Calculation of Pure Substance and Mixture Viscosities using PCP-SAFT and Entropy Scaling

Von der Fakultät Energie-, Verfahrens- und Biotechnik der
Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Ingenieurwissenschaften (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von
Oliver Lötgering-Lin
aus Stuttgart

Hauptberichter:	Prof. Dr.-Ing. Joachim Groß
Mitberichter:	Prof. Dr.-Ing. André Bardow

Tag der mündlichen Prüfung: 2019/11/18

Institut für Technische Thermodynamik und Thermische
Verfahrenstechnik der Universität Stuttgart
2020

Eidesstattliche Erklärung zu meiner Dissertation mit dem Titel:

Calculation of Pure Substance and Mixture Viscosities using PCP-SAFT and Entropy Scaling

Hiermit erkläre ich, dass ich die beigefügte Dissertation selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel genutzt habe. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet.

Ich versichere außerdem, dass ich die beigefügte Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.

Ort, Datum	Unterschrift

Abstract

Reliable and fast models for the description of transport properties of pure substances and mixtures are needed in many engineering disciplines; some require information about transport properties over wide ranges of temperature, pressure and concentration (like carbon capture and storage or the design of biofuels) while others require transport properties for a range of different substances (like the design of solvents or working fluids in process optimization). The Entropy Scaling approach has the potential to cover many of those needs in a charmingly simple framework originally introduced by Rosenfeld [Rosenfeld, Y. Phys. Rev. A 1977, 15, 2545-2549]. Applying the Entropy Scaling approach, dimensionless transport properties are described as a function of the residual entropy only. In this thesis, the needed residual entropy is calculated with the *Perturbed Chain Polar Statistical Associating Fluid Theory* (PCP-SAFT).

One aim of this thesis is the development of a model that describes the viscosity of pure substances in a predictive yet accurate manner. Rosenfeld's approach is extended beyond the originally regarded model fluids (like Lennard-Jones fluids): Non-polar, polar, self-associating (i.e. hydrogen-bonding) as well as strongly non-spherical substances are considered. To describe the dependence of the dimensionless viscosity on the residual entropy, a third-order polynomial is introduced. To investigate chain length dependencies and promote predictability, homologous series are investigated and a group contribution approach is established. In total, parameters of 22 functional groups are adjusted to viscosity data of 110 pure substances. Relative deviations between predicted viscosities and literature data are about 5% on average for non-associating substances. Self-associating substances like 1-alcohols and amines show higher deviations of about 10% as an average.

During the development of the proposed model, several empirical ansatz functions were investigated. To objectively rank competing models and, by that, guide the decision making process, a statistical approach (Bayesian Model Selection, BMS) based on Bayesian Statistics is applied. The goal is to select a model that provides a sound tradeoff between robustness (i.e. extrapolation) and accuracy (i.e. low deviations in reproducing experimental data), known as the bias-variance dilemma. To this end, a Markov Chain Monte Carlo method was used to determine parameter distributions based on given experimental data. These parameter distributions are then used to evaluate the competing

models for different applications: First, prediction of high pressure viscosities with parameters obtained from experimental data at low pressures and, second, prediction of viscosities of long n-alkanes based on experimental visosities of short n-alkanes. The results suggest that the proposed polynomial approach provides a solid tradeoff for both applications, though, the model ranking is dependent on the experimental error of the viscosity measurements.

The proposed model is extended to allow predictions of binary as well as multicomponent mixtures. To develop a suitable mixture model, results of simulations of mixtures of simple model fluids are used. The resulting model is able to predict real mixture viscosities, based only on pure component data (i.e. pure component parameters and no binary interaction parameters), accurately with deviations of 6.2% for mixtures of non-polar substance and 5% for mixtures with at least one polar but non-hydrogen-bonding component. Mixtures containing associating components, especially short alcohols like methanol or ethanol, show significantly higher deviations. Prediction of mixture viscosities with pure component parameters obtained via the proposed group contribution approach leads to slightly higher deviations. Overall, 566 mixtures with about 34500 experimental data points from literature were considered.

Due to the semi-empirical nature of the proposed model, experimental data is needed for parametrization. An optimal experimental design framework is applied to define scenario-specific optimal experimental conditions (i.e. substance, pressure and temperature) with the goal to minimize model uncertainty. Two example-applications are defined that each consist of initial knowledge (experimental viscosity that already exists), a designspace in which measurements are carried out and a predictionspace that defines the range for which model uncertainty is to be minimized. The first presented example-application focuses on pressure extrapolation while the second example-application makes use of the proposed group contribution approach. The applied Preposterior Data Impact Assessor (PreDIA) [Leube et al., Water Resour. Res. 2012, 48(2)] is based on Bayesian Statistics and serves as a purely predictive approach to Optimal Experimental Design (OED). Results suggest that PreDIA is capable to accurately predict the impact of additional data on model accuracy in the first example-application. In the second, more complex example-application, PreDIA is not able to accurately predict the absolute reduction of model variance. Though, trends are represented satisfyingly well, which is sufficient as a guideline for experimental design. The application of a Bayesian Statistics

based OED framework is shown to be a useful, purely predictive way to define experimental conditions optimal to determine either pure substance parameters or group contribution parameters. In the present case of a predictive viscosity model it could prove especially helpful to efficiently fine-tune pure substance parameters for systems of interest that were identified in a preceding screening study (e.g. with less accurate but purely predictive parameters stemming from the group contribution approach).

Zusammenfassung

Zuverlässige und schnelle Modelle zur Vorhersage von Transportgrößen von Reinstoffen und Mischungen werden in vielen Ingenieursdisziplinen benötigt; je nach Anwendung beispielsweise für einen großen Temperatur-, Druck- und Konzentrationsbereich (bspw. Carbon Capture and Storage oder Design von neuen Treibstoffen) oder zur Vorhersage von Transportgrößen bisher nicht untersuchter Substanzen (bspw. zur Identifikation geeigneter Lösungsmittel). Mittels des in dieser Arbeit untersuchten und weiterentwickelten Entropieskalierungsansatzes lassen sich viele dieser Anforderungen schnell und einfach erfüllen. Ursprünglich wurde der Entropieskalierungsansatz von Rosenfeld vorgeschlagen [Rosenfeld, Y. Phys. Rev. A 1977, 15, 2545-2549]. Mittels des Entropieskalierungsansatzes lassen sich Transportgrößen einzig als Funktion der residuellen Entropie ausdrücken. Zur Berechnung der benötigten residuellen Entropie wird hier die *Perturbed Chain Polar Statistical Associating Fluid Theory* (PCP-SAFT) verwendet.

Das vorgeschlagene Modell dient der prädiktiven Beschreibung der Viskosität von Reinstoffen. Dazu wird Rosenfeld's Ansatz, über die ursprünglich betrachteten Modellfluide (bspw. Lennard-Jones Fluide) hinaus, auf reale Systeme angewandt. Es werden unpolare, polare, selbstassoziierende sowie stark nicht-sphärische Substanzen betrachtet. Zur Beschreibung der Abhängigkeit der Viskosität von der residuellen Entropie wird ein Polynomansatz dritter Ordnung verwendet. Weiterhin wird ein Gruppenbeitragsmodell entwickelt, das es erlaubt, Viskositäten auch für Substanzen zu bestimmen, für welche keine experimentellen Daten vorliegen. Dazu wurden homologe Reihen von Substanzen untersucht (beispielsweise n-Alkane). Insgesamt wurden experimentelle Daten von 110 Substanzen verwendet um 22 funktionelle Gruppen zu parametrisieren. Relative Abweichungen zwischen den so prädiktiv bestimmten und experimentellen Viskositäten liegen bei nicht-selbstassoziierenden Substanzen im Mittel bei 5%. Selbst-assoziierende Substanzen wie 1-Alkohole und Amine werden mit einer Abweichung von ca. 10% im Mittel beschrieben.

Während der Modellentwicklung wurden neben dem vorgestellten Polynomansatz weitere empirische Ansatzfunktionen in Betracht gezogen. Mittels einem auf Bayes'scher Statistik basierenden Ansatz (*Bayesian Model Selection*, BMS) werden die unterschiedlichen Modelle objektiv verglichen. Ziel ist es, das Modell zu identifizieren, welches unter Berücksichtigung experimenteller Fehler den besten Kompromiss zwischen zuverlässigen (d.h. robuste Extrapolation

bzw. geringe Modellunsicherheit) und gleichzeitig exakten Vorhersagen (d.h. geringe Abweichungen zu experimentellen Daten) liefert. Hierfür werden Markov Chain Monte Carlo Berechnungen durchgeführt um Parameterverteilungen an experimentelle Daten anzupassen. Mittels dieser Parameterverteilungen werden unterschiedliche Modelle für zwei unterschiedlichen Anwendungsfälle verglichen: Erstens, eine Extrapolation von experimentellen Viskosiäten bei niedrigen Drücken zu Viskositäten bei hohen Drücken. Zweitens, die Vorhersage der Viskosität von langkettigen Molekülen basierend auf experimentellen Daten kurzkettiger Moleküle chemisch ähnlicher Substanzen. Die Ergebnisse zeigen, dass der gewählte Polynomansatz einen guten Kompromiss zwischen robusten sowie genauen Extrapolationen darstellt.

Weiterhin wird eine Modellerweiterung für Mischungsviskositäten entwickelt. Um die vorgestellten Modellgleichungen zur Berechnung der Viskositäten von Reinstoffen um Terme zur Berechnung der Viskosität von Mehrstoffsystemen zu erweitern, wurden mittels Molekularsimulation ermittelte Viskositäten binärer Mischungen unterschiedlicher Modellfluide verwendet. Mit dem so entwickelten Modell lassen sich Mischungsviskositäten einzig aus experimentellen Viskositäten der entsprechenden Reinstoffe und ohne weitere Binärparameter vorhersagen. Insgesamt wurden 566 Mischungen und 34500 experimentelle Datenpunkte (Literaturdaten) untersucht. Binäre Mischungen unpolarer Substanzen können mit einer mittleren relativen Abweichung von 6.2%, binäre Mischungen mit mindestens einer polaren Substanz mit einer mittleren relativen Abweichung von 5% vorhergesagt werden. Mischungen die assoziierende Substanzen (insbesondere Methanol und Ethanol) enthalten, zeigen signifikant höhere Abweichungen. Weiter wird gezeigt, dass die entwickelten Gruppenbeitragsparameter auch zur Vorhersage von Mischungsviskositäten verwendet werden können.

Für (Semi-) Empirische Modelle, wie das in dieser Arbeit entwickelte Modell, werden experimentelle Daten zur Parametrierung benötigt. Die Wahl der experimentellen Bedingungen (Substanz, Druck, Temperatur sowie bei Mischungen deren Zusammensetzung), bei welchen die entsprechenden Viskositäten bestimmt werden, hat einen signifikanten Einfluss auf die entsprechenden Parameter und damit auch auf die Güte der Modellvorhersagen. Es wird gezeigt, dass sich der Einfluss von experimentellen Daten (Data-Impact) auf Modellunsicherheiten (Varianz von Modellvorhersagen bei Verwendung unterschiedlicher Parametersätze) mittels dem *Preposterior Data Impact Assessor*

(PreDIA) [Leube et al., Water Resour. Res. 2012, 48(2)] abschätzen lässt und sich so optimale experimentelle Bedingungen zur Minimierung der Modellunsicherheit vorhersagen lassen. Es werden wiederum zwei Beispielanwendungen definiert und untersucht: Erstens, die Vorhersage von Hochdruckviskositäten auf Grundlage von Daten bei niedrigen Drücken und, zweitens, die Vorhersage von Viskositäten nicht untersuchter Substanzen basierend auf Daten chemisch ähnlicher Stoffe. Im ersten Szenario ist PreDIA in der Lage den Einfluss experimenteller Daten auf die Modellunsicherheit gut vorherzusagen. Im zweiten, komplexeren Anwendungsfall werden lediglich Trends gut vorhergesagt, nicht aber die absolute Verringerung der Modellunsicherheit. Insgesamt wird gezeigt, dass der auf Bayes'scher Statistik aufbauende Ansatz PreDIA zur rein prädiktiven Bestimmung von optimalen Versuchsbedingungen mit Blick auf Modellparametrisierung genutzt werden kann. Dies kann beispielsweise besonders hilfreich sein, um eine sich aus einer Screening-Studie ergebende Zielsubstanz effizient für weiterführende Betrachtungen neu zu parametrieren.

Journal publications

This thesis led to the following publications:

- Chapter 2: Lötgering-Lin, Gross: A group contribution method for viscosities based on entropy scaling using the perturbed-chain polar statistical associating fluid theory, Ind. Eng. Chem. Res., 54(32), 2015, 7942-7952
- Chapter 3: Lötgering-Lin, Schöniger, Nowak, Gross: Bayesian Model Selection Helps To Choose Objectively between Thermodynamic Models: A Demonstration of Selecting a Viscosity Model Based on Entropy Scaling, Ind. Eng. Chem. Res., 55(38), 2016, 10191-10207
- Chapter 4: Lötgering-Lin, Fischer, Hopp, Gross: Pure Substance and Mixture Viscosities Based on Entropy-Scaling and an Analytic Equation of State, Ind. Eng. Chem. Res., 57(11), 2018, 4095-4114

The chapters 2 to 4 present literal quotes of the published work. The Supporting Information to the single chapters are presented in the Appendix of this thesis.

Contents

Nomenclature

Abbreviations

AAD Average Absolute Relative Deviation. Used equivalently to MAD

f-theory Friction Theory

MAD Mean Absolute Relative Deviation. Used equivalently to AAD

$MBIAS$ Measurement bias, defined in eq. 3.27

$WMSE$ Weighted mean squared error, defined in eq. 3.28

$WMVAR$ Weighted mean variance, defined in eq. 3.29

AAD% Absolute average deviations in percent

BF Bayes factor, see eq .3.22

BF Bootstrap filter

BMA Bayesian Model Averaging

BME Bayesian Model Evidence

BMS Bayesian Model Selection

cfc chlorofluorohydrocarbons

D Designspace

DA Dymond and Assael viscosity model

DI Data impact

EoS Equation of State

FVT Free-Volume Theory

GA Genetic Algorithm

gc Group contribution

hfc Hydrofluorocarbons

IAPWS-95 Formulation from 1995 for properties of water and steam by the International Association for Properties of Water and Steam

LJ Lennard-Jones

MCMC Markov Chain Monte Carlo

MD Molecular Dynamic

OED Optimal Experimental Design

P Predictionspace

PC-SAFT Perturbed Chain Statistical Associating Fluid Theory

PCP-SAFT Perturbed Chain Polar Statistical Associating Theory

PreDIA Preposterior Data Impact Assessor

SAFT Statistical Associating Fluid Theory

std Relative standard deviation

VLE Vapor-Liquid-Equilibrium

VW Vesovic and Wakeham viscosity model

Greek Symbols

χ "Meta-parameters", i.e. parameters that are not modified in the respective study, see chapter 6

Δ_0 Vector containing residuals between a given data set 0 and model results

δ_i Relative deviation of a datapoint i

Δ_θ Maximum step size in parameter set θ in MCMC context

η Dynamic viscosity

η^* Dimensionless viscosity with Chapman-Enskog viscosity as reference

η^+ Generic dimensionless viscosity with reference to be specified, e.g. η_{R}

η_{R} Reference viscosity as defined by Rosenfeld[1]

$\eta_{\mathrm{CE,mix}}$ Chapman-Enskog viscosity for a mixture

η_{CE} Chapman-Enskog viscosity[2]

η_{ref} Generic reference viscosity

γ Adjustable parameter in eq. 2.13

γ Parameter in Isomorph Theory

κ^{AB} Association volume in SAFT EoS

μ Dipole moment in SAFT EoS

$\Omega^{(2,2)*}$ Reduced collision integral, (2,2) being the sonine polynomial expansion coefficients

Ω_k Parameter space of a model k

ϕ Intermolecular potential, here, Lennard-Jones Potential

ϕ Objective function in chapter 6

ϕ_{MCMC} Objective function in chapter 6, obtained by MCMC calculation

ϕ_{PreDIA} Objective function in chapter 6, obtained with PreDIA

ϕ_{wPost} Objective function in chapter 6, obtained by reweighting

ρ Number density

ρ_m Molar density

σ Particle diameter

σ Segment diameter in SAFT EoS

τ Shear Stress

θ Vector containing the viscosity model parameters, A_i to D_i

ε Dispersive energy parameter in SAFT EoS

ε Potential depth in the description of molecular potentials

$\varepsilon^{\mathrm{AB}}$ Association energy in SAFT EoS

ξ Dimensionless density

ζ Vector of substances in chapter 6

Latin Symbols

$\mathcal{R}$ Random number in MCMC context

$\overline{m}$ Averaged segment number in a mixture

A Area

A Helmholtz energy

a specific Helmholtz energy

A_i or A_α Viscosity model parameter of substance i or functional group α, respectively

