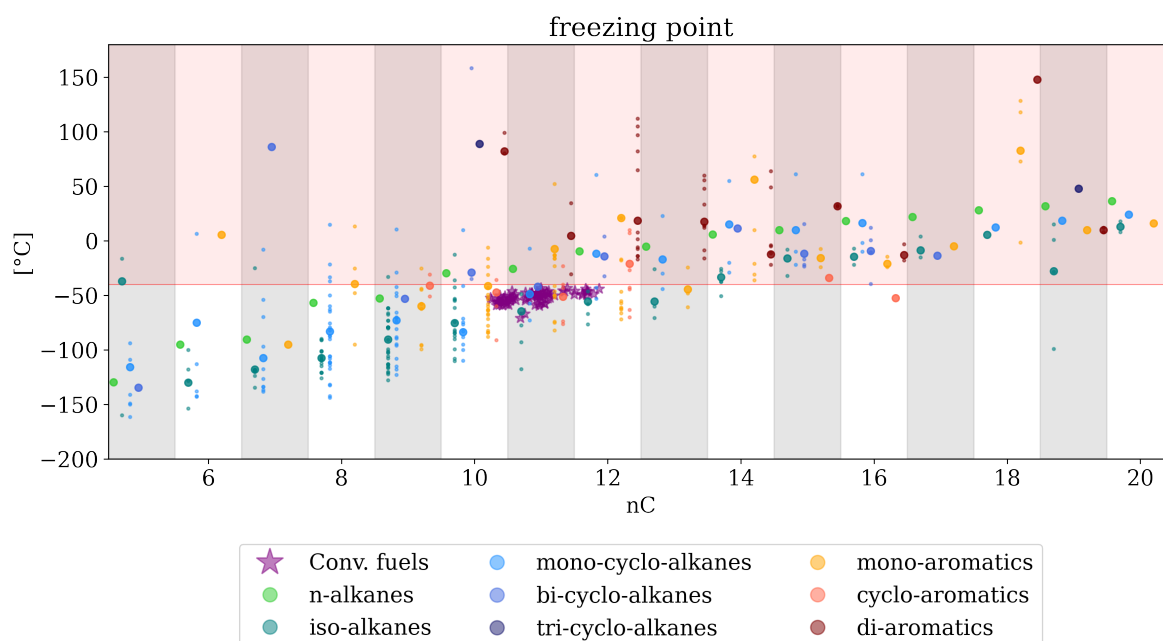


Figure 5.22: Freezing point values of the hydrocarbon families over the carbon number nC .

The observations are reflected by the correlation coefficients r for nC in Table 22. The positive correlation of the freezing point for alkanes and mono-aromatics was also stated by Wang et al. [107] and Cai and Zhang [135]. Using their QSPR model, Cai and Zhang furthermore predicted values for di-aromatics with varying numbers of carbon atoms, for which also no correlation was visible. This again coincides with the made observation. The comparison of the Δ_{HC} values in Table 22 shows the lowest relative differences for iso-alkanes, followed by cyclo-aromatics, mono-aromatics and mono-cyclo-alkanes. Higher freezing points are observed for n-alkanes and bi-cyclo-alkanes, di-aromatics and tri-cyclo-alkanes.

	Δ_{HC} [%]	nC		η_B		Annotation
		r	Δ_{nC} [%]	r	Δ_{η_B} [%]	
n-alkanes	12.8	9.0	4.7			
iso-alkanes	-5.2	7.0	5.4	6.0*	24.8*	No consistent trend over different carbon atoms, significant influence of different isomers
mono-cyclo-alkanes	6.2	8.0	6.3	4.0*	17.6*	No consistent trend over different carbon atoms, significant influence of different isomers
bi-cyclo-alkanes	25.2	1.0	8.0	∅	∅	
tri-cyclo-alkanes	63.8	∅	∅	∅	∅	
mono-aromatics	5.8	5.0	4.9	7.0	48.9	
cyclo-aromatics	2.9	∅	∅	∅	∅	
di-aromatics	35.3	-1.0	5.9	6.0	96.6	

Table 22: Summary influence of different hydrocarbon families and molecular descriptors on the freezing point, * indicates an annotation for the correlation, ∅ indicates insufficient data.

For the branching, the individual plots in Figure 5.23 to Figure 5.25 subscript (b) show great differences in the freezing point values for the different isomers. For iso-alkanes with 10 carbon atoms in Figure 5.23 (b) the values range from -110 °C to -36 °C, even crossing the ASTM specification limit of Jet A at -40 °C. The differences in the freezing point correlate in part with the branching index. Positive correlations over the branching index are visible for iso-alkanes, mono- and di-aromatics, as visible in the individual plots in Figure 5.23 (b), Figure 5.25 (b) and Figure G 18 (b), respectively. For mono-cyclo-alkanes no clear correlation is observed, while the low number of measurements for the other families prevents an investigation. Striking is the difference of property values between unbranched components e.g. n-decane and cyclohexane and branched components with a low branching index. To illustrate these striking differences in the freezing points, the unbranched components are indicated in the corresponding plots in Figure 5.23 to Figure 5.25 subscript (b), e.g. n-decane in Figure 5.23 (b) and cyclohexane in Figure 5.24. For all considered cases, the low-branched components have significantly lower freezing points compared to the unbranched components. Differences of -30 °C to -110 are visible for n-decane and low -branched iso-alkanes in Figure 5.23 (b) and differences of 6.45 °C to -110 °C for cyclohexane and low-branched mono-cyclo-alkanes in Figure 5.24 (b). This is even more remarkable since a positive correlation of the mono-cyclo-alkanes was observed with the carbon number and cyclohexane contains only 6 carbon atoms, 4 carbon atoms less than the investigated mono-cyclo-alkanes with 10 carbon atoms. The average difference for the branching Δ_{η_B} reflects the significant differences in the property values, with values ranging from 24.8 % to 96.6 % in Table 22. Depending on the hydrocarbon family, the isomers have the strongest influence on the freezing point value, followed by the influence of the different hydrocarbon families and the

number of carbon atoms. Similar to the flash point, the large variance in the freezing points of different isomers explains the large uncertainties in the predicted freezing points of fuels like the ATJ-SPK in Section 4.2. For the fuel design, the consideration of the branching of the components is therefore of utmost importance for the freezing point. Low freezing points can be achieved by increasing the fraction of low-branched alkanes, mono-cyclo-alkanes and mono-aromatics.

freezing point for n-alkanes & iso-alkanes

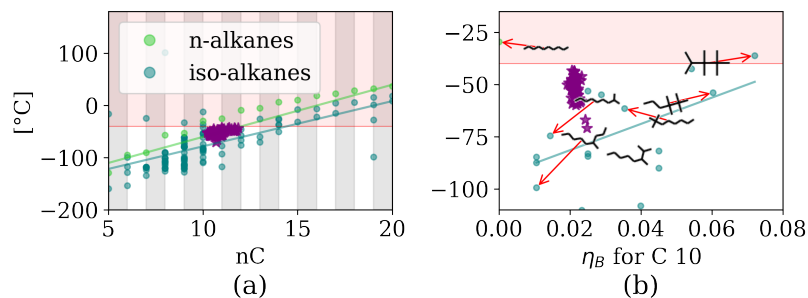


Figure 5.23: Freezing point values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

freezing point for mono-cyclo-alkanes

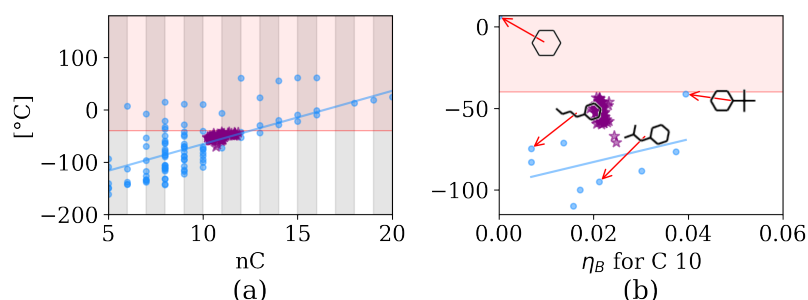


Figure 5.24: Freezing point values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

freezing point for mono-aromatics

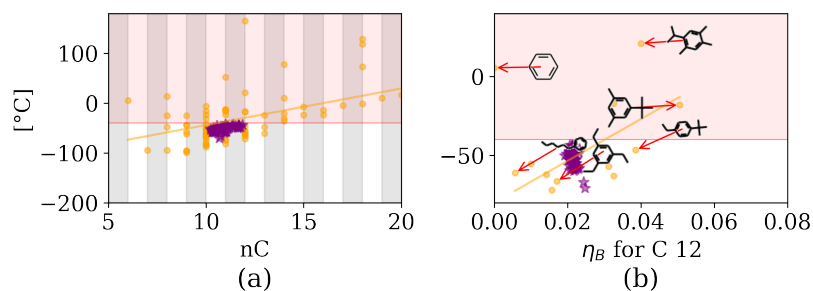


Figure 5.25: Freezing point values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

5.3.7 Cetane number

The “ignition quality” or autoignition behavior of a fuel is characterized by the cetane number. The property quantifies the ignition delay of the fuel relative to cetane (hexadecane n-alkane with 16 carbon atoms), a compound with good ignition quality with a set cetane number of 100. For jet fuel, the typical property range of cetane is between 40 to 50, with lower limit in the prescreening of 30. In the NJFCP program, it was identified as an important parameter to ensure similar ignition behavior of a new SAF compared to approved fuels and with it the safe operation in existing combustion systems.

For the cetane number in Figure 5.26, n-alkanes show significantly higher values than all other families. Iso-alkanes show the second highest average cetane numbers, followed by the mono-, bi- and tri-cyclo-alkanes as well as the aromatics. This order can be explained by the structural similarity of the components to hexadecane. N-alkanes and iso-alkanes have higher structural similarity to hexadecane and therefore similar ignition behavior and cetane numbers.

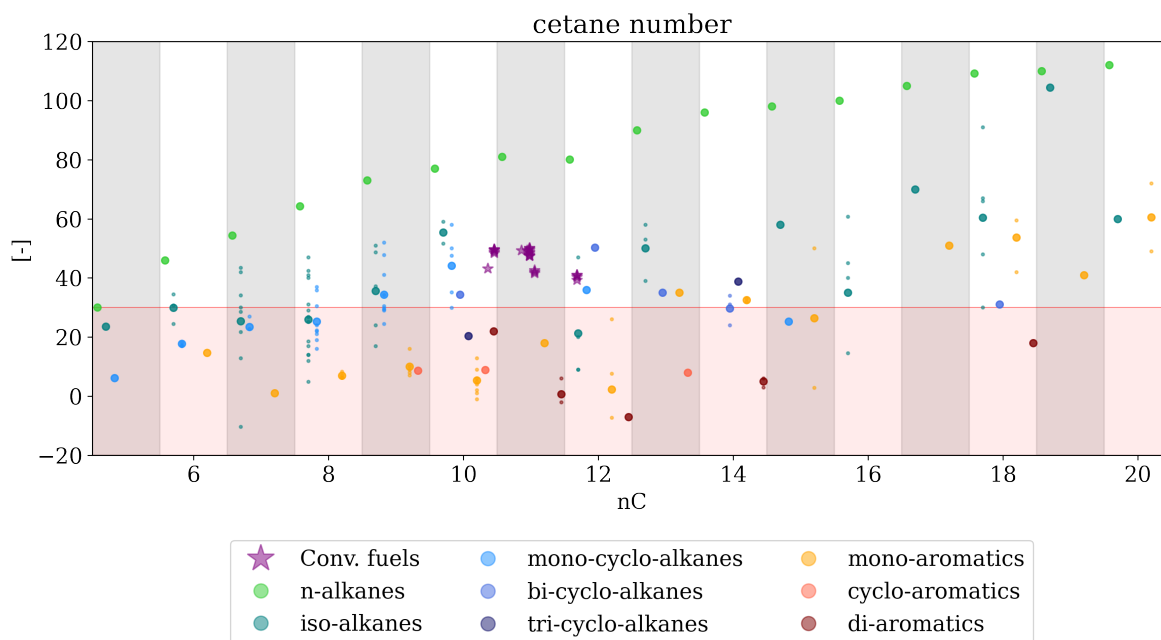


Figure 5.26: Cetane number values of the hydrocarbon families over the carbon number nC .

The identified order is reflected by the Δ_{HC} values, summarized in Table 23. With increasing carbon numbers, the cetane numbers of the n-alkanes, iso-alkanes, mono-cyclo-alkanes and mono-aromatics also increase, as visible in Figure 5.26. No clear correlations are visible for bi-cyclo-alkanes and di-aromatics. For tri-cyclo-alkanes and cyclo-aromatics the low number of measurement data prevents an investigation. The correlation of the property values with the number of carbon atoms can be seen more clearly in the individual plots of hydrocarbon families in Figure 5.27 to Figure 5.29 subscript (a). The correlation coefficients r for nC in Table 23

reflect the observations, with positive correlation values of 7 to 9 for n-alkanes, iso-alkanes and mono-aromatics in Table 23.

	Δ_{HC} [%]	nC		η_B		Annotation
		r	Δ_{nC} [%]	r	Δ_{η_B} [%]	
n-alkanes	76.0	9.0	6.5			
iso-alkanes	-15.8	7.0	19.5	-4.0	41.8	
mono-cyclo-alkanes	-31.2	5.0	12.2	-7.0	21.9	
bi-cyclo-alkanes	-19.2	-4.0	17.5	\emptyset	\emptyset	
tri-cyclo-alkanes	-36.0	\emptyset	\emptyset	\emptyset	\emptyset	
mono-aromatics	-65.5	8.0	84.3	-6.0	69.2	
cyclo-aromatics	-81.6	\emptyset	\emptyset	\emptyset	\emptyset	
di-aromatics	-88.4	2.0	194.1	\emptyset	\emptyset	

Table 23: Summary influence of different hydrocarbon families and molecular descriptors on the cetane number, * indicates an annotation for the correlation, \emptyset indicates insufficient data.

The r value of 5 for the mono-cyclo-alkanes is probably due to limited number of measurements for components with higher nC . For lower nC however, a clear correlation is visible, see Figure 5.28 (a). With respect to the influence of the branching, negative correlations and strong differences between the cetane numbers of the different isomers are visible, see Figure 5.27 to Figure 5.29 subscript (b). This is particularly striking for iso-alkanes in Figure 5.27 (a), where the specification limit is only met for low-branched isomers. This can again be explained by the greater structural similarity of the lower branched isomers with hexadecane. This correlation was also observed by Jameel for iso-alkanes [136]. For bi- and tri-cyclo-alkanes as well as cyclo- and di-aromatics no clear correlation between the carbon number and the branching index could be identified due to the insufficient number of measurements.

The computed average differences Δ_{nC} and Δ_{η_B} in Table 23 illustrate the significance of both the different sizes of the components and different branching. For iso-alkanes, the average difference Δ_{nC} of 19.5 % for the carbon number and 41.8 % for the branching Δ_{η_B} are calculated. For mono-cyclo-alkanes and mono-aromatics, the influences are even larger, with 84.3 % for Δ_{nC} and 69.2 % for Δ_{η_B} .

In summary, the cetane number is predominantly affected by the different hydrocarbon families, especially in comparison to n-alkanes. However, both the carbon number and branching have significant influence, especially for iso- and mono-cyclo-alkanes as well as mono-aromatics. More

measurements are needed for bi- and tri-cyclo-alkanes as well as cyclo- and di-aromatics to allow investigations for these families. For jet fuels, the cetane number can therefore be adjusted by first adjusting the fraction of the hydrocarbon families and then branching and carbon number. High cetane numbers can be achieved for fuels with a high content of n-alkanes or iso- and mono-cyclo-alkanes and a low branching index.

cetane number for n-alkanes & iso-alkanes

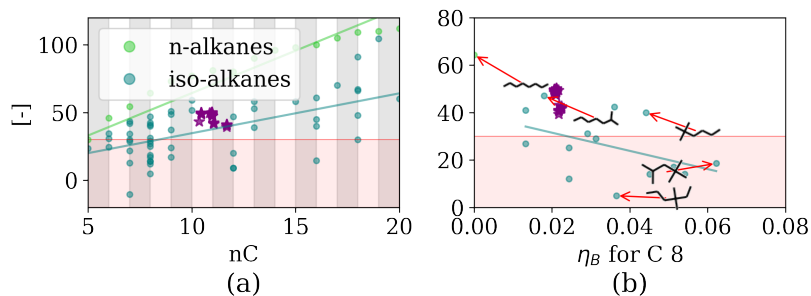


Figure 5.27: Cetane number values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

cetane number for mono-cyclo-alkanes

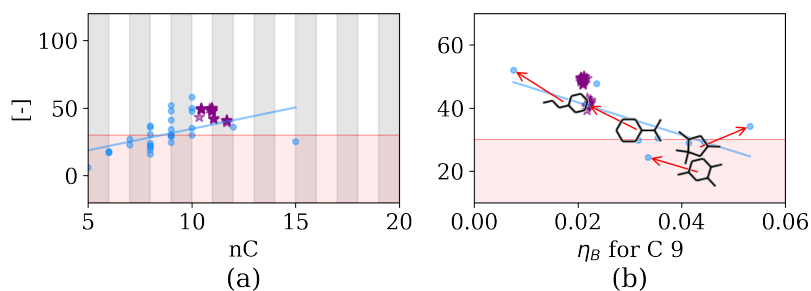


Figure 5.28: Cetane number values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b)..

cetane number for mono-aromatics

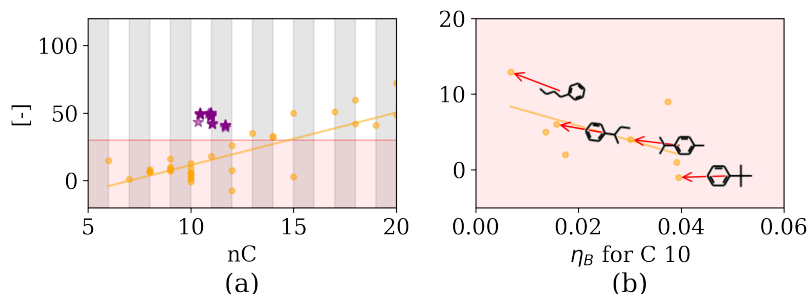


Figure 5.29: Cetane number values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

5.3.8 Boiling point

The boiling point is a characteristic measure of the volatility of the components and determines the distillation behavior of the jet fuel. The developed model for the prediction of the distillation line with the QSPR model after ASTM D2887 directly uses the boiling points of the components for their individual evaporation. The investigation of the boiling points of the components is therefore reasonable to assert the influence of the considered structural aspects on the distillation behavior of the jet fuels.

Figure 5.30 shows the boiling points for the components of the different hydrocarbon families over nC . Similar to the flash point, the values of the boiling points of the different families lie closer together compared to properties like the density or the net heat of combustion. For all hydrocarbon families, a positive correlation with increasing nC values can be observed. Since jet fuels have no “average distillation temperature”, reference boiling points are calculated by averaging the distillation temperatures at 10, 50 and 90 vol% for the conventional reference fuels. These average boiling points are shown in Figure 5.30 as purple stars and used as reference for the calculation of the metrics.

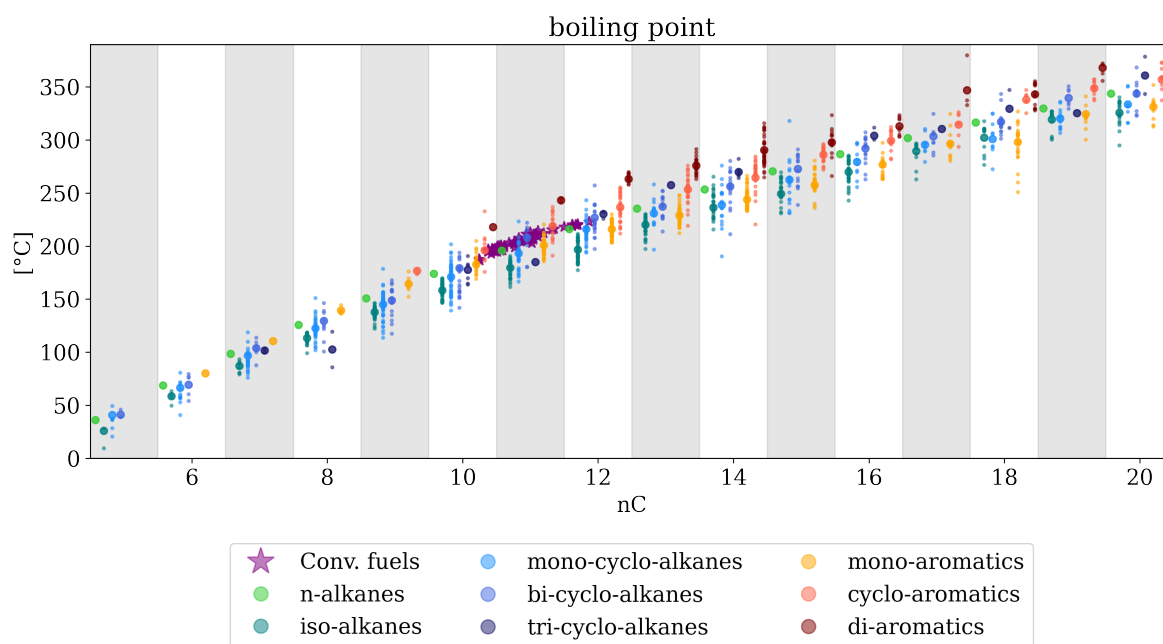


Figure 5.30: Boiling point values of the hydrocarbon families over the carbon number nC .

The correlation of boiling points with the carbon number is visualized even more clearly in the individual plots of the hydrocarbon families in Figure 5.31 to Figure 5.33 subscript (a). The correlation is furthermore reflected in the positive correlation coefficients in Table 24, with values of 10 for all hydrocarbon families. Comparing the differences of the different families Δ_{HC} in Table 24, iso-alkanes show the lowest average boiling point, 4.1 % lower than the average boiling point of the jet fuel reference. Mono-cyclo-alkanes show the second lowest boiling point, followed

by n-alkanes, mono-aromatics and bi-cyclo-alkanes. Tri-cyclo-alkanes as well as cyclo- and di-aromatics show the highest boiling points, with an average difference Δ_{nC} of 14.3 % to reference of the jet fuels for di-aromatics.

	Δ_{HC} [%]	nC		η_B		Annotation
		r	Δ_{nC} [%]	r	Δ_{η_B} [%]	
n-alkanes	-0.7	10.0	4.5			
iso-alkanes	-4.1	10.0	4.5	-4.0*	3.2*	No trend for molecules with tertiary carbon atoms
mono-cyclo-alkanes	-2.1	10.0	4.5	-8.0*	5.1*	Correlation with number of ring atoms
bi-cyclo-alkanes	-0.2	10.0	4.5	-6.0*	4.7*	Correlation with number of ring atoms
tri-cyclo-alkanes	-0.1	10.0	5.2	-9.0	0.9	
mono-aromatics	-0.4	10.0	3.8	3.0*	3.6*	Correlation with aromatic carbon atoms, connected only to other carbon atoms
cyclo-aromatics	8.0	10.0	3.5	-6.0	3.6	
di-aromatics	14.3	10.0	2.8	3.0*	3.6*	Correlation with aromatic carbon atoms, connected only to other carbon atoms

Table 24: Summary influence of different hydrocarbon families and molecular descriptors on the boiling point, * indicates an annotation for the correlation, \emptyset indicates insufficient data.

For the branching, decreasing boiling points over η_B are observed for iso-alkanes, cyclo-alkanes and di-aromatics in Figure 5.31, Figure 5.32, Figure G 19 and Figure G 21 subscript (b). For iso-alkanes in Figure 5.31 (b) components with tertiary carbon atoms³ strike out as outliers, with boiling points similar to the low-branched components. This also explains the lower correlation coefficient of -4.0 for η_B in Table 24 compared to the other hydrocarbon families. For mono- and di-aromatic components, the position of the side chain is decisive for the boiling point of the component, see Figure 5.33 (b) and Figure G 21 (b). The *cXHO* descriptor is again able to resolve this relationship further and shows a positive correlation for both families with the number of carbon atoms connected only to other carbon atoms in Figure 5.33 (c) and Figure G 21 (c). Similar to the density, the number of ring atoms strongly influences the property values for the mono-cyclo-alkanes and bi-cyclo-alkanes visible in Figure 5.32 (b) and Figure G 19 (b). Positive correlations are thereby observed with an increasing boiling point for an increasing number of ring atoms.

In comparison, the average differences in Table 24, the carbon number and the branching are equally important for the boiling point, with comparable values for Δ_{nC} and Δ_{η_B} . With the exception of cyclo- and di-aromatics the average values of the hydrocarbon families Δ_{HC} are on

³ Carbon atoms connected with three other carbon atoms

the other hand very similar. The distillation line of a jet fuel can therefore be mainly influenced by the selection of different fuel components with different carbon numbers and branching indices. Distillation temperatures can thereby be increased by increasing the carbon number and decreasing the average branching of the components and lowered vice versa.

boiling point for n-alkanes & iso-alkanes

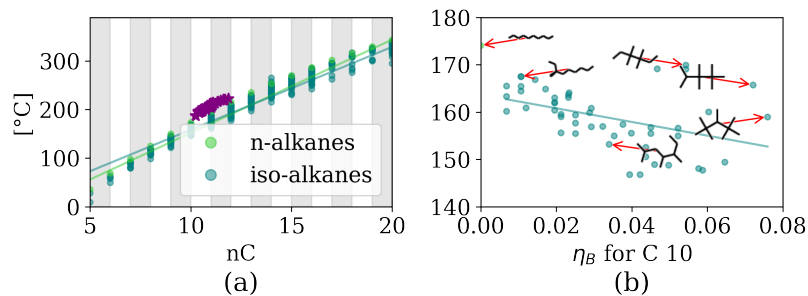


Figure 5.31: Boiling point values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

boiling point for mono-cyclo-alkanes

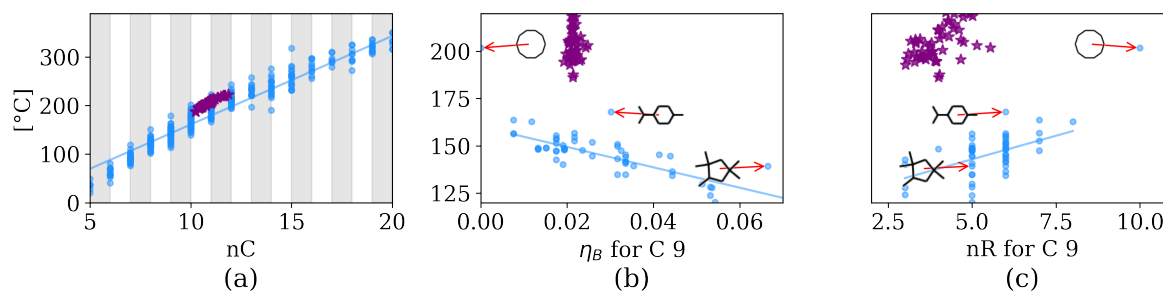


Figure 5.32: Boiling point values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b) and number of ring atoms nR (c).

boiling point for mono-aromatics

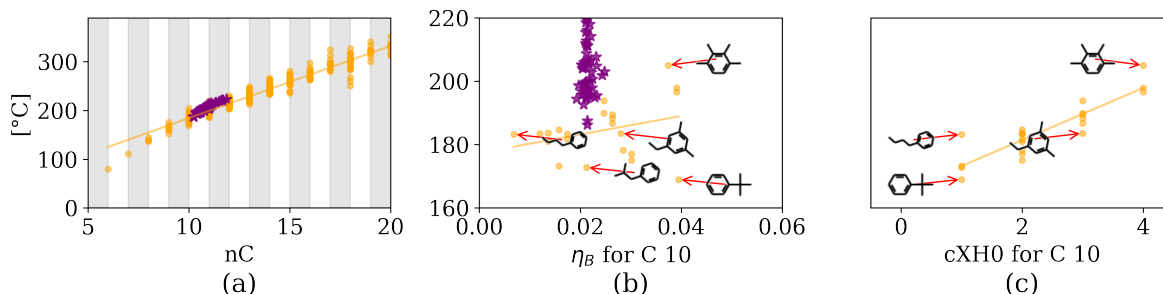


Figure 5.33: Boiling point values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

5.3.9 Yield sooting index

As mentioned in the introduction, the soot particle induced formation of contrail cirrus is of growing concern both for the scientific community and political policy makers. Lee et al. [7] as well as Voigt et al. [8] account contrail cirrus of aircrafts as the major share of the climate impact of the aircraft industry, due to back radiation. The importance of non-CO₂ effects of the aviation industry on the climate has been recognized in a recent report of the European Union Aviation Safety Agency, requested by the European Commission [137]. The sooting tendency is already an important property with specification limits for the smoke point of conventional fuels and blends in ASTM D1655 and ASTM D7566 and might become an important design parameter for the design of new jet fuels in the near future.

To quantify the sooting tendency of the fuel components, the yield sooting index (YSI) is utilized. The YSI was introduced by McEnally and Pfefferle in 2007 [128]. It allows the quantification of the soot tendency of pure components and mixtures with a significantly lower level of uncertainty compared to concepts like the smoke point after ASTM D1322 [138] or the threshold sooting index. A recent publication adjusted the yield sooting index to a unified scale and provided property values for pure compounds [139]. For the calculation of the YSI of mixtures and fuels, a linear molar-based mixing rule is recommended by Xiang et al. [140]. Since the data for the reference fuels from the CRC world fuel survey of 2006 does not list YSI measurements for the different fuels, the YSI values are predicted using the QSPR sampling model. Individual validation plots for the QSPR sampling model are provided for pure compounds in the Supplementary Material E in Figure E 39. The results of the YSI investigation are displayed in the following Figure 5.34 to Figure 5.37. Figure 5.34 shows the YSI values for components of the different hydrocarbon families with increasing carbon numbers.

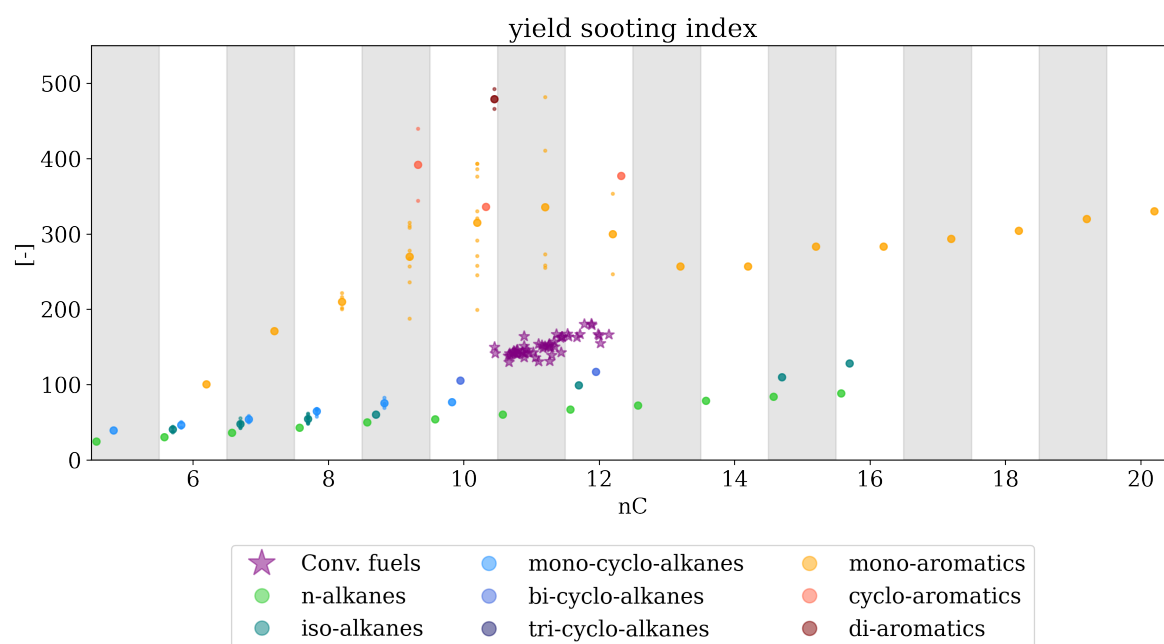


Figure 5.34: Yield sooting index values of the hydrocarbon families over the carbon number nC .

Since the YSI is a fairly recent concept, the number of available measurements is significantly smaller compared e.g. to the net heat of combustion. The displayed measurements are taken from the official database of the National Renewable Energy Laboratory [96]. Figure 5.34 shows a clear correlation between the YSI of all families and the carbon number of the components. N-alkanes show the lowest values, followed by iso-alkanes, mono-cyclo-alkanes, bi-cyclo-alkanes. For tri-cyclo-alkanes, no measurements are yet available. The order is reflected by the Δ_{HC} values and summarized in Table 25 with the introduced color scheme as background color. In comparison aromatic components show significantly higher YSI values in Figure 5.34 with values that excel those of non-cyclic alkanes multiple times.

The positive correlation of the YSI with the carbon number nC is reflected in the r values of 10 for n- and iso-alkanes. The r value of 3 for the mono-aromatics does not correspond to the identified trend in the individual plot for the mono-aromatics in Figure 5.37. This can be explained by the influence of the side chains, which result in significantly higher yield sooting indices for components with multiple side chains and lower yield sooting indices for components with only one. The “*” suffix in Table 25 indicate this important observation. For the branching positive correlations of the YSI with the η_B are visible, for the iso-alkanes cyclo-alkanes as well as mono-aromatics, see Figure 5.35 to Figure 5.37 subscript (b). For mono-aromatics, a clear relationship with the $cXH0$ descriptor is identified in Figure 5.37 subscript (c). For the other families, the low number of available measurements prevents an investigation.

	Δ_{HC} [%]	nC		η_B		Annotation
		r	Δ_{nC} [%]	r	Δ_{η_B} [%]	
n-alkanes	-58.0	10.0	10.1			
iso-alkanes	-44.7	10.0	11.9	9.0	7.5	
mono-cyclo-alkanes	-55.1	∅	∅	10.0	6.4	
bi-cyclo-alkanes	-26.1	∅	∅	∅	∅	
tri-cyclo-alkanes	∅	∅	∅	∅	∅	
mono-aromatics	78.0	3.0*	7.9*	6.0	17.9	Correlation with aromatic carbon atoms, connected only to other carbon atoms
cyclo-aromatics	144.5	∅	∅	∅	∅	
di-aromatics	218.1	∅	∅	∅	∅	

Table 25: Summary influence of different hydrocarbon families and molecular descriptors on the yield sooting index, * indicates an annotation for the correlation, ∅ indicates insufficient data.

The strong influence of both the carbon number and the branching is illustrated by the similar average differences Δ_{nC} and Δ_{η_B} Table 25. For the iso-alkanes, the influence of the carbon number is slightly higher on the YSI with a Δ_{nC} of 11.9 % compared to the Δ_{η_B} of 7.5 %. For mono-aromatics on the other hand a Δ_{η_B} of 17.9 % for the branching exceeds the one of the carbon number with of 7.9 % for Δ_{nC} .

In summary, the YSI is predominantly affected by the hydrocarbon family, however differences in the branching and carbon number of components are significant, especially for aromatics. The sooting tendency of a fuel can therefore be mainly lowered by increasing the fraction of n- and iso-alkanes with low carbon numbers and branching indices. Cyclic components, especially aromatics lead to exponential increases in the sooting tendency and should therefore be avoided if possible.

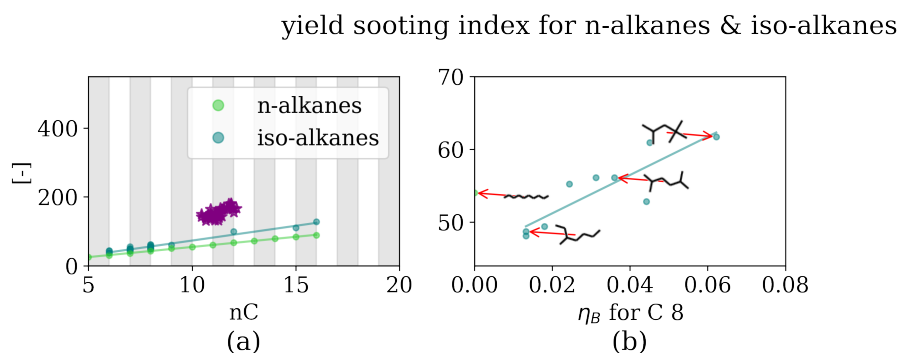


Figure 5.35: Yield sooting index values of n-alkanes and iso-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

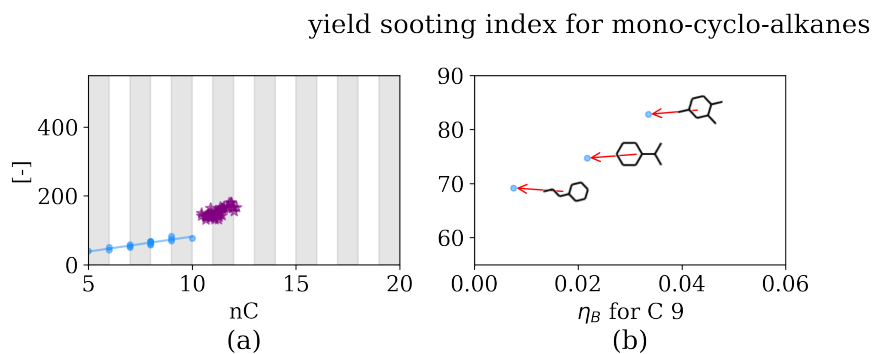


Figure 5.36: Yield sooting index values of mono-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

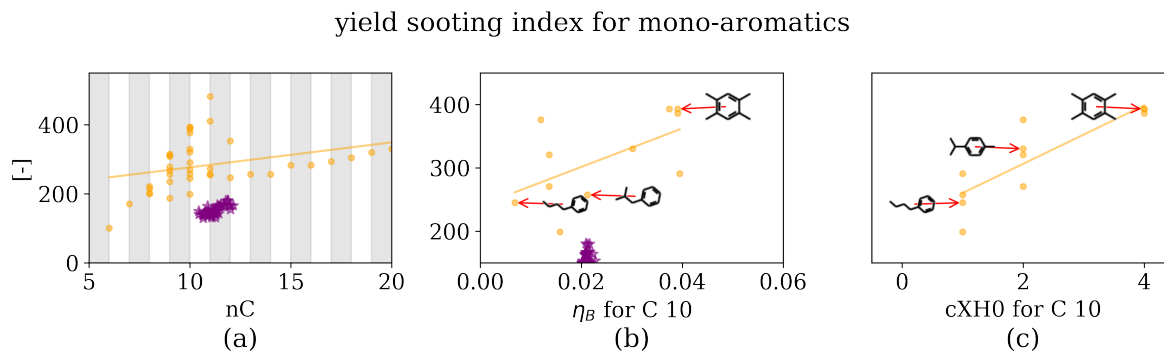


Figure 5.37: Yield sooting index values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

5.4 Summary table of the structure-property correlations

So far, the influences of the hydrocarbon family, the size and the branching of the fuel components have been investigated separately for each property. However, the considered fuel properties are interconnected. Increasing or decreasing the fraction of fuel components in the fuel affects all properties simultaneously. The overall number of interdependent correlations is thereby tremendous: Nine properties for components from eight different families and three structural aspects lead to 207 correlations, which were investigated using 333 metric values. For the fuel design however, interactions between the correlations have to be apparent to estimate the consequences of the design decision. Therefore, all metric values from the previous investigations are summarized in one table in Table 26.

Table 26 lists the correlation strength r and average differences Δ in % of the individual properties (rows) for each of the considered families (columns) for the three structural descriptors: hydrocarbon family (HC), carbon number (nC) and branching index of isomers (η_B). The numerical r values from the tables for the individual properties in Section 5.3 are thereby translated into symbols for clarity. The symbol ++ indicates a strong positive correlation $r \geq 8$, + a positive correlation $r > 5$, - a negative correlation $r < -5$ and -- strong negative correlation $r \leq -8$. For r with values between 5 and -5 the fields are left empty, indicating no observed correlation. The correlations of the property values with the hydrocarbon families HC are derived from the Δ_{HC} values of the previous investigations. A "*" as suffix of the average differences Δ indicates an annotation, e.g. correlation with another metric like the number of ring atoms or effects that were only observed visually, like the influence of the position of the side chains. Details about these annotations are found in the respective tables of the properties in the subsections of 5.3. The background color of the cells in Table 26 again indicates the average property value of a hydrocarbon family Δ_{HC} relative to the average value range of the conventional reference fuels. The symbol \emptyset indicates an insufficient number of measurements,

thereby pointing out the need for additional measurements. Empty fields for η_B of the n-alkanes are due to the nonexistence of branching.

		n-alkanes		iso-alkanes		mono-cyclo-alkanes		bi-cyclo-alkanes		tri-cyclo-alkanes		mono-aromatics		cyclo-aromatics		di-aromatics	
		r	Δ	r	Δ	r	Δ	r	Δ	r	Δ	r	Δ	r	Δ	r	Δ
density	HC	--	7.6	--	5.9	++	0.6	++	8.1	++	14.5	++	8.6	++	17.2	++	24.4
	nC	++	1.3	+	1.5	+	1.2		0.9		2.5		0.2	-	0.5		1.1
	η_B				1.7*	-	2.1*		1.4*		\emptyset		0.8*		1.1*		2.4*
surface tension	HC	--	5.6	--	10.8	++	2.3	++	23.2	++	43.1	++	19.5	++	28.8	++	43.4
	nC	++	3.3	+	4.0		3.3	++	3.3		\emptyset		2.7		\emptyset	-	3.2
	η_B				4.0	--	4.0		\emptyset		\emptyset		5.2*		\emptyset		\emptyset
kinematic viscosity	HC	--	18.6	--	11.8	++	97.6	++	95.0	++	176.2	++	15.8	++	195.3	++	219.1
	nC	++	39.7	++	37.1	+	42.3	+	43.5	+	104.9	+	41.1	+	30.9	+	54.6
	η_B			+	36.0*		45.9*		50.4*		\emptyset		39.1		46.3		34.0
net heat of combustion	HC	++	2.5	++	2.5	++	1.0	--	0.4	--	1.8	--	3.7	--	4.9	--	7.2
	nC	--	0.1	--	0.1		0.1		0.2	+	0.7	++	0.4	++	0.6	++	0.7
	η_B				0.1		0.4*	-	0.1		\emptyset	-	0.1*		0.1	-	0.2*
flash point	HC	+	5.7	-	0.2	-	1.9*	+	5.9	+	5.9	+	5.5	+	10.4	+	19.7
	nC	++	4.6	++	4.6	++	4.8	++	3.8		\emptyset	++	3.5	++	3.5	++	3.2
	η_B			--	14.0	-	9.7		\emptyset		\emptyset	--	5.1	--	10.5	++	2.3
freezing point	HC	+	12.8	-	5.2	+	6.2	+	25.2	+	63.8	+	5.8	+	2.9	+	35.3
	nC	++	4.7	+	5.4	++	6.3		8.0		\emptyset	+	4.9		\emptyset		5.9
	η_B			+	24.8*		17.6*		\emptyset		\emptyset	+	48.9		\emptyset	+	96.6
cetane number	HC	++	76.0	--	15.8	--	31.2	--	19.2	--	36.0	--	65.5	--	81.6	--	88.4
	nC	++	6.5	+	19.5	+	12.2		17.5		\emptyset	++	84.3		\emptyset		194.1
	η_B				41.8	-	21.9		\emptyset		\emptyset	-	69.2		\emptyset		\emptyset
boiling point	HC	-	0.7	-	4.1	-	2.1	-	0.2	-	0.1	-	0.4	+	8.0	+	14.3
	nC	++	4.5	++	4.5	++	4.5	++	4.5	++	5.2	++	3.8	++	3.5	++	2.8
	η_B				3.2*	--	5.1*	-	4.7*	--	0.9		3.6*	-	3.6		3.6*
yield soot-ing index	HC	--	58.0	--	44.7	--	55.1	--	26.1	--	\emptyset	++	78.0	++	144.5	++	218.1
	nC	++	10.1	++	11.9		\emptyset		\emptyset		\emptyset	*	7.9		\emptyset		\emptyset
	η_B			++	7.5	++	6.4		\emptyset		\emptyset	+	17.9		\emptyset		\emptyset

Table 26: Summary table for the relationship and influence of structural descriptors of fuel components on their physicochemical properties, average differences Δ are given in %.

Table 26 condenses the observations into one summary template. This table should serve as an overview and starting point for understanding the influence of the different structural aspects on the considered properties and consequently the fuel design. The strength of the correlations, expressed by the symbols of r , together with the average property differences Δ , allow an estimate of the change in the property value when an aspect is increased or decreased in a fuel. The arrangement of the structure-property correlations into one table furthermore illustrates the interconnections of the property correlations. To demonstrate the use of the summary table and illustrate the interconnection of the property correlations, two examples are given below.

5.4.1 Usage example of the structure-property correlations

Example: Increasing the density of a jet fuel candidate:

The first use case focuses on increasing the density of a jet fuel candidate with a high fraction of n- and iso-alkanes. Low densities at a similar heat of combustion can reduce the range of an aircraft due to the lower volumetric energy content and can lead to problems in the metering and gauging of sensors the aircraft. The ASTM D4054 therefore lists a lower limit of 775 kg/m³ to guarantee a safe application. This use case is relevant for the fuel design of new SAF candidates to reach the set specification limit or increase the blending fraction of SAF from already approved fuel production routes, e.g. FT-SPK.

Altering the hydrocarbon families

For both n- and iso-alkanes the correlation, symbol "– –" for *HC* in Table 26 states strong negative correlations. This indicates that large fractions of both families reduce the density of a jet fuel. The components of those two families have densities that are on average 7.6% and 5.9% lower respectively, than conventional fuels, as indicated by the Δ values. The lower average density of the two families is also indicated by the blue background colors of the cell. In contrast, cyclo-alkanes or aromatics have higher densities on average, increasing the fraction of those families would therefore increase the density. However, as Table 26 also shows, increasing the fraction of cyclo-alkanes and aromatics, especially families with multiple rings, would also increase the viscosity, freezing point and sooting tendency, properties that should be kept low in the fuel design. The best compromise would therefore be achieved by decreasing the fractions of n- and iso-alkanes and substituting them with mono-cyclo-alkanes.

Altering the average carbon number

Increasing the average carbon number *nC* for both n-alkanes and iso-alkanes would also increase the densities of n- and iso-alkanes. This can be derived from the "++" and "+" correlation symbols in the respective cells of Table 26. The Δ values indicate, that increasing the average carbon number of the fuel by 1 would increase the density by 1.3 and 1.5 % respectively. Compared to a change in the fractions of the family, this has significantly lower leverage. Changing the average carbon number would however also have unwanted effects. Viscosity, freezing point and sooting tendency, properties that should be kept low in the fuel design, would be increased. While the freezing point and sooting tendency would be less affected compared to a change of the hydrocarbon families, the viscosity would increase exponentially.

Altering the branching

For the branching η_B the missing correlation symbol and the Δ value of 1.7 % illustrate that no significant correlation of branching and density values was observed for iso-alkanes. A systematic change in the density can therefore not be expected by increasing or decreasing the branching of

the components. However, the “*” suffix indicates an existing correlation with the partial positive surface area, as outlined in Section 5.3.1. Higher densities could therefore be achieved by increasing the content of components with a low PPSA if the production process allows the tailoring of the fuel on a component level. For n-alkanes no values are listed due to non-existing branching.

Conclusion

In conclusion, an increase in density could best be achieved by reducing the mass fractions of n- and iso-alkanes and substituting them with mono-cyclo-alkanes. The density could further be increased by increasing the average carbon number of the fuel. Increasing the density by these measures would however also increase the freezing point, viscosity and sooting tendency. The change of these and other properties has to be considered to make sure that respective specification limits are not exceeded.

Example: Decreasing the sooting potential of a jet fuel

This use case seeks to lower the sooting tendency of a fuel candidate with a composition similar to conventional fuels. The sooting tendency is increasingly becoming an important factor in fuel design, as soot particles can act as nuclei for the formation of cirrus clouds and thus have a significant effect on the climate.

Altering the hydrocarbon families

The correlation symbols for *HC* in Table 26 show strong positive correlations and the highest average differences to conventional fuels Δ for aromatic families. The comparison of the average differences Δ illustrates that the sooting tendency of aromatics surpasses that of alkanes multiple times. To decrease the yield sooting index of the fuel candidate, the mass fraction of the aromatics could consequently be decreased and the fraction of the alkanes increased. Also, the substitution of cyclo- and di-aromatics by mono-aromatics would reduce the yield sooting index by more than half, comparing a Δ of 144 % and 218 % for diaromatics with a 78 % for mono-aromatics. A reduction of the aromatics would however also reduce the density and flash point, as visible in the corresponding rows in Table 26.

Altering the average carbon number

For the carbon number, investigations were only possible for n- and iso-alkanes as well as mono-aromatics, due to the low number of measurements for the other families. This is indicated by the \emptyset symbol in Table 26. Assuming similar correlations for cyclo- and di-aromatics, as for mono-aromatics, the yield sooting index would decrease with a lower average carbon number of the fuel. Table 26 lists a “*” symbol for mono-aromatics which indicates that a correlation does exist but that is not captured by the *r* value due to the branching effects. Similar to a change in the

hydrocarbon families, a lowering of the average carbon number would reduce density and freezing point.

Altering the branching

Likewise to the carbon number, correlations for the branching could only be identified for iso- and mono-cyclo-alkanes as well mono-cyclo-alkanes, due to low number of measurements for the other families. For all three families positive correlations were observed, indicated by “++” and “+” symbols in Table 26. The branching should therefore be reduced to lower the sooting tendency. The corresponding plot in Figure 5.37 (b) furthermore shows, that the position of the side chains also influences the YSI. Components with multiple side chains have on average higher YSI values compared to components with one longer side chain. Reducing the average branching only effects the density slightly, while the flash point on contrast is increased.

Conclusion

To reduce the sooting tendency of the fuel candidate, the mass fraction of aromatics should be reduced substituting them with n-, iso and cyclo-alkanes. The yield sooting index can be further reduced by lowering the average carbon number and branching of the fuels. While increasing the fraction of n-, iso- and cyclo-alkanes would also decrease density and the flash point, the reduction of the branching would lead to an increase of the flash point. The flash point could therefore be kept relatively similar. As for the first example, the changes in density and flash point have to be considered to make sure that their respective specification limits are not exceeded.

5.5 Summary and conclusion

The relationships between the composition of a jet fuel and its physicochemical properties are complex and manifold. Composed of hundreds of possible components from different hydrocarbon families the composition space of jet fuels is vast and difficult to comprehend. The design of jet fuels therefore requires tools that aid in understanding the relations of composition and properties and allow estimating the consequences of changes in the fuel composition. This chapter aimed to develop those design tools by systematically investigating the relationships between possible fuel components and all properties relevant for the fuel prescreening as well as the yield sooting index. For this, the molecular structures of over 1870 possible fuel components from eight hydrocarbon families were characterized by three main aspects: hydrocarbon family, carbon number and branching with corresponding structural descriptors. Investigations were conducted using both plots as well as metrics. While the plots allow a visual identification of the correlations, the metrics make a uniform comparison of the quantified differences possible for all hydrocarbon families and properties.

For most of the properties, the kind of hydrocarbon family of the component was found to affect the property values most significantly. Strong differences between the trends of the properties

were thereby observed for some properties. For example, while the average density of components changes gradually over the different families, for properties like the cetane number and the yield sooting index, the values change drastically, e.g. increasing exponentially from n-alkanes to diaromatics. Comparing the average property values of the hydrocarbon families to conventional fuels, mono-cyclo-alkanes were found to be the most similar. For the other hydrocarbon families, the property values are either higher or lower than the ones of the reference fuels, depending on the property.

The number of carbon atoms was found to be the second most important descriptor on average. For most properties, the values increased with the number of carbon atoms, with the exception of density, surface tension and net heat of combustion, where reverse trends were observed depending on the family. For kinematic viscosity and boiling point, the number of carbon atoms was found to be the most important or one of the most important structural aspects for the property value, with strong linear or even exponential trends.

The influence of the branching of the components on the property values was found to strongly depend on the hydrocarbon family and the respective property. For the freezing point, flash point and cetane number, the branching has a significant influence especially for iso-alkanes. The branching was observed to be often decisive for the compliance of a property value to the specification limits. For these properties, the branching is therefore an important parameter to consider for the fuel design, especially for fuels composed of just a few fuel components.

The findings of the observed trends and quantitative changes were summarized in a table listing all calculated metrics. The table illustrates the numerous intercorrelations of properties and structural descriptors, as well as the average differences in the property values. Together with the detailed findings of these investigations, the summary table will be the basis for the fuel design in the upcoming chapter.

6 Fuel Design and Blending Analysis

To demonstrate the full workflow of screening and designing a fuel candidate, a fuel assessment and reformulation is conducted for a real jet fuel candidate, to optimize its physicochemical properties and meet the approval specifications. This is a hypothetical scenario, with the goal of investigating the applicability of the developed tools by creating and screening new fuel variants designed from the original fuel candidate. The fuel variants are designed based on the screening results of the original fuel candidate, using findings from the composition-property relation investigations of Chapter 5 and rudimentary simulations of established refinery process operations, to ensure the reformulation could be carried out in a real production facility. Based on the prescreening results, the fuel variant with the highest chances for approval is subjected to a blending study with conventional oil-based fuels to examine if the fuel variant has chances as a synthetic blending component and if compliance to the respective specification limits can be achieved via blending. This hypothetical scenario investigates the applicability of the predictive models for jet fuel blending with the goal of estimating the maximum blending fraction of the synthetic jet fuels, under consideration of the jet fuel blending specifications. To illustrate the potential ecological benefits of the synthetic fuel, the reduction of the sooting tendency and CO₂ emissions of the blends are calculated, relative to the original conventional jet fuels. The blending analysis will be carried out twice: first without and second with the consideration of the uncertainties predicted by the models, to demonstrate the influence of uncertainties on the blending of the jet fuel and outline potential consequences if uncertainties are neglected.

6.1 Prescreening of the untreated jet fuel candidate

The fuel investigated in this fuel design, is a real jet fuel candidate, that was assessed by the Institute of Combustion Technology of the DLR. As visible in Figure 6.1, the fuel is composed solely of n-alkanes with a carbon number of 16 and 18 making up 97 mass%.

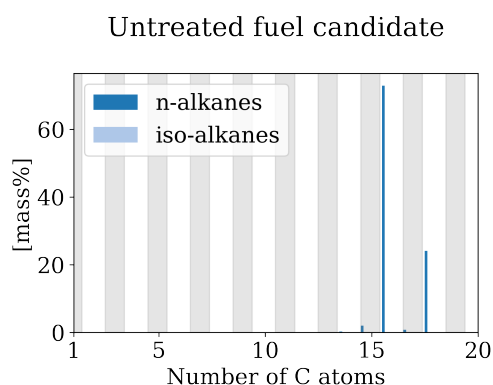


Figure 6.1: Plot of composition of the untreated jet fuel candidate.

In the first step of the fuel design process, the jet fuel candidate is assessed according to the Tier α prescreening protocol, described in Section 1.2.2. The critical fuel properties are thereby predicted using the three developed models and assessed in comparison to prescreening specification limits derived from ASTM D4054 and ASTM D7566. To additionally estimate the sooting tendency of the fuels, the yield sooting index is estimated using the QSPR sampling and M-QSPR model with a molar-fraction-based mixing rule see Section 2.2.1.

The results for the screening of the jet fuel candidate are displayed in Figure 6.2, using the same property screening plot introduced in Section 4.2. As an additional reference for the property values of the candidate, box plots are provided for each property in the figure. The box plots visualize the value range of conventional jet fuels based on the CRC jet fuel survey from 2006 [92]. Since no measurements of the yield sooting index are available for the conventional reference fuels, the values are predicted using the QSPR sampling model.

The prediction results for the three models (DC, QSPR sampling and M-QSPR) are displayed in blue, green and purple respectively in Figure 6.2. The displayed error bars indicate the estimated 95 % prediction intervals of the probabilistic models.

Comparing the model predictions, significant differences are visible for the DC and M-QSPR models. The DC model prediction for the net heat of combustion is 0.6 MJ/kg higher, the flash point, freezing point and distillation line are in part up to 50 K lower and the cetane number of 72.6 is one third lower compared to the predictions of the other models. For the M-QSPR model, deviations from the other models are visible for kinematic viscosity, which is 60 to 50 % lower compared to the predictions of the other models.

To determine if these outlying predictions of the M-QSPR and DC models are physically possible, the predictions are compared with measured property values of the pure compounds, that make up most of the fuel composition. The fuel is composed to 97 mass% of two n-alkanes with 16 and 18 carbon atoms: hexadecane and octadecane. The properties of the fuel are therefore mainly determined by those two components and must lie in between the property values of these two pure compounds. The properties of n-alkanes were illustrated as part of the fuel composition-property investigation in Section 5.3. From the respective property plots in Section 5.3, it can be concluded that the outlined predictions over- or underpredict the respective property value: The measured net heat of combustions of hexadecane and octadecane lie close to 44 MJ/kg, see Figure 5.15, the measured flash points are above 100 °C, see Figure 5.19, the measured freezing points are above 0 °C, Figure 5.23, the measured cetane numbers are close to 100, see Figure 5.26 and the measured boiling points are over 280 °C, see Figure 5.31. Furthermore, the estimated freezing points of the M-QSPR and QSPR sampling models well above 0 °C indicate, that the kinematic viscosity predictions of the M-QSPR model are estimated too low for this fuel. As a consequence, the predictions of the DC model for the net heat of combustion, flash point, freezing point, cetane number and distillation line as well as the prediction of the M-QSPR model for the kinematic viscosity, are not physically possible and are therefore not considered for the screening of the

fuel. These predictions are therefore framed by red box in Figure 6.2. The erroneous predictions can be explained by the dissimilarity of the screened fuel with the fuels available for training and testing of the DC and M-QSPR model. For the QSPR sampling models, the composition of the fuel is no issue since the distinct components can be directly selected for the modeling. This observation illustrates the necessity of critically evaluating the predictions of property models and comparing them with the predictions of other models as well as available measurements. The DC models, which showed the highest predictive capability in validation and screening in Chapter 4 cannot be used for large parts of this specific fuel screening.

Considering the trusted predictions, the results in Figure 6.2 clearly show that the estimated values for the kinematic viscosity, flash point, freezing point and distillation line exceed the specification limits significantly. These properties would therefore have to be modified in order to increase the chances of approval.

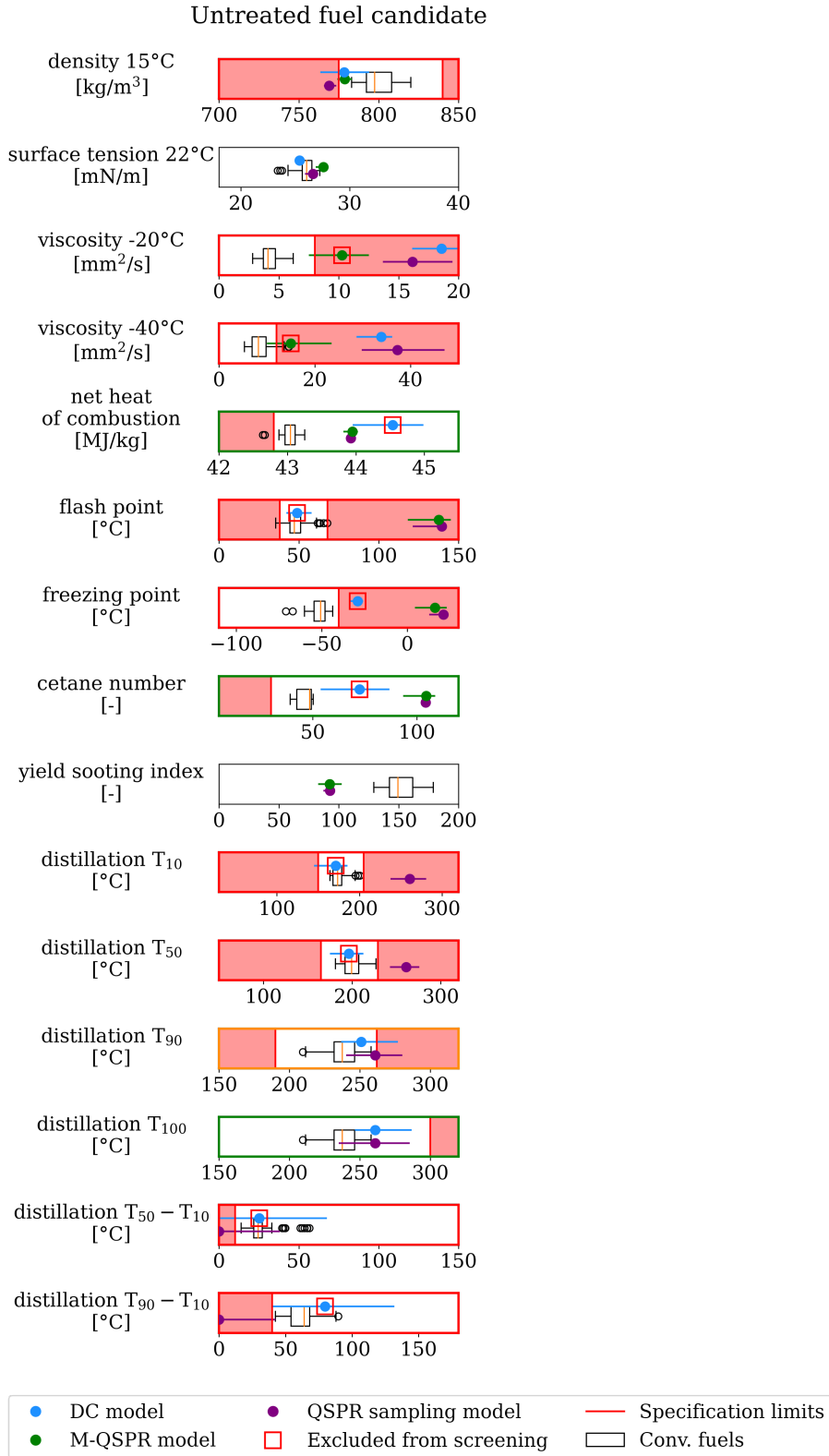


Figure 6.2: Results of the jet Tier α jet fuel prescreening for the untreated jet fuel candidate.

6.2 Fuel design and prescreening of fuel variants

To reduce the kinematic viscosity, flash point, freezing point and distillation line of the candidate fuel the findings of the composition-property investigations from Chapter 5 are utilized. From the summary in Table 26 it can be inferred that both an increase in iso-alkanes fraction as well as a reduction in the average carbon number of the components lead to a reduction of the exceeding properties. These compositional modifications can be achieved by refinery process operations, e.g. hydroisomerization and hydrocracking. Both hydroisomerization and hydrocracking are well-established process operations, utilized e.g. in the production of other sustainable aviation fuels like Alcohol-to-Jet and Fischer-Tropsch jet fuels [141]. The hydroisomerization process converts unbranched n-alkanes to branched iso-alkanes [142], while hydrocracking denotes the breaking of long chain hydrocarbons and the saturation of free valencies by hydrogen, both in the presence of hydrogen [143]. Both process operations are often carried out in one reactor, using the same functional catalyst and educt feed [141]. In a hydrocracking process, therefore, isomerization also takes place. The transition from hydroisomerization to hydrocracking is determined mainly by the reactor temperature: 190-230 °C for hydroisomerization and 230-280 °C for hydrocracking, depending on the functional catalyst [142]. In the scope of this work, hydroisomerization and hydrocracking are applied as two separate processes on the fuel candidate to produce two fuel variants. A third variant is formulated by treating the hydrocracked variant with a subsequent distillation.

Design of fuel variants

The hydroisomerization is modelled following the results of Weitkamp for the hydroisomerization of long-chain n-alkanes on a platinumium calcium zeolite catalyst of [142]. Weitkamp describes a conversion of up to 40 % of n-alkanes to iso-alkanes at a temperature of 200 °C, without the occurrence of hydrocracking. Based on these results, the first variant is formulated in which 40% of the n-alkanes are converted to iso-alkanes, with the carbon number remaining the same. To investigate the results for maximum possible isomerization without the occurrence of hydrocracking. Since information about the isomers formed during the isomerization and hydrocracking processes is scarce, it is assumed that all isomers with the respective carbon numbers listed in the utilized database are potentially present and the assumptions for the models are therefore valid.

The hydrocracking is also modelled following the results of Weitkamp, assuming ideal cracking on platinumium calcium zeolite catalyst at 255 °C with a conversion of 80 % by hydrocracking and 15 % by hydroisomerization, therefore converting 95 % n-alkanes to iso-alkanes in total [142]. For the cracked products, a broad distribution with a maximum yield at $C_{n/2}$ is assumed, based on the results for hexadecane of Weitkamp [143] and Coonradt and Garwood [144]. Figure 6.3 shows the schematic conversion curves for the five components: tetradecane, pentadecane, hexadecane, heptadecane and octadecane. The curves of tetradecane, pentadecane, heptadecane

and octadecane are modelled by transferring the conversion curve for hexadecane to the other fuel components.

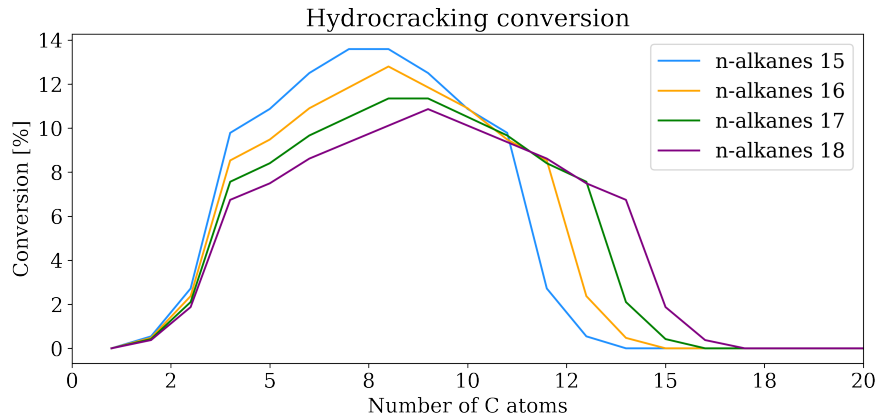


Figure 6.3: Conversion curves for hydrocracking of *n*-tetradecane, *n*-pentadecane, *n*-heptadecane and *n*-octadecane after Weitkamp [143] and Coonradt und Garwood [144].

In summary, three jet fuel variants are produced by using the described simulated process operations of hydroisomerization, hydrocracking and distillation. Table 27 lists the three variants and the original untreated fuel variant with the corresponding process operations, assumed process conditions and byproducts.