B_i or B_α Viscosity model parameter of substance i or functional group α, respectively

C_i or C_α Viscosity model parameter of substance i or functional group α, respectively

c_p Isobaric heat capacity

c_v Isochoric heat capacity

d Vector with design parameters

D_i or D_α Viscosity model parameter of substance i or functional group α, respectively

d_{opt} Vector with optimized design parameters

e Errors; Experimental or Model

$E(\cdot)$ Expected value

F Force

$f_\alpha^{(3)}$ Objective function with three adjustable parameters as defined in chapter 2.3.1

$f_\alpha^{(3,0)}$ Objective function with three adjustable parameters as defined in chapter 2.3.1

$f_\alpha^{(4)}$ Objective function with four adjustable parameters as defined in chapter 2.3.1

f_{lin} Generic linear function

k_B Boltzmann constant

k_{ij} Binary interaction parameter in SAFT EoS

L Likelihood

M Molar mass

m Segment number in SAFT EoS

M_{lin} Linear model in chapter 3

$M_{\text{pol,6p}}$, $M_{\text{pol,7p}}$ Polynomial model in chapter 3, with six or seven adjustable parameters, respectively

$M_{\text{sin,6p}}$, $M_{\text{sin,7p}}$ Sinusoidal model in chapter 3, with six or seven adjustable parameters, respectively

N Number of particles or molecules

n_α Number of functional groups α

N_A Avogadro constant

N_{MC} Number of realizations

N_{obs} Number of observations, i.e. hypothetical measurement outcomes

N_P Number of gridpoints in the Predictionspace

p	Pressure
$P(\cdot)$	Discrete probability distribution
$p(\cdot)$	Probability distribution of a continuous variable, in statistical context, e.g. chapter 3 or chapter 6
$p(\cdot\|\cdot)$	Conditional probability distribution
$p_{\text{trial}}^{\text{accept}}$	Acceptance probability in MCMC context
Q	Quadrupolar moment in SAFT EoS
R	Diagonal, covariance matrix used in the calculation of Likelihoods, see eq. 3.25
R	Ideal, universal gas constant
r_{ij}	Distance between two particles i and j
S	Augmented vector that contains s and χ, see chapter 6
s	Generic model realization, see chapter 6
s	Molar entropy
s^*	Dimensionless residual entropy as given in eq. 4.4 and eq. 4.7
s_{res}	Residual molar entropy
T	Temperature
v	Velocity
$Var(\cdot)$	Variance of the given distribution
W	Weight matrix in eq. 6.5
w_i	Vector of weights or probabilities
x_i	Mole fraction of component i
Z	Compressibility
z	Dimensionless residual entropy as given in eq. 2.11

Subscripts

α	Functional group
CE	Chapman-Enskog formulation
eval	Evaluation data set
nuis	Nuisances, i.e. "meta-parameters" or uncertain parameters of minor importance in the specific context
ref	Reference, to be specified
res	Residual property
train	Training data set
calc	Calculated value
crit	Critical state
exp	Experimental value
gc	Parameter obtained via a group contribution method
liq	Liquid state
sat	Saturation conditions, e.g. vapor pressure

Superscripts

assoc	Associative contribution
disp	Dispersive contribution
hc	Hard chain contribution
hs	Hard Sphere
ig	Ideal Gas contribution
meas	Measured values
polar	Polar contribution
pred	Predicted values

trial	Trial parameters in MCMC context
true	"True" values; hypothetical experimental value without experimental errors
D	Designspace
P	Predictionspace

1. Introduction

1.1 The dynamic shear viscosity of Newtonian fluids

In fluid dynamics, the motion of a fluid element can be described by three parts: Translation, strain (or deformation) and rotation. The rate of change of deformation relates the shear viscosity η to a stress. The total stress tensor $\bar{\bar{\sigma}}$ is linked to the pressure p and the deviatoric stress tensor $\bar{\bar{\tau}}$ by[3,4]

$$\bar{\bar{\sigma}} = -p\bar{\bar{I}} + \bar{\bar{\tau}} \tag{1.1}$$

where $\bar{\bar{I}}$ is the identity or unity matrix. Rotation does not contribute to deformation and is not part of the constitutive equation nor does it contribute to entropy production.

For Stokesian fluids, it is assumed that

- the deviatoric stress depends only on the local distribution of the velocity at the location of the considered fluid element.
- the fluid is homogeneous regarding the relation between stress and rate of strain.
- the fluid is isotropic.

Additionally, for fluids considered in this thesis, it is assumed that the relation between stress and rate of strain is linear, which defines a Newtonian fluid. Then, the deviatoric stress tensor can be written as

$$\bar{\bar{\tau}} = \eta \left[\nabla\bar{v} + (\nabla\bar{v})^T - \frac{2}{3}(\nabla\cdot\bar{v})\,\bar{\bar{I}} \right] + \lambda\,(\nabla\cdot\bar{v})\,\bar{\bar{I}} \tag{1.2}$$

where $\bar{v}$ describes the velocity field within the considered fluid and λ is the bulk viscosity *.

Focus of this thesis is the shear viscosity η. From a thermodynamicist's point of view what interests is the dependence of the viscosity on the temperature T and pressure p for pure substances and, for mixtures, also on composition. Industrial needs are similar and range from high pressure viscosity predictions, e.g. for Carbon Capture and Storage[5–7] applications or the design of fuels[8–11], to predictions of viscosities of various substances for screening or optimization studies[12,13]. The modeling of those dependencies and prediction of viscosities for pure Newtonian substances and mixtures is the main focus of this thesis. For brevity, in the remainder of this thesis, the term viscosity is used synonymously to Newtonian shear viscosity.

1.2 State of the art for correlating and predicting viscosities

There exist a wide variety of models for describing or predicting pure substance or mixture viscosities. An overview of case-specific, usually semi-empirical approaches (e.g. corresponding states methods[14]) to calculate viscosities are given in Poling et al.[15]. While the liquid phase is modeled differently in each model, many (in classic as well as „modern“ approaches) make in some way use of the Chapman-Enskog viscosity.

The Chapman-Enskog viscosity is an approximate solution of the Boltzmann equation[2]. In the Chapman-Enskog formulation, molecules are assumed to interact as hard spheres with uncorrelated binary collisions. Clearly, the assumption of uncorrelated binary collisions is only fulfilled for diluted gases, i.e. where the kinetic theory of gases provides a link between molecular and transport properties. Under these conditions, viscosities can be predicted accurately for simple gases[16,17]. Interestingly, the Enskog formalism can also be carried out for binary mixtures of rigid spheres, as done by Enskog and Thorne[2].

For dense fluids, the situation is less satisfying. One approximation for the viscosity of dense fluids is the modified Enskog rigid sphere theory[2]. Despite

*In literature, bulk viscosity is also referred to as volume viscosity, expansion viscosity or second viscosity

its success as basis for many theories and models, Enskog's rigid sphere theory is prone to the assumptions of instantaneous collisions (elastic collisions with immediate change of values, e.g. momenta) and uncorrelated motion of molecules, as it is basically an ad-hoc extension of the dilute-gas theory to dense systems[18]. An extension of Enskog's rigid sphere formulation to chains of tangentially bonded rigid spheres has been published rather recently by de Wijn et al.[19].

Two prominent semi-theoretical approaches that make use of Enskog's rigid-sphere formulation is the model by Dymond and Assael[20–27] and the model by Vesovic and Wakeham[28–31]. Both approaches are based on the idea that the viscosity of real dense fluids is dominated by the repulsive intermolecular interactions that can be connected to the equivalent hard sphere fluid. This hard sphere part is then interpreted with the Enskog formalism.

The model by Dymond and Assael describes the viscosity as a function of the reduced volume and treats the size of the rigid sphere as a temperature-dependant adjustable parameter. In its basic formulation, the model by Dymond and Assael is applicable for spherical molecules. It is then mostly referred to as smooth hard sphere model[22,32]. For non-spherical molecules a roughness factor is introduced that gives this modification its name, the rough hard sphere model[20,21,23–26]. Further extensions by Ciotta et al.[27] improved the model by Dymond and Assael such that it does not diverge for dilute gases and improved the accuracy at higher densities. Because the Chapman-Enskog viscosity can be formulated for mixtures, so can the model by Dymond and Assael, which was tested for mixtures of n-alkanes[33].

The model by Vesovic and Wakeham focuses on the description of mixtures. It extents the mixture formulation of the Chapman-Enskog viscosity by Thorne[2] (that is rigorously valid only in the dilute gas limit) to dense mixtures[28–30]. Recently, the model by Vesovic and Wakeham has been extended from rigid spheres to tangentially jointed chains of equal-sized rigid spheres[19,31].

The Free-Volume Theory (FVT), originally introduced by Allal et al.[34], aims to connect the viscosity to the molecular structure of the fluid considered[35]. The FVT uses the Chapman-Enskog Theory (including the extension by Chung et al.[36]) for the low density limit and adds a term for the dense phase region. For that dense phase term, a dumbbell model is used, in which molecules are treated as spheres connected by springs (as in the Rouse

model[37] for polymers). Furthermore, the free-volume fraction is introduced that describes the empty space between molecules. The original FVT is able to describe gaseous as well as liquid viscosities. Recent research enhances the model to also describe mixture viscosities[38–40].

Another recently developed approach to model viscosities is the friction theory (f-theory), originally introduced by Quiñones Cisneros et al.[41]. Similar to FVT, f-theory also makes use of the Chapman-Enskog viscosity (including Chung's extension[36]) in the low density limit and introduces a new approach for the dense fluid region. In f-theory, the viscosity is regarded rather as a „mechanical property“ than as a „transport property“. That is, the fluid is considered as layers of fluid moving relatively to each other. f-Theory then aims at describing the shear stress between these moving layers of fluid in order to obtain the viscosity by using Newton's law of viscosity[41,42]. While the original versions of f-theory were restricted to van der Waals type of EoS, later publications aim at modifications that allow the use of arbitrary EoS[43,44]. f-Theory is suitable for fluids at low densities, as well as for fluids at high densities, corresponding to liquids. To cover this wide range of densities, models based on f-theory require many substance-specific parameters.

Two other very promising concepts that can be found in recent publications are closely related; Isomorph Theory (with slight modifications also referred to as Thermodynamic Scaling) and Entropy Scaling. Entropy Scaling was first proposed in 1977 by Rosenfeld[1]. Rosenfeld found that dimensionless formulations of transport properties of pure liquids can be expressed as a function of the residual entropy only. This monovariable relation allows for easy and accurate correlations, as demonstrated in applications of Entropy Scaling to various pure fluids[45–51]. Quite recently, a series of publications by the Dyre group, based on simulations of simple, spherical model fluids[52–59], revealed the existence of so-called Isomorphs. Isomorphs are characterized by, to a good approximation, invariant dynamic properties (in reduced units) like the viscosity. Isomorphs can be described by a ρ_m^γ/T-relation, where ρ_m is the molar density, T the temperature and γ a material-specific and density-dependent parameter. These Isomorphs only exist in so-called *Roskilde* fluids that are characterized by a strong correlation between the internal energy and the virial of the system considered. The concept of Isomorphs was further investigated for Lennard-Jones chains[60], supercooled fluids[61,62] and real fluids[63,64]. The concept of Isomporphs was shown to fail for strongly associating substances[65].

Isomorphs are interesting in the context of Entropy Scaling, because besides invariant reduced dynamic properties, also the residual entropy is, to a good approximation, constant along an Isomorph[66]. Hence, Isomorph Theory seems to be related to Rosenfeld's Entropy Scaling. Though, Entropy Scaling is applicable to many systems that are not characterized as *Roskilde* fluids like gaussian-core liquids, polar substances or hard-sphere mixtures[58]. In the following chapter, focus will shift towards the Entropy Scaling approach as the main topic of this thesis.

1.3 Entropy Scaling as an approach for transport properties

Rosenfeld was the first to point out that for spherical monatomic liquids, there exists a monovariable relation $\eta^+(s_{\mathrm{res}})$ between a dimensionless formulation of the viscosity, η^+, and the residual entropy s_{res}. To formulate the dimensionless viscosity

$$\eta^+ = \frac{\eta}{\eta_{\mathrm{ref}}} = \frac{\eta}{\eta_{\mathrm{R}}} \ , \tag{1.3}$$

Rosenfeld suggests a reference viscosity η_{R} that originates from dimensional analysis[1,67] and will be introduced in more detail in following chapters. Interestingly, this monovariable relation appears to be universal for monatomic spherical liquids and can be described accurately by a simple linear ansatz function for the liquid domain. Rosenfeld's observation allowed to approximate the complex dependence of the viscosity on pressure and temperature $\eta(p, T)$ with a simple linear function f_{lin}

$$\ln(\eta^+) = f_{\mathrm{lin}}(s_{\mathrm{res}}) \ , \tag{1.4}$$

with one universal set of scaling parameters. In eq. 1.4, f_{lin} represents a linear function that will be discussed in later chapters. Goel et al.[47] applied the Entropy Scaling approach to chains of Lennard Jones spheres (LJ-chain). Viscosities of each individual LJ-chain still collapse onto a single curve, solely dependent on the residual entropy. LJ-chains of different chain length can not be expressed by one single set of scaling parameters but need to be parametrized individually. Dzugutov[68] investigated the Entropy Scaling approach from a more microscopic viewpoint and found a connection to the molecular radial

distribution function. Further works on Entropy Scaling of model fluids include dumbbell-shaped particles[46] and gaussian core fluids[69,70]. Results for the dumbbell-shaped particles imply a monovariable dependence of the dimensionless viscosity on the residual entropy just as in LJ-chains. Gaussian core fluids are known to exhibit anomalous behaviour in the density (i.e. expansion during isobaric cooling) as well as in transport properties (i.e. increased particle mobility upon isothermal compression) that correspond to the anomalies observed in water. In the anomalous as well as non-anomalous regime, a monovariable relation of the dimensionless viscosity on the residual entropy can be observed, though, can not be described by one single linear correlation.

Studies on real fluids confirm the monovariable, but fluid-specific behaviour for n-alkanes[50,51,71–73], CO_2[74], Nitrogen[75], refrigerants[76] and water (except the mentioned structurally anomalous region)[77–81].

The articles mentioned so far limit their investigations of the Entropy Scaling approach to the dense fluid regime. In the limit of low densities, Rosenfeld's definition of the dimensionless viscosity does still lead to a monovariable behaviour, though, shows a steep, diverging increase when approaching $s_{\mathrm{res}} \to 0$. This was pointed out by Rosenfeld[67] and others[51,71,82]. Several approaches were published to avoid this steep increase. Krekelberg et al.[82] formulated a different expression for the dimensionless diffusivity (which suggests, a similar approach would be possible for the viscosity as well) based on the second Virial coefficient and the Chapman-Enskog formulation of transport properties. The authors call that approach the Generalized Rosenfeld Scaling. Recently, the Generalized Rosenfeld Scaling was evaluated for the self-diffusion coefficients of real substances by Hopp et al.[83].

Other works apply the Chapman-Enskog viscosity, η_{CE}, to calculate a residual viscosity (i.e. the difference of the actual viscosity and its value in the diluted gas limit)[71,84,85]. A dimensionless form of the residual viscosity is then obtained by dividing the residual viscosity by Rosenfeld's reference viscosity, such that

$$\eta^{+}_{\mathrm{res}} = \frac{\eta - \eta_{\mathrm{CE}}}{\eta_{\mathrm{R}}}. \tag{1.5}$$

As a result the dimensionless residual viscosities, η^{+}_{res}, are no longer diverging for $s_{\mathrm{res}} \to 0$, though, still exhibit a pronounced curvature, such that a linear ansatz function like eq. 1.4 is not sufficient to cover a wide range of pressure and temperature. Novak[50] proposes to simply replace Rosenfeld's reference

viscosity by the Chapman-Enskog viscosity, such that

$$\eta^* = \frac{\eta}{\eta_{\mathrm{CE}}} \tag{1.6}$$

Novak[51] applied eq. 1.6 to several n-alkanes and finds a nearly linear relation (in terms of eq. 1.4) for the entire fluid region. This approach has several benefits: First, viscosity data (gaseous as well as liquid) can be correlated with a simple ansatz function that comes with few adjustable parameters. Second, due to the limited number of parameters, already few experimental data points suffice to parametrize the model well. Third, the nearly linear behaviour allows for robust predictions beyond training conditions.