	Process operations	Process conditions	By-products	Schematic illustration
Untreated candidate fuel	-	-	-	
Fuel variant -hydroisomerized	Hydroisomerization	200 °C, Pt/Y	-	
Fuel variant -hydrocracked	Hydrocracking	255 °C, Pt/Y	-	
Fuel variant -hydrocracked & distilled	Hydrocracking & distillation	255 °C, Pt/Y 120 °C	33.2 mass% light boiling fraction	

Table 27: Summary table of fuel design process operations

Plots of the resulting compositions of the upgraded fuels are shown in Figure 6.4. The composition of the hydroisomerized variant is shown on the upper right, the hydrocracked variant on the lower left and the hydrocracked and distilled variant is shown on the lower right. For comparison, the fuel composition of the untreated fuel variant is given on the upper left.

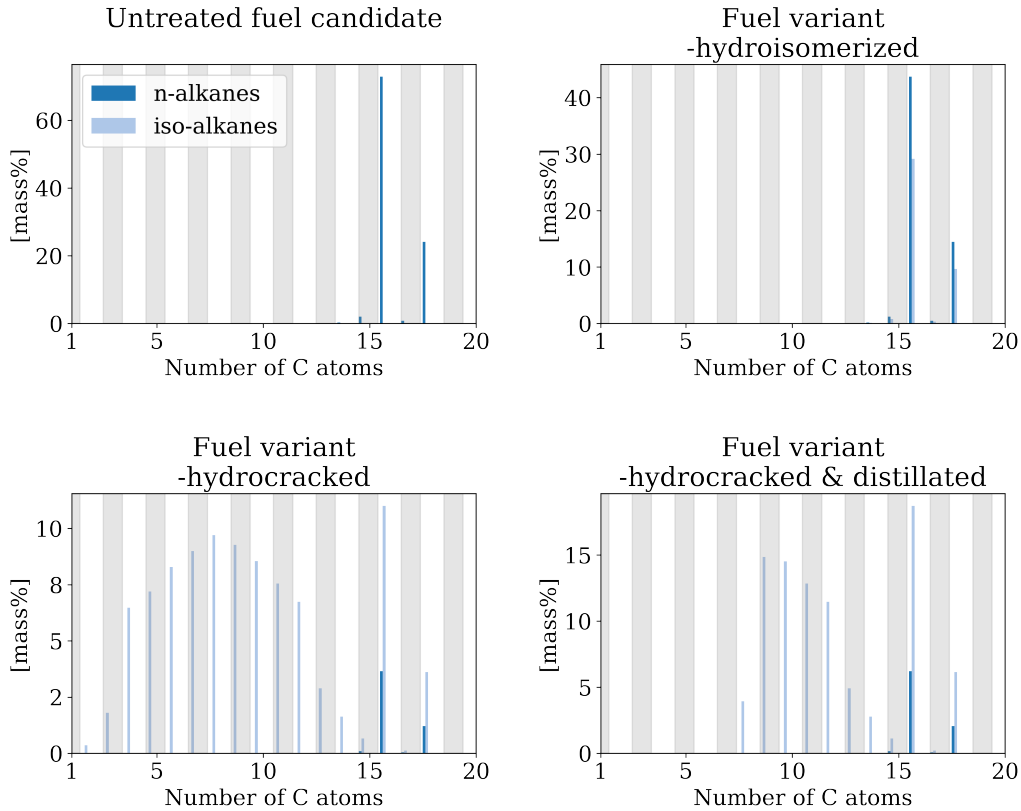


Figure 6.4: Composition plots of the untreated fuel candidate (upper left), hydroisomerized variant (upper right), hydrocracked fuel variant (lower left) and hydrocracked and distilled variant (lower right).

The composition plots in Figure 6.4 illustrate the effects of the process operations. The hydroisomerized variant on the upper right shows 40 % iso-alkanes. The effect of the hydrocracking is visible by the high fractions of iso- to n-alkanes and the bell-curve distribution of iso-alkanes with the carbon numbers below 16 on the lower left of Figure 6.4. The hydrocracking process also produces a significant fraction of components containing less than seven carbon atoms. As illustrated in Figure 5.31 in Chapter 5, these components have low boiling points, which lie outside of the typical distillation range of jet fuels. To reduce the fraction of low boiling components, an additional distillation is applied on the hydrocracked variant with a distillation temperature of 120 °C. To simulate the distillation, the QSPR sampling distillation model is utilized, distilling and thereby removing all fuel components with boiling points up to the defined temperature. For the distillation, instantaneous evaporation of the fuel components is assumed, when the respective boiling temperature of the component is reached. To consider all possible and unidentified isomers in the fuel, the distillation is repeated for 100 iterations, randomly sampling isomers at the different carbon numbers. The final composition of the fuel is computed by averaging the compositions of the distilled fuels of all iterations. A plot of the composition of the hydrocracked and distilled variant is given in the lower right of Figure 6.4. Compared to the composition of the hydrocracked variant, a large fraction of iso-alkanes containing seven carbon atoms or less is removed as an effect of the distillation. As byproduct,

the distillation produces 33.2 mass% of a light boiling fraction. This light boiling fraction is not further considered in this work, but could potentially be processed further and utilized for gasoline production or used in a blending operation.

Prescreening of fuel candidates

To review if the fuel design improves the fuel properties and therefore the chances of the variants for the approval, the Tier α prescreening is repeated for the three variants, using the same models as for the untreated fuel candidate. Figure 6.5 on the next page shows the prediction results for the variants, with individual rows for each fuel variant. The predictions for the untreated fuel variant of the QSPR sampling model are indicated as black triangles on the bottom of the respective property plots, as they were identified as trusted predictions in the previous screening.

Similar to the screening of the untreated fuel candidate, the predictions of the individual models are first checked for their plausibility for each fuel based on the property measurements from Chapter 5 and the comparison of the model predictions. The predictions are checked individually for each variant, starting with the hydroisomerized variant, then the hydrocracked variant and finally the hydrocracked and distilled variant.

For the hydroisomerized variant in the first column on the left, significant deviations of the predictions of the DC model colored in blue from those of the other models are visible. For the net heat of combustion, flash point, freezing point, cetane number and distillation temperatures at 10 (T_{10}) and 50 vol% (T_{50}), the values are predicted significantly lower than the M-QSPR and QSPR sampling models. Looking at the diagrams in Section 5.3, which illustrate the range of values of the outlined properties for n- and iso-alkanes with 15 and 18 carbon atoms, the property values predicted for the fuel by the DC model are found to be too low and physically implausible. These predictions are therefore not used for the assessment of this fuel variant and are again marked with a red box.

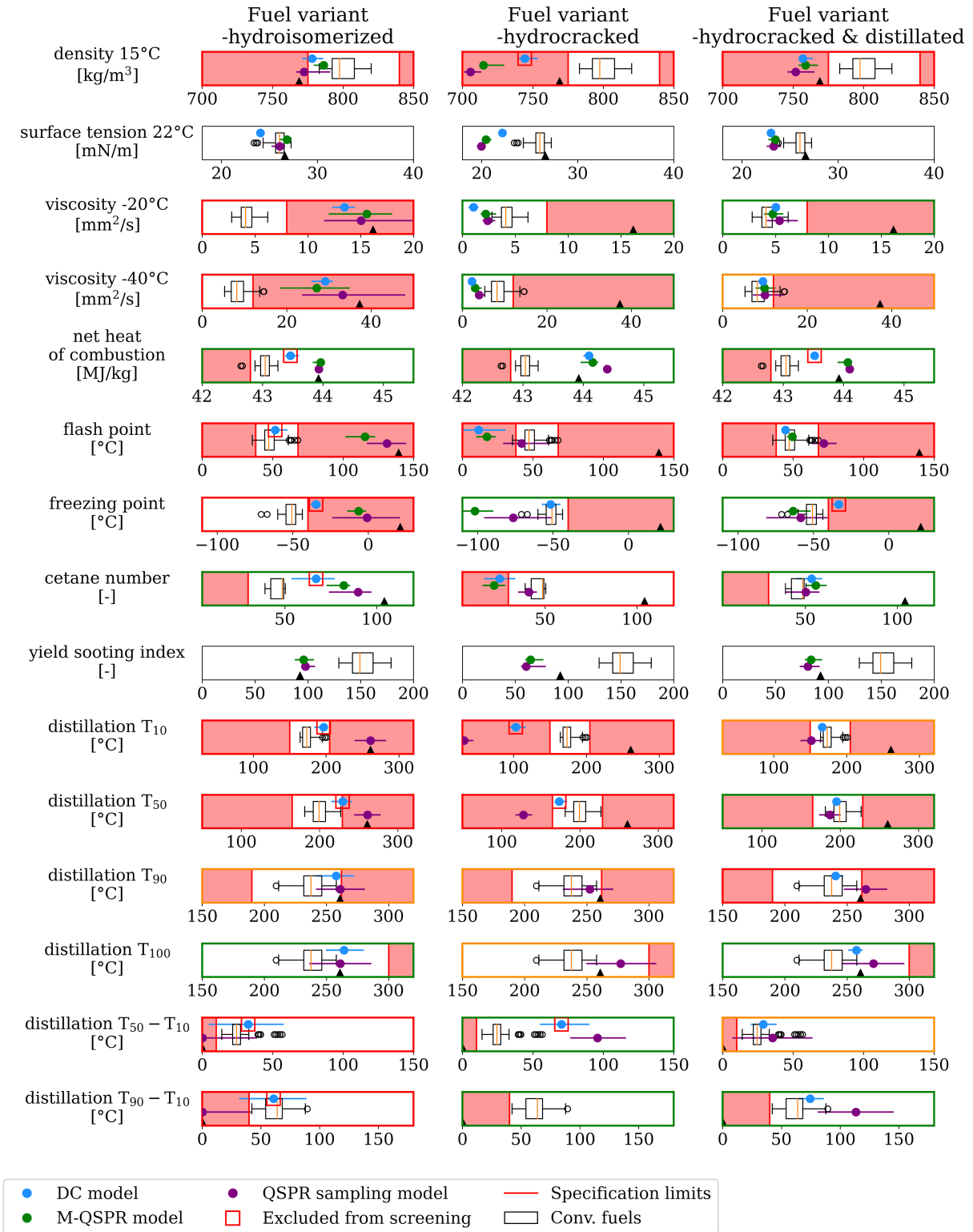


Figure 6.5: Results of the jet Tier α jet fuel prescreening for the three fuel variants: hydroisomerized (left), hydrocracked (middle) hydrocracked and distilled (right).

Considering the trusted predictions of the QSPR sampling model and M-QSPR model, reductions in the property values of the hydroisomerized variant are visible in comparison to the ones of the untreated fuel candidate, indicated as black triangles. The flash point is reduced from almost

150 °C to 100 °C, the freezing point from 25 °C close to 0 °C and the kinematic viscosities are reduced by around 25 %. For the distillation temperatures on the other hand, no significant changes are observed. Furthermore, the kinematic viscosity predictions of all models converge and now lie significantly closer together. A sign of higher agreement between the different modeling methods. Generally, the larger uncertainties of the predictions often overshadow the reduction. In summary, the predicted flash point, freezing point, kinematic viscosity and distillation line of the hydroisomerized variant are reduced, but still exceed the specification limits. The cetane number is significantly lower, while the remaining properties stay almost unchanged. These observations correspond to the observations in the Section 5.3, where the flash point, freezing point and kinematic viscosity are strongly affected by the chemical family.

Since pure hydroisomerization proved to be insufficient to improve the outlined properties, the second variant investigates the influence of components in the fuel produced by hydrocracking. Compared to the predictions of the hydroisomerized variant, the predictions of all models generally converge and lie closer together. This indicates, that the hydrocracked variant converges to known fuels utilized for the training of the DC models. Only the predictions of the DC model for the density and the distillation temperatures differ significantly from the predictions of the other models. Since the fuel composition with the broad distribution of possible components makes the identification of trustworthy results on the basis of pure compound measurements difficult, trusted results are now identified on the basis of the agreement of the models. Due to the high agreement between the predictions of the QSPR sampling and M-QSPR model, the density prediction of DC model is excluded. The distillation prediction of the DC model is furthermore excluded since the fuel variant contains over 20 mass% of iso-alkanes with a carbon number of 7 or lower, that have a boiling point below 100 °C as shown in Section 5.3.8. A T_{10} around 100 °C is therefore implausible.

Compared to the hydroisomerized variant, the flash point, freezing point, kinematic viscosity distillation values and cetane number are predicted significantly lower by all three models. For the freezing point and the kinematic viscosity, the predicted values now lie inside the specification limits. The predictions of the flash point and distillation line on the other hand either cross the specification limits with their prediction intervals or even completely fall below them. The most striking change in the predicted values is observed for the predicted T_{10} values of the QSPR sampling model. The predicted value significantly falls below the lower specification limit, along with the predicted density of all models. The reduction of the predicted properties can be explained by the presence of small fuel components with low carbon numbers as a result of the hydrocracking. Due to the low distillation temperatures, the variant produced solely by hydrocracking is inadequate and an additional process operation is necessary, to improve distillation properties and increase the flash point.

In order to improve flash point and distillation behavior of the fuel, a third variant is produced by hydrocracking and an additional distillation at 120 °C. The predicted properties for this

hydrocracked and distilled variant are shown in column three in Figure 6.5. The predictions of all models further converge. A significant disagreement of the model predictions is only observed for the DC model predictions for the net heat of combustion and freezing point. Similarly to the previous fuel variant, the fuel composition makes the identification of trustworthy results on the basis of pure compound measurements difficult. Trusted results are again identified on the basis of the agreement of the models. Due to the high agreement between the predictions of the QSPR sampling and M-QSPR model for the net heat of combustion and the freezing point, the predictions of the DC model are excluded from the assessment. For the other properties, the high agreement between the model predictions indicates an even greater similarity of the hydrocracked and distilled variants to the fuels used for the training and validation of the models. As seen in the validation of the different models on the hold-out fuels in Section 4.2, this is a strong indicator for correct predictions.

Compared to the predictions of the hydrotreated variant in the second column, the density, kinematic viscosity, flash point, cetane number and distillation line are all significantly increased by the distillation operation. The increase is especially visible for the distillation T_{10} which is increased by over 100 K. This clearly illustrates the effect of reducing the fraction of light-boiling components. With the exception of the density, flash point and the T_{90} , all predicted values lie either inside the specifications or cross them only with the uncertainty intervals. For the density, the predicted values still fall below the specification value, but are considerably higher compared to the non-distilled variant predictions in the third column. For the flash point, the predictions of the QSPR sampling model and the M-QSPR model disagree, without overlap. The M-QSPR and DC models predict the flash point inside the specification limits, while the prediction of the QSPR sampling model slightly exceeds the upper limit. The discordance of the predictions for the flash point is settled by preferring M-QSPR model on the basis of the validation metrics in Section 4.1.5 and the agreement with DC model predictions. If the variant would actually be formulated this decision could be verified. In comparison to the other three fuel variants, the hydrocracked and distilled variant shows the highest potential for a fuel candidate with respect to the considered specification limits. Further optimization of the fuel variant to meet the specification limits could be achieved by reformulating the fuel under other process parameters. Assuming that the hydrocracked and distilled variant would be approved as synthetic blending component, the low density and the high distillation temperature at T_{90} could be adjusted via blending. This option is investigated as part of this work in a blending study in the next section.

In summary, the simulated process operations result in a significant improvement in the prediction of the critical fuel properties. The kinematic viscosity, flash point, freezing point and distillation line, which exceeded the specification limits for the untreated fuel candidate, are reduced by the hydroisomerization and hydrocracking processes. The hydroisomerization reduces the exceeding properties, however they are still predicted outside the upper specification limits. Only hydrocracking, that breaks down the fuel components into smaller ones brings kinematic

viscosity and freezing point inside the specification range. However, due to the formation of too many light-boiling components in the cracking process, the lower limits of the flash point and the distillation line are not met. It requires an additional distillation operation to again increase distillation temperatures as well as flash point and bring predicted properties inside the specification limits. The distillation however comes with a separation of a light-boiling top fraction and therefore with a reduction of the desired product yield, which has to be considered in a potential economical assessment.

6.3 Blending study of fuel variant

The preceding fuel prescreening showed that the fuel variant processed by additional hydrocracking and distillation has the best chance of approval. However, the predicted density, the flash point and the distillation at 90 vol% still exceed the prescreening specification limits. To investigate whether this fuel variant has the potential to be used as a synthetic blending component for conventional oil-based jet fuels and whether the specification limits for jet fuel blends can be met by blending, a blending study is being conducted. The potential reduction in CO₂ emissions and sooting propensity of the blends will thereby be investigated to review the potential climate benefits of using this fuel variant.

This is a hypothetical scenario that assumes that the screened fuel is admissible, meaning that the production path using an additional hydrocracking and distillation would be approved and registered as an annex in ASTM D7566. At the time of writing, SAFs are allowed in blends with conventional jet fuels with fractions of up to 50 vol% after ASTM D7566. Similar to the screening, the blending study is conducted under consideration of specification limits, which ensure the safety and operability of critical fuel properties. ASTM D7566 defines specification limits for blends of conventional and synthetic fuels, which slightly differ from the limits used for the prescreening. Table 28 lists the utilized property limits for jet fuel blends after ASTM D7566. In contrast to the prescreening, ASTM D7566 lists a limit for the aromatic content of the blend.

The blending analysis is formulated as mathematical optimization, where the volumetric fraction of the synthetic fuel is maximized while the properties of the blend have to comply with the set specification limits as boundary conditions. Furthermore, the reduction of the sooting tendency and CO₂ emissions of the blends are calculated, relative to the conventional jet fuel used for blending. The consideration of these additional climate figures of merit should illustrate the benefit of utilizing the fuel variant as SAF. To quantify the sooting tendency, again the yield sooting index (YSI), introduced in Section 5.3.9 is used. The reduction in YSI and CO₂ are calculated on the basis of the molar fraction of synthetic fuel in the blend relative the unblended conventional fuel. For both the blend and the conventional equal fuel consumption is assumed. For the calculation of the CO₂ reduction, the synthetic fuel is assumed to be produced completely CO₂ neutral.

Property	Unit	Dependency	Min	Max
Density	kg/m ³	15 °C	775	840
Kinematic viscosity	mm ² /s	-20 °C		8
		-40 °C		12
Net heat of combustion	MJ/kg		42.8	
Flash point	°C			68
Freezing point	°C			-40
Aromatic content	vol%-		8	25
Distillation line	°C	10 vol%		205
		100 vol%		300
		$T_{50} - T_{10}$	10	
		$T_{90} - T_{10}$	40	

Table 28: Specification properties for jet fuels blends after ASTM D7566.

The blending study is conducted twice; first considering only the mean value of the property prediction and second considering also the predicted uncertainties of the properties with 95 % prediction intervals. In the first case, only the mean predicted property values have to comply with the specification limits. In the second case, also the predicted upper or lower bounds of the prediction intervals have to completely lie inside the specification limits. The first case is a deterministic optimization, where potential uncertainties in the predicted properties are neglected. The second case illustrates a risk informed case where 95 % of the possible value ranges, have to comply with the specification range. Similar to prescreening, a non-compliance with the specification limits may result in unacceptable consequences for the operability and safety of an aircraft. In both cases, the goal is the estimation of the maximum blendable volume fractions as well as the identification of the properties, that limit the blending fraction. The comparison of both cases will illustrate the impact of a risk-informed assessment compared to the deterministic one and outline potential risks, that would have been neglected.

For the blending study, jet fuels from the CRC world fuel survey of 2006 [92] are utilized to reflect the variance of conventional oil-based jet fuels. These conventional fuels are assumed to correctly represent the compositional variance in the jet fuel market and serve as reference for the potential reduction in sooting tendency and CO₂.

Since the CRC fuel survey also holds property measurements for the conventional fuels for all properties, except the density at 15 °C, the cetane number and the yield sooting index, the blending is conducted on the basis of the available measurements. Property values for the density at 15 °C, the cetane number and the yield sooting index of the conventional fuels are predicted. Since the blending analysis requires the use of only one model, all properties for the fuel variant and the missing properties of the conventional fuels are predicted using the M-QSPR model. The

M-QSPR model showed robust and trustworthy predictions in previous screening of the fuel variant and a high predictive capability in the testing of the models in Section 4.1.

The properties of the blends are calculated using mixing rules, with an exception for the distillation line. The distillation line of the blend cannot be calculated by a mixing rule, since it is a phase transition process, that requires the consideration of the respective fuel composition. The distillation line of the blends is therefore calculated using the QSPR sampling model. All properties of the blends, except the kinematic viscosity, are calculated with the same mixing rules, used for the QSPR sampling model in Section 2.2.1. Density, surface tension and net heat of combustion are calculated using a linear mixing rule weighted by the mass fraction. Flash point and freezing point are calculated using a linear mixing rule weighted by the volume fraction. Cetane number and yield sooting index are calculated using a linear mixing rule weighted by the molar fraction. For the kinematic viscosity, the mixing rule recommended in ASTM D7152 [145] is utilized, see Equation (6.1) to Equation (6.3). This mixing rule is the standard practice for the calculation of viscosity of blends and allows the estimation of the kinematic viscosity of blended fuels on the basis of their volume fractions. The kinematic viscosity of the blend ν_{mix} is calculated using a blending factor A_{mix} . The A_{mix} is the weighted sum of the individual blending factors of the fuels A_i , which are estimated from the viscosities of the individual fuels ν_i .

$$A_i = 14.534 * \ln(\ln(\nu_i) + 0.8) + 10.975 \quad (6.1)$$

$$A_{mix} = \sum_i w_i * A_i \quad (6.2)$$

$$\nu_{mix} = \exp\left(\exp\left(\frac{A_{mix} - 10.975}{14.534}\right)\right) - 0.8 \quad (6.3)$$

Since the blending analysis is conducted as an optimization problem, the SLSQP optimizer [146] from the scipy Python library is utilized [147]. The maximum number of possible iterations per blend is set to 40, which proved to be sufficient. If no blend could be formulated in the maximum number of iterations, the blend is marked as unsuccessful with a maximum volume fraction of 0 vol%. The specification limits are implemented as boundary conditions, with which the blend has to comply to be considered successful.

6.3.1 Case 1: blending study without uncertainty consideration

The first blending analysis represents the deterministic optimization, where the predicted uncertainties of the properties are neglected. The results of the blending analysis are shown in the plots of Figure 6.6. The left side of Figure 6.6 shows the estimated property ranges of the blends in blue, spanned between the minimum and maximum property values of the blends, as well as the mean property value of the blends in a darker blue. The mean prediction of the unblended fuel variant is shown as green line, without prediction uncertainties. As an additional reference, box plots illustrating the property range of the conventional fuels used for blending are shown. The ASTM 7566 specification limits and areas outside of these specifications are

indicated as red lines and red areas respectively. For the blending study the parallel line plot allows an easier comparison of the value range of multiple properties to the specification limits, compared to the screening plot in the previous section. The right side of Figure 6.6 shows histograms with the maximum successfully blended fraction of the fuel variant (top), as well as the reduction of the yield sooting index (middle) and the CO₂ emissions (bottom), relative to the unblended conventional fuel.

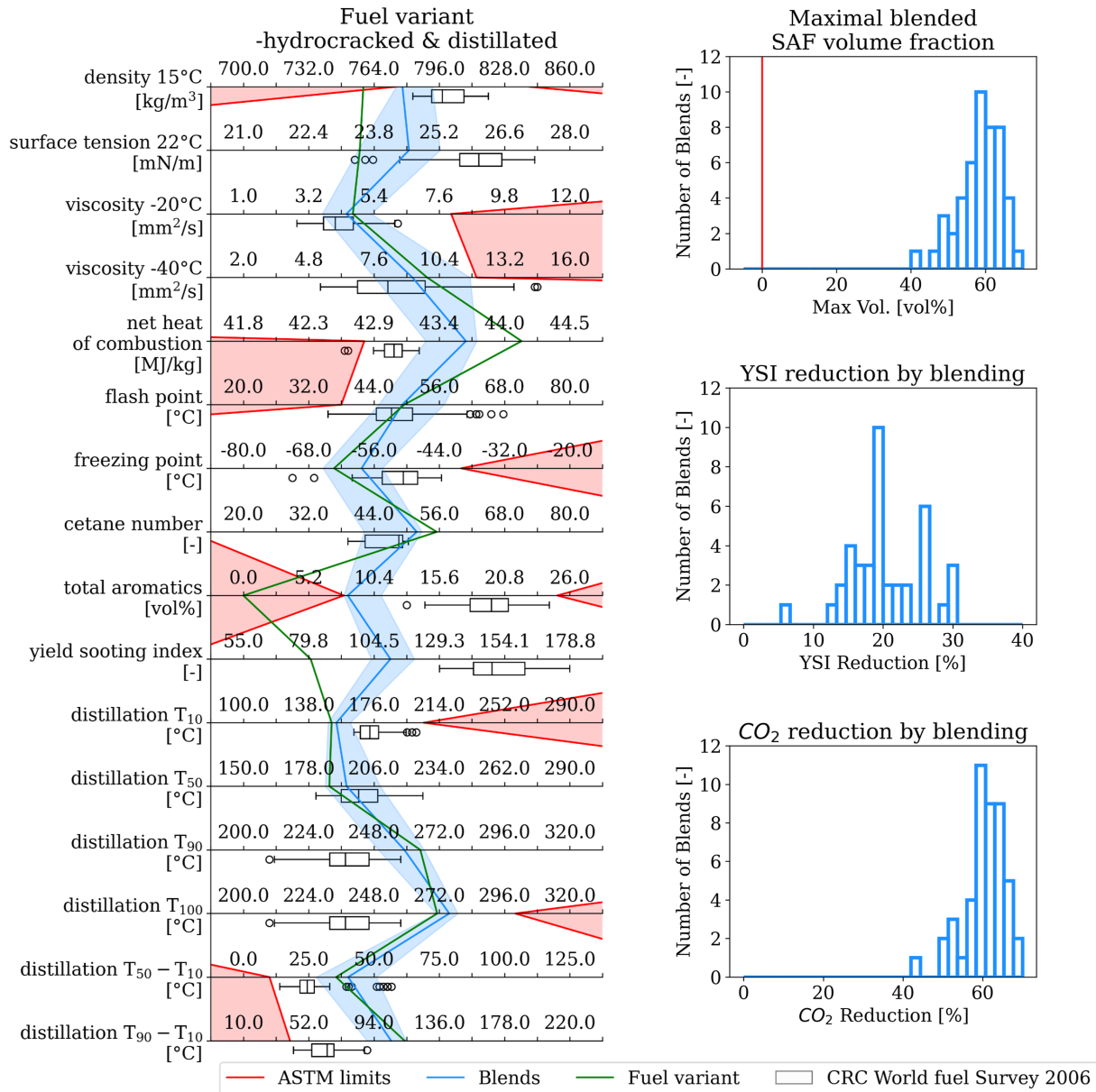


Figure 6.6: Results of deterministic blending analysis, left: property plot with property ranges of fuel variant and blends, right: maximum blending fraction (top) of fuel variant, reduction of yield sooting index of blends (middle), CO₂ reduction of blends (bottom).

The histogram for the maximum blending ratio of the fuel variant shows that successful blends are achieved with all considered conventional fuels. Estimated maximum blend fractions range from 40 to 70 vol% with an average of about 60 vol%, if blend percentages above the 50 vol%

specified in ASTM D7566 would be allowed. The blending also strongly affects the considered ecological aspects, sooting tendency and CO₂ emissions. The estimated yield sooting index of the blends is 15 to almost 40 % lower compared to the unblended conventional fuels and the estimated CO₂ reductions lie between 40 to 70 %, see Figure 6.6.

The estimated property range of the blends, shown in the parallel line property plot on the left of Figure 6.6, differs in part from the value range of the conventional fuels. The density, surface tension, freezing point, yield sooting index, aromatic content and distillation T_{10} are on average lower compared to the values of the conventional fuels. Net heat of combustion, the distillation temperatures T_{90} and T_{100} and therefore the differences of the distillation temperatures T_{50-10} and T_{90-10} are higher, but still inside the specification limits. The other properties are comparable. The blending would therefore improve aspects like the low temperature behavior indicated by the freezing point, the sooting tendency and the energy of combustion of the fuel.

Properties that limit further increases in volume fractions can be identified by the proximity of the property value range of the blends to the specification limits. For the density, kinematic viscosity at -40 °C and aromatic content, the value range of the blends lies close to the respective limits. To further increase the SAF fraction, those properties would therefore have to be increased by reformulating the fuel variant with other process conditions.

In summary, according to this blending study the utilization of the fuel variant would be unproblematic and highly ecologically beneficial. Relying on these results, the fuel variant could be blended with every conventional fuel with the minimum save fraction of ratios of up to 40 vol%, without the apprehension of critical implications. Knowledge about the exact property values and composition of the conventional fuel would thereby not be required. Whether this is also true when the prediction uncertainties are taken into account in the optimization, will be investigated in the next section.

6.3.2 Case 2: blending study with uncertainty consideration

In the second case, the blending analysis is repeated, this time considering the predicted uncertainties of the models in each iteration of the optimization. Thus, the predicted property is no longer a deterministic value, but a range of values spanning between the upper and lower prediction intervals. This second analysis reflects a risk-informed use case where the stakeholder requires 95% confidence that predicted values are within specification limits. Likewise to the first blending analysis, the results are displayed in plots in Figure 6.7. The estimated value ranges of the properties of the blends are again shown in the parallel line plot on the left. This time, the 95 % prediction intervals of the fuel variant properties are also indicated as green areas around the mean prediction. The histograms for the maximum blending fraction, the reduction in the yield sooting index and the CO₂ emissions relative to the unblended conventional fuel are shown again on the right of Figure 6.7.

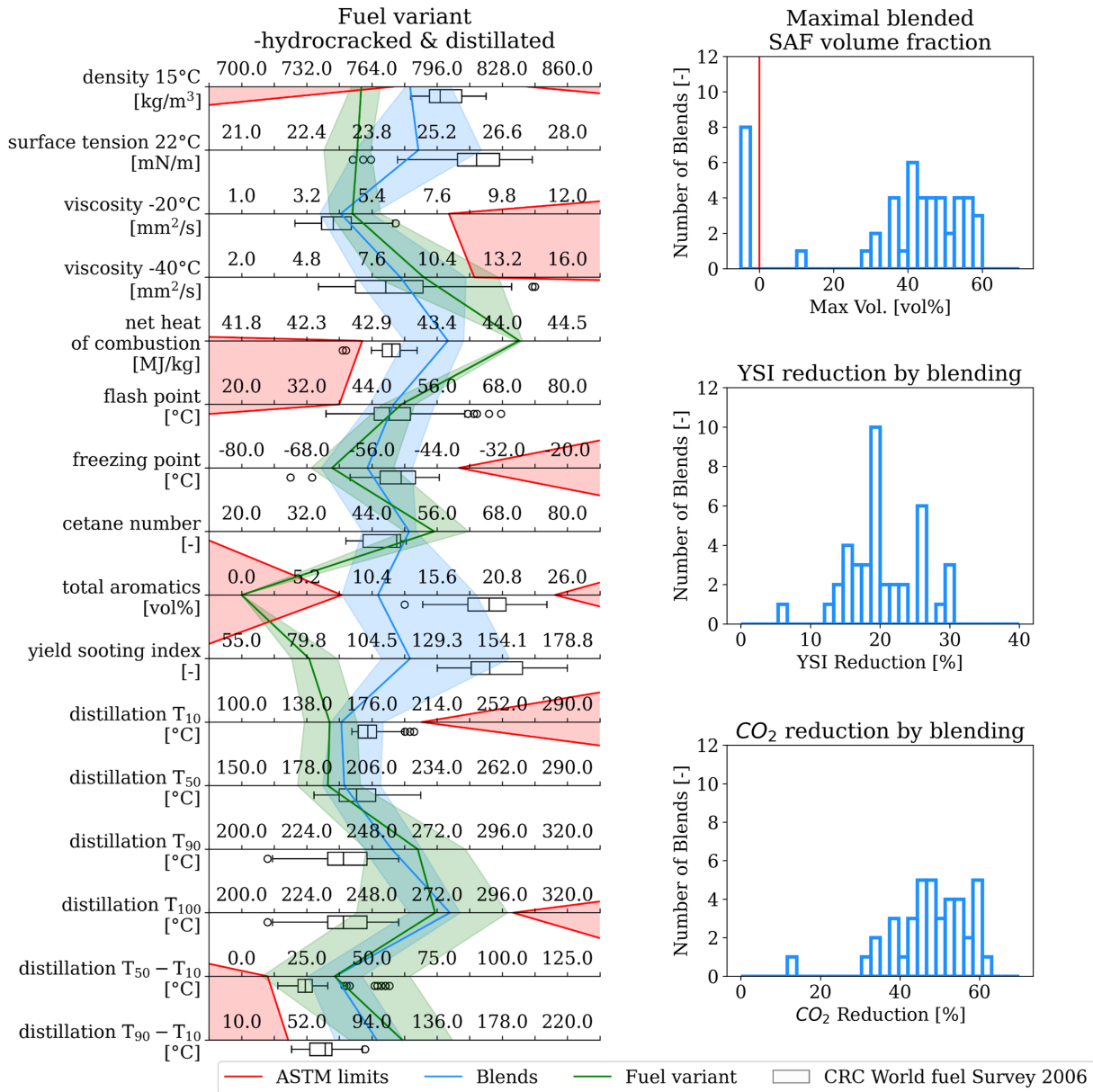


Figure 6.7: Results of blending analysis under consideration of uncertainties, left: parallel line property plot with property ranges of fuel variant and blends, right: maximum blending fraction (top) of fuel variant, reduction of yield sooting index of blends (middle), CO₂ reduction of blends (bottom).