In recent work, Novak[73] introduces a predictive corresponding-states model for n-alkanes that is based on Entropy Scaling. To this end, Novak divides the n-alkanes into equally sized „entities“ using the PCP-SAFT segment number m. Doing so, he obtains a highly predictive approach for viscosities. Even though a transfer of this entity-based viscosity to substances besides n-alkanes is debatable, his work suggests that the scaling parameters of the Entropy Scaling approach can be linked to the molecule's structure.

Voyiatzis et al.[86] discovered the interesting feature that Entropy Scaling can also be applied to long polymers, though, with the restriction that two sets of viscosity parameters are needed; one for the regime where no entanglement takes place, one for conditions at which the polymers are subject to entanglement.

Most of the articles cited so far applied the Entropy Scaling approach to the viscosity. Rosenfeld's original approach was not limited to the viscosity[1]. Works on the application of Entropy Scaling for other transport properties like self-diffusivity, based on model fluids like hard spheres[68,87], soft spheres and Lennard-Jones fluids[82,88–90], liquid metals[91], water-like fluids[70,78,92] as well as real fluids[50,72,83] exist. Insights gained from works on the viscosity seem to be applicable just as well (and vice-versa). Entropy Scaling was also successfully applied to thermal conductivity[84,93–95]. Prediction of the thermal conductivity via the Entropy Scaling approach currently faces two challenges: First, the critical enhancement (i.e. increase of transport properties in the proximity of the critical point) is more pronounced as compared to the critical enhancement in the viscosity and self-diffusion. Second, internal degrees of freedom (e.g. vibrational and rotational) have to be incorporated in the reference term.

Regarding the application of Entropy Scaling to the viscosity of mixtures, only few publications exist. Novak[96] applied his entity-based approach to a mixture of methane+ethane and natural gas mixtures with up to twelve components. For these mixtures, he obtains good results with absolute average deviations (AAD%) below 5% for all mixtures. While this suggests that the Entropy Scaling approach is well applicable to mixtures with rather straight forward mixing rules, a systematic investigation for more mixtures was missing prior to the studies documented in this thesis.

Delage-Santacreu et al.[85] investigated model mixtures of Mie n-6 fluids (i.e. fluids with LJ-like intermolecular potential but varying repulsive exponent). They did not apply Entropy Scaling directly but investigated these mixtures in the light of Isomorph Theory. Their successful prediction of mixture viscosities based only on pure substance viscosity parameters is encouraging for the Entropy Scaling framework, because Isomorph Theory is related to Entropy Scaling.

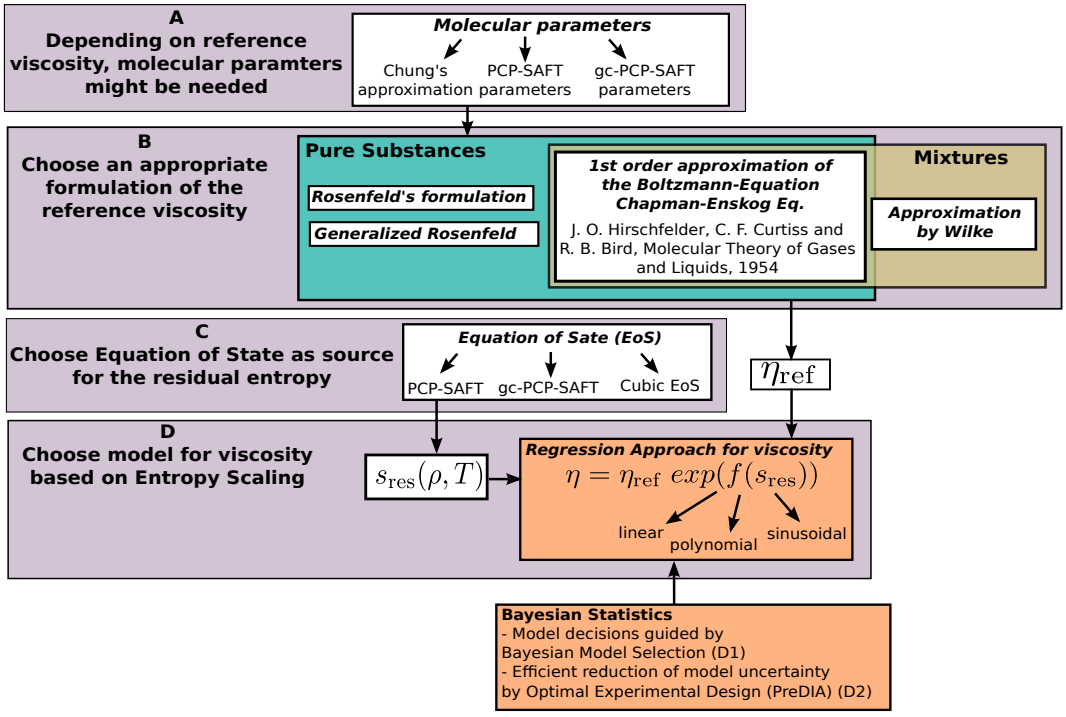


Figure 1.1: Flowchart illustrating the ingredients to the Entropy Scaling approach.

All Entropy Scaling approaches share certain steps for calculating or predicting transport properties (see fig. 1.1). First, molecular parameters (i.e. parameters describing the size, shape and the intermolecular potential) are

needed (step A) for calculating a reference viscosity (step B). Not all formulations of the reference viscosity necessarily need molecular parameters (e.g. Rosenfeld's original approach does not). Though, in terms of developing a predictive approach, formulations based on molecular parameters are beneficial, because parameters then have a physical meaning and therefore can be adopted from other models, like in this case, the PC-SAFT EoS. As step C, an equation of state is applied to calculate the residual entropy s_{res}. Finally, step D, one has to define a model for the relation between the dimensionless viscosity and residual entropy.

In this thesis, regarding step A and step C, the required residual entropy s_{res} is, in most cases, obtained via the PCP-SAFT EoS. Regarding the needed molecular parameters (as part of step A as well as for the EoS itself) two approaches are pursued: Most obviously, individually adjusted, substance-specific SAFT parameters. Alternatively I also apply a predictive EoS based on a group contribution approach. The next chapter will give more information about the applied PCP-SAFT EoS. Decisions made in step B and step D will be discussed in the main text of this thesis.

1.4 Perturbed Chain Polar Statistical Associating Fluid Theory - predicting the residual entropy

The central concept behind all SAFT Equations of State (EoS) was proposed more than 30 years ago by Wertheim[97]. Wertheim developed an analytical framework that allows to describe highly directional interactions (e.g. hydrogen bonding) within a perturbational formalism, called Thermodynamic Perturbation Theory. Applying Wertheim's theory in the limit of complete association, the formation of hard chains can be described[97–102]. This led Chapman et al.[103] to the first published SAFT EoS. From the very start, the SAFT EoS was designed to describe thermodynamic properties and phase behaviour of complex fluids on a sound statistical-mechanics basis. The fact that SAFT approaches are based on a molecular model is invaluable because molecular simulations can be conducted for the same molecular model, allowing to assess and improve the model.

As a perturbation theory, the SAFT EoS is based on the description of a

simple reference fluid to which further terms, describing additional molecular properties, are added. After the first SAFT EoS was published, different approaches how to further improve the SAFT EoS were developed and various versions of the SAFT EoS evolved. Some prominent representatives are PC-SAFT[104,105], soft-SAFT[106,107] and SAFT-VR[108,109]. Another development led to additional terms to also cover polar (i.e. dipolar as well as quadrupolar) substances[110–112]. The model is referred to as PCP-SAFT. In this work, the PCP-SAFT EoS was used to calculate the residual entropy s_{res} of pure substances and mixtures.

The PCP-SAFT EoS (like all other SAFT versions) is formulated in the specific Helmholtz energy a. Usually, these formulations are given as functions of Temperature T and number density ρ. The total Helmholtz energy is composed as sum of the Helmholtz energy of different contributions, as

$$a(\rho, T) = a^{\mathrm{ig}} + a^{\mathrm{hc}} + a^{\mathrm{disp}} + a^{\mathrm{assoc}} + a^{\mathrm{polar}} \tag{1.7}$$

where a^{ig} describes the ideal gas Helmholtz energy. Within the PCP-SAFT EoS, molecules are modeled as tangentially bonded identical segments. In the case of non-polar and non-associating substances, in addition to the ideal gas contribution, only the hard chain and dispersive contributions (a^{hc} and a^{disp}, respectively) need to be taken into account. These contributions require three adjustable parameters: the number of segments m, the segment size parameter σ and the dispersive energy parameter ε. If self-associating substances are considered, the contribution due to association, a^{assoc}, needs to be introduced. Association is described with two additional parameters: the association energy $\varepsilon^{\mathrm{AB}}$ and the association volume κ^{AB}. The term a^{polar} describes the specific Helmholtz energy due to polar interactions; introducing the dipole moment μ and quadrupolar moment Q as further parameters.

To predict mixture properties with PCP-SAFT, binary parameters σ_{ij} and ε_{ij} are required. These binary parameters can be approximated by combining rules. The most prominent and most widely used relations are the Lorentz-Berthelot combining rules[105,113], that are

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \tag{1.8}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{1.9}$$

where indices ii mark the pure substance parameters of the substance i, indices ij denote parameters defining the cross-wise binary interactions.

To improve the representation of non-ideal mixtures, the unlike energy parameter ε_{ij} is sometimes corrected by the introduction of an adjustable binary interaction parameter k_{ij}. Then, eq. 1.9 is modified such that

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}(1 - k_{ij}). \tag{1.10}$$

The main goal of this thesis is the prediction of pure substance and mixture viscosities. The binary interaction parameter k_{ij} is mixture specific. Most calculations in this work are performed simply assuming $k_{ij} = 0$, which leads to an easy, predictive model.

1.4.1 Group contribution PCP-SAFT formulation

One aim of this thesis is to develop a predictive viscosity model. Group contribution (gc) EoSs allow predicting thermodynamic properties of pure substances and mixtures even for cases where a pure constituent is not well-characterized by experimental data. For developing a predictive viscosity model, I here apply the gc PCP-SAFT EoS for predicting residual entropies of the Entropy Scaling concept.

gc approaches regard substances as a collection of functional groups α. For instance, n-alkanes are composed of CH_2 and CH_3 groups. Each group is defined through corresponding group contribution parameters (i.e. m_α, σ_α and so on) which are simultaneously adjusted to data of several substances carrying that chemical group. Properties of other substances can then be predicted by composing the substance under investigation from the functional groups.

For the family of SAFT-like EoS, there exist two conceptually different approaches to group contribution frameworks: the homosegmented and the heterosegmented approaches. Homosegmented approaches use mixing rules that allow to compose pure substance parameters from functional group parameters. For example in the case of n-alkanes, the PCP-SAFT pure substance parameters m_i, σ_i and ε_i are predicted from group parameters m_α, σ_α and ε_α. The resulting, predicted pure substance parameters are then used in the EoS just like parameters individually adjusted for pure substances. The benefit of predicted parameters comes with the price of a less accurate prediction of the

pure substance and mixture properties. Various approaches to define those mixing rules exist[114–117].

In heterosegmented group contribution approaches, molecules are composed of non-identical segments. Group parameters are not combined to pure substance parameters but the EoS itself is reformulated in terms of the group parameters. That bears the possibility of taking additional information into account, e.g. in what sequence groups are connected. Where the homosegmented approach is invariant for different chemical isomers, that information can be included in the heterosegmented approach (albeit only through neighbor-connectivities of functional groups). The benefit of heterosegmented approaches is the increased accuracy while maintaining the same predictive character as the homosegmented approaches[118]. The disadvantage is the more complex implementation due to the need to reformulate the EoS. Prominent representatives are the SAFT-γ method[119], the gc-SAFT-VR method[120] and a heterosegmented version of the PCP-SAFT EoS[118,121,122].

Due to its simplicity, in this thesis, I applied the homosegmented PCP-SAFT approach when working with group contributions. The mixing rules applied in this thesis are adopted from Vijande et al.[114] and are given in the Appendix 1.B. More detail will be provided in chapter 2.

1.4.2 Calculation of the residual entropy s_{res}

The specific residual entropy s_{res} is defined as the difference between the entropy at given density ρ and temperature T and the ideal gas entropy at the same density and temperature, as

$$s_{\mathrm{res}}(\rho, T) = s(\rho, T) - s^{\mathrm{ig}}(\rho, T) \tag{1.11}$$

This quantity can directly be derived from the specific residual Helmholtz energy a_{res} as

$$s_{\mathrm{res}}(\rho, T, \boldsymbol{x}) = -\left(\frac{\partial a_{\mathrm{res}}(\rho, T, \boldsymbol{x})}{\partial T}\right)_{\rho, x} \tag{1.12}$$

where the bold x represents the vector of molar fractions of all components of the mixture. The residual Helmholtz energy a^{res}, according to eq. 1.7, is

expressed as the sum of various contributions

$$a^{\mathrm{res}}(\rho, T) = a(\rho, T) - a^{\mathrm{ig}} = a^{\mathrm{hc}} + a^{\mathrm{disp}} + a^{\mathrm{assoc}} + a^{\mathrm{polar}} \ . \tag{1.13}$$

It is important to note that for a given state point with temperature T, pressure p and density ρ one has to decide whether the ideal gas entropy at that temperature T and pressure p or at that temperature T and density ρ needs to be calculated, because

$$s^{\mathrm{ig}}(p, T) = s^{\mathrm{ig}}(\rho^{\mathrm{ig}}(p, T), T) \neq s^{\mathrm{ig}}(\rho, T) \tag{1.14}$$

and therefore

$$s_{\mathrm{res}}(p, T) \neq s_{\mathrm{res}}(\rho, T) \tag{1.15}$$

where

$$s_{\mathrm{res}}(p, T) = s(p, T) - s^{\mathrm{ig}}(p, T) \tag{1.16}$$

and $s_{\mathrm{res}}(\rho, T)$ as defined in eq. 1.11. $\rho^{\mathrm{ig}}(p, T)$ describes the density of an ideal gas at given pressure p and temperature T whereas $\rho(p, T)$ describes the density of a real substance at that given pressure and temperature. For the Entropy Scaling approach, it is necessary to obtain the residual entropy at a given density ρ and temperature T (i.e. $s_{\mathrm{res}}(\rho, T)$). In case the EoS is not already formulated as a function of density ρ and temperature T, $s_{\mathrm{res}}(\rho, T)$ can be obtained from $s_{\mathrm{res}}(p, T)$ by

$$s_{\mathrm{res}}(\rho, T) = s_{\mathrm{res}}(p, T) - R \ln Z \ , \tag{1.17}$$

with Z being the compressibility.

A derivation of eq. 1.17 can be found in the Appendix 1.A.

1.5 Development of a model for the viscosity and model evaluation with Bayesian Model Selection

The functional form of the $\eta^*\left(s^{\mathrm{res}}\right)$-relation is not prescribed by the Entropy Scaling. Therefore, any ansatz function (here referred to as „model candidate“) is empirical. Competing ansatz functions are usually ambiguous to compare.

Even more so, when the different candidates not only vary in the type of function but also in the number of adjustable parameters. Decisions then often rely on the choice of experimental data used for assessing the competing, empirical models.

Most frequently, models are evaluated by comparing the accuracy of the model in representing a certain set of data points. That approach would only lead to a clear decision if differences in the accuracy of the competing models are large. If accuracies of the competing models are of the same magnitude, differences might be within the experimental accuracy of datapoints and hence a model choice would be ambiguous. Also, after accepting that data points are never a perfect representation of reality but subject to experimental errors, a modeler has to wonder: How many adjustable parameters are reasonable to represent the given data, i.e. might the model be „under- or overfitted“? An „underfitted“ model (too few adjustable parameters) will have uniquely identifiable parameter values and a small variance (i.e. high confidence) in its predictions. Though, it will also not be able to represent all relevant features of the data points the model was adjusted to and will therefore show a significant error (also called bias) in its predictions. Contrarily, an „overfitted“ model (too many adjustable parameters) will be able to capture all relevant features of the data points but will also react sensitively to experimental errors (noise). Predictions, especially beyond the training conditions (i.e. conditions beyond the conditions of the data points the parameters have been adjusted to) will show larger variance.

This problem of finding a sound compromise between bias and variance is often referred to as bias-variance dilemma[123]. The principle of parsimony (also referred to as „Occam's razor“[124]) suggests to chose the model candidate of least complexity (here, number of adjustable parameters) that still represents the experimental data „reasonably well“. The vague formulation „reasonably well“ already denotes the ambiguity of intuitive (and at least to a certain part subjective) model decisions. Therefore, I applied an objective statistical method to question and guide the decision making process in the development of the proposed viscosity model.

The procedure applied is called Bayesian Model Selection (BMS)[125,126]. Based on Bayes' Theorem[127], BMS offers a quantitative comparison of candidate models that also takes experimental errors into account[128,129]. The result of BMS is the Bayesian Model Evidence (BME). The BME-values can

then be averaged (this step is called Bayesian Model Averaging, BMA) to obtain a ranking of model weights[130,131]. The probabilistic information about the models gathered for obtaining the BME (and effort invested to obtain these) can be further exploited in data-worth analyses. These data-worth analyses can serve as a tool to identify structural errors of candidate models[132] or to plan future experiments[133–135] with the aim of reducing model uncertainty.

In summary, the key advantages of applying a probabilistic approach to model evaluation are as follows: First, the importance of rigorously taking experimental errors into account in thermodynamic model building has been recognized but is not yet widely applied[136] and is an inherent part of BMS. Second, by accepting that the underlying experimental data are not free of errors leads to the concept of uncertain parameters. Opposed to ending up with one parameter set that offers the best „fit“ to a given number of experimental data, BMS gives probability distributions of parameter values. These parameter distributions reflect the limits of parameter identifiability in the light of experimental errors and limited number of available experimental data. Furthermore, using parameter distributions to obtain model predictions gives a predicted distribution of viscosities instead of a single value. The predicted distributions carry valuable information about the uncertainty of predictions from a candidate model; which is of great value for analyzing thermodynamic models by statistical means[136] or for guiding in optimal design strategies for future experiments[137].

1.6 Outline and structure of thesis

The aim of this thesis is to develop a model for correlating and predicting viscosities of pure substances and mixtures in a wide range of pressure and temperature. To further emphasize predictability and to combine this model with other group contribution based technqiues developed in our group, the viscosity model should also be applicable as a group contribution method.

The chosen framework for the modeling of viscosities is the Entropy Scaling approach. In all chapters, the required residual entropy s_{res} is obtained via the PCP-SAFT EoS.

In chapter 2, I illustrate the group contribution approach to predict viscosities. Besides the principles of Entropy Scaling, I introduce the group contribution based version of the EoS and demonstrate how the group contribution

framework can be applied to transport properties like the viscosity.

In chapter 3, I present BMS, the statistical tool that helped to evaluate and guide decisions in the development of the viscosity model (illustrated as step D1 in fig. 1.1). I first give an overview of the considered candidate viscosity models (i.e. regression approaches in fig. 1.1). Afterwards, the statistical (Bayesian) framework is introduced. Two scenarios are investigated: a pressure extrapolation and an application of the group contribution approach (i.e. adjusting the functional group viscosity parameters to one set of substances and predicting the viscosity of another set of substances of the same chemical family).

After the model evaluation, in chapter 4 and chapter 5, an extension of the viscosity model to binary and multicomponent mixtures is introduced. The extension to mixtures does not need any further adjustable parameters; mixture viscosities are calculated purely predictive. Besides binary mixtures, multicomponent mixtures of up to five components are investigated.

In chapter 6, an optimal experimental design procedure based on Bayesian Statistics (which was not earlier used in the field of thermodynamics) is applied within Entropy Scaling. The optimal experimental design study was carried out for two different examples reflecting two different possible applications of the model developed in this thesis.

Appendix

1.A Calculation of of the residual entropy $s_{\mathrm{res}}(\rho, T)$ from $s_{\mathrm{res}}(p, T)$

It is

$$s_{\mathrm{res}}(\rho, T) - s_{\mathrm{res}}(p, T) = \underbrace{s(\rho, T) - s(p, T)}_{=0} + s^{\mathrm{ig}}(p, T) - s^{\mathrm{ig}}(\rho, T) \tag{1.18}$$

With

$$\mathrm{d}s^{\mathrm{ig}}(p, T) = \frac{c_p^{\mathrm{ig}}}{T}\mathrm{d}T - R\ \mathrm{d}(\ln p) \tag{1.19}$$

and

$$s^{\mathrm{ig}}(\rho, T) = s^{\mathrm{ig}}(p^{\mathrm{ig}}(\rho, T), T) \tag{1.20}$$

eq. 1.18 reads

$$s_{\mathrm{res}}(\rho, T) - s_{\mathrm{res}}(p, T) = \underbrace{\int_T^T \frac{c_p^{\mathrm{ig}}}{T}\mathrm{d}T}_{=0} - \int_{p^{\mathrm{ig}}}^{p} R\mathrm{d}\ln p \tag{1.21}$$

$$= -R\ln\frac{p}{p^{\mathrm{ig}}} = -R\ln\frac{p}{\rho RT} = -R\ln Z \tag{1.22}$$

with the compressibility factor

$$Z = \frac{p}{\rho RT} \tag{1.23}$$

1.B Mixing rules of the homosegmented PCP-SAFT group contribution approach

For the generation of pure component parameters from group contribution parameters, the following mixing rules, proposed by Vijande et al. [114], are used in this work, that are

$$m_i = \sum_{\alpha_i} n_{\alpha_i} m_{\alpha_i} \tag{1.24}$$

$$m_i \sigma_i^3 = \sum_{\alpha_i} n_{\alpha_i} m_{\alpha_i} \sigma_{\alpha_i}^3 \tag{1.25}$$

$$m_i \varepsilon_i = \sum_{\alpha_i} n_{\alpha_i} m_{\alpha_i} \varepsilon_{\alpha_i} \tag{1.26}$$

in which the summations run over all types of functional groups α_i of the regarded substance i and where n_{α_i} denotes the number of functional groups of type α_i contained in molecule i.

2. A Group Contribution Method for Viscosities Based on Entropy Scaling Using the Perturbed-Chain Polar Statistical Associating Fluid Theory

This chapter is a literal quote of the publication: O. Lötgering-Lin, J. Gross; Ind. Eng. Chem. Res., 2015, 54 (32), pp 7942-7952.

In this work, we propose a new predictive entropy-scaling approach for Newtonian shear viscosities based on group contributions. The approach is based on Rosenfeld's original work [Y. Rosenfeld, Phys. Rev. A 1977, 15, 2545 - 2549]. The entropy scaling is formulated as third order polynomial in terms of the residual entropy as calculated from a group-contribution perturbed chain polar statistical associating fluid theory (PCP-SAFT) equation of state. In this study we analyze the course of entropy scaling parameters within homologous series and suggest suitable mixing rules for the parameters of functional groups. The viscosity of nonpolar, of polar, and of self-associating (hydrogen bonding) components are considered. In total, 22 functional groups are parametrized to viscosity data of 110 pure substances, from 12 different chemical families. The mean absolute relative deviations (*MADs*) to experimental viscosity data are typically around 5%. For three chemical families, namely, branched alkanes, 1-alcohols, and aldehydes, we obtain higher *MADs* of about 10%. Water is correlated with a *MAD*-value of 3.09%.

2.1 Introduction

Detailed knowledge of the fluid viscosity is of importance in many engineering disciplines. High pressure applications like carbon capture and storage[5–7] or the design of new biofuels[9–11,138,139], for example, often lack viscosity data. Furthermore, the design and optimization of solvents or working fluids relies on predictions of fluid viscosities[13,140–143].

In many of those applications, the working fluid undergoes process steps in a wide range of temperature and pressure, involving phase transition. Due to the strong temperature and pressure dependence of the viscosity (see fig. 2.1), many measurements are needed for characterizing even pure substances by experiments. Predictive models, applicable to the complete range of conditions of technical relevance are therefore desirable[144].

Several promising approaches to model viscosities can be found in the literature. While some mainly focus on high accuracy for a specific component[145,146], others focus on robust extrapolations or even predictions, for instance Free-Volume Theory (FVT)[38,39,147], Friction Theory[41,43] or the Dymond and Assael (DA) model[19,22,23,27,148] based on the Enskog equation[2]. An intriguingly elegant way of capturing the complex behavior of the viscosity is Rosenfeld's entropy-scaling[1,67], where the viscosity is regarded as a function of the residual entropy s_{res} only. In his original work, Rosenfeld was the first to point out that for spherical monatomic fluids, there exists a quasiuniversal, monovariable relation between a dimensionless form of the viscosity η^+ and the residual entropy. This means that the complex temperature and pressure dependence of the viscosity can be reduced to one simple function $\eta^+(s_{\mathrm{res}})$. This entropy scaling was rediscovered and confirmed from a more microscopic viewpoint by Dzugutov[68].

Over the years, it was observed that the entropy scaling does not only hold for spherical monatomic fluids. For non-monatomic fluids, the Rosenfeld-scaling does not obey a universal relation; for molecular substances the entropy scaling leads to component-specific, but still monovariable relations between the dimensionless viscosity and the residual entropy[45–47,71,72,74,75,77–79,81,84]. Using molecular simulations Goel et al.[47] demonstrated a monovariable behavior for Lennard-Jones chains. This monovariable relation was described by a linear approach, introducing two scaling parameters. Goel et al.[47] showed, that these scaling parameters can be correlated to the molecules chain length.

Subsequent research on increasingly complex substances confirmed the monovariable relation for further n-alkanes[71,72,84], for dumbbell-shaped particles[46] and water[78,79,81]. Limitations to the concept of transferable parameters were pointed out by Chopra et al.[45], which suggests that the entropy-scaling as a monovariable relation between dimensionless viscosity and residual entropy is a powerful, but component-specific concept. Excellent scaling with the residual entropy was confirmed for experimental viscosities of real systems[74,75,77]. Recent investigations revealed scaling invariances of dynamic and structural properties of many liquids along so called isomorphs. Besides transport properties also the residual entropy s_{res} was shown to be constant along an isomorph[54,58,60]. Therefore, isomorph theory might give new insights for the scaling of the viscosity with the residual entropy.

As pointed out by Rosenfeld[67] and others[51,71,82], the dimensionless viscosity shows a steep increase in the low density region (i.e. $s_{\mathrm{res}}/k_B \geq -0.5$), as shown in fig. 2.2. Novak[51] proposed a different expression for the reduced viscosity by relating the viscosity of several n-alkanes to the corresponding Chapman-Enskog viscosity and obtained a nearly linear relation for the entire fluid region. This result yields the possibility to correlate and in particular extrapolate experimental (or simulation) data, so that already a small data set is sufficient to determine the viscosity η for the entire fluid phase region. Recently[73], Novak modified his approach by additionally relating the reduced viscosity to entities, representing a predictive corresponding-states model for n-alkanes (and some further substances). Even though the „entities“ have no direct structural correspondence, his results suggest that the scaling parameters can be correlated with the molecular structure.

In this work we propose a new predictive approach for viscosities based on entropy-scaling. To calculate the residual entropy we use a group contribution method based on the PCP-SAFT EoS which was originally suggested by Vijande et al.[114] and was recently analysed and reparameterised by Sauer et al.[118]. The proposed model is entirely based on functional groups.

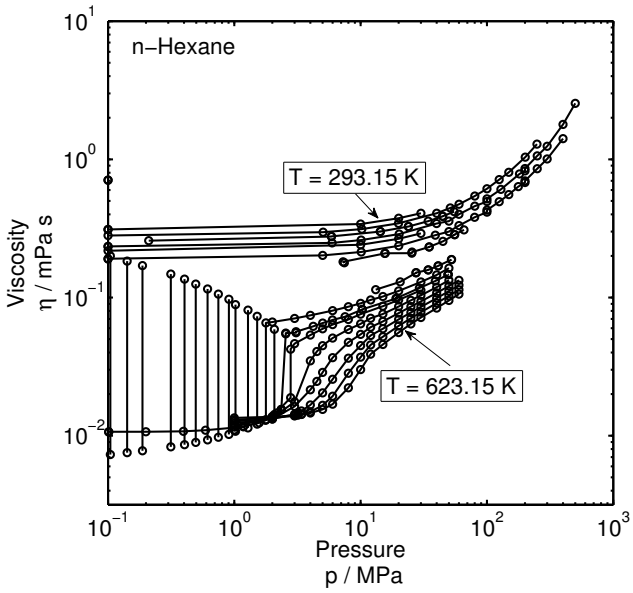


Figure 2.1: Experimental viscosities of n-hexane. Circles represent the experimental data; lines are a guide to the eye connecting viscosities at the same temperature.

2.2 Theoretical Background

2.2.1 Entropy-Scaling

According to Rosenfeld's entropy-scaling approach, transport properties of pure substances are related to the molar residual entropy,

$$s_{\mathrm{res}}(\rho, T) = \left(s(\rho, T) - s^{ig}(\rho, T)\right), \tag{2.1}$$

where ρ is the number density, T the temperature and the superscript ig indicates the ideal gas contribution. Rosenfeld's argument starts with a hard-sphere fluid (superscript hs), for which the viscosity η^{hs} is entirely determined by the dimensionless density $\xi^{\mathrm{hs}} = \frac{1}{6}\pi\rho\sigma^3$, with σ as the particle diameter. Because the dimensionless density ξ^{hs} of a hard-sphere fluid is a monotone function of the residual entropy, this means that the viscosity of a purely repulsive fluid can be expressed as a unique function of the residual entropy, i.e.

$$\eta^{\mathrm{hs}}\left(\xi^{\mathrm{hs}}(\rho)\right) = \eta^{\mathrm{hs}}\left(s^{\mathrm{hs}}_{\mathrm{res}}(\rho)\right) \tag{2.2}$$

By using a first order perturbation theory he pointed out that the residual entropy of a simple attractive fluid s_{res} can be approximated by the residual entropy of a hard-sphere fluid $s_{\text{res}}^{\text{hs}}$ of the same dimensionless density. Rosenfeld then showed by simulations that the dimensionless viscosity η^+ of an attractive fluid is also entirely determined by the residual entropy s_{res}, so that

$$\eta^+ (\rho, T) = \eta^+ (s_{\text{res}} (\rho, T)) \tag{2.3}$$

The last step in this chain of arguments has no proof and is empirical, although statistical mechanical approaches such as mode coupling theory may provide arguments and assumptions on the validity for the last hypothesis.

The dimensionless viscosity η^+ in Rosenfeld's original approach is obtained by dividing the viscosity by a reference viscosity η_{ref} given as

$$\eta^+ (s_{\text{res}} (\rho, T)) = \frac{\eta}{\eta_{\text{ref}}} = \frac{\eta}{\rho^{\frac{2}{3}} \sqrt{M k_B T / N_A}} \tag{2.4}$$

Applying this approach to experimental viscosity data of n-hexane (fig. 2.1) results in the expected monovariable behavior (fig. 2.2). The reduced viscosity is confirmed to be a function of the residual entropy, only. However, in the low density region (i.e. $s_{\text{res}}/k_B \geq -0.5$) the model shows a strong increase of the reduced viscosity due to the breaking down of the cubic lattice assumption as pointed out by Novak[50].

It can be shown[51], that the reference viscosity η_{ref} of Rosenfeld's approach is a simplified version of the Chapman-Enskog viscosity. As a first-order approximation of the Boltzmann-equation, the Chapman-Enskog viscosity is given as[149,150]

$$\eta_{\text{CE}} = \frac{5}{16} \frac{\sqrt{M k_B T / (N_A \pi)}}{\sigma^2 \Omega^{(2,2)*}} \tag{2.5}$$

Eq. 2.5 expresses the pure component Chapman-Enskog viscosity as a function of the molar mass M, temperature T, characteristic particle diameter σ and reduced collision integral $\Omega^{(2,2)*}$. The reduced collision integral is defined as

$$\Omega^{(2,2)*} (T^*) = \frac{\Omega^{(2,2)}}{\Omega^{(2,2)}_{\text{hard sphere}}} \tag{2.6}$$

with the reduced temperature $T^* = \frac{k_B T}{\varepsilon}$ and potential depth ε. The coefficients $(2, 2)$ are the sonine polynomial expansion coefficients and depend on

the order of approximation of transport properties in the Chapman-Enskog formalism[2,149].

By approximating the volume per particle by a cube of edge length σ (cubic lattice structure) and applying the hard-sphere model ($\Omega^{(2,2)*} = 1$) to eq. 2.5 the reduced viscosity η^+ is obtained as given in eq. 2.4.

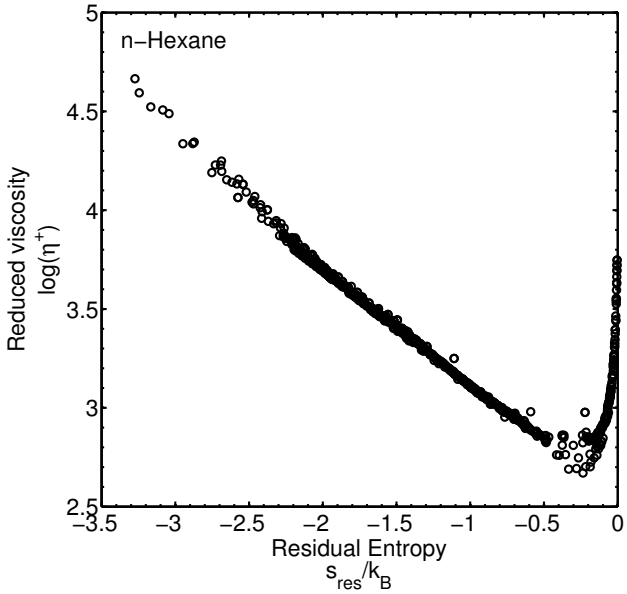


Figure 2.2: Reduced experimental viscosities of n-hexane according to Rosenfeld's approach[1], i.e. eq. 2.4.