The histogram for the maximum blended fraction on the top right of Figure 6.7 illustrates, that the result under consideration of prediction uncertainties differs significantly from the previous deterministic case. This time, eight conventional fuels could not be successfully blended with the fuel variant, as indicated by the bar behind the red 0 vol% line in the histogram. For these conventional fuels, all blends exceed the specification limits by at least one property. This is mainly due to the fact that the properties of the conventional fuels themselves do not meet the ASTM D7566 blending requirements. On the left of Figure 6.7 this is visible for the kinematic viscosity at -40 °C for the conventional fuels, where the boxplot exceeds the specification limit.

For the conventional fuels which could be blended, the volume fraction of the fuel variant is reduced by almost 20 %, with a minimum of 12 to 60 % and an average of around 40 %. This also influences the ecological aspects. The average reduction of the yield sooting index and the CO₂ are reduced by around 15 % and now lies at 20 % for the yield sooting index and 45 % for the CO₂ reduction.

In the parallel line property plot on the left of Figure 6.7 the value range of the blends is on average broader compared to the previous deterministic case. The properties of the successful blends therefore vary more. The broader range of the property values can be explained by the uncertainties considered now for the fuel variant predictions and the wider distribution of blending fractions compared to the previous case. Due to the larger fractions of conventional fuels in the blends and the uncertainty of the fuel variant property predictions, the resulting range of values for the blends is closer to the range of values for the conventional fuels. This is visible for the freezing point, net heat of combustion and yield sooting index, where the value range of the blends in large parts overlaps the boxplots of the conventional fuels. Compared to the previous deterministic case, the improvements for the low temperature behavior, the energy content and the sooting tendency of the blends would therefore be smaller.

To identify the property specifications, that limit the blending fraction of the synthetic fuels, exceeding property values are counted during the blending and displayed in Figure 6.8.

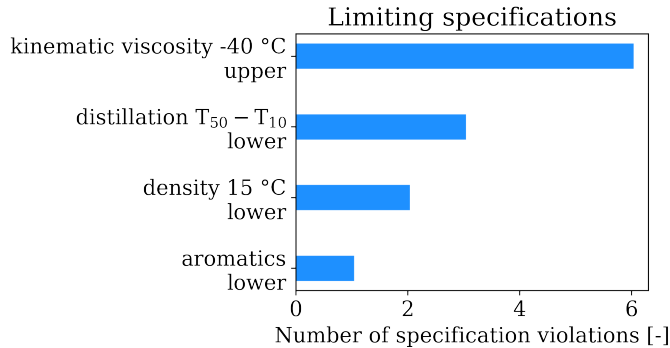


Figure 6.8: Number of properties preventing blends for the blending study under consideration of uncertainties.

In Figure 6.8 the upper kinematic viscosity limit at -40 °C followed by the lower limit of the distillation difference $T_{50} - T_{10}$ and density at 15 °C are identified as limiting properties. Since the synthetic fuel variant does not contain any aromatics, and only successful blends were observed in the deterministic blending study, the lower aromatic limit cannot be a limiting factor. It must have occurred together with another exceeding property.

In summary, under consideration of uncertainties the fuel variant cannot be blended with every conventional fuel without exact knowledge of its composition and properties. If the fuel variant would be blindly blended in ratios estimated in the previous blending study, the viscosity and distillation behavior of the blends with some conventional fuels could exceed the specification.

This poses a potential safety risk. To achieve successful blends, the property values of the conventional fuels must be known. Overall, even under considerations of the uncertainties the possible blends have improved ecological behavior with reduced sooting behavior and CO₂ emissions of 20 % and 45 % on average, depending on the fuel.

To increase the blendable fraction of the synthetic fuel variant based on this study, the fuel variant would either need to be modified or the uncertainties in the properties predictions reduced. The former could be achieved by redesigning the fuel under other process conditions to e.g. reduce the kinematic viscosity. To reduce uncertainty in properties predictions such as viscosity and distillation, the fuel variant would have to be formulated and the properties be measured, or more information about the unidentified isomers would be needed as input for the modeling.

6.4 Summary and conclusion

The design of jet fuels is the second major application of the tools developed in this work. It not only evaluates the suitability of a jet fuel as a potential candidate, but also suggests options for composition and property optimization. Fuel design requires the utilization of both the developed models and the design tools from the composition-property relation investigation.

In this chapter, a fuel design was conducted for a real fuel candidate. In a first step, the fuel candidate was screened using the model-based Tier α prescreening procedure to assess the chances for approval and determine properties that exceed the specification limits. In the second step, the composition of the base fuel was optimized to meet the specification limits, according to the findings of the composition-property relations. As exceeding properties in the prescreening, the flash point, freezing point, kinematic viscosity and distillation line were identified. To reduce the values of the identified properties, simulated refinery operations were applied to convert n-alkanes to iso-alkanes and generally reduce the size of fuel components. As refinery processes hydroisomerization, hydrocracking and distillation are used to produce three respective fuel variants. The three fuel variants were then screened to assess their chances for fuel approval. Of the three variants, the hydrocracked and distilled variant was found to have the highest chances, with mean predicted property values inside all specifications except for the density and the final boiling point of the distillation line.

To investigate the potential of the most promising fuel variant as a synthetic blending component, a blending study was conducted using conventional oil-based fuels. It was thereby investigated if the fuel variant could be blended with a range of representative conventional fuels under consideration of the specification for jet fuel blends in ASTM D7566. The blending study was performed twice, once without the consideration of uncertainties in the predicted properties and once considering these uncertainties. The two cases compared a risk-uninformed case, where

the influence of potential uncertainties on the property values are neglected and a risk-informed case where the uncertainties and the associated risks are taken into account.

According to the uninformed blending study, the fuel variant could be mixed with conventional fuels in ratios of up to 40 vol% without the necessary knowledge of the composition and properties of the conventional fuels and without any potential associated risk. The blends were thereby estimated to be highly ecologically beneficial, with reductions in the sooting tendency of 15 to 30 % and CO₂ 40 to 70 % relative to the unblended conventional fuels. The second blending study however revealed, that a safety risk might exist due to potentially high viscosity values at -40 °C. As a consequence, the properties of the conventional fuels have to be known to formulate blends that comply with the specification limits. The blendable volume fraction of the fuel variant is predicted to be only 20 vol% on average, which also lowers the reductions of the sooting tendency and the CO₂ emissions by 15 % on average. To increase the blendable fraction of the fuel variant under consideration of uncertainties, the fuel would either have to be redesigned or the uncertainties of the property values would have to be reduced. The first could be achieved by improving the properties of the fuel variant even further and coupling the utilized process operations with the developed models in a co-optimization to adapt both the fuel and the fuel production process. A greater holistic optimization that incorporates the developed models and the process simulation in a technoeconomic optimization of both fuel and production process would therefore be the next step. The reduction of the uncertainty in the property values would require the actual production of the fuel and the measurement of its properties or a reduction of the uncertainties by implicitly identifying or constraining possible isomers in the fuel composition. In a future work, the blending study could therefore be repeated if analytical methods or knowledge about the production process allow the identification or selection of isomers.

7 Summary and Outlook

7.1 Summary

To achieve the ambitious goals of climate neutrality in the civil aviation sector, expert panels such as the International Civil Aviation Organization (ICAO) have deemed an increased utilization of sustainable aviation fuels (SAF) as indispensable. However, current SAF production capacities will be insufficient to meet the growing demand, necessitating further research of novel SAF production routes and the development of concepts that streamline the fuel development process, from upscaling to the approval. Fast and inexpensive concepts that enable the assessment and design of new SAF candidates based on minimal fuel volumes are considered key enablers by the scientific community and industry. Such concepts, like the prescreening approach introduced by Heyne and Rauch, involve modern compositional analytics combined with predictive methods to assess critical jet fuel properties based on minimal volume samples of the jet fuel candidates. However, the successful implementation of this concept hinges on the availability of highly accurate and reliable predictive models, as well as fuel design tools, to provide feedback on fuel optimization to the producers.

This work investigated the question whether and how data-based models can be used to support the outlined processes of fuel screening and design. Both models for the prediction of eight critical jet fuel properties as well as tools for fuel design were developed and applied to relevant use cases. To reduce the need for extensive measurement campaigns, the tools were built on a database, which was systematically built from various DLR internal and external sources. This allowed the focus to be placed on the development of new Machine Learning based property models. Further limitations of the developed tools were investigated to demonstrate the potential and the need for further research.

For the prediction of the critical properties, three different modeling methods were developed that model jet fuels in fundamentally different ways: the direct correlation method (DC), the Quantitative Structure-Property Relationship Modeling method with sampling (QSPR sampling), and the Mean Quantitative Structure-Property Relationship method (M-QSPR). All three modeling methods are probabilistic, meaning they do not predict one deterministic value for the property but a distribution of possible values. Using probabilistic models enables estimating uncertainties arising from factors like unidentified isomers, measurement noise, and differences between training and test data. Accounting for these uncertainties is indispensable, as the outlined factors can critically impact the possible value range. In order to account for these uncertainties, the Monte-Carlo Dropout Neural Network algorithm, a probabilistic Machine Learning regression algorithm, was utilized for all three models.

For the development of the property models, a comprehensive database was built that holds data on over 1870 pure fuel components, 75 conventional and 56 synthetic jet fuels and blends with

over 63 000 measurements for eight critical properties. The collected data was unified and stored in a database using a standardized schema, which allows the sustainable use of the database beyond the scope of this work.

The developed probabilistic models were validated for the prediction of all eight critical properties on 75 conventional crude oil-based jet fuels and 56 synthetic fuels and blends. The predictive results were assessed using specially developed metrics and compared with deterministic results from the literature. The predictive results of the deterministic models were outperformed in terms of both accuracy and the additionally estimated uncertainty intervals due to the influence of isomers. The estimated uncertainty intervals were found to effectively illustrate the possible value range of the property, which is information that deterministic models cannot provide. The developed models accurately and precisely predicted fuel properties when the composition matched the underlying assumption of a broad isomer distribution. However, if the fuel composition deviated from this default, greater uncertainties and even deviations could arise, especially for properties with high variance across isomers.

The adequacy of the developed models for jet fuel screening was demonstrated through a simulated preliminary prescreening of three fuels not included in the test set: one conventional and two synthetic fuels. The results of the model-based screening were consistent with the results of the model testing. Adequate results were directly achieved for the conventional fuel and the synthetic fuel produced via the Fischer-Tropsch route, due to their broad distribution of isomers in their composition. However, for the synthetic fuel from the Alcohol-to-Jet route, composed almost solely of a few distinct components, the number of isomers needed to be constrained to predict adequate results. This was achieved by directly constraining the isomeric selection for the QSPR sampling models, as the isomeric composition of the fuel was largely known. The use of multiple competing models to compare and cross-check predictions proved particularly helpful in identifying trustworthy predictions. If predictions from multiple different models were in high accordance with each other, the predicted values were observed to be close to the true measurement. On the other hand, if predictions disagreed significantly and had large associated uncertainties, significant errors could occur, and additional property measurements are recommended.

To optimize the fuel composition for the jet fuel approval, design tools were developed on the basis of the pure compound database. The influence of the chemical family, size and topology of fuel components on the critical jet fuel properties was investigated for 1870 fuel components from eight hydrocarbon families. Relations were assessed both visually as well as quantitatively by metrics and structural descriptors. The quantified metrics of the correlations were summarized in tables that, along with the detailed plots, served as tools for the following fuel design.

To demonstrate the capability of the developed tools for a combined workflow of fuel screening and design, a jet fuel candidate was screened and optimized to increase its chances of passing the fuel approval process. The fuel candidate was thereby screened in a first step using the model-based prescreening procedure and optimized in the second step, creating fuel variants according to the developed design tools. The variants were thereby formulated using the simulated process operations of established refinery operations to ensure that compositional optimization could actually be carried out in a real production facility. From the three produced variants, one complied with all but two critical properties, which could potentially be resolved by blending the fuel variant with conventional fuel. This was investigated in a subsequent blending study, where the most promising fuel variant was mixed with a representative selection of conventional jet fuels. The blends were estimated based on the maximum blended volume fraction, to consider the economic aspect of the fuel producer as well as the potential reduction of CO₂ and soot emissions of the blend. The blending analysis was carried out both with and without the consideration of uncertainties, to illustrate the influence of the predicted uncertainties for a risk-informed use case.

In summary, all set objectives were fulfilled, and the stated research question of whether data-based methods can support the development process of sustainable aviation fuels can be answered with a resounding "yes". The data-based tools developed in this work are capable of supporting both the screening and the design of fuels. The predictions of the models can substitute costly and fuel-intensive measurements, reducing the need for extensive testing while illustrating the uncertainties in the predictions and estimating potential risks as part of their output. The design tools illustrate the relationship between fuel composition and properties, allowing for the optimization of fuel candidates to increase their chances of approval. Combining both the models and design tools enables the exploration of different fuel candidate options to find a fuel candidate with high chances of approval. This integrated approach can accelerate the fuel development process and reduce the risk of costly iterations or dropouts during the fuel approval process.

Nevertheless, the developed tools have limitations that must be considered and further investigated in future research. For the property models, unidentified isomers in GCxGC composition measurements can lead to high uncertainties and errors in the predictions, potentially preventing the use of the predictions due to the associated risk. This limitation becomes especially prominent for fuels with distinct compositions likely dominated by only a few isomers. To increase the predictive capability for such cases, more detailed information is needed, either by identifying individual fuel components or by constraining the range of possible isomers. Another strong limitation is the availability of fuel and pure compound data for the training of the property models and the development of the fuel design tools. Smaller datasets can increase the chances of erroneous predictions and uncertainties of the property models. For the design tools, missing measurement data prevented the analysis of structural aspects for some

hydrocarbon families during the design tool investigations. This was especially observed for the branching of fuel components, where investigations were not possible for some hydrocarbon families. Hence, more data is needed for both fuels and fuel components, especially for properties with a currently low number of available measurements, e.g., flash point and cetane number. The developed tools have been built with this in mind and can be easily updated and even extended if additional data becomes available.

7.2 Outlook

As the concept of screening and design for the jet fuel approval is itself a fairly recent one, there exist multiple possible next steps to further improve the developed tools and continue the started research. The next goal should be to overcome the identified limitations and apply the developed tools to new use cases.

Further increase the predictive capability of fuel property models

For the fuel property models, two major limitations were identified that constrain their predictive capability: 1) High uncertainty due to unidentified isomers and 2) Small training datasets with low variability for properties like the flash point. To overcome the first limitation, more detailed compositional information about the fuels is necessary, which could be achieved by further developing the analytical methods, e.g., combining GCxGC with vacuum ultraviolet spectroscopy to identify individual isomers [148,149]. Apart from advancing analytics, a more detailed consideration of the respective fuel production path might be useful for the modeling, as the selection of possible isomers could be narrowed down based on the underlying reaction mechanism. In both cases, the developed modeling methods have the flexibility to process more detailed information, with the QSPR sampling methods considering only the identified isomers and the M-QSPR creating mean occurrence maps with constrained isomer selections. The second limitation can be overcome by gathering more data and repeating the model training and validation, especially with the addition of fuels with new distinctly different compositions to further improve the predictive capability of the developed models. Future research could also consider the implementation of new algorithms for the correlation of the input representation and the fuel property, such as sequential structure modeling [150] or graph neural networks [122, 123, 151-154].

Further investigation of component structure-property relations

This work identified and marked gaps of missing measurement data, that prevented correlations of structural aspects of fuel components of some hydrocarbon families with the considered physical properties. The identified gaps should be closed by systematic measurement campaigns and further gathering of data. The enlarged database can be the basis for future investigations of the influence of structural aspects, like size and branching of components and allow the development of more accurate and extensive fuel design tools.

Consideration of additional properties for Jet fuel screening design

The next step for the jet fuel screening and design could be an application of the developed models to other parts of fuel assessment, as well as the modeling of additional properties. This could include the modeling of additional physicochemical properties relevant from later parts of the fuel assessment, such as the specific heat capacity or density over a set temperature range described, for example, in Tier 2 of ASTM D4054. The transferability of the modeling methods was successfully demonstrated for the modeling of the sooting tendency on the basis of the yield sooting index, in Chapter 6. Furthermore, additional metrics characterizing environmental aspects, such as the demonstrated yield sooting index, as well as metrics describing the behavior in engine or spray tests, could be simulated and considered. For the jet fuel screening, this would allow a more comprehensive assessment of a fuel candidate, reducing the necessary cost, time and fuel volume even further. For the fuel design, holistic optimization loops for the evaluation of the production routes of fuel candidates are possible. In the Chapter 6 this was outlined for the evaluation of fuel variants produced by different refinery process operations based on screening results and sooting behavior. This concept could be extended to also include economic aspects of the fuel production, allowing the comprehensive evaluation of a new jet fuel candidate from its production to the chances of its approval.

Holistic optimization and technoeconomic assessment of new SAF production paths

As outlined in summary of Chapter 6, the developed tools could be leveraged to holistically assess the production pathways for prospective sustainable aviation fuel (SAF) candidates. In collaboration with SAF producers, these pathways could be optimized by exploring the influence of various process operations, such as hydro-isomerization and hydrocracking, and fine-tuning process parameters to yield SAF formulations with high approval prospects. Beyond process design, a techno-economic assessment of the production process should be conducted to evaluate the feasibility of fuel manufacturing. Fuel design concepts for low-compound gasoline fuels already incorporate production feasibility into their optimization frameworks, providing a model that could be emulated for SAF development [122,125,155,156]. This integrated approach, combining process optimization, economic analysis, and approval considerations, would enable the creation of viable, sustainable aviation fuels aligned with industry standards and market demands.

References

- [1] Fetting C. The European Green Deal 2020:17.
- [2] United States of America Government. Memorandum of Understanding Sustainable Aviation Fuel Grand Challenge. [August 17, 2022]; Available from: https://www.energy.gov/sites/default/files/2021-09/S1-Signed-SAF-MOU-9-08-21_0.pdf.
- [3] Scheelhaase J, Maertens S, Grimme W. Synthetic fuels in aviation – Current barriers and potential political measures. *Transportation Research Procedia* 2019;43:21–30. <https://doi.org/10.1016/j.trpro.2019.12.015>.
- [4] Pörtner H-O, Roberts DC, Tignor M, Poloczanska ES, Mintenbeck K, Alegría A et al. *Climate Change 2022: Impacts, Adaptation, and Vulnerability: Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change 2022*.
- [5] International Civil Aviation Organisation. ICAO SAF facilities map. [November 10, 2022]; Available from: https://datastudio.google.com/reporting/2532150c-ff4c-4659-9cf3-9e1ea457b8a3/page/p_2sq3qol5nc?s=.
- [6] International Civil Aviation Organisation. Report on the feasibility of long-term aspirational goal (LTAG) for international civil aviation CO2 emission reduction 2022.
- [7] Lee DS, Fahey DW, Skowron A, Allen MR, Burkhardt U, Chen Q et al. The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018. *Atmos Environ* (1994) 2021;244:117834. <https://doi.org/10.1016/j.atmosenv.2020.117834>.
- [8] Voigt C, Kleine J, Sauer D, Moore RH, Bräuer T, Le Clercq P et al. Cleaner burning aviation fuels can reduce contrail cloudiness. *Communications Earth & Environment* 2021;2(1). <https://doi.org/10.1038/s43247-021-00174-y>.
- [9] Faber J, Király J, Lee D, Owen B, O’Leary a. Potential for reducing aviation non-CO2 emissions through cleaner jet fuel 2021.
- [10] White D. Feasibility Study on the use of Sustainable Aviation Fuels 2018.
- [11] European Commission. Fit for 55': Delivering the EU' S 2030 Climate Target on the Way to Climate Neutrality. Brussels: European Commission; 2021.
- [12] Commercial Aviation Alternative Fuels Initiative. CAAFI Fuel Qualification. [June 29, 2022]; Available from: https://www.caafi.org/focus_areas/fuel_qualification.html.
- [13] Shell Aviation. Clearing the runway for sustainable aviation fuels. [August 18, 2022]; Available from: <https://www.airlines.iata.org/reports/sponsored-clearing-the-runway-for-sustainable-aviation-fuels>.
- [14] Neste. Neste MY Sustainable Aviation Fuel. [August 18, 2022]; Available from: <https://www.neste.com/products/all-products/saf#722b64b5>.
- [15] World Energy. World Energy CEO Shares Net-Zero Carbon Strategy During Virtual White House Roundtable on Sustainable Aviation Fuel 2021, 10 September 2021; Available from: <https://www.prnewswire.com/news-releases/world-energy-ceo-shares-net-zero-carbon-strategy-during-virtual-white-house-roundtable-on-sustainable-aviation-fuel-301373076.html>. [August 18, 2022].

-
- [16] International Air Transport Association. Net zero 2050: sustainable aviation fuels 2022.
- [17] World Economic Forum. Guidelines for a Sustainable Aviation Fuel Blending Mandate in Europe: Insight Report 2021.
- [18] O'Malley J, Pavlenko N, Searle S. Estimating sustainable aviation fuel feedstock availability to meet growing European Union demand 2021:1.
- [19] European Union. DIRECTIVE (EU) 2018/ 2001 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - of 11 December 2018: - on the promotion of the use of energy from renewable sources 2018.
- [20] D02 Committee. Practice for Evaluation of New Aviation Turbine Fuels and Fuel Additives. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D4054-21A>.
- [21] Hileman, James. Fuel Approval Process & Status 2019.
- [22] Colket M, Heyne J, Rumizen M, Gupta M, Edwards T, Roquemore WM et al. Overview of the National Jet Fuels Combustion Program. *AIAA Journal* 2017;55(4):1087–104. <https://doi.org/10.2514/1.J055361>.
- [23] B. Rauch et al. JET fuel SCREENing and optimization (JETSCREEN) project: European Union's Horizon 2020 research and innovation program under grant agreement No 723525. [March 03, 2022]; Available from: <https://cordis.europa.eu/project/id/723525>, checked on 3/3/2022.
- [24] Heyne J, Rauch B, Le Clercq P, Colket M. Sustainable aviation fuel prescreening tools and procedures. *Fuel* 2021;290:120004. <https://doi.org/10.1016/j.fuel.2020.120004>.
- [25] D02 Committee. Specification for Aviation Turbine Fuel Containing Synthesized Hydrocarbons. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D7566-21>.
- [26] D02 Committee. Specification for Aviation Turbine Fuels. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D1655-21C>.
- [27] Edwards JT. Jet Fuel Properties 2020.
- [28] Shell. Civil Aviation Fuel Jet Fuel Specifications Shell Global. [August 18, 2022]; Available from: <https://www.shell.com/business-customers/aviation/aviation-fuel/civil-jet-fuel-grades.html>.
- [29] Klerk A de. Fischer-Tropsch Refining. Wiley; 2011.
- [30] Rauch B. Systematic accuracy assessment for alternative aviation fuel evaporation models. Universität Stuttgart; 2018.
- [31] Rumizen M. ASTM D4054 Users Guide 2018.
- [32] Rumizen MA. Qualification of Alternative Jet Fuels. *Front. Energy Res.* 2021;9. <https://doi.org/10.3389/fenrg.2021.760713>.
- [33] Moses C. Airbreathing Propulsion Fuels and Energy Exploratory Research and Development: Review of Materials Compatibility Tests of Synthesized Hydrocarbon Kerosenes And Blends 2017.
- [34] Alves Fortunato M (ed.). Jetscreen Program Quantitative Assessment of the Jet Fuel Physical, Chemical and Thermophysical Properties and development of low and High Fidelity Models; 2019.

- [35] D02 Committee. Test Method for Hydrocarbon Types in Middle Distillates by Mass Spectrometry. West Conshohocken, PA: ASTM International.
<https://doi.org/10.1520/D2425-21>.
- [36] D02 Committee. Test Method for Determination of Aromatic Hydrocarbon Types in Aviation Fuels and Petroleum Distillates High Performance Liquid Chromatography Method with Refractive Index Detection. West Conshohocken, PA: ASTM International.
<https://doi.org/10.1520/D6379-21E01>.
- [37] Vozka P, Modereger BA, Park AC, Zhang WTJ, Trice RW, Kenttämaa HI et al. Jet fuel density via GC × GC-FID. *Fuel* 2019;235:1052–60.
<https://doi.org/10.1016/J.FUEL.2018.08.110>.
- [38] Coburn A, Yang Z, Boehm R, Heyne JS. Determination of a Freeze Point Blend Prediction Model for Jet Fuel Range Hydrocarbons. In: AIAA SCITECH 2022 Forum. Reston, Virginia: American Institute of Aeronautics and Astronautics; 2022.
- [39] Mondello L, Tranchida PQ, Dugo P, Dugo G. Comprehensive two-dimensional gas chromatography-mass spectrometry: a review. *Mass Spectrom Rev* 2008;27(2):101–24.
<https://doi.org/10.1002/mas.20158>.
- [40] Shi X, Li H, Song Z, Zhang X, Liu G. Quantitative composition-property relationship of aviation hydrocarbon fuel based on comprehensive two-dimensional gas chromatography with mass spectrometry and flame ionization detector. *Fuel* 2017;200:395–406.
<https://doi.org/10.1016/J.FUEL.2017.03.073>.
- [41] Kehimkar B, Hoggard JC, Marney LC, Billingsley MC, Fraga CG, Bruno TJ et al. Correlation of rocket propulsion fuel properties with chemical composition using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry followed by partial least squares regression analysis. *J Chromatogr A* 2014;1327:132–40.
<https://doi.org/10.1016/j.chroma.2013.12.060>.
- [42] Le Clercq P (ed.). Emission and Climate Impact of Alternative Fuels ECLIF - I: First Measurement Campaign; 2017.
- [43] Vozka P, Kilaz G. A review of aviation turbine fuel chemical composition-property relations. *Fuel* 2020;268:117391. <https://doi.org/10.1016/j.fuel.2020.117391>.
- [44] Le Clercq P, Doué N, Rachner M, Aigner M. Validation of a Multicomponent-Fuel Model for Spray Computations. In: 47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition. Reston, Virginia: American Institute of Aeronautics and Astronautics; 2009.
- [45] Harstad K, Le Clercq P, Bellan J. A Robust Statistical Model for the Evaporation of Multicomponent-Fuel Drops. In: 41st Aerospace Sciences Meeting and Exhibit. Reston, Virginia: American Institute of Aeronautics and Astronautics; 2003.
- [46] Ruoff S, Stöhr M, Rauch B, Eckel G, Le Clercq P, Meier W et al. Numerical Simulation and Uncertainty Quantification of a Generic Droplet Evaporation Validation Test Case. *Atomization and Sprays* 2020;30(12):861–79.
<https://doi.org/10.1615/AtomizSpr.2020035295>.
- [47] Hall C, Rauch B, Bauder U, Le Clercq P, Aigner M. Application of Machine Learning algorithms for the prediction of fuel properties in comparison with state-of-the-art physical

- models. Proceedings of the 16th International Conference on Stability, Handling and Use of Liquid Fuels 2019 2019.
- [48] Yang Z, Kosir S, Stachler R, Shafer L, Anderson C, Heyne JS. A GC \times GC Tier α combustor operability prescreening method for sustainable aviation fuel candidates. *Fuel* 2021;292:120345. <https://doi.org/10.1016/j.fuel.2021.120345>.
- [49] Bell D, Heyne JS, Won SH, Dryer F, Haas FM, Dooley S. On the Development of General Surrogate Composition Calculations for Chemical and Physical Properties. In: 55th AIAA Aerospace Sciences Meeting.
- [50] Cookson DJ, Latten JL, Shaw IM, Smith BE. Property-composition relationships for diesel and kerosene fuels. *Fuel* 1985;64(4):509–19. [https://doi.org/10.1016/0016-2361\(85\)90086-9](https://doi.org/10.1016/0016-2361(85)90086-9).
- [51] Cookson DJ, Lloyd CP, Smith BE. Investigation of the chemical basis of kerosene (jet fuel) specification properties. *Energy Fuels* 1987;1(5):438–47. <https://doi.org/10.1021/ef00005a011>.
- [52] Cookson DJ, Smith BE. Calculation of jet and diesel fuel properties using carbon-13 NMR spectroscopy. *Energy Fuels* 1990;4(2):152–6. <https://doi.org/10.1021/ef00020a004>.
- [53] Cookson DJ, Iliopoulos P, Smith BE. Composition-property relations for jet and diesel fuels of variable boiling range. *Fuel* 1995;74(1):70–8. [https://doi.org/10.1016/0016-2361\(94\)P4333-W](https://doi.org/10.1016/0016-2361(94)P4333-W).
- [54] Cookson DJ, Smith BE. Observed and predicted properties of jet and diesel fuels formulated from coal liquefaction and Fischer-Tropsch feedstocks. *Energy Fuels* 1992;6(5):581–5. <https://doi.org/10.1021/ef00035a007>.
- [55] Morris, Robert E. Johnson, K., Hammond, Mark H., Rose-Persson, S. The Development of Advanced Sensor Technologies to Measure Critical Navy Mobility Fuel Properties. Naval Research Laboratory 2006.
- [56] Morris RE, Hammond MH, Cramer JA, Johnson KJ, Giordano BC, Kramer KE et al. Rapid Fuel Quality Surveillance through Chemometric Modeling of Near-Infrared Spectra. *Energy Fuels* 2009;23(3):1610–8. <https://doi.org/10.1021/ef800869t>.
- [57] Oberkampf WL, Roy CJ. Verification and validation in scientific computing. Cambridge: Cambridge Univ. Press; 2010.
- [58] Murphy KP. Machine learning: A probabilistic perspective; 2012.
- [59] Mogk G, Mrziglod T, Schuppert A. Application of Hybrid Models in Chemical Industry. In: European Symposium on Computer Aided Process Engineering-12, 35th European Symposium of the Working Party on Computer Aided Process Engineering. Elsevier; 2002, p. 931–936.
- [60] European Symposium on Computer Aided Process Engineering-12, 35th European Symposium of the Working Party on Computer Aided Process Engineering. Elsevier; 2002.
- [61] Yang S, Navarathna P, Ghosh S, Bequette BW. Hybrid Modeling in the Era of Smart Manufacturing. *Computers & Chemical Engineering* 2020;140:106874. <https://doi.org/10.1016/j.compchemeng.2020.106874>.
- [62] Schweidtmann AM, Esche E, Fischer A, Kloft M, Repke J-U, Sager S et al. Machine Learning in Chemical Engineering: A Perspective. *Chemie Ingenieur Technik* 2021;93(12):2029–39. <https://doi.org/10.1002/cite.202100083>.

-
- [63] Kersten P, Wagner J, Tipler PA, Mosca G. Mechanik von Massepunkten. In: Tipler PA, Mosca G, Kersten P, Wagner J, editors. Physik. Berlin, Heidelberg: Springer Berlin Heidelberg; 2019, p. 29–82.
- [64] Kersten P, Wagner J, Tipler PA, Mosca G. Die kinetische Gastheorie. In: Tipler PA, Mosca G, Kersten P, Wagner J, editors. Physik. Berlin, Heidelberg: Springer Berlin Heidelberg; 2019, p. 555–577.
- [65] Stephan P, Kabelac S, Kind M, Mewes D, Schaber K, Wetzel T (eds.). VDI-Wärmeatlas. Berlin, Heidelberg: Springer Berlin Heidelberg; 2019.
- [66] Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. Beijing, Boston, Farnham, Sebastopol, Tokyo: O'Reilly; 2019.
- [67] Jobac KG, REID RC. Estimation of Pure-Component Properties from Group-Contributions. *Chemical Engineering Communications* 1987;57(1-6):233–43. <https://doi.org/10.1080/00986448708960487>.
- [68] Fredenslund A, Jones RL, Prausnitz JM. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J.* 1975;21(6):1086–99. <https://doi.org/10.1002/aic.690210607>.
- [69] National Institute of Standards and Technology. NIST Standard Reference Database 103a. [January 17, 2022]; Available from: <https://www.nist.gov/mml/acmd/trc/thermodata-engine/srd-nist-tde-103a>.
- [70] Design Institute for Physical Properties. DIPPR 801 Database. [January 18, 2022]; Available from: <https://www.aiche.org/dippr/events-products/801-database>.
- [71] Saldana DA, Starck L, Mougin P, Rousseau B, Pidol L, Jeuland N et al. Flash Point and Cetane Number Predictions for Fuel Compounds Using Quantitative Structure Property Relationship (QSPR) Methods. *Energy Fuels* 2011;25(9):3900–8. <https://doi.org/10.1021/ef200795j>.
- [72] Saldana DA, Starck L, Mougin P, Rousseau B, Ferrando N, Creton B. Prediction of Density and Viscosity of Biofuel Compounds Using Machine Learning Methods. *Energy Fuels* 2012;26(4):2416–26. <https://doi.org/10.1021/ef3001339>.
- [73] Saldana DA, Starck L, Mougin P, Rousseau B, Creton B. On the rational formulation of alternative fuels: melting point and net heat of combustion predictions for fuel compounds using machine learning methods. *SAR QSAR Environ Res* 2013;24(4):259–77. <https://doi.org/10.1080/1062936X.2013.766634>.
- [74] Daylight Theory: SMILES. [March 04, 2022]; Available from: <https://www.daylight.com/>.
- [75] RDKit. [March 04, 2022]; Available from: <https://www.rdkit.org/>.
- [76] Hall C, Creton B, Rauch B, Bauder U, Aigner M. Probabilistic Mean Quantitative Structure–Property Relationship Modeling of Jet Fuel Properties. *Energy Fuels* 2022;36(1):463–79. <https://doi.org/10.1021/acs.energyfuels.1c03334>.
- [77] Flora G, Kosir ST, Behnke L, Stachler RD, Heyne JS, Zabarnick S et al. Properties Calculator and Optimization for Drop-in Alternative Jet Fuel Blends. In: AIAA Scitech 2019 Forum.