A more well-behaved function $\eta^+(s_{\text{res}})$ in the region of low densities can be obtained by using the Chapman-Enskog viscosity η_{CE} according to eq. 2.5 as reference viscosity η_{ref}.

In the context of a group-contribution (gc) approach, it is necessary to express eq. 2.5 in terms of gc parameters. Similar to Novak[50], we relate the Chapman-Enskog viscosity to PCP-SAFT segments, with

$$\eta_{CE,gc} = \frac{5}{16} \frac{\sqrt{M k_B T / (m_{gc} N_A \pi)}}{\sigma_{gc}^{\,2}\, \Omega_{gc}^{(2,2)*}} \tag{2.7}$$

The index gc indicates pure component parameters that are calculated with the homosegmented group contribution method based on PCP-SAFT EoS which will be addressed in the next subsection. The reduced viscosity is then defined as

$$\eta^* = \frac{\eta}{\eta_{CE,gc}} \tag{2.8}$$

To calculate the reduced collision integral $\Omega_{gc}^{(2,2)*}$, we use the empirical approximation by Neufeld et al.[151].

2.2.2 Group Contribution PCP-SAFT EoS

Within the PCP-SAFT EoS molecules are considered to be linear chains of tangentially bonded identical segments[105,110]. These molecules are described by the pure component parameters m_i, σ_i and ε_i. The homosegmented group contribution method proposed by Vijande et al.[114] suggests mixing rules to calculate pure component parameters by averaging the group parameters m_α, σ_α and ε_α. Tihic et al.[115] introduced first order and second order groups[116] for a set of mixing rules similar to those of Vijande et al.[114]. A group contribution method with a new set of mixing rules was proposed by Tamouza et al.[117]. Besides homosegmented *gc* approaches there exist heterosegmented *gc* approaches that consider molecules as chain molecules composed of non-identical segments. Two important representatives of the heterosegmented approach are the SAFT-γ method by Lymperiadis et al.[119] and the *gc*-SAFT-VR method by Peng et al.[120]. A heterosegmented *gc* approach based on the PC-SAFT EoS was presented by Paduszyński and Domańska[121], Peters et al.[122] and Sauer et al.[118].

An earlier study of our group showed the heterosegmented approach to be superior to the homosegmented model for phase equilibria. While liquid densities are described about equally well by both approaches, vapor pressures are described more accurately by the heterosegmented approach[118]. For the here considered case of predicting viscosities the heterosegmented *gc* approach did not lead to significant improvements. Therefore, the pure component parameters from the combining rules by Vijande et al.[114] are used to calculate the residual entropy s_{res} as well as the Chapman-Enskog viscosity. Group parameters are taken from Sauer et al.[118] which have been adjusted to vapor pressure and saturated liquid density data. For methane, ethane and methanol we used the pure component parameters as published by Gross and Sadowski[104,105]. This group contribution method based on the PCP-SAFT EoS will henceforth be referred to as homosegmented GC-PCP-SAFT.

2.2.3 Viscosity Based on Group Contributions

In his original work, Rosenfeld[1] applied the principle of entropy scaling using a linear ansatz-function to describe the linear, monovariable behavior of dimensionless logarithmic transport coefficients (for the liquid phase region), according to

$$\ln \eta_i^+ = A_i + B_i \left(\frac{s_{\text{res}}}{k_B} \right) \tag{2.9}$$

Deviations from this linear relation were already reported in supercooled liquids[46,61,152], alkanes at high densities[45] and water[77,81]. In this work, we use a third order polynomial, where the second and third order terms account for deviations from the linear behavior that we found when applying eq. 2.8 and eq. 2.9 to experimental viscosities. The empirical correlation function we use to determine the viscosity of a pure substance i is then

$$\ln \eta_i^* = A_i + B_i z + C_i z^2 + D_i z^3 \tag{2.10}$$

with

$$z = \left(\frac{s_{\text{res}}}{k_B \; m_{gc,i}} \right) \tag{2.11}$$

The division by $m_{gc,i}$ ensures that molecules of very different molecular mass have a similar range of values for the dimensionless entropy z. The viscosity parameters A_i to D_i of pure substances are obtained from the parameters A_α to D_α of functional groups α, respectively. We propose the following empirical relations for mixing group-contribution parameters

$$A_i = \sum_\alpha n_{\alpha,i} \; m_\alpha \; \sigma_\alpha^3 \; A_\alpha \tag{2.12}$$

$$B_i = \sum_\alpha \frac{n_{\alpha,i} \; m_\alpha \; \sigma_\alpha^3}{V_{tot,i}^\gamma} \; B_\alpha \tag{2.13}$$

$$C_i = \sum_\alpha n_{\alpha,i} \; C_\alpha \tag{2.14}$$

$$D_i = D \sum_\alpha n_{\alpha,i} \tag{2.15}$$

with

$$V_{tot,i} = \sum_\alpha n_{\alpha,i} \; m_\alpha \; \sigma_\alpha^3 \tag{2.16}$$

where $n_{\alpha,i}$ denotes the number of functional groups of type α in the substance i. The exponent γ and the parameter D were kept constant for all investigated substances and were optimized for n-alkanes. In summary, our approach for predicting viscosities applies a group contribution equation of state (GC-PCP-SAFT) with its own group contribution parameters. In order to predict viscosities, we adjusted group contribution parameters A_α, B_α, and C_α of all groups α, as well as model constants $D = -0.01245$ and $\gamma = 0.45$.

2.2.4 Procedure for Adjusting Model Parameters

The group contribution parameters related to the viscosity were adjusted to experimental data of pure substances. The experimental data was taken from the Dortmund Database[153] and is listed in the Supporting Information together with the temperature and pressure ranges and with references to the original publications.

Because the accuracy of the GC-PCP-SAFT EoS declines in the vicinity of the critical point, data points within the condition range

$$(0.9\ p_{crit} < \ p \ < 1.1\ p_{crit}) \wedge (0.9\ T_{crit} < \ T \ < 1.1\ T_{crit})$$

were neglected.

The group contribution parameters related to the viscosity were determined with the following procedure: First, all available experimental viscosity data was read from data-files and the corresponding Chapman-Enskog viscosity and reduced viscosity of each data point were calculated. Then, the dimensionless residual entropy $z = s_{\text{res}} \cdot (k_B\ m_{gc,i})^{-1}$ was calculated from the GC-PCP-SAFT EoS using the analytical solution of

$$s_{\text{res}}\,(\rho, T) = -\left(\frac{\partial a_{\text{res}}}{\partial T}\right)_\rho \tag{2.17}$$

with the specific Helmholtz energy $a_{\text{res}} = A_{\text{res}}/N$ given by Gross and Sadowski[105].

The functional group parameters of the mixing rules eq. 2.12 to eq. 2.14 were then optimized by minimizing the squared relative deviations of calculated to experimental viscosities for all data points of the considered substances using a Levenberg-Marquardt algorithm[154].

2.3 Results and Discussion

The homologous series of n-alkanes is important for developing a predictive approach based on group contributions. First, the effect of increasingly non-spherical molecular shape can be investigated, without superpositioning specific interactions due to (local) polar groups. Second, the PC-SAFT equation of state is well-suited to correlate and predict thermodynamic properties of n-alkanes. Third, there is comparably plentiful experimental data for n-alkanes. We first analyzed this homologous series in detail.

2.3.1 Parameter Behavior

The work of Novak[51] showed that the entropy scaling, with an appropriate ideal gas limit, leads to roughly linear scaling behavior of the logarithmic dimensionless viscosity and the residual entropy. In fig. 2.3 (symbols) we illustrate the scaling behavior of the viscosity for n-alkanes. In this study we strive to approximately capture also the nonlinear part of the scaling behavior using a third order polynomial ansatz function. In order to propose suitable mixing rules for the group parameters (such as A_α for a group of type α) in eq. 2.12 to eq. 2.15, we first adjusted the scaling parameters (such as A_i) individually for all n-alkanes i and analyzed how these parameters develop with increasing carbon number. The residual entropy s_{res} as well as the Chapman-Enskog viscosity was still calculated using the homosegmented GC-PCP-SAFT EoS.

We observed, that the correlation of parameters, especially of C_i and D_i is strong, due to the relatively small deviation of the scaling behavior from a linear relation. For obtaining meaningful parameters C_i and D_i, experimental data needs to cover a large portion of the range of residual entropy. The individually adjusted parameters depend strongly on the number of data points as well as the pressure- and temperature range. For ensuring a robust group contribution concept we reduced the number of adjustable parameters by defining the parameter D_i with a constant (i.e. group-independent) coefficient, according to eq. 2.15. The rational for this choice was the following: first, we observed that the D_i values that were individually optimized for each substance increase with length of the n-alkane species. Second, we tentatively made the Ansatz $D_i = \sum_\alpha n_{\alpha,i} D_\alpha$, which represented the n-alkanes well with $D_{\text{CH3}} = D_{\text{CH2}}$, i.e. with the same coefficient for the CH_3 and the CH_2-functional groups. We then

adjusted all coefficients A_α to D_α to our entire set of data and observed that the values of D_α scattered around the value of $D = -0.01245$ and we adopted this value as a model constant in eq. 2.15.

To support this decision we defined a parameter $(f_\alpha^{(3)}/f_\alpha^{(4)})$ that measures, how the result of the model with constant D deteriorates compared with the model where D is adjusted for each group. Specifically, $f_\alpha^{(4)}$ is the objective function of a model where parameters A_α to D_α where adjusted to each group α, and $f_\alpha^{(3)}$ is the objective function of the model where parameters A_α to C_α are adjusted and parameter D is set to a constant value of $D = D_\alpha = -0.01245$. Fig. 2.4 illustrates this measure for the chemical families considered in this study. The diagram shows that $f_\alpha^{(3)}/f_\alpha^{(4)}$ is close to unity (always smaller than 1.3) for all cases, which means that setting D to a constant value only mildly weakens the regression result. The dashed line in fig. 2.4 represents the analogous measure, $(f_\alpha^{(3,0)}/f_\alpha^{(4)})$, but now for the case, where $f_\alpha^{(3,0)}$ is the model with parameters A_α to C_α adjusted and parameter D is set to zero. It is shown, that the objective function for this case is significantly higher and the correlation result is thus weaker compared to the proposed parametrization. We note, that for ethers the objective function $f_\alpha^{(3,0)}$ is lower than $f_\alpha^{(4)}$. That is possible, because the functional group parameters A_α to D_α were not adjusted simultaneously for all functional groups but successively in order as given in fig. 2.4.

For devising a robust group-contribution approach, it is important to verify that the viscosity parameters according to the mixing rules, eq. 2.12 to eq. 2.13, are well-behaved. To evaluate the proposed mixing rules we examine the behavior of the viscosity correlation parameters A_i and B_i with increasing number of carbon atoms. The basis of this consideration is a set of parameters A_i and B_i individually optimized for n-alkanes (with $D = -0.01245$ in eq. 2.15).

For comparison, this analysis was also performed for the case $C_i = 0$ and $D_i = 0$, which corresponds to eq. 2.9. The resulting parameter behavior is shown in fig. 2.5. The slope-parameter B_i does scale very well with the chain length using either of the described correlation functions. The parameter A_i does not correlate well when using eq. 2.9. Applying eq. 2.10 the parameter A_i decreases monotonically with increasing chain length illustrating the suitability as a group contribution method.

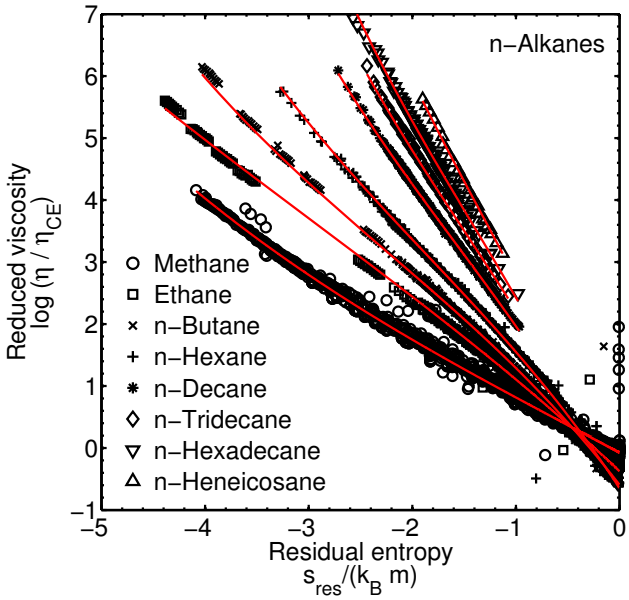


Figure 2.3: Reduced viscosity of a selection of the investigated n-alkanes. Symbols represent reduced experimental viscosities for residual entropies calculated with GC-PCP-SAFT, and red lines show the result of our correlation.

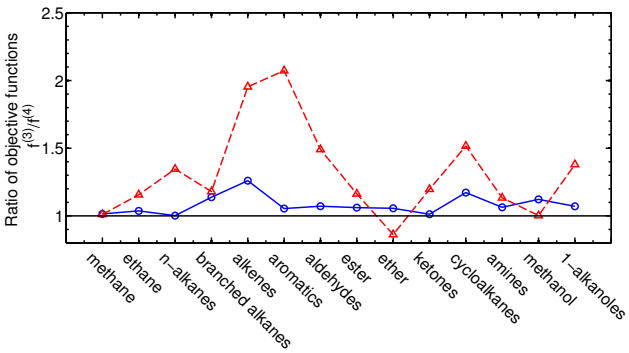


Figure 2.4: Ratio between the objective function of the proposed model $f_\alpha^{(3)}$ ($D_\alpha = -0.01245$) and the objective function of the same model, $f_\alpha^{(4)}$, but with 4 adjustable parameters, A_α to D_α (solid line with open circles). Furthermore, the ratio between the objective functions $f_\alpha^{(3,0)}$ and $f_\alpha^{(4)}$ is shown, where $f_\alpha^{(3,0)}$ is for a model with $D_\alpha = 0$ (dashed line with triangles).

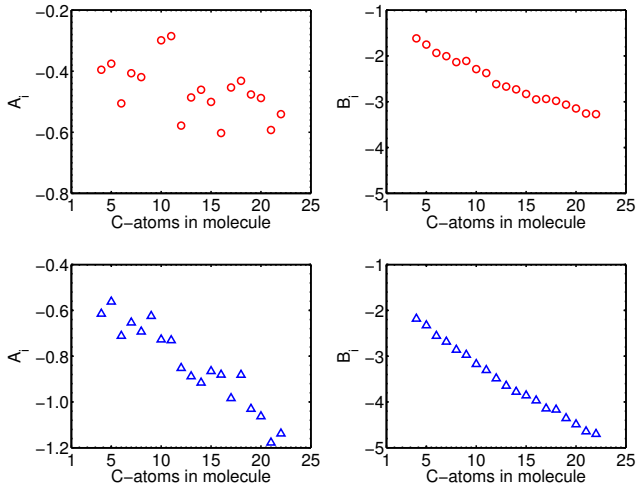


Figure 2.5: Viscosity correlation parameters A_i and B_i resulting from individual regressions to several n-alkanes. Reduced experimental viscosity was adjusted to either eq. 2.9 (○) or eq. 2.10 (Δ).

2.3.2 Predicting the Viscosity of Substances Outside the Training Set

To assess the predictive ability of the group contribution method, we divided the available data set into a training set and an evaluation set. In a first step, we considered the homologous n-alkane series. Only the data of n-butane to n-nonane was used to determine the alkane viscosity group contribution parameters A_α - C_α (training set). Those parameters were then used to predict the viscosity of the remaining n-alkanes up to n-hexatriacontane (evaluation set). The resulting deviations of calculated to experimental viscosities are shown in fig. 2.6.