-
- [78] Grunberg L., Nissan A. H. Mixture law for viscosity GRUNBERG, L; NISSAN, A. H. *Nature* 1949;164(4175):799. <https://doi.org/10.1038/164799b0>.
- [79] D02 Committee. Practice for Viscosity-Temperature Equations and Charts for Liquid Petroleum or Hydrocarbon Products. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D0341-20E01>.
- [80] Ford DC, Miller WH, Thren RC, Wertzler R. Correlation of ASTM Method D 2887-73 Boiling Range Distribution Data with ASTM Method D 86-67 Distillation Data. In: Green LE, Albert DK, editors. Correlations of ASTM D 2887-73 with Physical Properties of Petroleum Fractions. 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959: ASTM International; 1975, 20-20-11.
- [81] Ajmani S, Rogers SC, Barley MH, Livingstone DJ. Application of QSPR to mixtures. *J Chem Inf Model* 2006;46(5):2043–55. <https://doi.org/10.1021/ci050559o>.
- [82] Gal Y, Ghahramani Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning; 2015.
- [83] Buduma N, Locascio N. Fundamentals of deep learning: Designing next-generation machine intelligence algorithms. Beijing, Boston, Farnham, Sebastopol, Tokyo: O’Reilly; 2017.
- [84] Mitchell TM. Machine learning. New York, NY: McGraw-Hill; 20]10.
- [85] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization; 2014.
- [86] PyTorch. [March 09, 2022]; Available from: <https://pytorch.org/>.
- [87] scikit-learn. 3.1. Cross-validation: evaluating estimator performance. [July 22, 2022]; Available from: https://scikit-learn.org/stable/modules/cross_validation.html.
- [88] scikit-optimize: sequential model-based optimization in Python — scikit-optimize 0.8.1 documentation. [October 26, 2021]; Available from: <https://scikit-optimize.github.io/stable/>.
- [89] Hall C, Rauch B, Bauder U, Le Clercq P, Aigner M. Predictive Capability Assessment of Probabilistic Machine Learning Models for Density Prediction of Conventional and Synthetic Jet Fuels. *Energy Fuels* 2021;35(3):2520–30. <https://doi.org/10.1021/acs.energyfuels.0c03779>.
- [90] Pang J, Liu D, Peng Y, Peng X. Optimize the Coverage Probability of Prediction Interval for Anomaly Detection of Sensor-Based Monitoring Series. *Sensors (Basel)* 2018;18(4). <https://doi.org/10.3390/s18040967>.
- [91] Pearce T, Zaki M, Brintrup A, Neely A. High-Quality Prediction Intervals for Deep Learning: A Distribution-Free, Ensembled Approach 2018. <https://doi.org/10.48550/arXiv.1802.07167>.
- [92] Coordinating Research Council. World Fuel Sampling Program Final Report with CRC 2006.
- [93] PubChem. PubChem. [March 04, 2022]; Available from: <https://pubchem.ncbi.nlm.nih.gov/>.
- [94] ChemSpider. [March 04, 2022]; Available from: <https://www.chemspider.com/>.
- [95] J. Yanowitz, M.A. Ratcliff, R.L. McCormick, and J.D. Taylor, M.J. Murphy. Compendium of Experimental Cetane Numbers 2017.

- [96] National Renewable Energy Laboratory. YSI Estimator. [April 06, 2022]; Available from: <https://ysi.ml.nrel.gov/>.
- [97] Li R, Herreros JM, Tsolakis A, Yang W. Machine learning regression based group contribution method for cetane and octane numbers prediction of pure fuel compounds and mixtures. *Fuel* 2020;280:118589. <https://doi.org/10.1016/j.fuel.2020.118589>.
- [98] Blakey S, Rauch B, Oldani A, Lee T. Advanced Fuel Property Data Platform: Overview and Potential Applications. *Front. Energy Res.* 2022;10. <https://doi.org/10.3389/fenrg.2022.771325>.
- [99] Handbook of Pharmaceutical Analysis by HPLC. Elsevier; 2005.
- [100] National Institute of Standards and Technology. 1.3.5.17. Detection of Outliers. [July 26, 2022]; Available from: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm>.
- [101] Blakey S, Novelli P, Costes P, Bringtown S, Christensen D, Sakintuna B et al. State of the Art on Alternative Fuels in Aviation. SWAFEA. Sustainable Way for Alternative Fuels and Energy in Aviation. Trends in Plant Science - TRENDS PLANT SCI 2010.
- [102] molgen group. MOLGEN. [March 10, 2022]; Available from: <https://www.molgen.de/>.
- [103] Henze HR, Blair CM. The Number of Isomeric Hydrocarbons of the Methane Series. *J. Am. Chem. Soc.* 1931;53(8):3077–85. <https://doi.org/10.1021/ja01359a034>.
- [104] van der Maaten L, Hinton G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 2008;9:2579–605.
- [105] Liu G, Wang L, Qu H, Shen H, Zhang X, Zhang S et al. Artificial neural network approaches on composition–property relationships of jet fuels based on GC–MS. *Fuel* 2007;86(16):2551–9. <https://doi.org/10.1016/j.fuel.2007.02.023>.
- [106] Thom MA. CRC Project No. AV-23-15/17: Review of Existing Test Methods Used for Aviation Jet Fuel and Additive Property Evaluations with Respect to Alternative Fuel Compositions. CRC; 2018.
- [107] Wang X, Jia T, Pan L, Liu Q, Fang Y, Zou J-J et al. Review on the Relationship Between Liquid Aerospace Fuel Composition and Their Physicochemical Properties. *Trans. Tianjin Univ.* 2021;27(2):87–109. <https://doi.org/10.1007/s12209-020-00273-5>.
- [108] Mehrotra AK. A generalized viscosity equation for pure heavy hydrocarbons. *Ind. Eng. Chem. Res.* 1991;30(2):420–7. <https://doi.org/10.1021/ie00050a021>.
- [109] D02 Committee. Test Method for Distillation of Petroleum Products and Liquid Fuels at Atmospheric Pressure. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D0086-20B>.
- [110] Ding Y, Wang S, Hanson RK. Sensitive and interference-immune formaldehyde diagnostic for high-temperature reacting gases using two-color laser absorption near 5.6 μm . *Combustion and Flame* 2020;213:194–201. <https://doi.org/10.1016/j.combustflame.2019.11.042>.
- [111] Braun-Unkhoff M, Kathrotia T, Rauch B, Riedel U. About the interaction between composition and performance of alternative jet fuels. *CEAS Aeronaut J* 2016;7(1):83–94. <https://doi.org/10.1007/s13272-015-0178-8>.

-
- [112] Kosir S, Stachler R, Heyne J, Hauck F. High-performance jet fuel optimization and uncertainty analysis. *Fuel* 2020;281:118718. <https://doi.org/10.1016/j.fuel.2020.118718>.
- [113] Kosir ST, Behnke L, Heyne JS, Stachler RD, Flora G, Zabarnick S et al. Improvement in Jet Aircraft Operation with the Use of High-Performance Drop-in Fuels. In: *AIAA Scitech 2019 Forum*. Reston, Virginia: American Institute of Aeronautics and Astronautics; 01072019.
- [114] Muldoon JA, Harvey BG. Bio-Based Cycloalkanes: The Missing Link to High-Performance Sustainable Jet Fuels. *ChemSusChem* 2020;13(22):5777–807. <https://doi.org/10.1002/cssc.202001641>.
- [115] Harvey BG, Harrison KW, Davis MC, Chafin AP, Baca J, Merriman WW. Molecular Design and Characterization of High-Cetane Alkyl Diamondoid Fuels. *Energy Fuels* 2016;30(12):10171–8. <https://doi.org/10.1021/acs.energyfuels.6b01865>.
- [116] Lu X, Han D, Huang Z. Fuel design and management for the control of advanced compression-ignition combustion modes. *Progress in Energy and Combustion Science* 2011;37(6):741–83. <https://doi.org/10.1016/j.peccs.2011.03.003>.
- [117] Saldana DA, Creton B, Mougou P, Jeuland N, Rousseau B, Starck L. Erratum to: Rational Formulation of Alternative Fuels using QSPR Methods: Application to Jet Fuels. *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* 2014;69(3):499. <https://doi.org/10.2516/ogst/2014012>.
- [118] Creton B, Dartiguelongue C, Bruin T de, Toulhoat H. Prediction of the Cetane Number of Diesel Compounds Using the Quantitative Structure Property Relationship. *Energy Fuels* 2010;24(10):5396–403. <https://doi.org/10.1021/ef1008456>.
- [119] Cai G, Liu Z, Zhang L, Zhao S, Xu C. Quantitative Structure–Property Relationship Model for Hydrocarbon Liquid Viscosity Prediction. *Energy Fuels* 2018;32(3):3290–8. <https://doi.org/10.1021/acs.energyfuels.7b04075>.
- [120] Gani R, Zhang L. Editorial overview: Frontiers of Chemical Engineering: Chemical Product Design. *Current Opinion in Chemical Engineering* 2020;27:A1–A3. <https://doi.org/10.1016/j.coche.2020.03.005>.
- [121] Zhang L, Mao H, Liu Q, Gani R. Chemical product design – recent advances and perspectives. *Current Opinion in Chemical Engineering* 2020;27(11):22–34. <https://doi.org/10.1016/j.coche.2019.10.005>.
- [122] Dahmen M, Marquardt W. Model-Based Design of Tailor-Made Biofuels. *Energy Fuels* 2016;30(2):1109–34. <https://doi.org/10.1021/acs.energyfuels.5b02674>.
- [123] Dahmen M, Marquardt W. Model-Based Formulation of Biofuel Blends by Simultaneous Product and Pathway Design. *Energy Fuels* 2017;31(4):4096–121. <https://doi.org/10.1021/acs.energyfuels.7b00118>.
- [124] Burkardt P, Ottenwalder T, Konig A, Viell J, Mitsos A, Wouters C et al. Toward co-optimization of renewable fuel blend production and combustion in ultra-high efficiency SI engines. *International Journal of Engine Research* 2023;24(1):29–41. <https://doi.org/10.1177/14680874211040995>.
- [125] Konig A, Siska M, Schweidtmann AM, Rittig JG, Viell J, Mitsos A et al. Designing production-optimal alternative fuels for conventional, flexible-fuel, and ultra-high

- efficiency engines. *Chemical Engineering Science* 2021;237:116562.
<https://doi.org/10.1016/j.ces.2021.116562>.
- [126] Ackermann P, Braun KE, Burkardt P, Heger S, König A, Morsch P et al. Designed to Be Green, Economic, and Efficient: A Ketone-Ester-Alcohol-Alkane Blend for Future Spark-Ignition Engines. *ChemSusChem* 2021;14(23):5254–64.
<https://doi.org/10.1002/cssc.202101704>.
- [127] Fleitmann L, Ackermann P, Schilling J, Kleinekorte J, Rittig JG, Vom Lehn F et al. Molecular Design of Fuels for Maximum Spark-Ignition Engine Efficiency by Combining Predictive Thermodynamics and Machine Learning. *Energy Fuels* 2023;37(3):2213–29.
<https://doi.org/10.1021/acs.energyfuels.2c03296>.
- [128] McEnally C, Pfefferle L. Improved sooting tendency measurements for aromatic hydrocarbons and their implications for naphthalene formation pathways. *Combustion and Flame* 2007;148(4):210–22. <https://doi.org/10.1016/j.combustflame.2006.11.003>.
- [129] Montgomery MJ, Das DD, McEnally C, Pfefferle L. Analyzing the robustness of the yield sooting index as a measure of sooting tendency. *Proceedings of the Combustion Institute* 2019;37(1):911–8. <https://doi.org/10.1016/j.proci.2018.06.105>.
- [130] Kunal and Ghosh. Introduction of Extended Topochemical Atom (ETA) Indices in the Valence Electron Mobile (VEM) Environment as Tools for QSAR/QSPR Studies 2003;2.
- [131] Todeschini R, Consonni V. *Molecular Descriptors for Chemoinformatics*. Wiley; 2009.
- [132] Roy K, Saha A. QSPR with TAU Indices: Water Solubility of Diverse Functional Acyclic Compounds. *Internet Electron. J. Mol. Des.* 2003;2:475–49.
- [133] Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 2011;32(7):1466–74. <https://doi.org/10.1002/jcc.21707>.
- [134] Travis Kessler. GitHub - ecrl/padelpy: A Python wrapper for PaDEL-Descriptor software. [April 20, 2022]; Available from: <https://github.com/ecrl/padelpy>.
- [135] Cai G-Q, Zhang L-Z. Systematic diesel molecular performance evaluation based on quantitative structure-property relationship model. *Petroleum Science* 2021.
<https://doi.org/10.1016/j.petsci.2021.10.022>.
- [136] Abdul Jameel AG. A functional group approach for predicting fuel properties. KAUST Research Repository; 2019.
- [137] European Union Aviation Safety Agency. Updated analysis of the non-CO2 climate impacts of aviation and potential policy measures pursuant to the EU Emissions Trading System Directive Article 30(4). European Commission; 2020.
- [138] D02 Committee. Test Method for Smoke Point of Kerosene and Aviation Turbine Fuel. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D1322-22>.
- [139] Das DD, St. John PC, McEnally CS, Kim S, Pfefferle LD. Measuring and predicting sooting tendencies of oxygenates, alkanes, alkenes, cycloalkanes, and aromatics on a unified scale. *Combustion and Flame* 2018;190:349–64.
<https://doi.org/10.1016/j.combustflame.2017.12.005>.
- [140] Xiang Z, Chen K, McEnally C, Pfefferle L. Sooting tendencies of diesel fuel component mixtures follow a linear mixing rule; 2022.

-
- [141] Wang W-C, Tao L. Bio-jet fuel conversion technologies. *Renewable and Sustainable Energy Reviews* 2016;53:801–22. <https://doi.org/10.1016/j.rser.2015.09.016>.
- [142] Weitkamp J. Isomerization of long-chain n-alkanes on a Pt/CaY zeolite catalyst. *Ind. Eng. Chem. Prod. Res. Dev.* 1982;21(4):550–8. <https://doi.org/10.1021/i300008a008>.
- [143] Weitkamp J. Catalytic Hydrocracking—Mechanisms and Versatility of the Process. *ChemCatChem* 2012;4(3):292–306. <https://doi.org/10.1002/cctc.201100315>.
- [144] Coonradt HL, Garwood WE. Mechanism of Hydrocracking. Reactions of Paraffins and Olefins. *Ind. Eng. Chem. Proc. Des. Dev.* 1964;3(1):38–45. <https://doi.org/10.1021/i260009a010>.
- [145] D02 Committee. Practice for Calculating Viscosity of a Blend of Petroleum Products. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D7152-11R16E01>.
- [146] Kraft DA. A software package for sequential quadratic programming. Tech. Rep. DFVLR-FB 1988:88-28.
- [147] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17(3):261–72. <https://doi.org/10.1038/s41592-019-0686-2>.
- [148] Feldhausen J, Bell DC, Yang Z, Faulhaber C, Boehm R, Heyne J. Synthetic aromatic kerosene property prediction improvements with isomer specific characterization via GCxGC and vacuum ultraviolet spectroscopy. *Fuel* 2022;326:125002. <https://doi.org/10.1016/j.fuel.2022.125002>.
- [149] Heyne J, Bell D, Feldhausen J, Yang Z, Boehm R. Towards fuel composition and properties from Two-dimensional gas chromatography with flame ionization and vacuum ultraviolet spectroscopy. *Fuel* 2022;312:122709. <https://doi.org/10.1016/j.fuel.2021.122709>.
- [150] Hall C, Rauch B, Bauder U, Le Clercq P, Aigner M. Probabilistic sequential Neural Networks for the modelling of jet fuel compound properties 2022;Proceedings of the 17th International Conference on Stability, Handling and Use of Liquid Fuels 2022.
- [151] Schweidtmann AM, Rittig JG, König A, Grohe M, Mitsos A, Dahmen M. Graph Neural Networks for Prediction of Fuel Ignition Quality. *Energy Fuels* 2020;34(9):11395–407. <https://doi.org/10.1021/acs.energyfuels.0c01533>.
- [152] Rittig JG, Gao Q, Dahmen M, Mitsos A, Schweidtmann AM. Graph Neural Networks for the Prediction of Molecular Structure–Property Relationships. In: Zhang D, del Río Chanona EA, editors. *Machine Learning and Hybrid Modelling for Reaction Engineering*. Royal Society of Chemistry; 2023, p. 159–181.
- [153] Rittig JG, Ritzert M, Schweidtmann AM, Winkler S, Weber JM, Morsch P et al. Graph machine learning for design of high-octane fuels. *AIChE J.* 2023;69(4):1993. <https://doi.org/10.1002/aic.17971>.
- [154] Schweidtmann AM, Rittig JG, Weber JM, Grohe M, Dahmen M, Leonhard K et al. Physical pooling functions in graph neural networks for molecular property prediction. *Computers & Chemical Engineering* 2023;172(15):108202. <https://doi.org/10.1016/j.compchemeng.2023.108202>.

- [155] Ulonska K, Voll A, Marquardt W. Screening Pathways for the Production of Next Generation Biofuels. *Energy Fuels* 2016;30(1):445–56.
<https://doi.org/10.1021/acs.energyfuels.5b02460>.
- [156] Ulonska K, König A, Klatt M, Mitsos A, Viell J. Optimization of Multiproduct Biorefinery Processes under Consideration of Biomass Supply Chain Management and Market Developments. *Ind. Eng. Chem. Res.* 2018;57(20):6980–91.
<https://doi.org/10.1021/acs.iecr.8b00245>.
- [157] Department of Defense. Handbook Aerospace Fuels Certification. United States: Department of Defence, United States of America; 2014.
- [158] Colket M, Heyne J, Lieuwen TC. Fuel Effects on Operability of Aircraft Gas Turbine Combustors. Reston, VA: American Institute of Aeronautics and Astronautics, Inc; 2021.
- [159] D02 Committee. Test Method for Density, Relative Density, and API Gravity of Liquids by Digital Density Meter. West Conshohocken, PA: ASTM International.
<https://doi.org/10.1520/D4052-22>.
- [160] 10.1520/D0445-21E02. Test Methods for Surface and Interfacial Tension of Solutions of Paints, Solvents, Solutions of Surface-Active Agents, and Related Materials. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D1331-20>.
- [161] D02 Committee. Test Method for Kinematic Viscosity of Transparent and Opaque Liquids (and Calculation of Dynamic Viscosity). West Conshohocken, PA: ASTM International.
<https://doi.org/10.1520/D0445-21E02>.
- [162] D02 Committee. Test Method for Estimation of Net Heat of Combustion of Aviation Fuels. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D4529-17>.
- [163] D02 Committee. Test Method for Estimation of Net Heat of Combustion of Aviation Fuels. West Conshohocken, PA: ASTM International.
https://doi.org/10.1520/D3338_D3338M-20A.
- [164] D02 Committee. Test Method for Heat of Combustion of Liquid Hydrocarbon Fuels by Bomb Calorimeter (Precision Method). West Conshohocken, PA: ASTM International.
<https://doi.org/10.1520/D4809-18>.
- [165] D02 Committee. Test Method for Flash Point by Tag Closed Cup Tester. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D0056-22>.
- [166] D02 Committee. Test Methods for Flash Point by Pensky-Martens Closed Cup Tester. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D0093-20>.
- [167] Ogston AR. A Short History of Aviation Gasoline Development, 1903-1980. In: SAE Technical Paper Series. SAE International 400 Commonwealth Drive, Warrendale, PA, United States; 1981.
- [168] Edwards T. Advancements in Gas Turbine Fuels From 1943 to 2005. *Journal of Engineering for Gas Turbines and Power* 2007;129(1):13–20.
<https://doi.org/10.1115/1.2364007>.
- [169] D02 Committee. Test Method for Freezing Point of Aviation Fuels (Automatic Phase Transition Method). West Conshohocken, PA: ASTM International.
<https://doi.org/10.1520/D5972-16>.

-
- [170] D02 Committee. Test Method for Freezing Point of Aviation Fuels (Automatic Laser Method). West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D7153-22AE01>.
- [171] D02 Committee. Test Method for Freezing Point of Aviation Fuels. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D2386-19>.
- [172] D02 Committee. Test Method for Cetane Number of Diesel Fuel Oil. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D0613-18AE01>.
- [173] D02 Committee. Test Method for Determination of Derived Cetane Number (DCN) of Diesel Fuel Oils Fixed Range Injection Period, Constant Volume Combustion Chamber Method. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/D7170-16>.
- [174] White RD. Refining and blending of aviation turbine fuels. *Drug Chem Toxicol* 1999;22(1):143–53. <https://doi.org/10.3109/01480549909029728>.
- [175] Richter S, Braun-Unkhoff M, Naumann C, Riedel U. Paths to alternative fuels for aviation. *CEAS Aeronaut J* 2018;9(3):389–403. <https://doi.org/10.1007/s13272-018-0296-1>.
- [176] Meurer A, Kern J. Fischer–Tropsch Synthesis as the Key for Decentralized Sustainable Kerosene Production. *Energies* 2021;14(7):1836. <https://doi.org/10.3390/en14071836>.
- [177] Koepf E, Zoller S, Luque S, Thelen M, Brendelberger S, González-Aguilar J et al. Liquid fuels from concentrated sunlight: An overview on development and integration of a 50 kW solar thermochemical reactor and high concentration solar field for the SUN-to-LIQUID project. In: *SOLARPACES 2018: International Conference on Concentrating Solar Power and Chemical Energy Systems*. AIP Publishing; 2019, p. 180012.
- [178] Hu J, Yu F, Lu Y. Application of Fischer–Tropsch Synthesis in Biomass to Liquid Conversion. *Catalysts* 2012;2(2):303–26. <https://doi.org/10.3390/catal2020303>.
- [179] Zschoke A, Scheuermann, Sebastian, Ortner, Jens. High Biofuel Blends in Aviation (HBBA): Final Report 2014.
- [180] Geleynse S, Brandt K, Garcia-Perez M, Wolcott M, Zhang X. The Alcohol-to-Jet Conversion Pathway for Drop-In Biofuels: Techno-Economic Evaluation. *ChemSusChem* 2018;11(21):3728–41. <https://doi.org/10.1002/cssc.201801690>.
- [181] Li L, Coppola E, Rine J, Miller JL, Walker D. Catalytic Hydrothermal Conversion of Triglycerides to Non-ester Biofuels. *Energy Fuels* 2010;24(2):1305–15. <https://doi.org/10.1021/ef901163a>.
- [182] Kurzawska P. Overview of Sustainable Aviation Fuels including emission of particulate matter and harmful gaseous exhaust gas compounds. *Transportation Research Procedia* 2021;59:38–45. <https://doi.org/10.1016/j.trpro.2021.11.095>.
- [183] Bauen A, Bitossi N, German L, Harris A, Leow K. Sustainable Aviation Fuels. *Johnson Matthey Technology Review* 2020. <https://doi.org/10.1595/205651320X15816756012040>.
- [184] Marker TL, Felix LG, Linck MB, Roberts MJ. Integrated hydrolysis and hydroconversion (IH2) for the direct production of gasoline and diesel fuels or blending components from biomass, part 1: Proof of principle testing. *Environ. Prog. Sustainable Energy* 2012;31(2):191–9. <https://doi.org/10.1002/ep.10629>.

- [185] Virent, Inc. |. BioForming. [June 30, 2022]; Available from:
<https://www.virent.com/technology/bioforming/>.
- [186] Dally B, Cortright R. Virent Thunderbird Close Out Meeting: Virent is replacing crude oil 2014.
- [187] Marker TL, Felix LG, Linck MB, Roberts MJ, Ortiz-Toral P, Wangerow J. Integrated hydropyrolysis and hydroconversion (IH 2[®]) for the direct production of gasoline and diesel fuels or blending components from biomass, Part 2: continuous testing. *Environ. Prog. Sustainable Energy* 2014;33(3):762–8. <https://doi.org/10.1002/ep.11906>.
- [188] Poling BE, Prausnitz JM, O'Connell JP. *The properties of gases and liquids*. 5th ed. New York: McGraw-Hill; 2001.
- [189] Katti PK, Chaudhri MM. Viscosities of Binary Mixtures of Benzyl Acetate with Dioxane, Aniline, and m -Cresol. *J. Chem. Eng. Data* 1964;9(3):442–3.
<https://doi.org/10.1021/je60022a047>.

A. Descriptions of Jet Fuel Screening Properties

Research projects like JETSCREEN or the NJFCP identified fuel properties that are particularly critical for the assessment of a jet fuel at an early stage of development. All properties appear in the property specifications of ASTM D4054 and ASTM D7655 or in the Fit-for-Purpose testing Tier 2 of the ASTM D4054 either as specification or as reporting property. The importance of the individual properties for the safe application of a jet fuel and its performance is explained in the following paragraphs.

Density

The density is a basic fuel property that directly influences the loadable weight of a fuel on the aircraft and the injected quantity of the fuel in the combustion chamber. Together with the combustion energy, described by the net heat of combustion, the density therefore defines the overall range and influences the performance of an aircraft [27,107,157]. It furthermore affects the thermal expansion of the fuel [157], influences the ability to remove heat generated by the fuel injection system [27] and is an important measured control parameter for metering and balancing e.g. fuel gauging [158]. For conventional oil-based jet fuels and blends containing synthetic aviation fuels, the density is set in a value range between 775-840 kg/m³ at 15 °C in both of the corresponding jet fuel specification norms ASTM D1655 and D7566 for blends. The property is generally measured by the ASTM D4052 [159] or an equivalent measurement protocol listed in the jet fuel specification norms.

Surface tension

The surface tension is an important property that influences the atomization characteristics, together with viscosity and density, and hence the fuel combustion, ignition and reignition [27,157]. Higher surface tensions and viscosities make a liquid harder to atomize [158]. The property has no set limits in the relevant jet fuel specifications, but is measured and recorded as part of the Fit-for-Purpose property testing for new synthetic jet fuel candidates after ASTM D4054. In the Fit-for-Purpose testing, the property has to comply with a value range of approved fuels. The surface tension is generally measured using the ASTM D1331 [160].

Kinematic viscosity

The viscosity measures the internal resistance to motion of a fluid, caused by cohesive forces among the fluid components [157]. For jet fuels, the kinematic viscosity is utilized, which relates the dynamic viscosity to the density of the fuel. The viscosity of fuels varies inversely and exponentially with temperature. Together with the freezing point, the specification limits of the kinematic viscosity regulate the low temperature behavior of the fuels. The specification limit is set to an upper limit of 8 mm²/s at -20 °C for conventional oil-based jet fuels in ASTM D1655 and additionally 12 mm²/s at -40 °C for synthetic fuels after ASTM D4054 and blends after

ASTM D7566. Sufficiently low viscosities are of great importance for the operability of pumps and nozzles as well as the atomization process at high altitudes and low temperatures [27,157]. The limit for $-40\text{ }^{\circ}\text{C}$ is more stringent, and corresponds to an average kinematic viscosity of $6.5\text{ mm}^2/\text{s}$ to ensure the operability of synthetic fuels [27]. Together with the surface tension and the density, the viscosity strongly impacts the atomization process. Fuels with higher viscosities lead to larger droplets and uniform droplet distributions, which hinders the evaporation process and negatively influences the ignition and reignition process. At low temperatures or reduced fuel supply in the combustor, this can have critical consequences and lead to extinction and impeded high altitude reignition [158]. The jet fuel specification states the ASTM D445 [161] or equivalent norm for the measurement of the kinematic viscosity.

Net heat of combustion

The net heat of combustion quantifies the energy of a fuel released by burning at constant pressure, with all products including water being gaseous. Together with the density, it determines the range and influences the performance of an aircraft [157]. The net heat of combustion has a lower limit of 42.8 MJ/kg in the ASTM D1655 for conventional oil-based jet fuels and ASTM D7566 for blends with synthetic fuels. Both specifications list ASTM D4529 [162], D3338 [163] and D4809 [164] as possible measurement methods.

Flash point

The flash point measures the temperature at which the fuel must be heated to emit a flammable vapor that flashes when brought into contact with an external flame [157]. Likewise, to the distillation line the flash point characterizes the flammability of a jet fuel and therefore the ignition and reignition behavior [158]. It is furthermore a leading factor determining fire safety in fuel handling [157]. Depending on the specification, the flash point has a lower and upper limit between $38\text{ }^{\circ}\text{C}$, $60\text{ }^{\circ}\text{C}$ for military jet fuel JP-5 and $68\text{ }^{\circ}\text{C}$. The upper limit is thereby only set in the specification for the evaluation of new jet fuel candidates in ASTM D4054. The flash point of a fuel is generally determined by ASTM D56 [165] or D93 [166].

Freezing point

Together with the kinematic viscosity, the freezing point characterizes the low-temperature behavior of the fuel. The freezing point states the temperature at which a clear crystallization is visible in a fuel and the fuel starts to solidify [157]. The minimum freezing point is set to $-40\text{ }^{\circ}\text{C}$ for Jet A, synthetic fuels after ASTM D4054 and blends after ASTM D7566. For JP-5, the freezing point is set to $-46\text{ }^{\circ}\text{C}$ for Jet A-1 and JP-8 to $-47\text{ }^{\circ}\text{C}$. Specification limits for the freezing point were initially introduced in the 1940s as requirement for high-altitude operations [167,168]. At high altitudes low atmospheric temperatures could lead to fuel crystallization and freezing, which can cause plugging and malfunction in pumps and nozzle operations [157]. As measurement methods the jet fuel specification list ASTM D5972 [169], D7153 [170], D2386 [171] or equivalent methods.

Cetane number

The cetane number characterizes the autoignition behavior, or ignition quality of a fuel in a relative ranking to cetane (n-hexane), a pure compound with fast autoignition and good ignition quality [95]. It is based on the ignition delay, the time between the fuel injection and ignition. The cetane number is not listed in the relevant jet fuel specifications, however it was found to be highly relevant for the combustion extinction by the NJFCP [158]. For conventional oil-based jet fuels, the cetane number ranges between 39-50 [158]. The NJFCP determined a lower limit of 30 due to observations made for synthetic jet fuels that showed poor ignition quality. The cetane number can be determined by the mentioned single-cylinder engine using ASTM D613 [172] or an ignition quality tester after ASTM D7170 [173].

Distillation line

The distillation line is the oldest specified property of an aviation fuel standard. Already in 1918 temperature limits for the recovered volume were set for aviation gasoline [167]. In current specifications, the distillation line measures temperatures at set evaporated volume fractions of jet fuel. Depending on the specifications, temperatures at 10 vol% (T_{10}), 50 vol%, 90 vol% and 100 vol% evaporated are specified. Furthermore, differences between temperatures of the evaporated fractions are specified, e.g. $T_{90}-T_{10}$ and $T_{50}-T_{10}$ [25,26]. For conventional jet fuels and blends, only the T_{10} and T_{100} with upper limits at 205 and 300 °C are specified after ASTM D4954 and ASTM D7566. Annexes for certified synthetic jet fuels FT-SPK [25] as well as the evaluation of new jet fuel candidates in ASTM D4054 also specify temperature differences. The temperature differences prevent “narrow” boiling compositions, where large fractions of the fuel evaporate at once, which could occur for synthetic jet fuels [157]. The distillation line is strongly influenced by the fuel composition [157]. The different fuel components evaporate from the fuel based on their vapor pressure, which differs based inter alia by chemical family and the overall size of the component. The specification thereby indirectly limits the possible fuel composition within the range of experience since only fuel with certain compositional distributions comply to the set specification limits [157]. Due to the strong relationship between distillation and evaporation, the distillation line relates to the way the fuel combustion initiates and sustains. The distillation behavior therefore also influences the extinction behavior of a fuel at reduced fuel supply, so-called lean blow-out, and the extinction and re-ignition at high altitudes [158]. Both scenarios are absolutely critical for the safe operation of the jet fuel. The distillation line is usually measured by the ASTM D86 [109].

B. Approved and Pending Jet Fuel Production Routes

Production from crude oil and oil sands

For the production of conventional jet fuel, crude oil is typically distilled into different product streams such as naphtha, gasoline and kerosene, refined using e.g. hydrotreatment and subsequently blended to formulate the final jet fuel product [174]. If oil sands or shale is used as feedstock, a crude oil like product is previously derived using the respective operations, oil extraction and pyrolysis & thermal dissolution [174]. The hydrotreatment is necessary to remove metals and heteroatoms in the presence of hydrogen, such as sulfur, oxygen and nitrogen from the products. The blending combines the different product streams into the final jet fuels by adjusting the fractions to meet the fuel specifications [174]. Due to the large variety in the composition of the crude oil and the degrees of freedom in the blending, the final composition of the jet fuel as well as the fuel properties can vary substantially. Typical conventional fuels are composed of broad and varying distributions of alkanes, cyclo-alkanes and aromatics. One can therefore not speak of “one” default conventional oil-based jet fuel.

Fisher-Tropsch hydroprocessed synthesized paraffinic kerosene (FT-SPK)

The Fisher-Tropsch hydroprocessed synthesized paraffinic kerosene process (FT-SPK) is the oldest approved pathway for the production of synthetic jet fuel. Standardized by ASTM in 2009, the production pathway was developed to produce synthetic jet fuel from syngas, a mixture of H_2 and CO using the name giving Fischer-Tropsch reaction. The syngas can thereby be produced either from many different sources: gasification of municipal waste or coal, reformation of bio or natural gas [175] or carbon capture and hydrogen production via a power-to-liquid [176] or sun-to-liquid [177] process. The Fisher-Tropsch reaction produces a distribution of long-chained hydrocarbons from the syngas in the presence of a metal catalyst [178]. Decisive for the reaction is the ratio of H_2 of CO in the syngas, which is managed by a pre-pending water-gas-shift reaction process. The long-chained hydrocarbons are hydroisomerized in the presence of hydrogen to increase the fraction of branched hydrocarbons and subsequently separated using a distillation column. The final FT-SPK fuel is often a product of the blending of the different distillation streams to meet the corresponding specification requirements of the fuel type. Synthetic fuels produced by the FT production process can be used in fractions of up to 50 vol% after ASTM D7566 and are typically mainly composed of n-alkanes and iso-alkanes containing 5-20 carbon atoms. FT-SPK fuel can also be co-produced with a maximum volume fraction of 5 vol% after the ASTM D1655 [12].

Synthesized paraffinic kerosene from hydro-processed esters and fatty acids (HEFA-SPK)

In contrast to the Fischer-Tropsch process, the fuel produced by the hydroprocessed esters and fatty acids pathway (HEFA-SPK) always relies on refined oil from biomaterial. The exact sources of the refined oil may vary, from oil from crop plants e.g. rapeseed or soy to used cooking oils [179]. Depending on the source, different pretreatments are necessary [175]. The most important components for the refining are free fatty acids and triglycerides, which are the base material of any natural fats and oils [175]. The fatty acids and triglycerides are converted into long-chained hydrocarbons with H₂ in the hydrogenation and deoxygenation reactions under the elimination of H₂O and CO₂ respectively. Since the process produces mostly unbranched alkanes and alkenes, isomerization and cracking in the presence of hydrogen are applied to convert alkenes to alkanes and unbranched alkanes to branched iso-alkanes. The crude product is distilled and the resulting streams are often blended to produce a product that matches the specifications. HEFA-SPK fuels are allowed in blends with a maximum fraction of 50 vol% after ASTM D7566 and 5 vol% for a co-processing route after ASTM D1655. Similar to the FT-SPK fuels, HEFA-SPK fuels are typically composed mainly of n-alkanes and iso-alkanes containing 5-20 carbon atoms. The production route was approved in 2011 by the ASTM [12].

Synthesized isoparaffins from hydroprocessed fermented sugars (SIP)

The route for the production of Synthesized isoparaffins from hydroprocessed fermented sugars produces, in contrast to the other processes, not a distribution of hydrocarbons from different families with different carbon numbers but one single component, farnesane [179]. Farnesane is a branched iso-alkane containing 15 carbon atoms. The process uses sugars as its main feedstock, which are converted by genetically modified yeast to farnesene, an iso-alkene precursor. The precursor is hydrotreated to remove unsaturated bonds. The stream is finally distilled to separate farnesane and unconverted farnesene [175,179]. SIP fuel is allowed in blends with a maximum volume fraction of 10 vol% after ASTM D7566, it was approved in 2014 by the ASTM [12].

Synthesized kerosene with aromatics derived by alkylation of light aromatics from nonpetroleum sources (SPK/A)

This is an extension of the FT production processes, which refines and blends the product streams of the FT production process further. The light boiling top fraction of the FT product distillation, predominantly propene and butene, is used for the alkylation of a benzene rich co-product stream from another production process [32]. The benzene rich stream is often available as a co-product from a coal gasification process. In the alkylation, the alkenes react in the presence of a catalyst to branched mono-aromatics, which are then blended into the FT jet fuel product [32]. The resulting synthetic jet fuel, called FT-SPK/A is considered compositionally identical to conventional oil-based jet fuel, containing alkanes, cyclo-alkanes and aromatics. However, the maximum blending limit is set to 50 vol% in ASTM D7566 to allow the accumulation of service

experience prior to permitting the unblended usage [32]. The production route was approved in 2015 by ASTM [12].

Alcohol-to-jet synthetic paraffinic kerosene (ATJ-SPK)

As the name implies, this production route uses alcohol as feedstock for the fuel production. The alcohols, mostly ethanol and butanol are thereby produced by fermentation of sugars and starch [175]. The alcohols are dehydrated in the presence of catalyst e.g. zeolites, forming ethylene and butylene, and oligomerized to long-chained hydrocarbons [180]. Unsaturated alkenes are subsequently hydrogenated and fractionated using a distillation column [175]. ATJ-SPK are composed typically only of iso-alkanes containing a distinct number of carbon atoms. The SAF can be blended in volume fractions of up to 50 vol% according to ASTM D7566. The production route was approved in 2016 [12].

Catalytic hydrothermolysis synthesized kerosene (CHJ)

The production process for the synthetization Catalytic hydrothermolysis jet fuel (CHJ) uses refined oil from bio-material similar to the HEFA process. In contrast to the HEFA process however, the fatty acids are converted into n- and iso-alkanes, cyclo-alkanes as well as aromatics using catalytic hydrothermolysis [179]. In this hydro thermolysis, the ester bonds of the triglycerides are broken in the presence of water. The free fatty acids are converted into n- and iso-alkanes as well as cyclo-alkanes and aromatics in a complex interplay of cracking, isomerization, decarboxylation and cyclization [181]. The product stream is mildly hydrogenated to convert possible alkenes to alkanes and subsequently fractionated in a distillation column [179]. The typical CHJ is composed of n-alkanes, cyclo-alkanes and aromatics, similarly to conventional jet fuel. ASTM D7566 regulates the maximum fraction of CHJ fuels to 50 vol%. The process was approved in 2020 [12].

Hydroprocessed hydrocarbons, esters, and fatty acids synthetic paraffinic kerosene (HHC-SPK)

This is the latest certified production pathway, standardized in 2020 and also the first pathway to receive expedited review under ASTM's D4054 "fast track" approval process. The production uses the same process operations as the HEFA-SPK process, however only on the basis of refined oils from the botryococcus braunia algae, known as botryococcenes [32,182]. Since the fuel was approved under the "fast-track" review process, the maximum fraction is limited to 10 vol% in ASTM 7566.

Fuel production routes in review

According to CAAFI four additional fuel production pathways are currently actively pursuing certification after ASTM D4054: Hydro-deoxygenation Synthetic Kerosene (HDO-SAK), High Freeze Point Hydroprocessed Esters and Fatty Acids Synthetic Kerosene (HFP HEFA-SK), Integrated Hydrolysis and Hydroconversion (IH²) and Alcohol-to-Jet Synthetic Kerosene

with Aromatics (ATJ-SKA) [12]. The new production pathways rely predominantly on cellulosics and lignocellulosics as feedstock, which have a higher availability and a smaller ILUC and potential effect on the food market [183,184]. The processes of HFP HEFA-SK and ATJ-SKA are similar or identical to the already presented production routes and differ on the utilized feedstock and additional up- or downstream processes. HDO-SAK and IH² in contrast use new process technologies and are therefore presented shortly.

The HDO-SAK process is developed by Virent and uses hydrogenolysis in a first step to break the glucosidic bonds of the cellulose to form oxygenated intermediates like alcohols, ketones and sugars at high hydrogen pressures [179,185,186]. In a second step, intermediates are catalytically condensed, in the name giving hydro-dexoygenation, to form long-chained hydrocarbons [186]. These products are hydrogenated to remove residual oxygen and subsequently fractionated in a distillation column [179,186]. The process is currently in Tier 3 and Tier 4 of the test program [12].

The IH² pathway is developed by Shell and designed to convert lignocellulosic biomass from crop plants and trees to fuels composed mostly of cyclo-alkanes [110]. The lignocellulosic biomass is first broken into smaller hydrodeoxygenated polymers in the presence of H₂ and a proprietary catalyst and secondly treated by a hydroconversion to remove the residual oxygen under formation of cyclic-alkanes, water and off-gas. Water and off-gas are separated and the alkanes are fractionated using a distillation column [184,187]. The process is also currently in Tier 3 and Tier 4 of the approval process [12].

C. Utilized Structural Molecular Features

Table C 1: Description of utilized structural molecular features.

Feature	Description
[CX4H3]	Aliphatic C with 1 further total connection with 3 further hydrogen
[CX4H2]	Aliphatic C with 2 further total connections with 2 further hydrogen
[CX4H1]	Aliphatic C with 3 further total connections with 1 further hydrogen
[CX4H0]	Aliphatic C with 4 further total connections with 0 further hydrogen
[CX4H1]	Aliphatic C with 1 further total connection with 2 further hydrogen
[CX3H1]	Aliphatic C with 2 further total connections with 1 further hydrogen
[CX3H0]	Aliphatic C with 3 further total connections with 0 further hydrogen
[CX2H1]	Aliphatic C with 1 further total connection with 1 further hydrogen
[CX2H0]	Aliphatic C with 2 further total connections with 0 further hydrogen
[CX4H2R]	Aliphatic C with 2 further total connections 2 further hydrogen, in a ring
[CX4H1R]	Aliphatic C with 3 further total connections 1 further hydrogen, in a ring
[CX4H0R]	Aliphatic C with 4 further total connections 0 further hydrogen, in a ring
[CX3H1R]	Aliphatic C with 2 further total connections 1 further hydrogen, in a ring
[CX3H0R]	Aliphatic C with 3 further total connections 0 further hydrogen, in a ring
[cX3H1](:*):*	Aromatic C with 0 further total connections with 1 further hydrogen
[cX3H0](:*)(:*)*	Aromatic C with 0 further total connections with 0 further hydrogen

<chem>[cX3H0](.*)(.)*.*</chem>	Aromatic C with 0 further total connections with 0 further hydrogen
<chem>[cX3H0]-[cX3]</chem>	Aromatic C with 2 further total connections with 0 further hydrogen
<chem>[cX3H0](.*)(.*)(.)*</chem>	Aromatic C with 0 further total connections with 0 further hydrogen
<chem>[cX3H0](.*)(.)*(-[CX4H2R])</chem>	Aromatic C with 0 further total connections with 0 further hydrogen with connection to an aliphatic C with 1 further total connection with 2 further hydrogen, in a ring
<chem>[CX4H2]-[CX4H1]-[CX4H2]</chem>	Bound of aliphatic C with 1 further total connection with 1 further hydrogen
<chem>[C][C](!CX1)(!CX1)!CX1</chem>	Bound of aliphatic C with one further aliphatic C and 3 non C atoms with 0 further total connections
<chem>!C][C]([C])([C])[C]</chem>	Bound of aliphatic C with two 3 further aliphatic C and one non C atom
<chem>[C][C]([C])([C])[C]</chem>	Bound of aliphatic C with 4 further aliphatic C
<chem>[C][CR](!C)(!C)[C]</chem>	Bound of aliphatic C in a ring with two further aliphatic C and two non C atoms
<chem>[C][CR](!C)([C])[C]</chem>	Bound of aliphatic C in a ring with three further aliphatic C and one non C atoms
<chem>[C][CR]([C])([C])[C]</chem>	Aliphatic C in a ring with 4 further aliphatic C
<chem>[cH1]</chem>	Aromatic C with 1 further connection
<chem>[cH0]</chem>	Aromatic C with 0 further connections
<chem>[C]=[C](!CX1)!CX1</chem>	Bound of 2 aliphatic C with branched double bound and 2 non C atoms with 0 further connections
<chem>[C]=[C]([C])!C</chem>	Bound of 3 aliphatic C and one non C atom, with branched double bound between first and second C
<chem>[C]=[C]([C])[C]</chem>	Bound of 4 aliphatic C, with branched double bound between second and third C
<chem>[C]=[C]=[C]</chem>	3 aliphatic C with double bounds
<chem>!CX1][C]#[C]</chem>	Bound of aliphatic C with triple bound to aliphatic C and C with 0 further total connections
<chem>!X1][C]#[C]</chem>	Bound of aliphatic C with triple bound to aliphatic C and one atom with 0 further connections
<chem>!#1][CH2]!#1</chem>	Bound of aliphatic C with two further hydrogen and two atoms that are not hydrogen atoms

[CX3H1]=[CX3H0]	Bound of aliphatic C with 1 further total connection with 1 further hydrogen and aliphatic C with 2 further total connections with 0 further hydrogen
[CX3H0]=[CX3H0]	Bound of two aliphatic C with 2 further total connections with 0 further hydrogen
[CX3H2]=[CX3H1]	Bound of aliphatic C with 0 further total connections with 2 further hydrogen with aliphatic C with 1 further total connection with 1 further hydrogen
[CX3H1]=[CX3H1]	Bound of two aliphatic C with 1 further total connection with 1 hydrogen
[CX3H2]=[CX3H0]	Bound of aliphatic C with 0 further total connections with 2 further hydrogen with aliphatic C with 2 further total connections with 0 further hydrogen
[c][CX4H3]	Bound of aromatic C with aliphatic C with 0 further total connections with 3 further hydrogen
[c][CX4H2]	Bound of aromatic C with aliphatic C with 1 further total connection with 2 further hydrogen
[c][CX4H1]	Bound of aromatic C with aliphatic C with 2 further total connections with 1 further hydrogen
[CX2H1]#[CX2H0]	Triple bound of aliphatic C with 0 further total connections with 1 further hydrogen and aliphatic C with 1 further total connection with 0 further hydrogen
[CX2H0]#[CX2H0]	Aliphatic C with triple bond and 1 further total connection 0 further hydrogen
[R]	Number of atoms in ring

D. Fuel Database Schema

```
"property": {  
  "density": [  
    {  
      "test_method": "D4052",  
      "unit": "kg/m3",  
      "value": 852.59,  
      "temperature_unit": "C",  
      "temperature_value": -35.3  
    },  
    {  
      "test_method": "D4052",  
      "unit": "kg/m3",  
      "value": 840.97,  
      "temperature_unit": "C",  
      "temperature_value": -19.6  
    },  
    {  
      "test_method": "D4052",  
      "unit": "kg/m3",  
      "value": 827.73,  
      "temperature_unit": "C",  
      "temperature_value": -0.9  
    }  
  ]  
}
```

Figure D 1: Example for the storage of density values using the utilized database schema.

E. Cross-validation Results of Models for Training and Testing

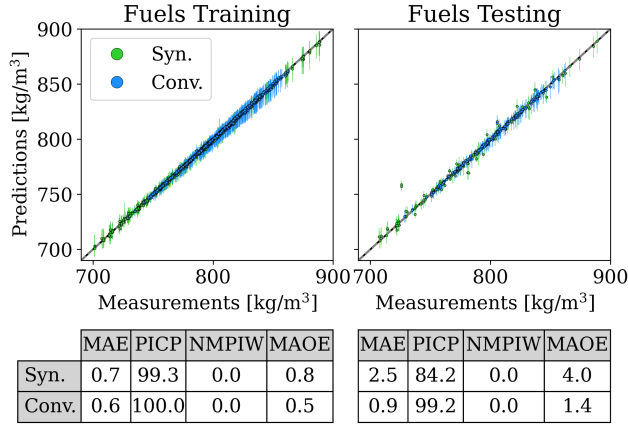


Figure E 1: Cross-validation results training and testing for density prediction of the DC model, MAE [kg/m^3], PICP [%], NMPIW [%], MAOE [kg/m^3].

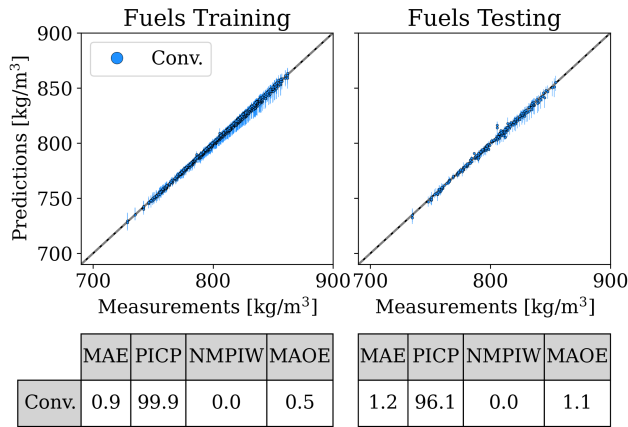


Figure E 2: Cross-validation results training and testing for density prediction of the DC -no Syn. model, MAE [kg/m^3], PICP [%], NMPIW [%], MAOE [kg/m^3].

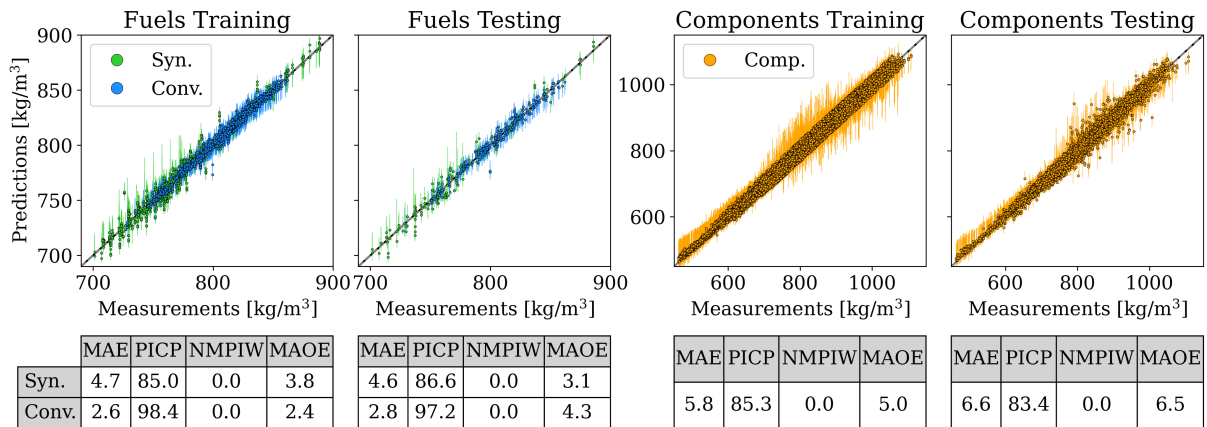


Figure E 3: Cross-validation results training and testing for density prediction of the M-QSPR model, MAE [kg/m^3], PICP [%], NMPIW [%], MAOE [kg/m^3].

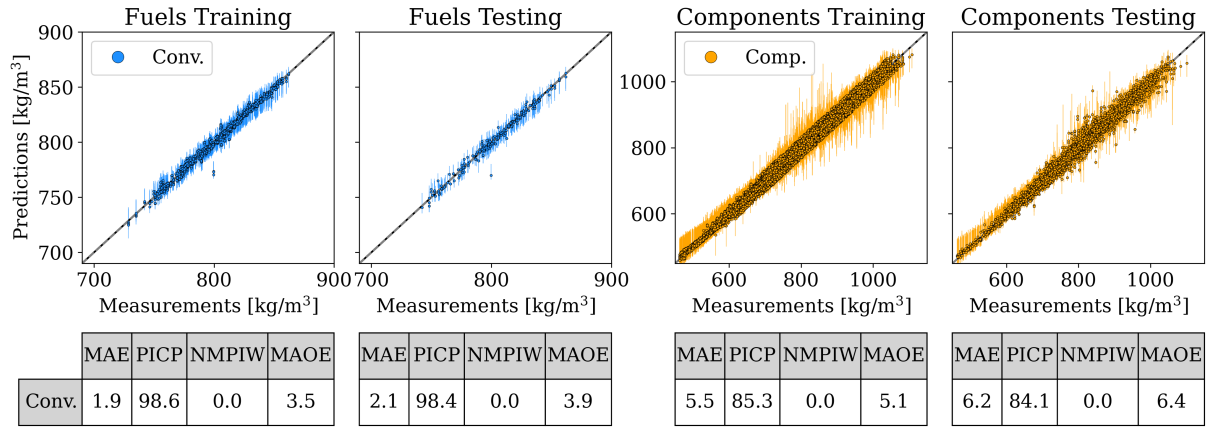


Figure E 4: Cross-validation results training and testing for density prediction of the M-QSPR - no Syn. model, MAE [kg/m^3], PICP [%], NMPIW [%], MAOE [kg/m^3].

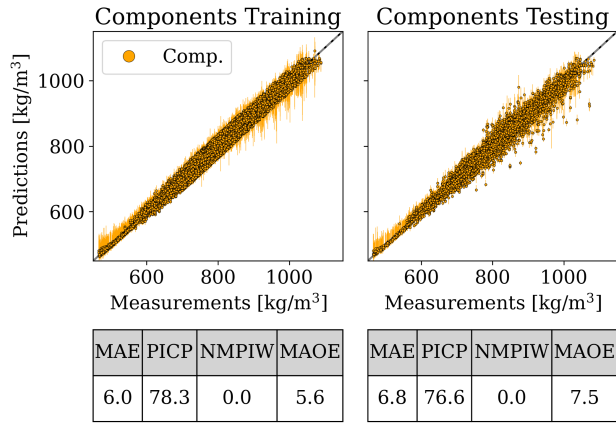


Figure E 5: Cross-validation results training and testing for density prediction of the QSPR model, MAE [kg/m^3], PICP [%], NMPIW [%], MAOE [kg/m^3].

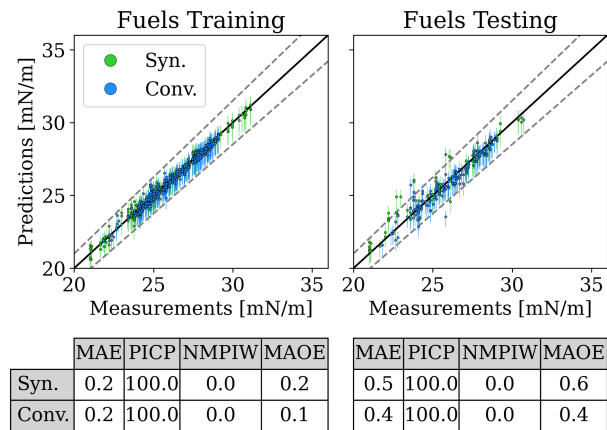


Figure E 6: Cross-validation results training and testing for surface tension prediction of the DC model, MAE [mN/m], PICP [%], NMPIW [%], MAOE [mN/m].

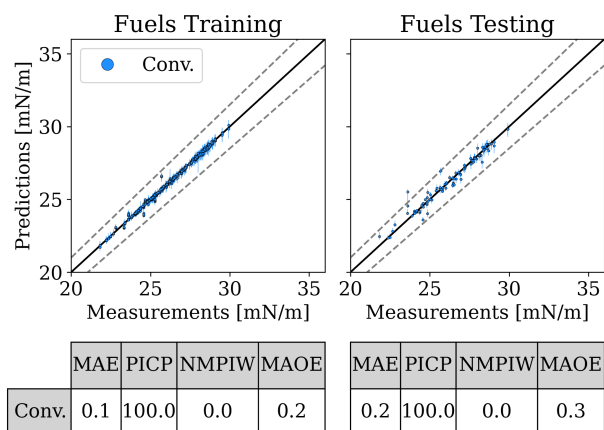


Figure E 7: Cross-validation results training and testing for surface tension prediction of the DC -no Syn. model, MAE [mN/m], PICP [%], NMPIW [%], MAOE [mN/m].

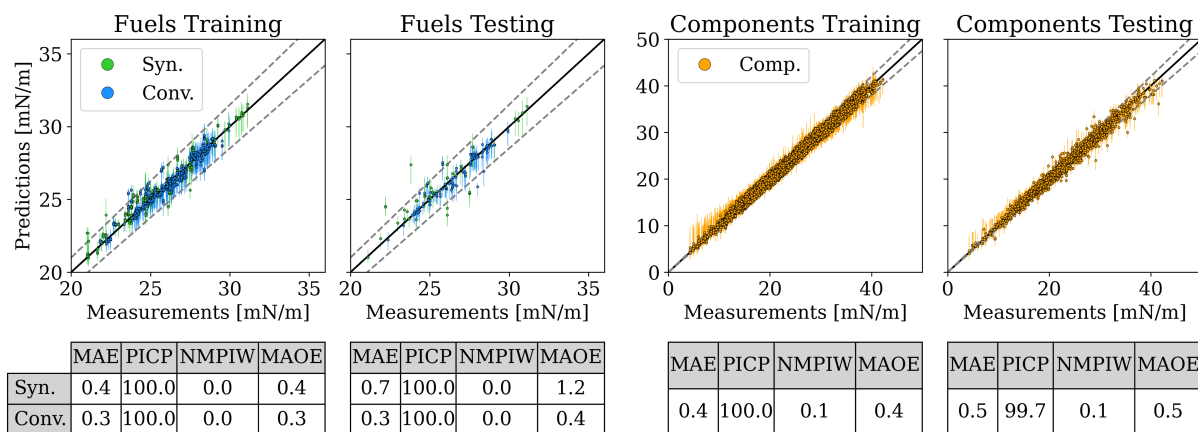


Figure E 8: Cross-validation results training and testing for surface tension prediction of the M-QSPR model, MAE [mN/m], PICP [%], NMPIW [%], MAOE [mN/m].

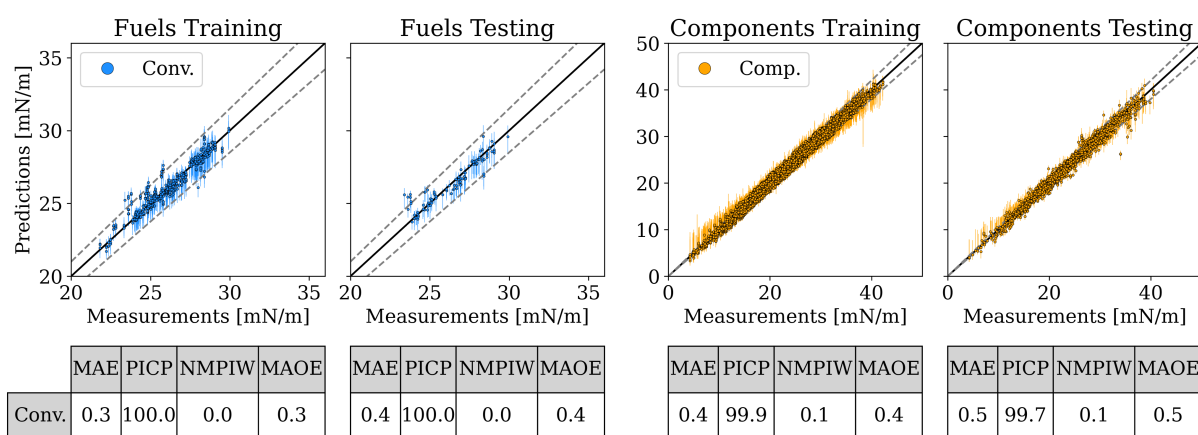


Figure E 9: Cross-validation results training and testing for surface tension prediction of the M-QSPR -no Syn. model, MAE [mN/m], PICP [%], NMPIW [%], MAOE [mN/m].

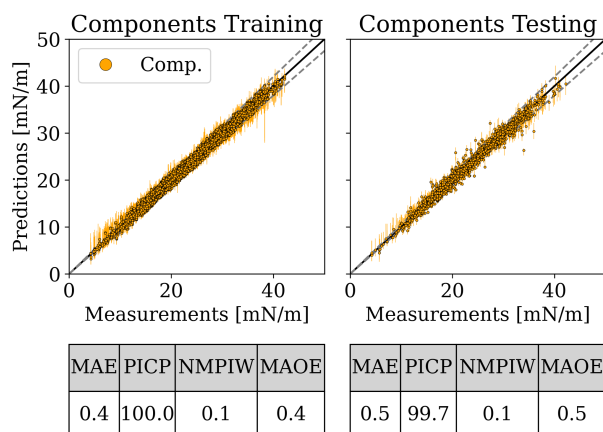


Figure E 10: Cross-validation results training and testing for surface tension prediction of the QSPR model, MAE [mN/m], PICP [%], NMPIW [%], MAOE [mN/m].

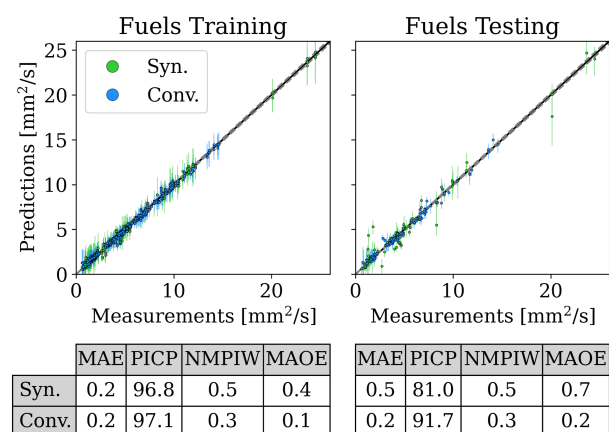


Figure E 11: Cross-validation results training and testing for kinematic viscosity prediction of the DC model, MAE [mm²/s], PICP [%], NMPIW [%], MAOE [mm²/s].

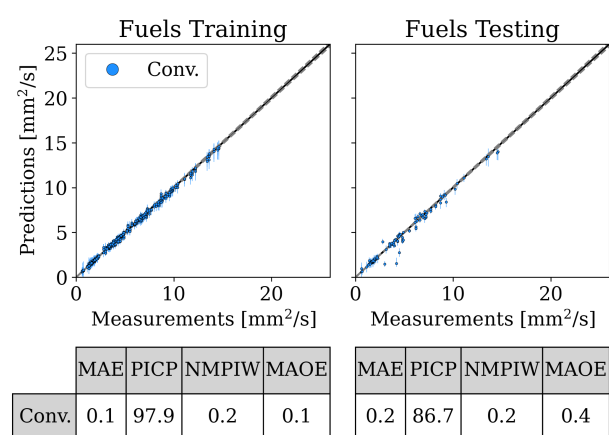


Figure E 12: Cross-validation results training and testing for kinematic viscosity prediction of the DC -no Syn. model, MAE [mm²/s], PICP [%], NMPIW [%], MAOE [mm²/s].

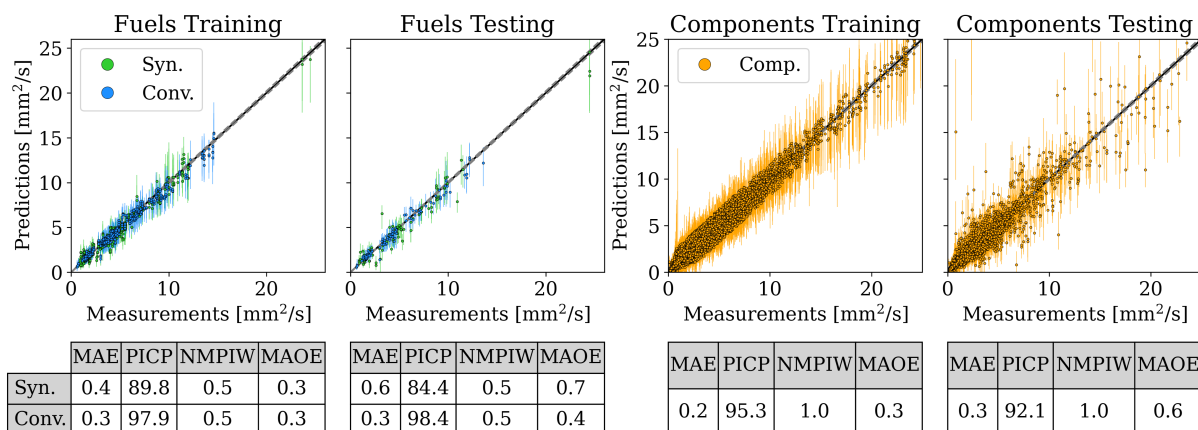


Figure E 13: Cross-validation results training and testing for kinematic viscosity prediction of the M-QSPR model, MAE [mm^2/s], PICP [%], NMPIW [%], MAOE [mm^2/s].

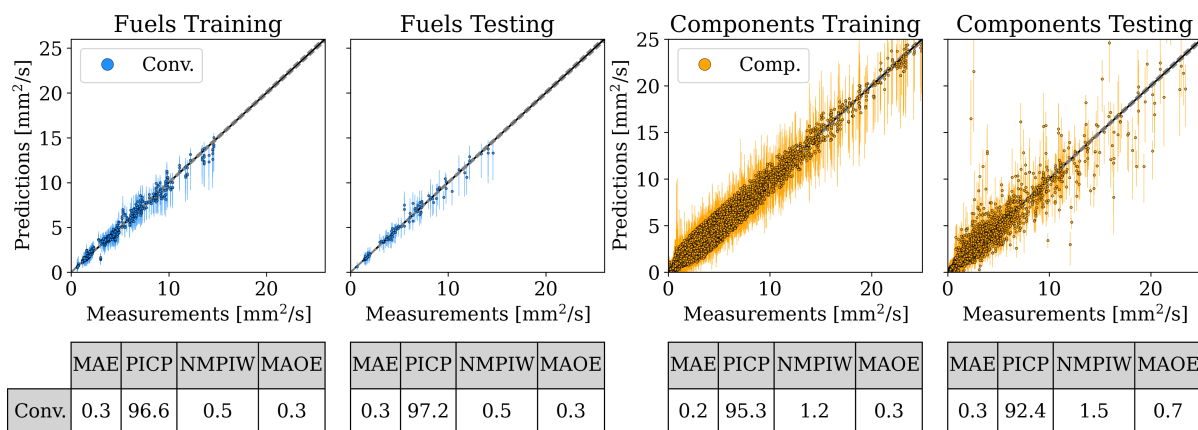


Figure E 14: Cross-validation results training and testing for kinematic viscosity prediction of the M-QSPR -no Syn. model, MAE [mm^2/s], PICP [%], NMPIW [%], MAOE [mm^2/s].