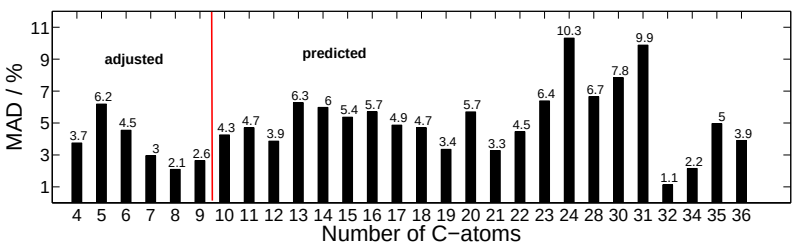


Figure 2.6: Deviations of the predicted viscosity of 22 n-alkanes with parameters adjusted to experimental data of six n-alkanes.

The deviations reported in fig. 2.6 are given as mean absolute relative deviations, $MAD\% = \frac{100}{N_{data}} \sum_1^{N_{data}} \frac{|\eta_{calc} - \eta_{exp}|}{\eta_{exp}}$, because the $MAD\%$-values are more

intuitive than squared deviations. Fig. 2.6 shows that the deviations remain nearly constant for the complete range of substances. The viscosity of the longest investigated n-alkane, n-hexatriacontane, is predicted with a deviation of 3.9%, implying that the parameters of the training set can reliably be extrapolated. As an example, the predicted viscosity of n-octadecane as well as the experimental results from literature are shown in fig. 2.7.

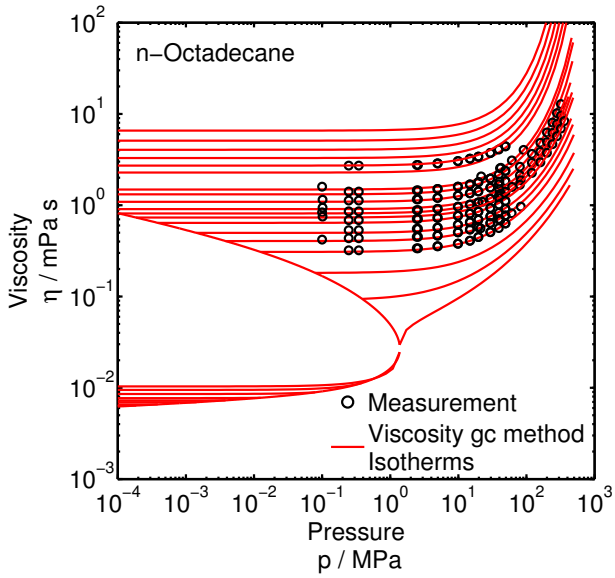


Figure 2.7: Predicted viscosity of n-octadecane.

The parameters adjusted to the six n-alkanes were then used to also predict the viscosity of amines. Out of the 12 amines (see Table 3 of the Supporting Information) only three primary and one secondary amine were used to determine the additional viscosity group contribution parameters of amines (training set). The resulting *MADs* are given in fig. 2.8. Within the data set not considered for the parametrization (evaluation set), the deviations of the primary amines are only slightly higher than the correlated results in tab. 2.4. The prediction for dipropylamine as a secondary amine is noticeable, because the *MAD* is three times higher than the correlated result. Given the small database in our training set, we still assess this result as sufficient. In summary our analysis shows that the group-contribution method is suitable for predicting viscosities. More specifically the parameters are well-behaved, so that the method can be applied to substances outside the training data set. We do not have sufficient data for substances with multiple functional groups

to systematically assess predictions for such substances.

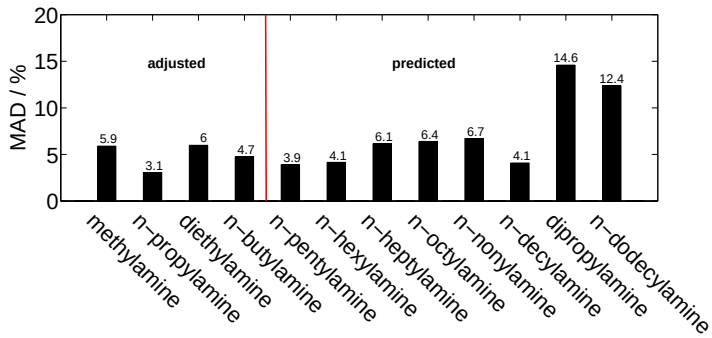


Figure 2.8: Deviations of the predicted viscosity of eight amines with parameters adjusted to experimental data of four amines.

2.3.3 Correlation Result for the Group-Contribution Method

For the final parametrization of the group contribution model, we included the data of all substances considered in our work. For the 31 n-alkanes the parametrization resulted in a mean deviation MAD of 4.81%. Within the class of n-alkanes the maximum deviation is 6.93%, which confirms the proposed mixing rules eq. 2.12 to eq 2.14. Propane was omitted during parameter adjustment, therefore the result for propane shown in tab. 2.2 is predictive. The reduced viscosity $\eta^*(s_{\mathrm{res}})$ of a selection of the investigated n-alkanes along with our calculation is shown in fig. 2.3

Comparing fig. 2.1 and fig. 2.3, the advantage of the entropy-scaling approach becomes obvious: The complex dependence of viscosity on pressure and temperature simplifies to a monovariable dependence of the viscosity on the residual entropy s_{res}. The substances mainly differ in slope and intercept, which corresponds to our parameters B_i and A_i, respectively. The slope as well as the intercept decrease with increasing chain length which is taken into account by our mixing rules. Note, that if the Chapman-Enskog viscosity according to our parametrization, eq. 2.7, were quantitatively correct for all fluids in the limit of ideal gases, then all intercepts would be zero and there would not be a (non-zero) A_i parameter. Fig. 2.9 exemplarily shows the calculated viscosity $\eta(p, T)$ of n-octane applying the parameters in tab. 2.1.

For branched alkanes, we obtain a mean deviation MAD of 10.1% with

Table 2.1: Viscosity Group Parameters Obtained by Adjusting eq. 2.10 with Mixing Rules eq. 2.12 to eq. 2.14 to All Available Data of the Corresponding Group. All the Members of a Chemical Group are listed in the Supporting Information.

Functional group α	$A_\alpha \cdot 10^3$	$B_\alpha \cdot 10$	$C_\alpha \cdot 10^2$
	Methane		
CH_4(methane)	-1.5097	-1.0346	-1.7662
	Ethane		
-CH_3(ethane)	-5.3556	-1.5252	-11.0450
	n-Alkanes		
-CH_3	-8.6878	-1.7951	-12.2359
-CH_2-	-0.9194	-1.3316	-4.2657
	Branched alkanes		
-$CH <$	12.8159	-0.3416	5.5752
$> C <$	152.5629	12.3998	13.8406
	Alkenes		
$=CH_2$	-6.3736	42.4367	-10.8726
$=CH$-	-5.0637	-43.5678	-5.4247
	Aromatics		
-CH- (aromatic)	-4.7664	-1.6842	-6.5606
-C-R (aromatic)	14.4280	0.1104	-0.0699
	Aldehydes		
-R-CH=O	8.8675	-1.4782	-9.1325
	Ester		
-COO-	-0.3295	-1.1893	6.9576
	Ether		
-OCH_3	-4.6001	-1.6433	-5.3118
-OCH_2-	-5.5435	-2.2345	-9.9967
	Ketones		
$> C$=O	-2.2137	-1.4362	1.0417
	Cycloalkanes		
-CH_2- (C5)	-3.8105	-1.6280	-6.7905
-CH_2- (C6)	-3.7896	-1.4796	-4.5017
	Amines		
-NH_2	-4.4048	-0.6089	-8.9017
$> NH$	17.3783	-0.Dyre20145	-11.5122
	Methanol		
H_3C-OH (Methanol)	-11.8859	-0.1253	13.4714
	1-Alkanoles		
-OH	-15.7583	-2.5654	-23.1537
	Water		
$H2O$ *	-14.7515	-1.8512	-29.0646

* Individually adjusted parameter $D_{H2O} = -0.04059$

Table 2.2: Correlation Result for Viscosity Data: Mean Absolute Relative Deviations *MAD* of Viscosities for all *n*-Alkanes.

Substance	*MAD* / %	Number of data points
n-Alkanes	4.81	12010
Methane	4.13	2551
Ethane	4.45	2132
Propane*	10.92	1190
n-Butane	3.78	602
n-Pentane	6.12	1186
n-Hexane	4.55	894
n-Heptane	3.41	473
n-Octane	2.26	491
n-Nonane	1.65	123
n-Decane	2.35	624
n-Undecane	2.81	184
n-Dodecane	2.46	136
n-Tridecane	3.40	84
n-Tetradecane	4.36	82
n-Pentadecane	3.81	160
n-Hexadecane	4.15	107
n-Heptadecane	3.93	276
n-Octadecane	4.33	300
n-Nonadecane	3.72	243
n-Eicosane	5.36	29
n-Heneicosane	3.82	30
n-Docosane	5.15	24
n-Tricosane	7.32	11
n-Tetracosane	7.44	15
n-Octacosane	6.03	14
n-Triacontane	4.82	3
n-Hentriacontane	6.93	5
n-Dotriacontane	1.78	8
n-Tetratriacontane	1.19	2
n-Pentatriacontane	4.66	12
n-Hexatriacontane	4.11	19

* Prediction

Table 2.3: Correlation result for viscosity data: mean absolute relative deviations *MAD* of viscosities for hydrocarbons (other than *n*-alkanes).

Substance	*MAD* / %	Number of data points
Branched alkanes	10.10	1846
Isobutane	8.88	549
Isopentane	5.44	305
Neopentane	16.65	68
2-Methylpentane	29.81	21
3-Methylpentane	7.21	7
2,2-Dimethylbutane	30.83	19
2,3-Dimethylbutane	7.34	6
3-Ethylpentane	16.54	6
2,4-Dimethylpentane	23.27	6
2,2,4-Trimethylpentane	7.16	387
2,3,4-Trimethylpentane	7.36	16
Squalane	19.84	282
2,2,4,4,6,8,8-Heptamethylnonane	5.46	174
Alkenes	7.95	883
1-Propene	12.41	276
1-Hexene	6.51	254
1-Heptene	4.54	107
1-Octene	3.45	106
1-Nonene	8.20	87
1-Decene	7.16	53
Aromatics	4.88	4251
Benzene	7.06	990
Toluene	4.31	1735
Ethylbenzene	6.86	364
1,3-Dimethyl-benzene	2.11	572
1,4-Diemthyl-benzene	3.30	459
Mesitylene	13.73	29
p-Cymene	12.68	15
n-Heptylbenzene	6.07	42
n-Nonylbenzene	1.53	15
n-Dodecylbenzene	6.06	30
Cycloalkanes	3.85	1089
Cyclopentane	1.86	306
Cyclohexane	4.62	783

Table 2.4: correlation result for viscosity data: mean absolute relative deviations *MAD* of viscosities for substances containing oxygen or amine groups.

Substance	*MAD* / %	Number of data points
Ether	4.43	292
Dimethyl ether	3.19	117
Dibutyl ether	3.63	76
Methyl-tert-butyl ether	9.44	6
Dipentyl ether	10.24	16
Dipropyl ether	5.51	77
Ester	5.45	504
n-Propyl acetate	6.45	105
n-Butyl acetate	6.12	114
Isopentyl acetate	7.84	78
Ethenyl acetate	5.53	15
n-Pentyl acetate	3.29	102
n-Heptyl acetate	3.81	90
Aldehydes	10.10	297
Acetaldehyde	34.79	11
Propanal	6.07	7
Butanal	8.91	221
2-Methyl-propanal	9.83	50
3-Methyl-butyraldehyde	14.08	8
Ketones	6.27	680
Acetone	5.24	189
3-Pentanone	5.39	116
Methyl-Isobutyl-ketone	13.41	68
4-Heptanone	4.89	116
2-Hexanone	4.00	86
2-Octanone	7.86	105
Amines	5.10	631
Methylamine	5.82	107
Diethylamine	6.35	105
Propylamine	2.96	44
Dipropylamine	4.54	25
Butylamine	5.04	122
Pentylamine	4.10	61
Hexylamine	4.66	60
Heptylamine	6.12	47
Octylamine	4.91	31
Nonylamine	4.75	14
Decylamine	1.90	14
Dodecylamine	10.08	1
1-Alcohols	10.98	3070
Methanol	5.87	1428
1-Propanol	16.00	372
1-Butanol	18.04	290
1-Pentanol	16.68	64
1-Hexanol	16.78	66
1-Heptanol	15.57	16
1-Octanol	14.34	153
1-Nonanol	7.96	129
1-Decanol	7.34	136
1-Undecanol	12.44	194
1-Dodecanol	25.64	187
1-Tetradecanol	18.03	8
1-Octadecanol	5.81	27
Miscellaneous		
Water	3.09	1406

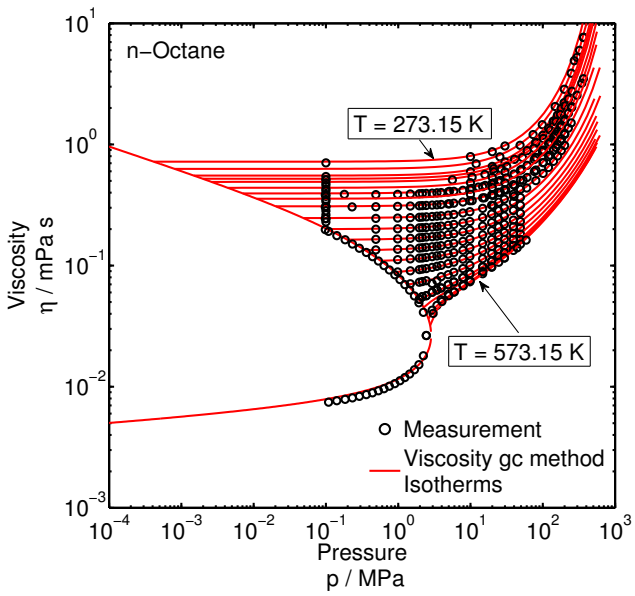


Figure 2.9: Calculated viscosities of n-octane using the parameters given in tab. 2.1.

MADs for various branched alkanes ranging from 5.44% to 30.83%. The resulting *MADs* show the tendency to increase with decreasing number of available data points. In the case of 2-methylpentane and 3-methylpentane, our proposed approach can not differentiate between those two isomers. Both substances have little influence during the parameter regression and strongly depend on the quality of the experimental data: For instance, when the oldest of four references with experimental data of 2-methylpentane is omitted, we find a lower *MAD*-value of 13.44%. We note that other isomers differ with regard to the combined pure substance parameters. Examples are isomeric pairs of isopentane/neopentane; 2, 2-dimethylbutane/2, 3-dimethylbutane; and 2, 2, 4-trimethylpentane/2, 3, 4-trimethylpentane.

The investigated 1-alkenes and aromatics provide a good indication for the applicability of the alkane parameters to other classes of substances. Our correlation results (see tab. 2.3) are in good agreement with experimental data with averaged *MAD*-values of 7.95% for 1-alkenes and 4.88% for aromatic substances, respectively. For both classes, the *MAD*s are not increasing with increasing alkyl moiety. That suggests that the parameters obtained from the n-alkanes are well suited for these types of substances. Mesitylene and p-cymene show significantly higher *MADs*.

As polar substances we investigated esters, ethers, aldehydes and ketones.

Overall the calculated viscosity is in good agreement with the experimental data (see tab. 2.3 and tab. 2.4). For ketones the reduced viscosity is shown in fig. 2.10. Similar to alkanes we find a nearly linear, monovariable dependence of the reduced viscosity on the residual entropy. [...]

Ethenyl acetate is a good indicator for the transferability of the alkene parameters. Only 15 data points were available for ethenyl acetate, so that its impact during the parameter adjustment step is small and the good result (i. e. *MAD* of 5.53%) suggests sufficient transferability of the group-contribution parameters.

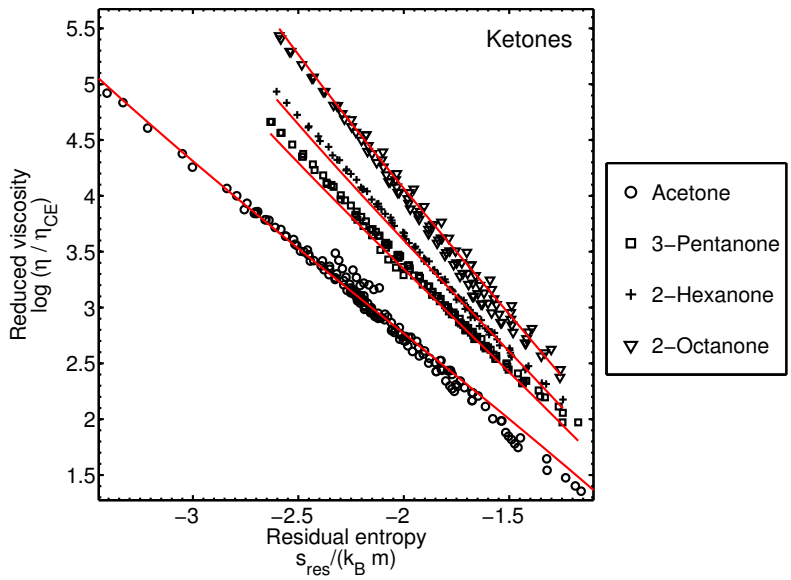


Figure 2.10: Reduced viscosity of the investigated ketones. Symbols represent reduced experimental viscosities for residual entropies calculated with GC-PCP-SAFT, and red lines show the result of our correlation.