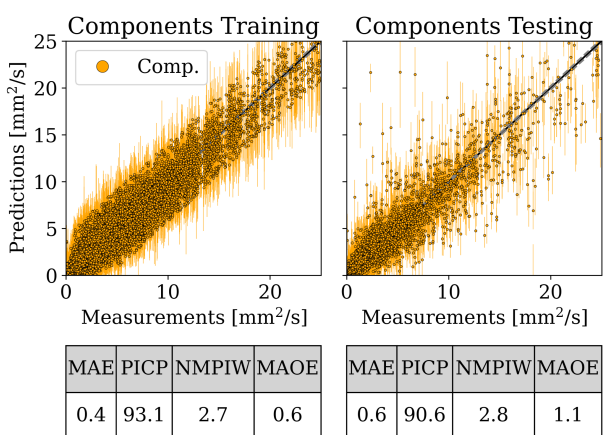


Figure E 15: Cross-validation results training and testing for kinematic viscosity prediction of the QSPR model, MAE [mm^2/s], PICP [%], NMPIW [%], MAOE [mm^2/s].

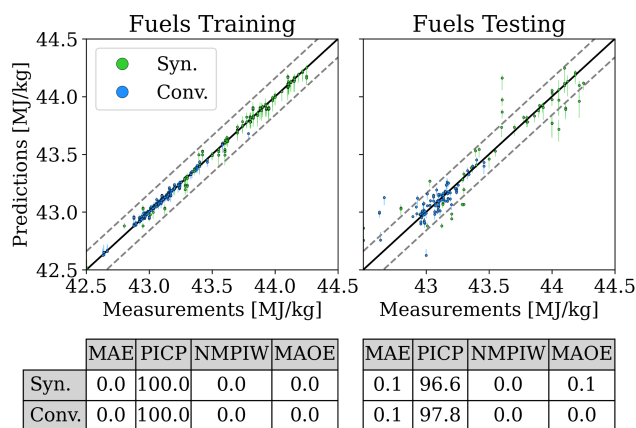


Figure E 16: Cross-validation results training and testing for net heat of combustion prediction of the DC model, MAE [MJ/kg], PICP [%], NMPIW [%], MAOE [MJ/kg].

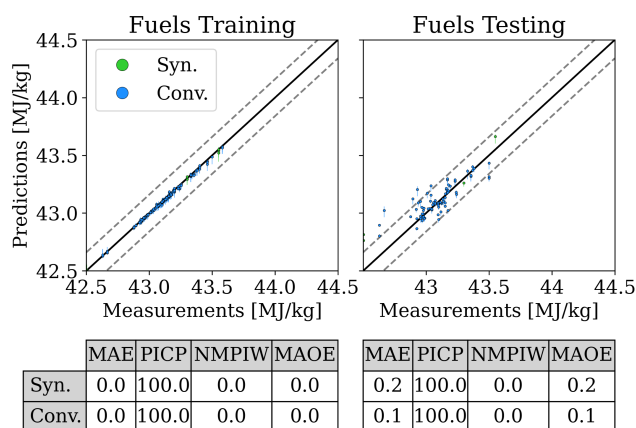


Figure E 17: Cross-validation results training and testing for net heat of combustion prediction of the DC -no Syn. model, MAE [MJ/kg], PICP [%], NMPIW [%], MAOE [MJ/kg].

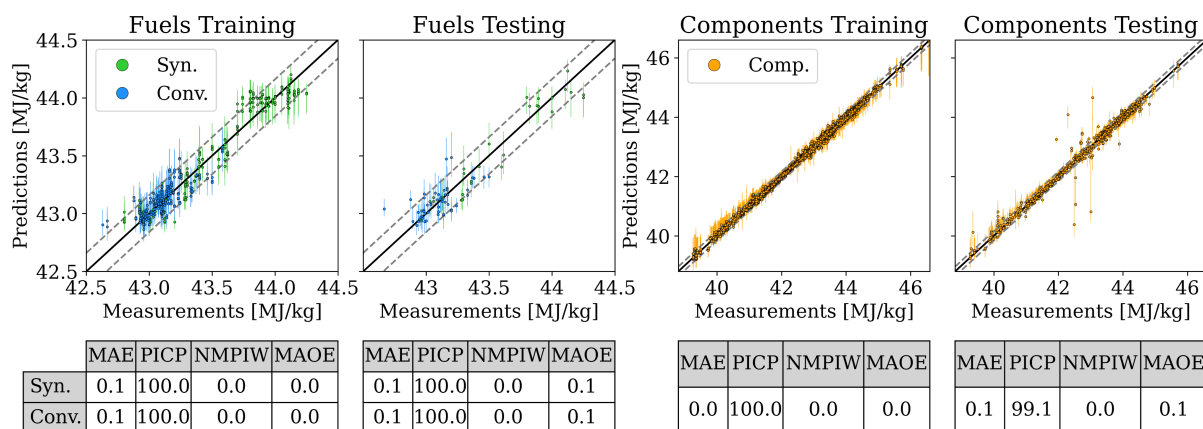


Figure E 18: Cross-validation results training and testing for net heat of combustion prediction of the M-QSPR model, MAE [MJ/kg], PICP [%], NMPIW [%], MAOE [MJ/kg].

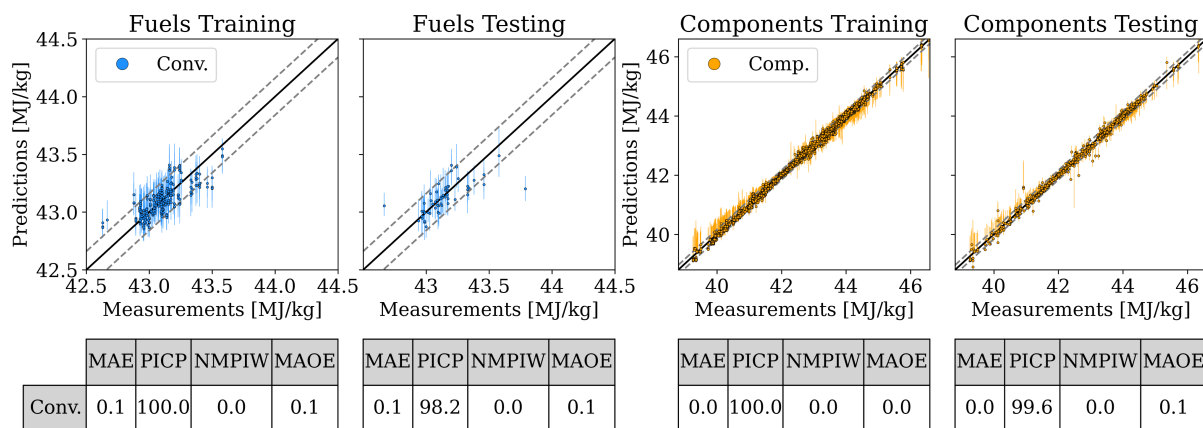


Figure E 19: Cross-validation results training and testing for net heat of combustion prediction of the *M-QSPR -no Syn.* model, MAE [MJ/kg], PICP [%], NMPIW [%], MAOE [MJ/kg].

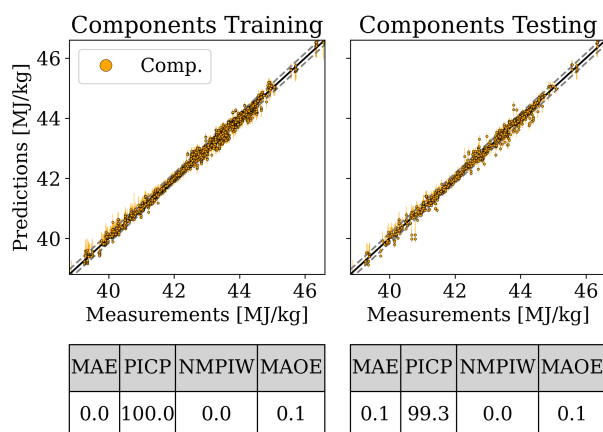


Figure E 20: Cross-validation results training and testing for flash point prediction of the *QSPR* model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

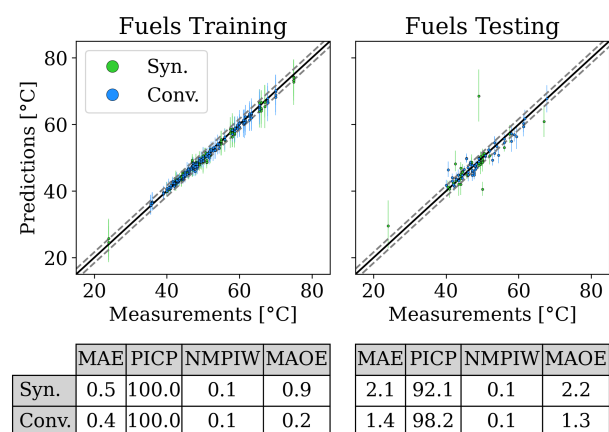


Figure E 21: Cross-validation results training and testing for flash point prediction of the *DC* model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

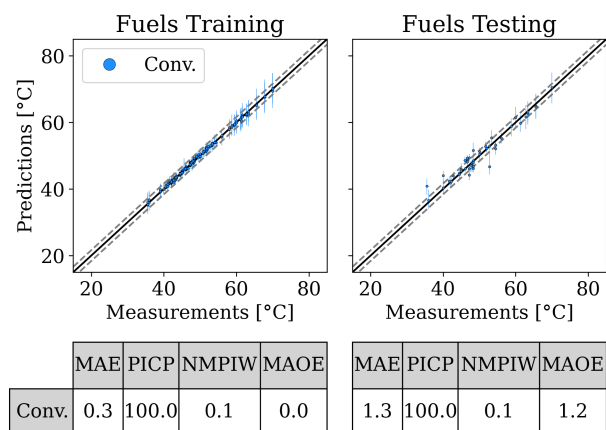


Figure E 22: Cross-validation results training and testing for flash point prediction of the DC - no Syn. model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

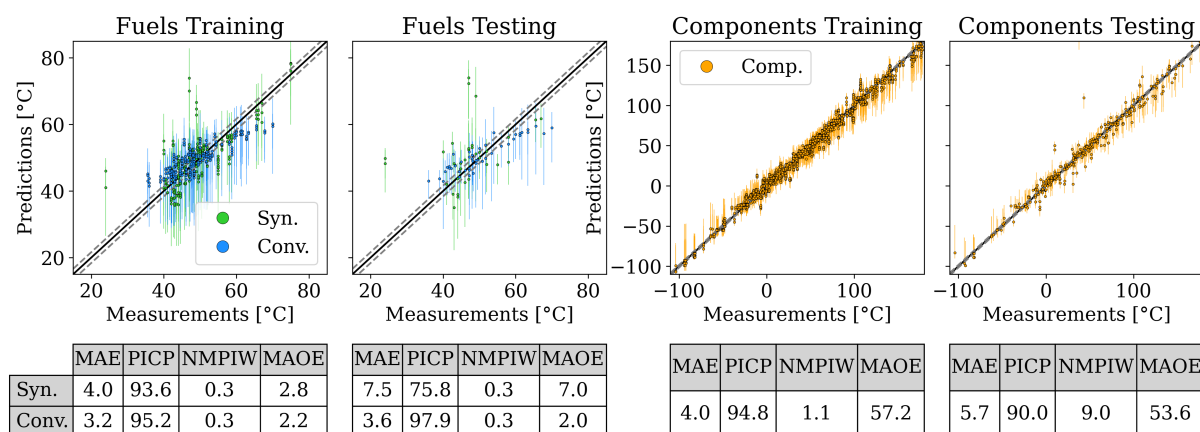


Figure E 23: Cross-validation results training and testing for flash point prediction of the M-QSPR model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

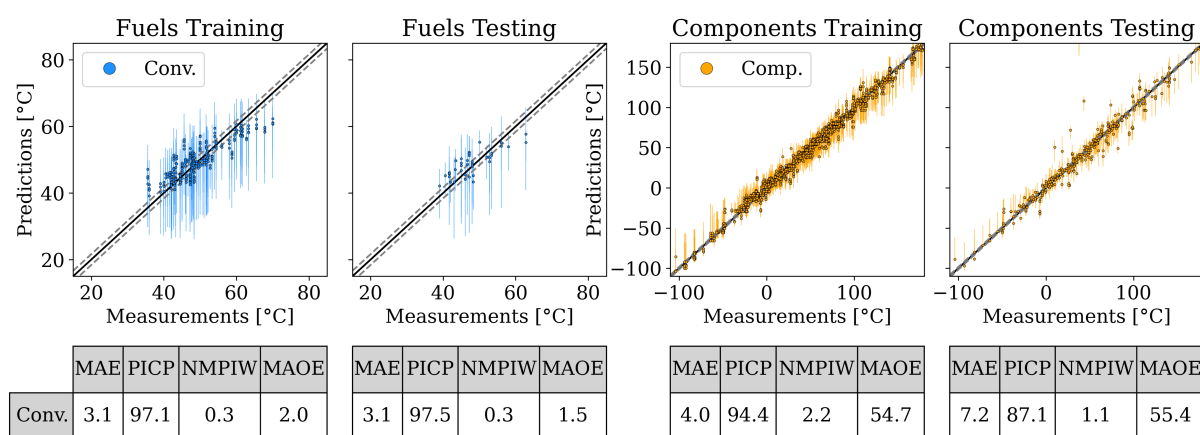


Figure E 24: Cross-validation results training and testing for flash point prediction of the M-QSPR -no Syn. model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

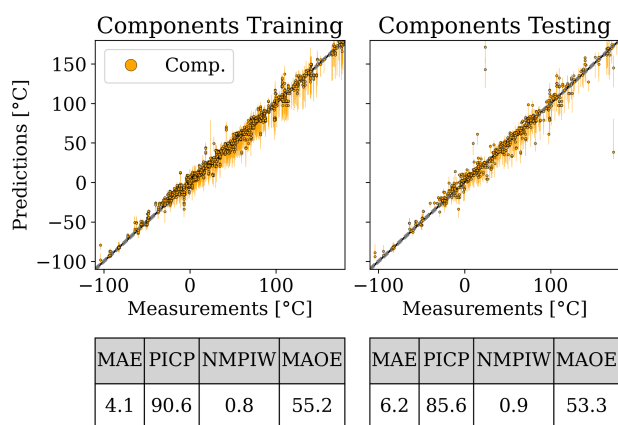


Figure E 25: Cross-validation results training and testing for flash point prediction of the QSPR model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

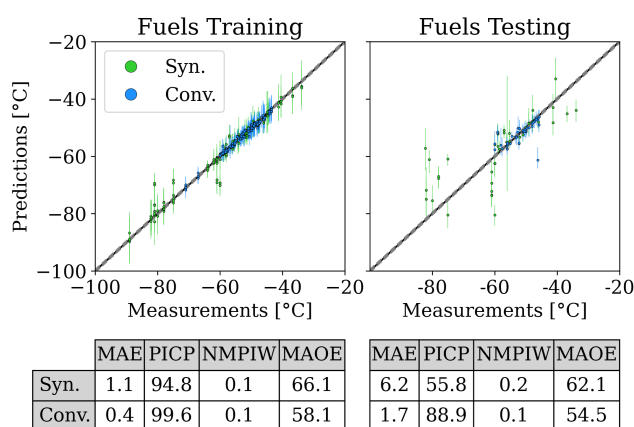


Figure E 26: Cross-validation results training and testing for freezing point prediction of the DC model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

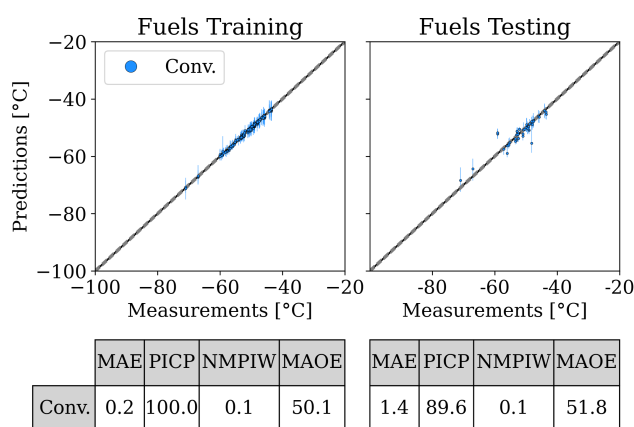


Figure E 27: Cross-validation results training and testing for freezing point prediction of the DC -no Syn. model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

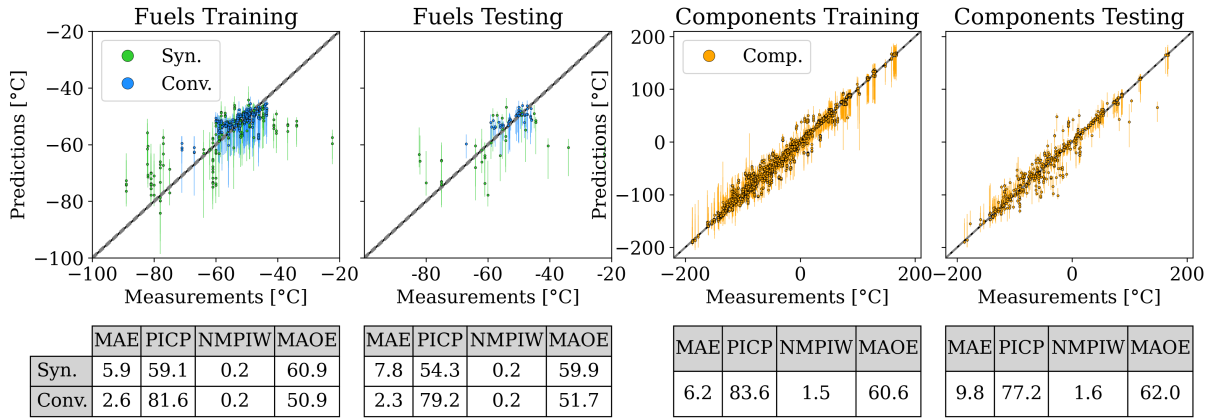


Figure E 28: Cross-validation results training and testing for freezing point prediction of the M-QSPR model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

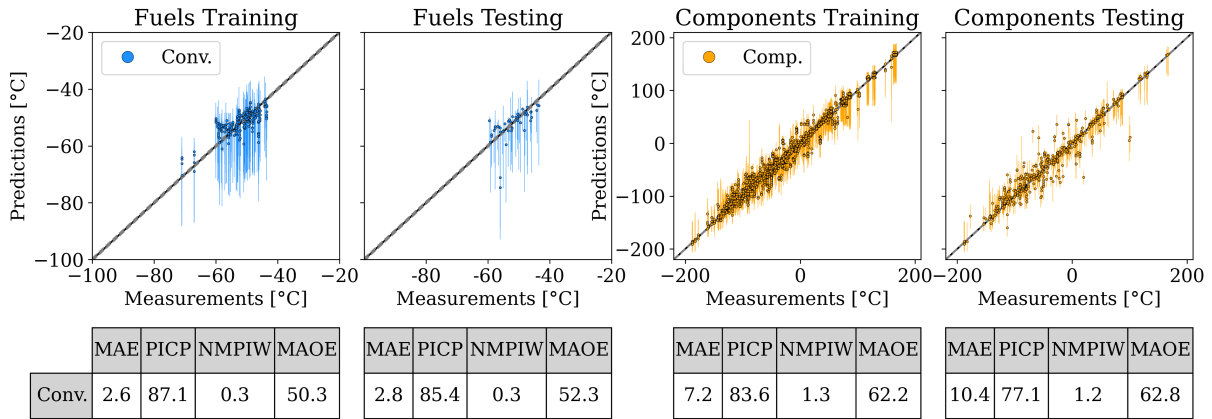


Figure E 29: Cross-validation results training and testing for freezing point prediction of the M-QSPR -no Syn. model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

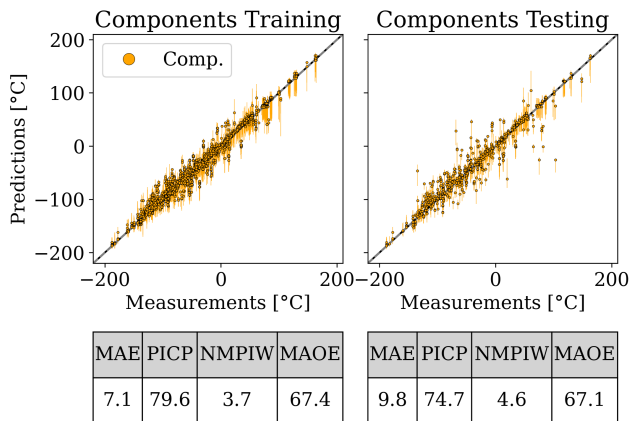


Figure E 30: Cross-validation results training and testing for freezing point prediction of the QSPR model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

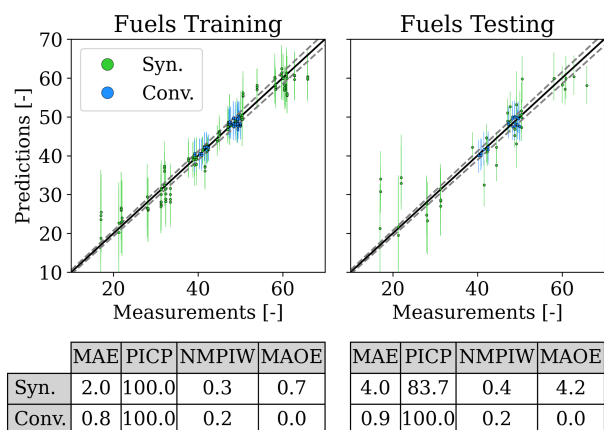


Figure E 31: Cross-validation results training and testing for cetane number prediction of the DC model, MAE [-], PICP [%], NMPIW [%], MAOE [-].

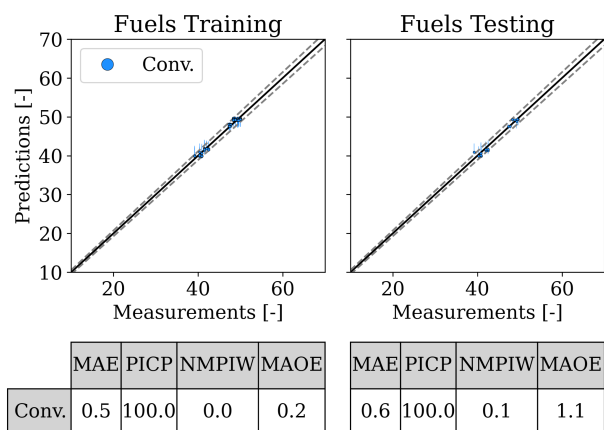


Figure E 32: Cross-validation results training and testing for cetane number prediction of the DC -no Syn. model, MAE [-], PICP [%], NMPIW [%], MAOE [-].

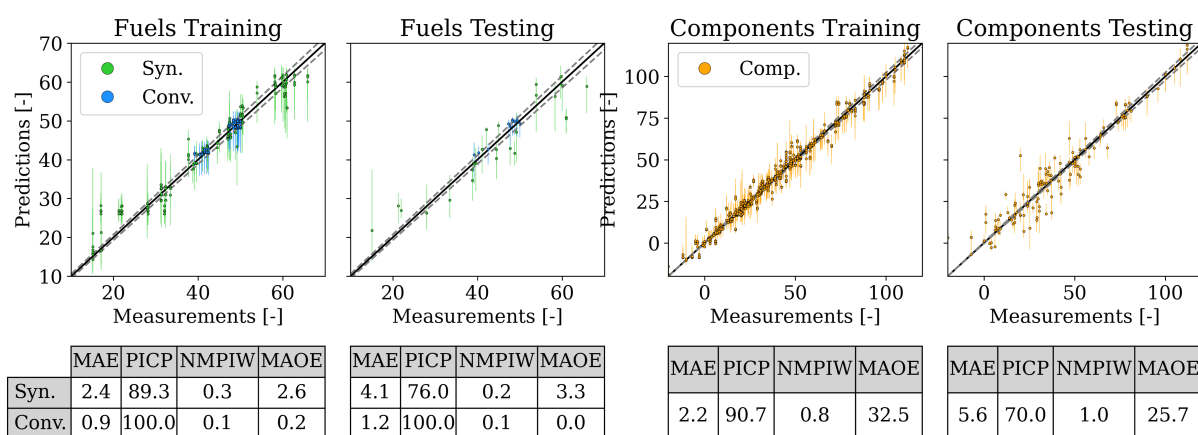


Figure E 33: Cross-validation results training and testing for cetane number prediction of the M-QSPR model, MAE [-], PICP [%], NMPIW [%], MAOE [-].

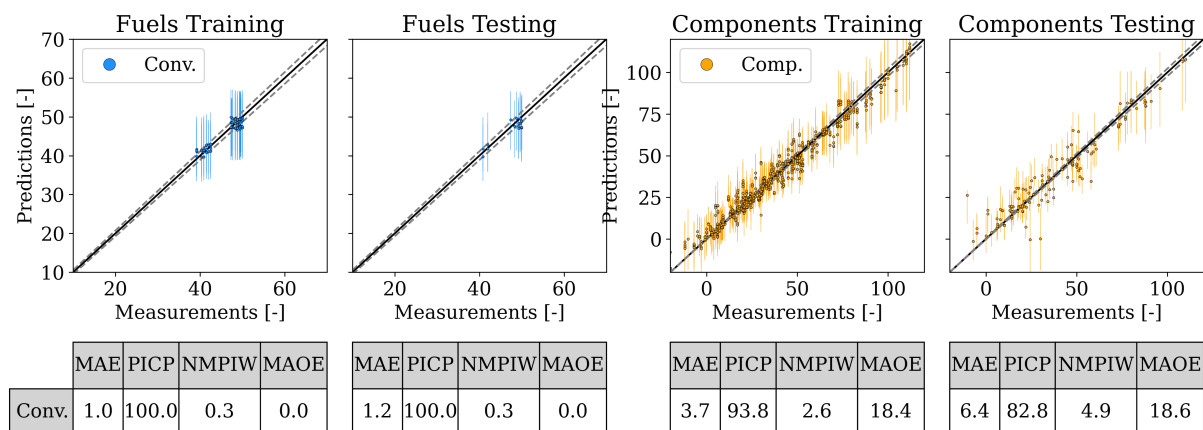


Figure E 34: Cross-validation results training and testing for cetane number prediction of the *M-QSPR -no Syn.* model, MAE [-], PICP [%], NMPIW [%], MAOE [-].

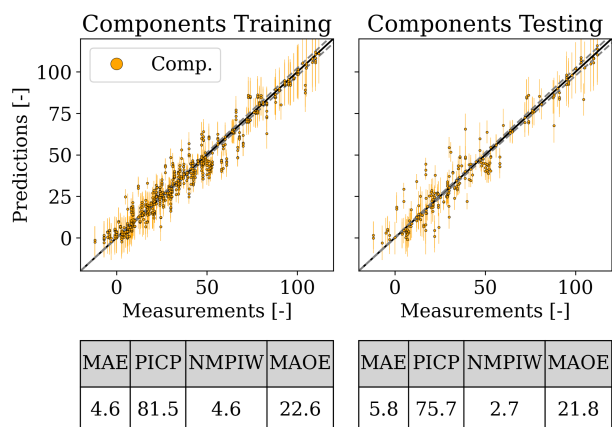


Figure E 35: Cross-validation results training and testing for cetane number prediction of the *QSPR* model, MAE [-], PICP [%], NMPIW [%], MAOE [-].

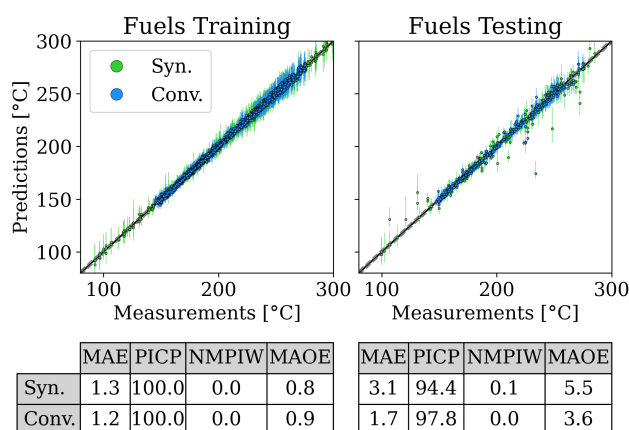


Figure E 36: Cross-validation results training and testing for distillation prediction of the *DC* model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

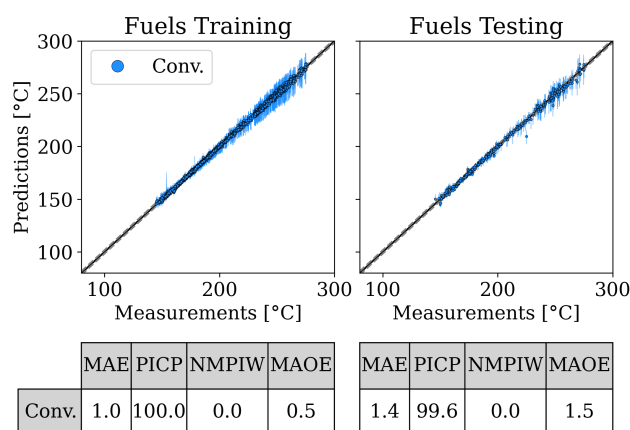


Figure E 37: Cross-validation results training and testing for distillation prediction of the DC - no Syn. model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

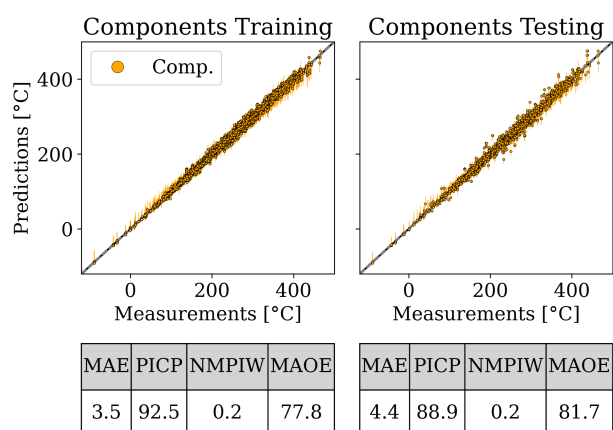


Figure E 38: Cross-validation results training and testing for distillation / boiling point prediction of the QSPR model, MAE [°C], PICP [%], NMPIW [%], MAOE [°C].

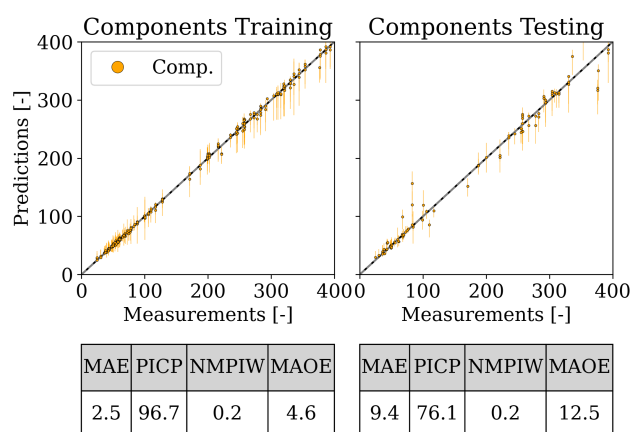


Figure E 39: Cross-validation results training and testing for yield sooting index point prediction of the QSPR model, MAE [-], PICP [%], NMPIW [%], MAOE [-].

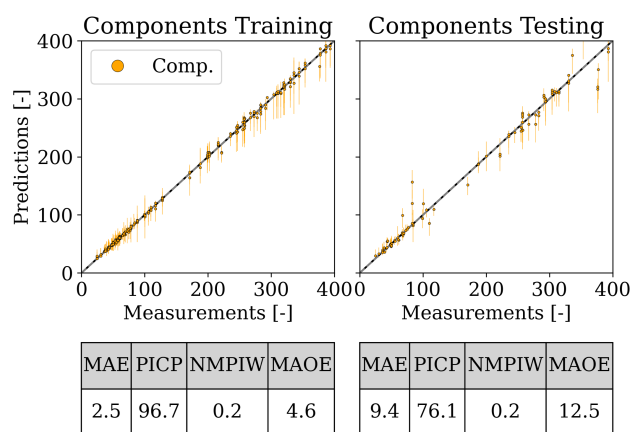


Figure E 40: Cross-validation results training and testing for yield sooting index point prediction of the M-QSPR model, MAE [-], PICP [%], NMPIW [%], MAOE [-].

F. Reference Models

Discrete Component Model by Le Clercq

The Discrete Component Model by Le Clercq [45,47] models a fuel as a mixture of components with one representative species for every hydrocarbon family and carbon number. A summary of the considered components is given in Table F.1. In a first step, the individual properties of the representative species are calculated on the basis of critical thermodynamic constants: critical temperature T_{ci} , critical pressure P_{ci} and acentric factor ω_i . The bulk property of the fuel is subsequently calculated using mixing rules. For components, where the necessary critical constants are unavailable, the constants are predicted using the Group-Contribution methods of Joback as well as Constantinou and Gani [188].

For the density, the molar volume V_m is calculated using Equation (F.1) to Equation (F.3). The molar volume V_m is first estimated on the basis of the Racket's compressibility factor Z_{RAM} for the fuel, which is given by the weighted sum of the Racket's compressibilities Z_{RAi} of each component, see Equation (F.3).

$$V_m = R \left(\sum_i \frac{x_i T_{ci}}{P_{ci}} \right) Z_{RAM}^{[1+(1-T_r)^{0.2857}]} \quad (\text{F.1})$$

$$Z_{RAM} = \sum x_i Z_{RAi} \quad (\text{F.2})$$

$$Z_{RAi} = 0.29056 - 0.0877\omega_i \quad (\text{F.3})$$

The kinematic viscosity is calculated from the dynamic viscosity μ based on the relation of Mehrotra with Equation (F.4) [108]. Exponent b is regressed using the coefficients B_0 , B_1 and B_2 together with the molar mass M according to Mehrotra in [108]. The coefficients thereby depend on the chemical families of the components. As mixing rule for the calculation of the bulk viscosity of the fuel, the equation after Katti and Chaudhri is utilized, see Equation (F.6) [189]

$$\log(\mu + 0.8) = 100(0.01T)^b \quad (\text{F.4})$$

$$b = B_0 + B_1[\log M] + B_2[\log M]^2 \quad (\text{F.5})$$

$$\ln \mu_m V_m = \sum x_i \ln \mu_i V_i \quad (\text{F.6})$$

Multilinear Regression (MLR) models by Liu et al.

The multilinear regression models (MLR) by Liu et al. are direct correlation models that correlate the summed up mass fraction w in % of the respective families with the properties net heat of combustion (HOC), flash point (T_{flash}) and freezing point (T_{freeze})[105]. The models are given in Equation (F.7) to Equation (F.9).

Net heat of combustion

$$\begin{aligned} HOC = & 2.18 * w_{n-alkanes} + 1.23 * w_{iso-alkanes} \\ & + 2.18 * w_{cyclo-alkanes} + 2.18 * w_{aromatics} \\ & + 41.76 \end{aligned} \quad (F.7)$$

Flash point

$$\begin{aligned} T_{flash} = & 53.88 * w_{n-alkanes} - 32.53 * w_{iso-alkanes} \\ & - 10.91 * w_{cyclo-alkanes} - 3.79 * w_{aromatics} \\ & - 50.86 \end{aligned} \quad (F.8)$$

Freezing point

$$\begin{aligned} T_{freeze} = & -46.14 * w_{n-alkanes} - 55.21 * w_{iso-alkanes} \\ & - 41.77 * w_{cyclo-alkanes} - 98.96 * w_{aromatics} \\ & + 126.25 \end{aligned} \quad (F.9)$$

Table F 1: Selected components for the DCM modeling.

Carbon number	N-alkanes	Iso-alkanes	Mono-cyclo-alkanes	Bi-cyclo-alkanes	Mono-aromatics	Cyclo-aromatics	Di-aromatics
1	methane						
2	ethane						
3	propane						
4	butane						
5	pentane	neopentane (2,2-dimethylpropane)	cyclopentane				
6	hexane	3-methylpentane	cyclohexane		benzene		
7	heptane	neoheptane (2,2-dimethylpentane)	ethylcyclopentane (and NOT cycloheptane)	norbornane (bicyclo[2.2.1]heptane)	toluene (methylbenzene)		
8	octane	isooctane (2,2,4-trimethylpentane)	t-1,4-dimethylcyclohexane (and NOT cyclooctane)	2-methylnorbornane (2-methylbicyclo[2.2.1]heptane)	ethylbenzene		
9	nonane	isononane (2-methyl octane)	propylcyclohexane	2-ethylbicyclo[2.2.1]heptane	propylbenzene	indane (2,3-dihydro-1H-indene)	
10	decane	2,2,5,5-tetramethylhexane	pentylcyclopentane	2,6,6-trimethylbicyclo[3.1.1]heptane	butylbenzene	tetralin (1,2,3,4-tetrahydronaphthalene)	naphthalene

11	undecane	2,2,3,4-tetramethylheptane	hexylcyclopentane	1-methyldecalin	pentylbenzene	1,1-dimethylindane (2,3-dihydro-1,1-dimethyl-1H-indene)	2-methylnaphthalene
12	dodecane	2,2,4,6,6-pentamethylheptane	heptylcyclopentane	cis-dimethyldecalin	hexylbenzene	cyclohexylbenzene	1,6-dimethylnaphthalene
13	tridecane	2,2,3-trimethyldecane	octylcyclopentane	4a-methyl-2-ethyldecalin	heptylbenzene	1-isopropyl-6-methylindane	1-n-propylnaphthalene
14	tetradecane	isotetradecane (2-methyltridecane)	nonylcyclopentane	alpha-sec-butyldecalin	octylbenzene	1-butyltetralin	1-butyl-naphthalene
15	pentadecane	2,5-dimethyltridecane	decylcyclopentane	2-(1-methylethyl)-1,1-bicyclohexyl	nonylbenzene	6-(tert-butyl)-1,1-dimethylindane	1-pentyl-naphthalene
16	hexadecane	2,6,11-trimethyltridecane	decylcyclohexane	2-butyl-1,1-bicyclohexyl	pentaethylbenzene	(trans-4-butylcyclohexyl)benzene	
17	heptadecane	2,6,10-trimethyltetradecane	dodecylcyclopentane	1,1-dicyclohexylpentane	undecylbenzene	1-ethyl-4-(trans-4-propylcyclohexyl)benzene	

18	octadecane	2,6,10-trimethylpentadecane	tridecylcyclopentane	2,4-dicyclohexyl-2-methylpentane	hexaethylbenzene		
19	nonadecane	pristane (2,6,10,14-tetramethylpentadecane)	tridecylcyclohexane		tridecylbenzene		
20	eicosane	phytane (2,6,10,14-tetramethylhexadecane)	tetradecylcyclohexane				
21	heneicosane		hexadecylcyclopentane				
22	docosane		hexadecylcyclohexane				
23	tricosane		octadecylcyclopentane				
24	tetracosane		octadecylcyclohexane				
25	pentacosane						

G. Pure Compound Descriptor Plots

Density

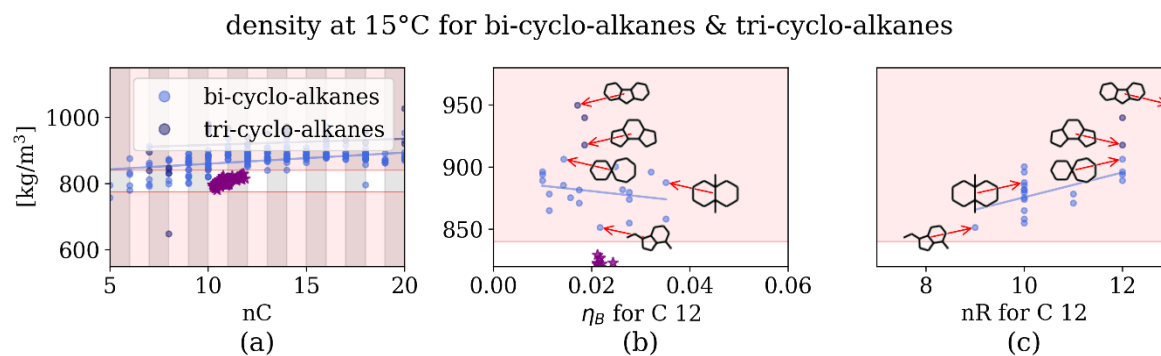


Figure G 1: Density values at 15 °C of bi-cyclo-alkanes and tri-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b) and number of ring atoms nR (c).

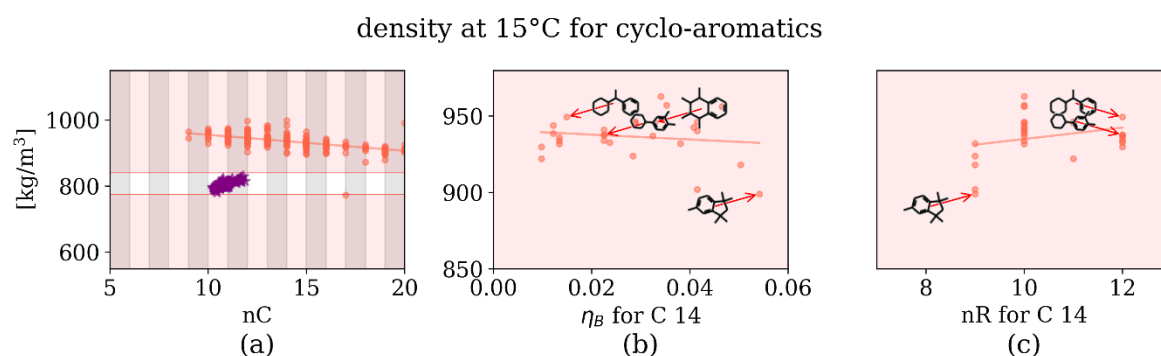


Figure G 2: Density values at 15 °C of cyclo-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b) and number of ring atoms nR (c).

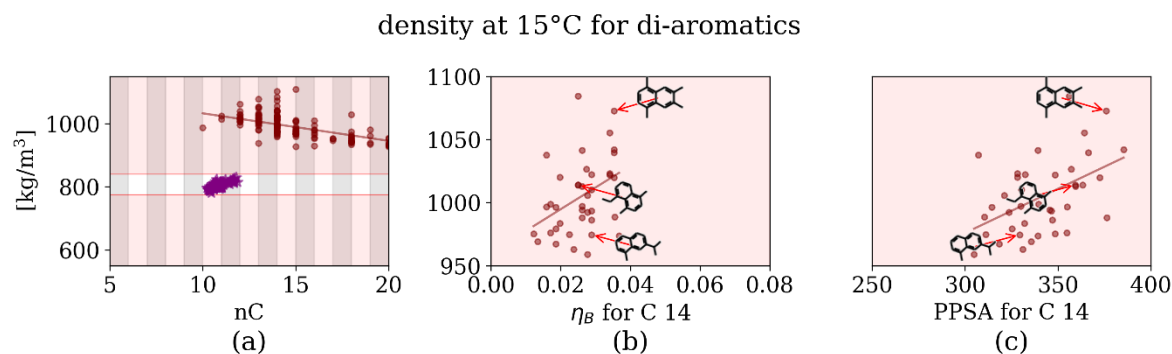


Figure G 3: Density values at 15 °C of di-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b) and number of ring atoms nR (c).

Surface tension

surface tension at 22°C for bi-cyclo-alkanes & tri-cyclo-alkanes

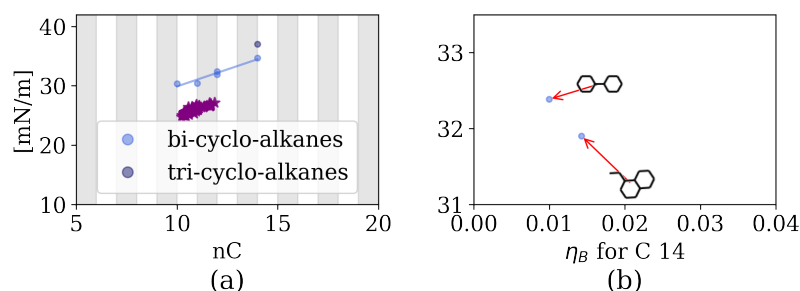


Figure G 4: Surface tension values of bi-cyclo-alkanes and tri-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

surface tension at 22°C for cyclo-aromatics

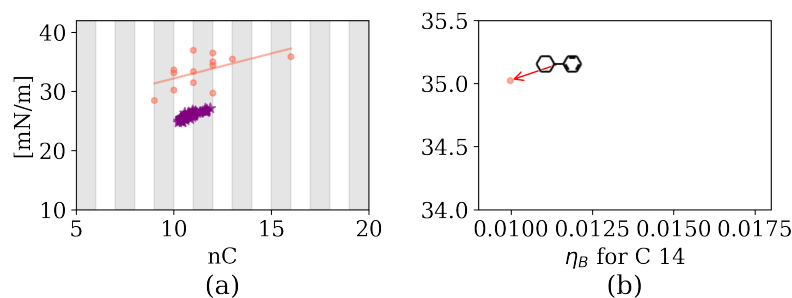


Figure G 5: Surface tension values of cyclo-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

surface tension at 22°C for di-aromatics

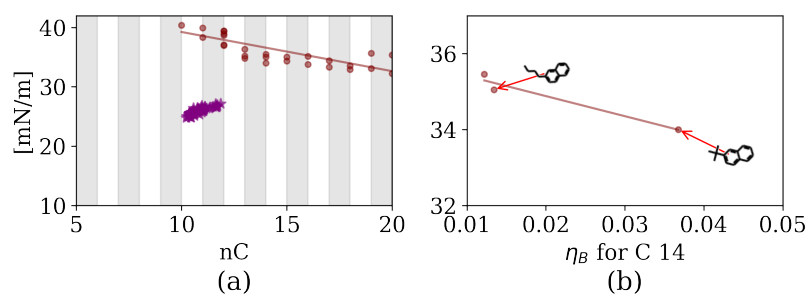


Figure G 6: Surface tension values of di-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

Kinematic viscosity

kinematic viscosity at 0°C for bi-cyclo-alkanes & tri-cyclo-alkanes

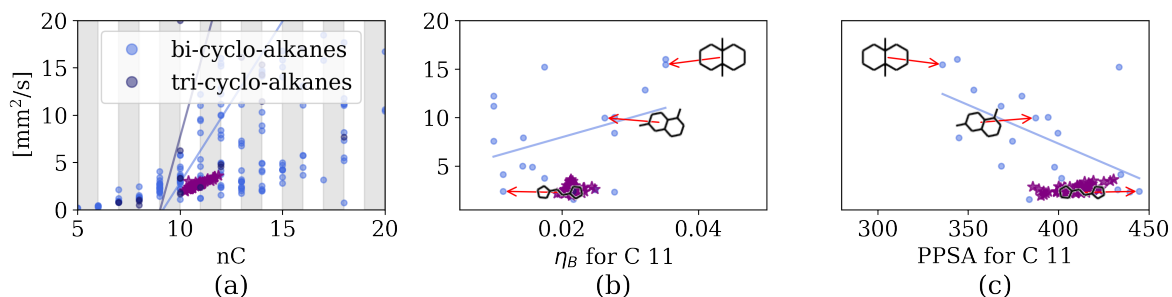


Figure G 7: Kinematic viscosity values of bi-cyclo-alkanes and tri-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b) and partial positive surface area PPSA (c).

kinematic viscosity at 0°C for cyclo-aromatics

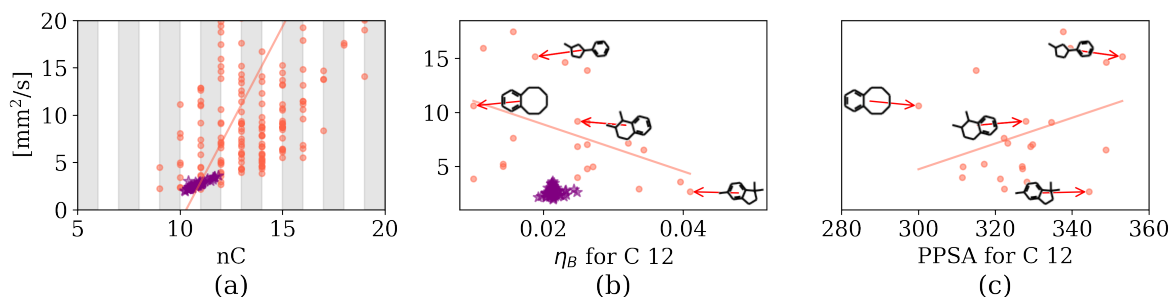


Figure G 8: Kinematic viscosity values of cyclo-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b) and partial positive surface area PPSA (c).

kinematic viscosity at 0°C for di-aromatics

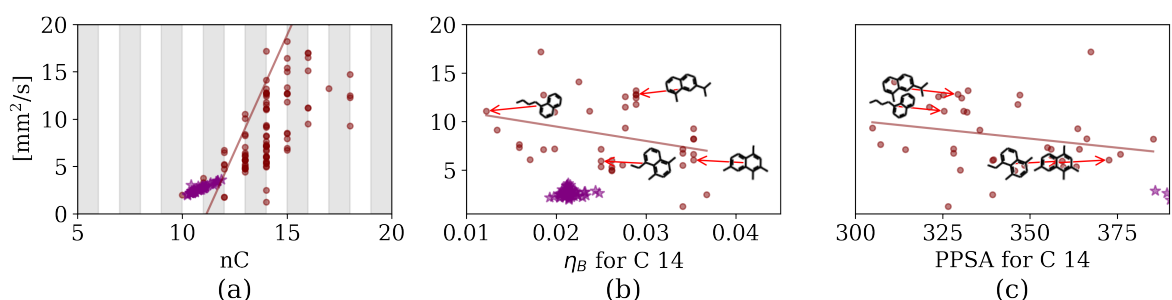


Figure G 9: Kinematic viscosity values of di-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b) and partial positive surface area PPSA (c).

Net heat of combustion

net heat of combustion for bi-cyclo-alkanes & tri-cyclo-alkanes

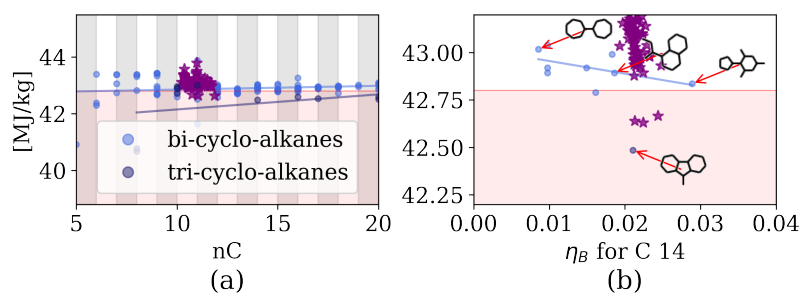


Figure G 10: Net heat of combustion values of bi-cyclo-alkanes and tri-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

net heat of combustion for cyclo-aromatics

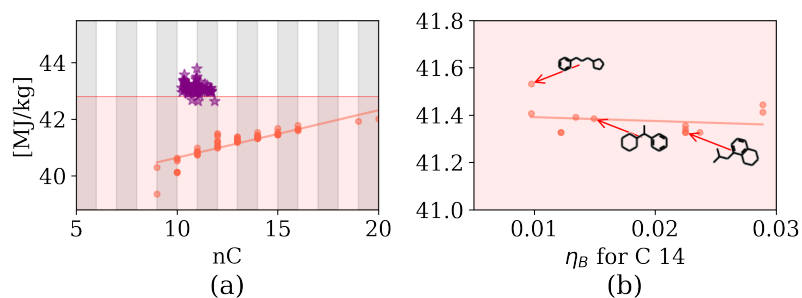


Figure G 11: Net heat of combustion values of mono-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

net heat of combustion for di-aromatics

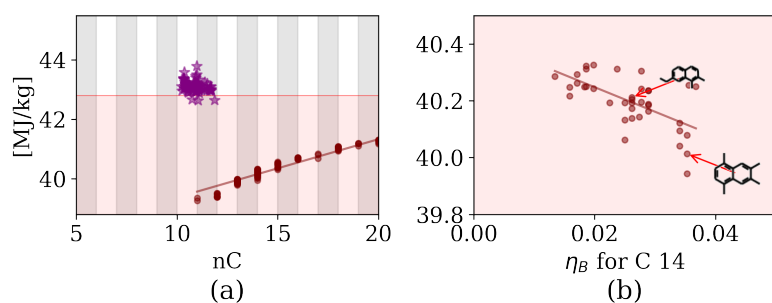


Figure G 12: Net heat of combustion values of di-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

Flash point

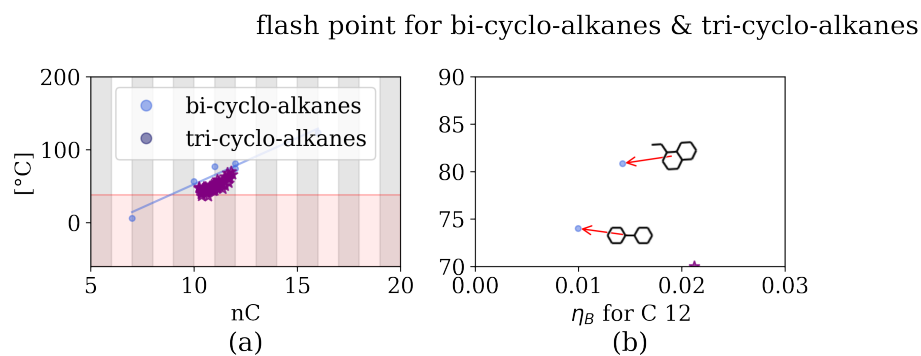


Figure G 13: Flash point values of bi-cyclo-alkanes and tri-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

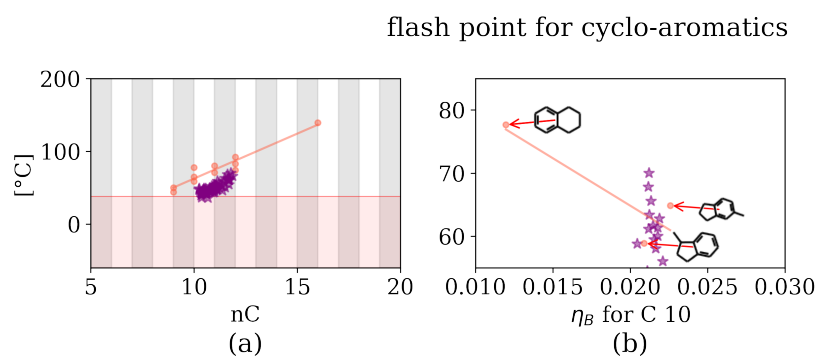


Figure G 14: Flash point values of cyclo-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

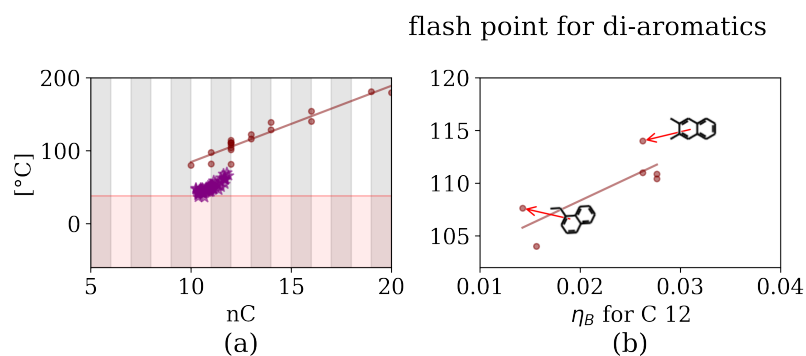


Figure G 15: Flash point values of di-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

Freezing point

freezing point for bi-cyclo-alkanes & tri-cyclo-alkanes

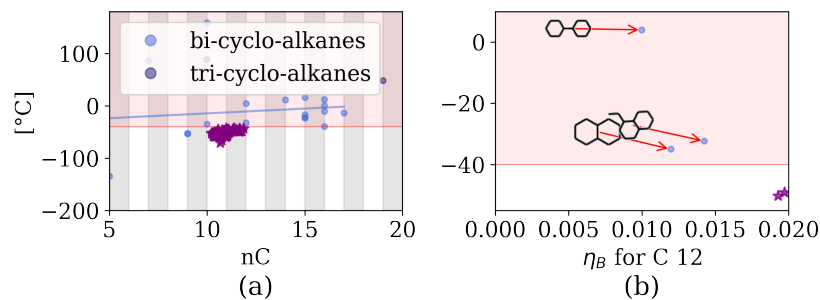


Figure G 16: Freezing point values of bi-cyclo-alkanes and tri-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

freezing point for cyclo-aromatics

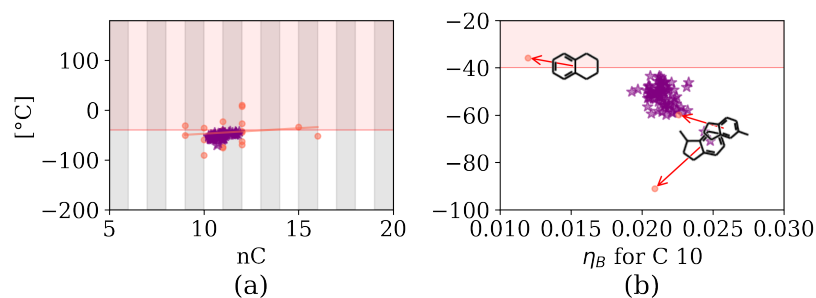


Figure G 17: Freezing point values of cyclo-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

freezing point for di-aromatics

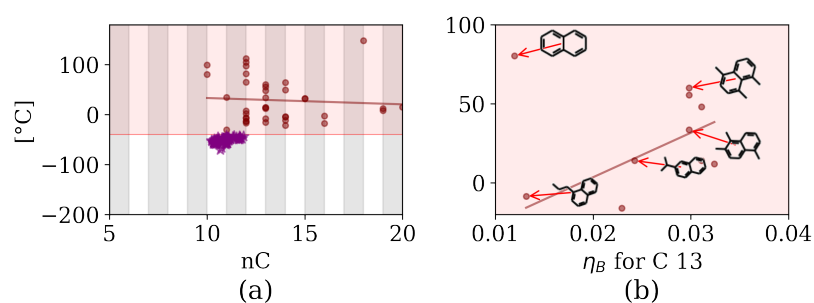


Figure G 18: Freezing point values of di-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

Boiling point

boiling point for bi-cyclo-alkanes & tri-cyclo-alkanes

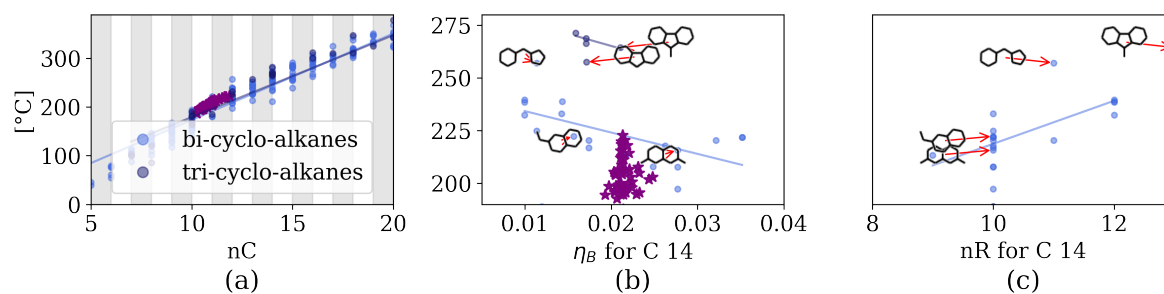


Figure G 19: Boiling point values of bi-cyclo-alkanes and tri-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b) and number of ring atoms nR (c).

boiling point for cyclo-aromatics

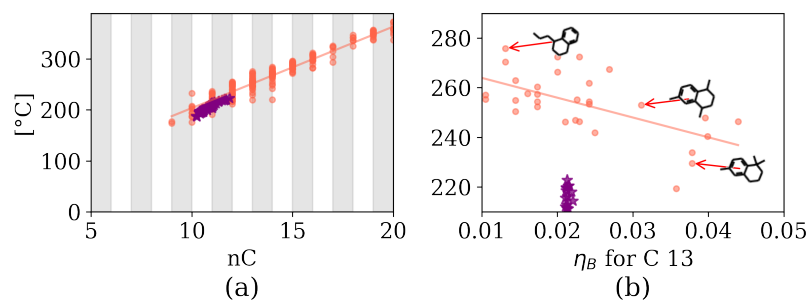


Figure G 20: Boiling point values of cyclo-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

boiling point for di-aromatics

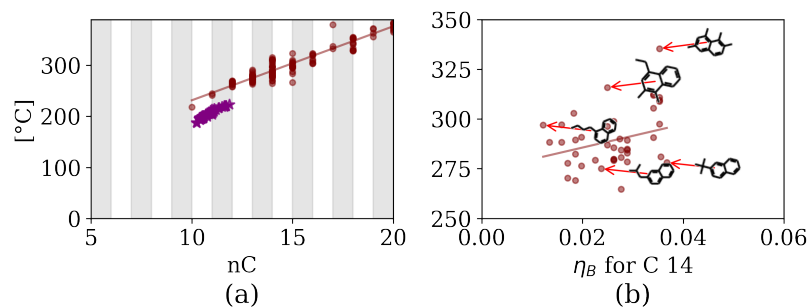


Figure G 21: Boiling point values of di-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

Cetane number

cetane number for bi-cyclo-alkanes & tri-cyclo-alkanes

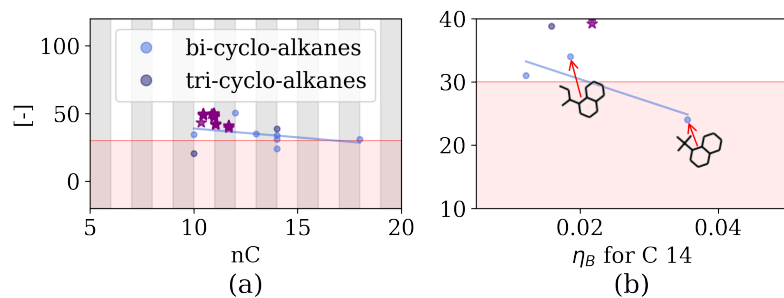


Figure G 22: Cetane number values of bi-cyclo-alkanes and tri-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b)

cetane number for cyclo-aromatics

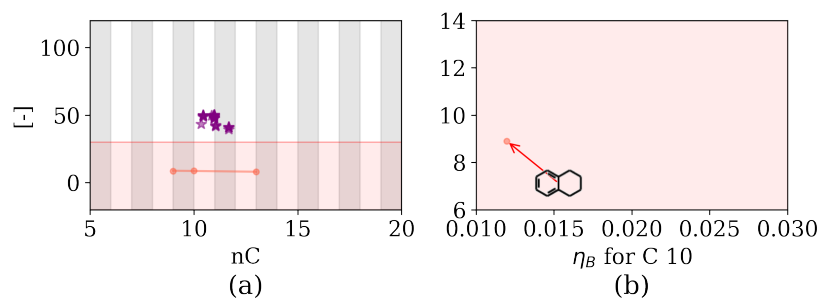


Figure G 23: Cetane number values of cyclo-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

cetane number for di-aromatics

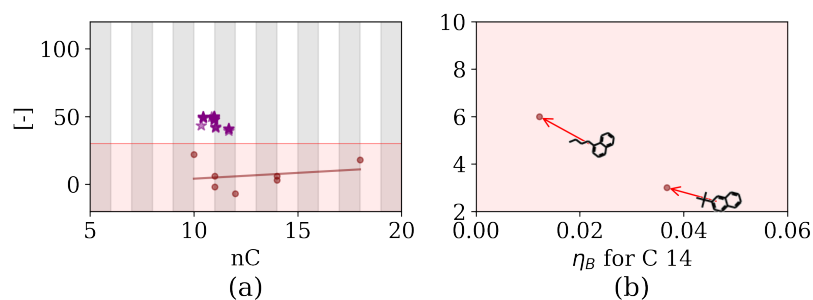


Figure G 24: Cetane number values of di-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

Yield sooting index

yield sooting index for bi-cyclo-alkanes & tri-cyclo-alkanes

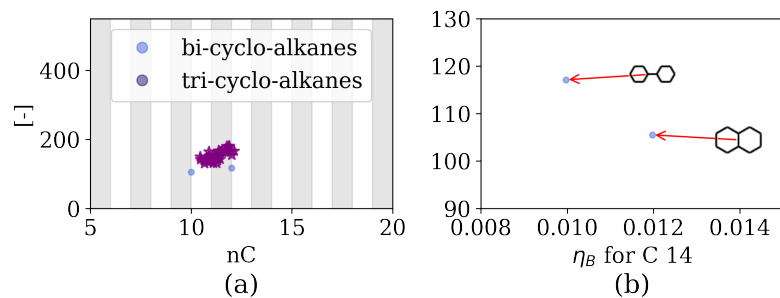


Figure G 25: Yield sooting index values of bi-cyclo-alkanes and tri-cyclo-alkanes over molecular descriptors: carbon number nC (a), branching index η_B (b).

yield sooting index for cyclo-aromatics

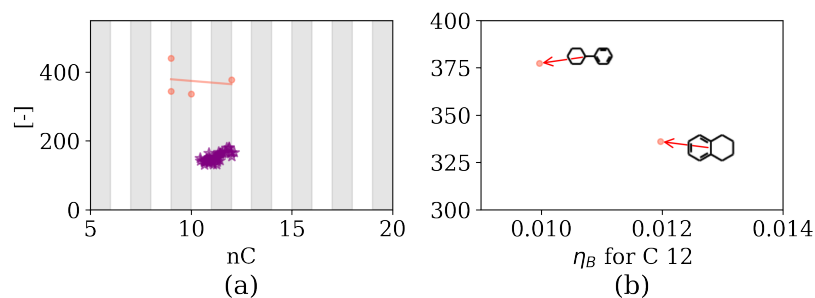


Figure G 26: Yield sooting index values of cyclo-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).

yield sooting index for di-aromatics

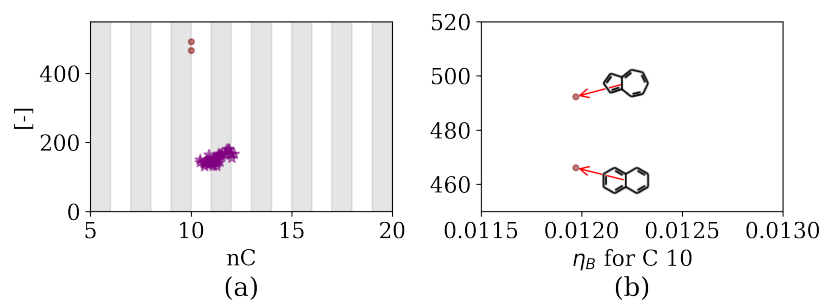


Figure G 27: Yield sooting index values of di-aromatics over molecular descriptors: carbon number nC (a), branching index η_B (b).