As self-associating substances we investigate amines and 1-alcohols. The reduced viscosities are shown in fig. 2.11 and fig. 2.12 respectively. In both classes of substances the entropy-scaling approach results in a monovariable behavior. While the behavior of the amines can still be considered nearly linear, the alcohols show a strong curvature. Especially for longer alcohols like 1-decanol our approach can not represent the strong curvature because the A_i - C_i parameters are dominated by the alkane parameters. Given the limited number of adjustable parameters, we assess the overall representation of the 1-alcohols as still satisfactory. We identify the need for a wider range of experimental data for higher alcohols and amines (towards the gas phase and to

highly dense liquid phases) in order to refine our model or the parametrization of the model.

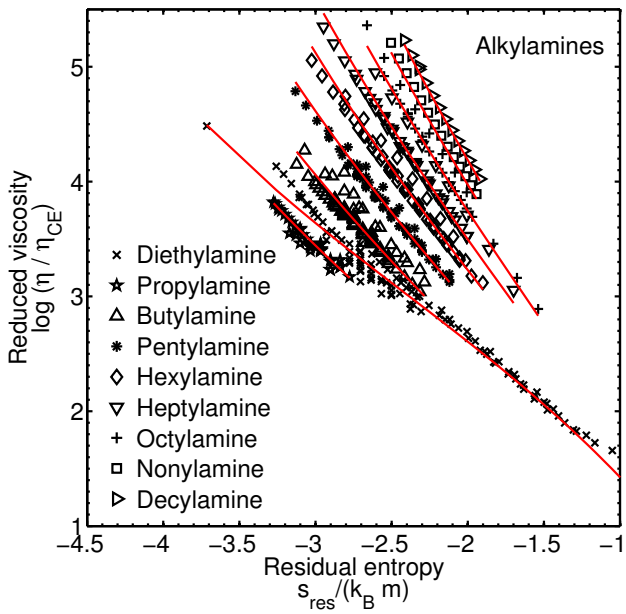


Figure 2.11: Reduced viscosity of amines. Symbols represent reduced experimental viscosities for residual entropies calculated with GC-PCP-SAFT, and red lines show the result of our correlation.

In view of the pronounced curvature of the $\ln(\eta^*)$-s_{res}-relation for alcohols, it is interesting to consider water as a strongly associating substance. For water it is possible to eliminate any ambiguity about the quality in which s_{res} is determined. For pure water, the IAPWS-95 model is a highly accurate equation of state[155,156], which gives s_{res} from correlation of best experimental data. Fig. 2.13 shows the reduced viscosity η^* of water versus the residual entropy, calculated with the IAPWS-95 equation of state. We limited the experimental database for the viscosity of water to literature from the year 1990 or more recent. From fig. 2.13 it is possible to conclude, first, that water obeys the principle of entropy-scaling over the entire range of fluid conditions, with good accuracy. Secondly, the entropy scaling shows marked deviations from a linear relation, with a strong positive curvature at high densities (i.e. low s_{res} values), as well as in the vicinity of the critical point, around $s_{\mathrm{res}}/k_B = -1.9$. Inspired by the question of an anonymous reviewer, we investigated subcooled liquids (as shown in the Supporting Information) and find that the principle of entropy scaling holds for the metastable region (with pronounced curvature

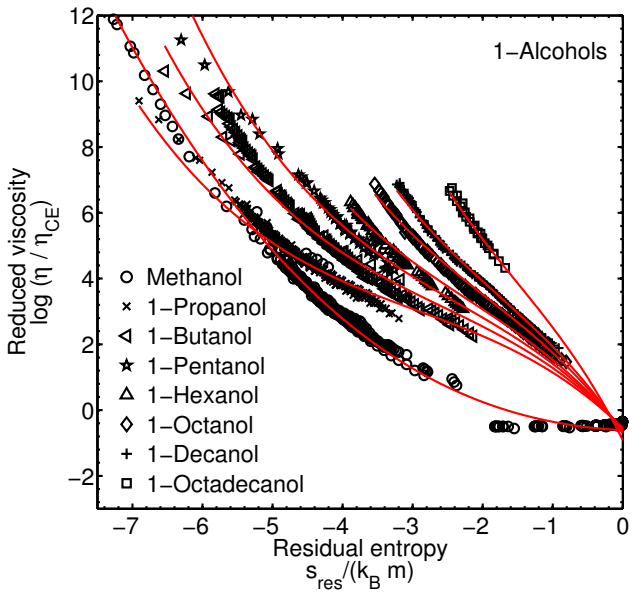


Figure 2.12: Reduced viscosity of 1-alcohols. Symbols represent reduced experimental viscosities for residual entropies calculated with GC-PCP-SAFT, and red lines show the result of our correlation.

of the scaling relation for associating substances).

We now consider s_{res} of water as calculated from the PCP-SAFT model. Water has individual group-contribution parameters in the PCP-SAFT equation of state. We here use pure component parameters that were adjusted to vapor pressure data, to liquid density data and to supercritical isotherms, as $m = 1.271$, $\sigma = 2.82$ Å, $\varepsilon/k = 281.94$ K, $\mu = 1.855$ D, $Q = 3.44$ D with four association sites according to scheme 4C as defined by Huang and Radosz [157] with association parameters $\varepsilon^{AB}/k = 952.66$ K and $\kappa^{AB} = 4.35 \cdot 10^{-3}$. Comparison of the scaling behavior according to the PCP-SAFT equation of state and the IAPWS-95 model, in fig. 2.13, shows good agreement in the dense phase, whereas for lower densities (i.e. $s_{\text{res}}/k_B > -2.5$), the results differ significantly. The most different s_{res}-values from both equations of state occur in the vicinity of the critical point. Despite the different course of the data according to PCP-SAFT compared with IAPWS-95, we find a good, nearly monovariable scaling relation for PCP-SAFT. All four viscosity parameters A_α to D_α were adjusted for water. The correlation leads to average deviations (MAD) of 3.09% over the entire range of conditions, namely 1406 data points, temperature from 248.15 K to 715.57 K and pressure from 0.0522 MPa to

60 MPa.

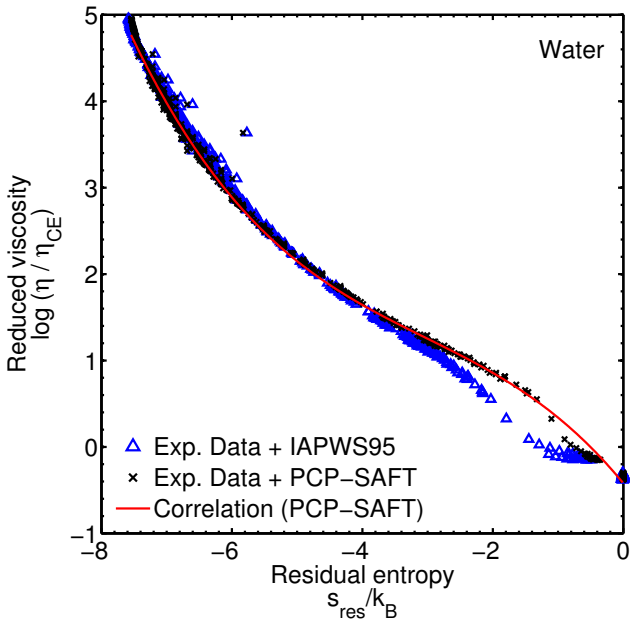


Figure 2.13: Reduced viscosity of water. Symbols represent reduced experimental viscosities versus residual entropies calculated with either the PCP-SAFT EoS (black crosses) or the IAPWS-95 formulation (blue triangles) respectively. The red line represents our correlation.

2.4 Conclusion

In this work, we present a new group contribution method for the correlation and prediction of pure component shear viscosities. The method applies a group contribution equation of state (PCP-SAFT) and uses the entropy scaling according to Rosenfeld. We formulate the entropy scaling in terms of three group contribution parameters per functional group. The method thus requires no pure component parameters. We observe entropy scaling to hold for strongly elongated substances, such as n-hexatriacontane and for hydrogen bonding species, such as 1-alcohols.

While the viscosity of linear molecules can be correlated and predicted very accurately, highly branched molecules show more significant deviations to experimental data. Members of the 1-alcohol family show the strongest hydrogen bonding of the substances we considered. For 1-alcohols we observe pronounced deviations from a linear entropy scaling. Strongly hydrogen bonding fluids thus

justify a further investigation. Moreover, our study is limited to substances with single functional groups. As a subsequent step, we evaluate the applicability of a group contribution approach to predict viscosities of mixtures and the prediction of other transport properties, such as thermal conductivity and diffusion.

3. Bayesian Model Selection helps to Choose Objectively between Thermodynamic Models: A Demonstration of Selecting a Viscosity-Model based on Entropy Scaling

This chapter is a literal quote of the publication: O. Lötgering-Lin, A. Schöniger, W. Nowak, J. Gross; Ind. Eng. Chem. Res., 2016, 55 (38), pp 10191-10207.

Additions implemented in this chapter compared to the original journal-publication are indicated by square brackets.

Specific contributions of the authors of the original publication are as follows: Anneli Schöniger provided the code for the Bayesian Statistics based calculations applied in this chapter. Both Wolfgang Nowak and Anneli Schöniger advised as experts on Bayesian Statistics, especially in the analysis of the calculated results.
I, under the supervision of Joachim Gross, implemented the Bayesian Statistics code blocks into the thermodynamic framework, developed the code for the Markov Chain Monte Carlo calculations, defined the investigated scenarios and the competing thermodynamic models and also conducted the calculations presented in this chapter.

Objective measures to compare the adequacy of models can be very useful to guide the development of thermodynamic models. Thermodynamicists are fre-

quently faced with the so-called bias-variance dilemma where one model may be less accurate in correlating experimental data but more robust in extrapolations than another model. In this work, we use Bayesian model selection (BMS) to identify the optimal balance between bias and variance. BMS is a statistically rigorous procedure that elegantly accounts for experimental errors and implicitly performs a bias-variance trade-off in model selection. We present a first-time application of BMS to thermodynamic model selection. As an example, we consider modeling approaches to predict viscosities using Rosenfeld's entropy scaling approach [Y. Rosenfeld, Phys. Rev. A 1977, 15, 2545 - 2549]. Our goal is to objectively rank the adequacy of three competing model variants that all describe the functional dependence of viscosity on residual entropy: the well established linear regression approach, a recently introduced polynomial approach [Lötgering-Lin, O., Gross, J. Ind. Eng. Chem. Res. 2015, 54, 7942 - 7952] and a sinusoidal approach. We investigate the suitability of the models for extrapolating viscosities to different pressures and to different carbon chain lengths. Technically, we implement in a first step a Markov chain Monte Carlo algorithm to train the competing models on a common data set. In a second step, we determine the statistical evidence for each model in light of an evaluation data set with brute-force Monte Carlo integration. We call this way of implementation *two-step BMS*. Results show that, in general, both non-linear models outperform the linear model, with the polynomial approach being much more reliable for carbon chain length extrapolation. However, assumptions about experimental error influence the choice of the most appropriate model candidate. Hence, we point out the benefits of applying two-step BMS based on specific data sets and specific error assumptions as a situation-specific guide to model selection.

3.1 Introduction

For problems where model development is not guided by physical principles, i.e. for problems that require an empirical ansatz, it is usually ambiguous to compare and evaluate competing models. The reason lies in a limited experimental database that does not allow to clearly discriminate between competing candidate model approaches, including uncertainty about the number of adjustable parameters that are justified for the considered problem. This study is motivated by the uneasy awareness, that the selection of one model among

several candidate models in our previous work[48] [as given in chapter 2] was not guided by a clear methodology. We present a method to make a well-founded and objective comparison between several candidate models. The candidate models differ in the mathematical ansatz functions, but also differ in their numbers of adjustable parameters.

The most frequently used criterion for choosing a model among several model candidates is the accuracy of the proposed models. However, differences in model accuracy might be within the expected range of experimental error (i.e., measurement errors in the data set). The noisy data should then neither be 'over- nor underfitted' in order to obtain reliable model predictions. A simple model with only few adjustable parameters will have uniquely identifiable parameter values and a small prediction variance (a high confidence in predictions). However, it will not be able to capture all relevant features in the training data set (it will 'underfit' the data) and is therefore expected to show a significant bias (error) in predictions; that bias will occur both in the training phase as well as beyond the training conditions. In contrast, a highly flexible model with a large number of adjustable parameters will achieve a significantly lower bias in the training phase, but it tends to overfit noise in the training data set and leads to degeneracy in parameters (i.e. correlated parameters that are not uniquely identifiable.) Subsequently, this leads to a larger variance (lower confidence) in predictions especially beyond training conditions. This dilemma is known as the bias-variance tradeoff in statistics[123]. The principle of parsimony or "Occam's razor"[124] suggests to approach this dilemma by choosing the most parsimonious model from a set of alternative models that is still able to fit the training data reasonably well. This choice can either be guided by intuition, or objective statistical measures can be used to identify an optimal balance between accuracy and parsimony.
Bayesian model selection (BMS)[125,126] is a statistical multi-model approach that yields a quantitative comparison of alternative models based on Bayes' theorem. Measurement noise in the available data (for training) is accounted for on a statistical basis. The outcome of BMS is a set of model weights that represent posterior probabilities for each model to be the most adequate one in the set of considered models. With these weights, the competing models can be ranked and the most adequate model can be selected[128,129]. BMS thus serves as a framework for Bayesian hypothesis testing, in which several alternative models can be tested against each other in a probabilistic manner.

If none of the competing models turns out to be clearly the best one in the set, model weights can be used to combine the individual model predictions into a weighted average. This procedure is known as Bayesian model averaging (BMA)[130,131]. In this case, the contribution of model choice uncertainty (at least the fraction of uncertainty which is captured by the differences between the available competing model candidates) to the overall prediction uncertainty can be quantified[158,159]. Finally, data-worth analyses based on BMS can help to identify model structural errors[132] and to plan future experiments to increase the confidence in model selection[133,134]. BMS and BMA have been successfully applied in many disciplines, such as sociology[128], ecology[160], hydrogeology[161], or contaminant hydrology[162]. In this study, we present to the best of our knowledge a first-time application of BMS to thermodynamic model selection.

There are several key advantages of using a statistical (here: Bayesian) approach to model evaluation: First, experimental errors can be accounted for. The rigorous treatment of experimental errors in thermodynamic model building and selection has been recognized, but has not been widely applied yet[136]. Second, acknowledging that the experimental data are not a perfect (i.e. complete and error-free) representation of the true underlying system leads to the concept of uncertain parameters. Instead of identifying a single best parameter set which achieves the closest agreement with the measured data, we obtain a probability distribution of parameter values. Third, this distribution reflects the limits of parameter identifiability due to the presence of measurement noise and due to a limited amount of data. Fourth, when propagating parameter uncertainty to model predictions, we obtain predictive probability distributions instead of single fixed predictions. This allows us to equip our model predictions with credible intervals, which is of great value to analyse thermodynamic models by statistical means[136] or guide e.g. optimal design strategies for future experiments[137]. Finally, model ranking might turn out differently when also accounting for predictive uncertainty instead of solely focusing on the performance of a best-fit parameter set. Therefore, BMS is well-suited for robust and reliable model selection.

The impact of the assumed level of measurement noise on model ranking results has been discussed and investigated by Schöniger et al.[163]. Here, we consider five different noise levels in order to cover a realistic range of experimental error corresponding to experiments carried out with various techniques and under various conditions.

An elegant approach for developing predictive models for transport properties, such as shear viscosity η, is Rosenfeld's entropy scaling principle[1,67]. In the entropy scaling framework, a dimensionless form of the viscosity η^* is recognized as a function of the residual entropy s_{res} only. In most previous works on entropy scaling, the relation between the dimensionless viscosity and the residual entropy $\eta^*(s_{\mathrm{res}})$ was described by a linear approach[1,50,51,71,73], even though deviations from a linear relation were reported[45,46,48,61,81,152]. Recently, some of the authors of this study proposed a group-contribution approach for predicting $\eta^*(s_{\mathrm{res}})$ using a 3rd-order polynomial ansatz[48] [described in chapter 2], with s_{res} from the perturbed-chain statistical associating fluid theory (PC-SAFT) equation of state. In the process of developing the 3rd-order polynomial (cubic) model, other alternative candidate models were considered as well.

The original idea of scaling transport properties with the residual entropy s_{res} goes back to Rosenfeld[1] in 1977. He was the first to show that there exists a monovariable relation between a dimensionless form of the viscosity η^* and the residual entropy s_{res}[1,67]. This entropy-scaling approach was recently applied and further developed[1,45–47,50,51,67,68,71–75,77–79,81,84]. While Rosenfeld investigated spherical, monoatomic liquids only[1,67,68], the entropy-scaling approach was shown to hold also for non-spherical model fluids (Lennard-Jones chains)[47] and dumbbell-shaped particles[46], for non-spherical real substances, n-alkanes[71,72,84], and other real systems[45,74,75], including water[77–79,81] as a substance with highly orientation-dependent interactions. Recent works introduce the Chapman-Enskog viscosity for making the viscosity dimensionless, which leads to a very well-behaved functional form of entropy scaling[50,51,71,73].

As main contributions of this study, we (1) present a first-time application of the BMS approach to thermodynamic model selection, (2) investigate the bias-variance trade-off behavior of three distinct models that predict viscosities via entropy scaling, and (3) identify the most appropriate modeling approach to predict viscosities over large pressure ranges and for carbon chain length extrapolation, taking into account realistic levels of measurement noise in the viscosity data.

Our case study solely covers a viscosity model. We emphasize however, that the BMS method is applicable to arbitrary models, including models for static properties (e.g. the calculation of vapor-liquid equilibria[136]). Our case study is suited to provide guidance on how to apply BMS and how to interpret its

results.

3.2 Theoretical Background

In this section we first provide details about correlating and predicting shear viscosities using entropy scaling. Subsequently, we review BMS as a method for discriminating and choosing between a set of candidate models.

3.2.1 Calculation of Viscosity via Entropy Scaling

Entropy Scaling

In entropy scaling, the dimensionless viscosity $\eta^*(\rho, T)$, like self-diffusion coefficients and thermal conductivity, is expressed as a function of the residual entropy $s_{\text{res}}(\rho, T)$ only[1,67], so that

$$\eta^*(\rho, T) = \eta^*(s_{\text{res}}(\rho, T)) \tag{3.1}$$

where ρ is the density and T is temperature. To obtain the dimensionless viscosity η^*, Novak[51] proposed to use the Chapman-Enskog viscosity η_{CE}

$$\eta_{\text{CE}} = \frac{5}{16} \frac{\sqrt{M k_B T / (N_A \pi)}}{\sigma^2 \, \Omega^*} \tag{3.2}$$

so that

$$\eta^* = \frac{\eta}{\eta_{\text{CE}}} \tag{3.3}$$

Here, M is the molar mass, k_B and N_A denote Boltzmann's constant and Avogadro's constant, respectively, σ is the particle diameter, and Ω^* is the reduced collision integral[2,149]. The residual entropy is defined as the difference of the entropy at given density ρ and temperature T and the ideal gas entropy at the same density and temperature: $s_{\text{res}}(\rho, T) = s(\rho, T) - s^{\text{ig}}(\rho, T)$. By applying eq. 3.3 to experimental viscosity data of three n-alkanes (see fig. 3.1a - c), the complex temperature and pressure dependence (i.e. $\eta(p, T)$) collapses to a monovariable, nearly linear component-specific curve, $\eta^*(s_{\text{res}})$, as shown in fig. 3.1d. Therefore, entropy scaling is one of the few approaches that is able to cover a wide range of pressure and temperature (including liquid, gaseous as well as supercritical conditions) with a very simple, well-behaved functional

form. Once the function $\eta^*(s_{\text{res}})$ is available, one can easily back-calculate $\eta\,(\rho, T)$ from eq. 3.1 and 3.3, using the analytic equation of state.

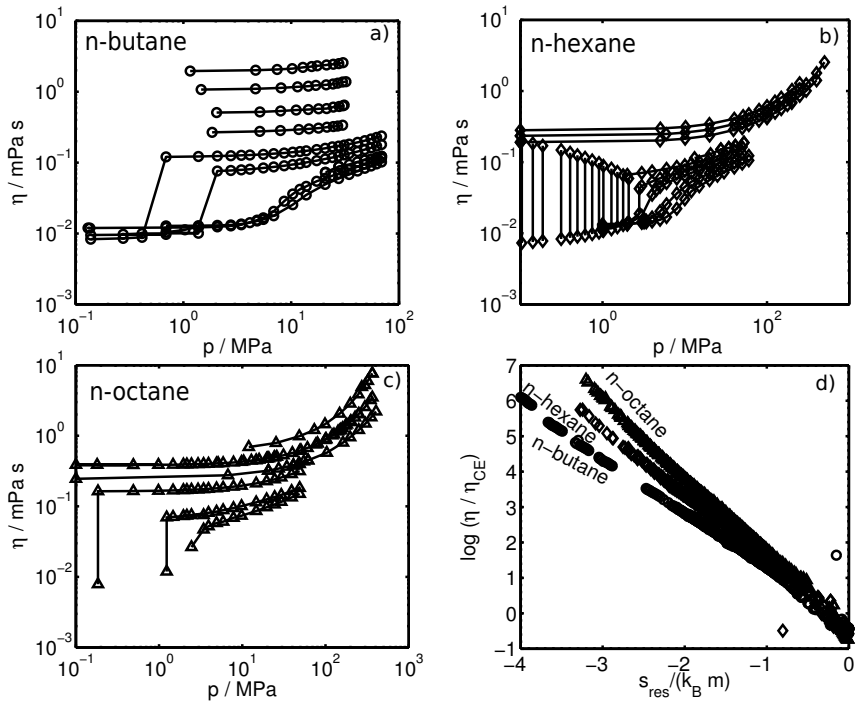


Figure 3.1: Experimental viscosities of n-butane (a), n-hexane (b) and n-octane (c) as a function of pressure p and temperature T. Lines are a guide to the eye connecting viscosities at the same temperature. d) Dimensionless experimental viscosities η_i^* of n-butane, n-octane and n-hexane as a function of the residual entropy s_{res}.

Regression Approaches

The functional dependence $\eta^*(s_{\text{res}})$ is specific to each substance. Ansatz functions for $\eta^*(s_{\text{res}})$ need to be used. In this section we present selected entropy scaling models that will be evaluated in the remainder of this work.

In line with Rosenfeld's original approach[1], several authors[50,51,68,71,73,75] applied a linear function of the form

$$\ln(\frac{\eta}{\eta_{\text{CE}}}) = A + Bz \qquad \text{linear model, } M_{\text{lin}} \tag{3.4}$$

with

$$z = \frac{s_{\text{res}}}{k_B m} \tag{3.5}$$

and segment number m to determine the viscosity of a pure substance. Henceforth, we will call this model the linear model M_{lin}.

Recently, Lötgering-Lin and Gross[48] [presented in chapter 2] proposed an approach that adds a second and third order polynomial to the linear part of M_{lin}. Deviations from the linear behavior as found by, e.g., Chopra et al.[45],[46], Lötgering-Lin and Gross[48], Ingebrigtsen et al.[61], Mittal et al.[152] and Johnson and Head-Gordon[81] are then better represented. In this work, this non-linear, polynomial model will be referred to as M_{pol}, given as

$$\ln(\frac{\eta}{\eta_{\text{CE}}}) = A + Bz + Cz^2 + Dz^3 \qquad \text{polynomial model, } M_{\text{pol}} \tag{3.6}$$

During the process of developing the polynomial model M_{pol}[48], [as described in chapter 2,] the possibility to describe the non-linear part of $\eta^*(s_{\text{res}})$ using a sinusoidal term was considered. Therefore, the third regression approach we investigate here consists of the linear part of M_{lin} and an added sinusoidal term. We henceforth refer to this model as M_{sin}, with

$$\ln(\frac{\eta}{\eta_{\text{CE}}}) = A + Bz + C\sin(Dz) \qquad \text{sinusoidal model, } M_{\text{sin}} \tag{3.7}$$

These three different regression approaches will be evaluated and compared against each other in this study. The polynomial model M_{pol} equals the sinusoidal model M_{sin} in terms of complexity (i.e. number of adjustable parameters). The linear model M_{lin} is simpler, but represents the most established approach. In engineering applications, one usually selects models of a certain complexity (number of adjustable parameters) by intuition. We wish to rationalize such decisions.

Calculating s_{res} with the Perturbed Chain Statistical Associating Fluid Theory

To evaluate the proposed regression approaches, we need access to the residual entropy s_{res} as well as molecular parameters to calculate the Chapman-Enskog viscosity η_{CE}. We use the perturbed chain statistical associating fluid theory (PC-SAFT) Equation of State (EoS) for this purpose[105]. In the SAFT framework, molecules are described as linear chains of tangentially bonded identical segments. The corresponding pure-component parameters are the

segment number m, the segment size parameter σ and the energy parameter ε describing the Van der Waals attraction.

The specific residual Helmholtz energy $a_{\text{res}} = A_{\text{res}}/N$ can be calculated as given by Gross and Sadowski[105], with N for the number of molecules. For the specific residual entropy, we have

$$s_{\text{res}}(\rho, T) = -\left(\frac{\partial a_{\text{res}}}{\partial T}\right)_{\rho} \tag{3.8}$$

The same set of pure-component parameters is used to also calculate the corresponding Chapman-Enskog viscosity η_{CE} from eq. 3.2.

3.2.2 Group Contribution-Based Approach

We work with a group contribution approach, where molecules are decomposed into functional groups α, such as CH_3. Parametrization is limited to the functional group parameters. Pure-substance parameters are then obtained by combining the functional group parameters with appropriate mixing rules.

To obtain the PC-SAFT parameters m, σ and ε from the group parameters m_α, σ_α and ε_α, we use the mixing rules suggested by Vijande et al.[114]. These group parameters are adjusted to liquid densities $\rho_{liq,i}$ and vapor pressures $p_{sat,i}$ as published by Sauer et al.[118]. As part of the group contribution method, we use a slightly modified Chapman-Enskog expression that is related to the segment number m[48,51],

$$\eta_{\text{CE}} = \frac{5}{16}\frac{\sqrt{M k_B T/(m N_A \pi)}}{\sigma^2\,\Omega^*} \tag{3.9}$$

According to the idea of group-contribution approaches, the parameters A and B of the three models M_{lin}, M_{pol}, and M_{sin} (eqs 3.4 to 3.7) are obtained from mixing rules, given byLötgering-Lin and Gross[48] [, see chapter 2, as]

$$A = \sum_\alpha n_\alpha\, m_\alpha\, \sigma_\alpha^3\, A_\alpha \tag{3.10}$$

$$B = \sum_\alpha \frac{n_\alpha\, m_\alpha\, \sigma_\alpha^3}{V_{tot}^{0.45}}\, B_\alpha \tag{3.11}$$

with

$$V_{tot} = \sum_\alpha n_\alpha\, m_\alpha\, \sigma_\alpha^3 \tag{3.12}$$

where n_α denotes the number of functional groups of type α in the considered substance. When targeting the viscosity of n-alkanes, for example, the linear model M_{lin} requires four adjustable parameters (A_{CH2}, B_{CH2}, A_{CH3} and B_{CH3}).

The additional mixing rules for the parameters C and D of polynomial model M_{pol} are[48], [as given chapter 2,]

$$C = \sum_\alpha n_\alpha \, C_\alpha \tag{3.13}$$

$$D = D^{univ} \sum_\alpha n_\alpha \tag{3.14}$$

Note that we introduced a universal parameter $D^{univ} = -0.01245$ for all functional groups α (i.e. $D^{univ} = D_\alpha$, that means for n-alkanes: $D^{univ} = D_{CH2} = D_{CH3}$). For n-alkanes the model has six adjustable parameters A_{CH2} to C_{CH2} and A_{CH3} to C_{CH3}. We use the abbreviation $M_{\text{pol,6p}}$ for this model. To compare two structurally similar models with different numbers of adjustable parameters, we analyze a second version of the polynomial model M_{pol}, where the parameter D^{univ} is treated as adjustable (but still universal for all functional groups α). For n-alkanes, this results in seven adjustable parameters (A_{CH2} to C_{CH2}, A_{CH3} to C_{CH3} and the universal D^{univ} parameter). We refer to this version as the seven-parameter polynomial model $M_{\text{pol,7p}}$.

In preparation of an earlier publication[48][, see chapter 2], some of the authors of the current study had developed promising mixing rules for the sinusoidal model M_{sin}, given as

$$C = \sum_\alpha \frac{n_\alpha}{\sum_\alpha n_\alpha} \, C_\alpha \tag{3.15}$$

$$D = D^{\text{univ}} \sum_\alpha \frac{n_\alpha}{\sum_\alpha n_\alpha} \tag{3.16}$$

Analogously to $M_{\text{pol,6p}}$ and $M_{\text{pol,7p}}$, we investigate two sinusoidal models; one with seven adjustable group parameters in the case of n-alkanes ($M_{\text{sin,7p}}$), and one with six parameters ($M_{\text{sin,6p}}$) and a constant D^{univ}-parameter, $D^{\text{univ}} = 1.6434$. The value of D^{univ} in $M_{\text{pol,6p}}$ as well as in $M_{\text{sin,6p}}$ was obtained by adjustment to experimental viscosities of n-alkanes, ranging from n-butane up to n-hexatriacontane.

In the remainder of this manuscript, parameters of model M_k (i.e. M_{lin},

$M_{\text{pol,6p}}$, $M_{\text{pol,7p}}$, $M_{\text{sin,6p}}$, or $M_{\text{sin,7p}}$) are denoted as $\boldsymbol{\theta}_k$. The dimension of vector $\boldsymbol{\theta}_k$ depends on the model M_k, as

$$\boldsymbol{\theta}_k = (\underbrace{\underbrace{\underbrace{A_{CH3,k}, B_{CH3,k}}_{M_{\text{lin}}}, C_{CH3,k}}_{M_{\text{pol,6p}} \text{ or } M_{\text{sin,6p}}}, D^{\text{univ}}}_{M_{\text{pol,7p}} \text{ or } M_{\text{sin,7p}}}, \quad A_{CH2,k}, ...)^T$$

3.2.3 Bayesian Framework for Model Evaluation

The scientific jargon used in the field of thermodynamics differs from that of scientific texts in the field of statistics. Here, we use the expressions that belong to thermodynamic texts. However, in order to harmonize with literature from the field of statistics, we add the corresponding synonym used in the statistical community in curly brackets (at the first occurrence).

Model Parametrization via Bayesian Updating

In this chapter, we explain the principles of Bayesian Updating that is used to parametrize {calibrate} a model. Here, we are concerned with a number N_m of competing models M_k. All of them predict a set of viscosity values $\boldsymbol{\eta}^{\text{pred}}$ as a function of residual entropy $\boldsymbol{s}_{\text{res}}$ and of model parameters $\boldsymbol{\theta}_k$, as

$$\boldsymbol{\eta}_k^{\text{pred}} = M_k\left(\boldsymbol{\theta}_k\right) = f_k\left(\boldsymbol{\theta}_k, \boldsymbol{s}_{\text{res}}\right) \tag{3.17}$$

The uncertainty in model parameters is defined by a probability distribution $p\left(\boldsymbol{\theta}_k|M_k\right)$, where $(\cdot|\cdot)$ symbolizes a conditional probability distribution and vectors are denoted by bold symbols. We read $p\left(\boldsymbol{\theta}_k|M_k\right)$ as the probability density of a set of model parameters $\boldsymbol{\theta}_k$ for a given model M_k. Because the uncertainty in parameters propagates via the model to the predictions, for each model we obtain a predictive distribution $p\left(\boldsymbol{\eta}_k^{\text{pred}}|M_k\right)$ instead of deterministic values $\boldsymbol{\eta}_k^{\text{pred}}$. This viewpoint is an essential difference to conventional approaches of thermodynamic model building.

Viscosity measurements $\boldsymbol{\eta}^{\text{meas}}$ are available at various values of pressure and temperature to parametrize the models. Through the available data, the true

viscosity values are assumed known up to an experimental error $\boldsymbol{e}_\eta$, that is

$$\boldsymbol{\eta}^{\text{meas}} = \boldsymbol{\eta}^{\text{true}} + \boldsymbol{e}_\eta \tag{3.18}$$

Assumptions on experimental errors are formulated as a probability distribution $p(\boldsymbol{e}_\eta)$, which is typically an independent Gaussian distribution for each measurement with zero mean (i.e., without systematic error) and an appropriately chosen value for the variance.

Our objective is now to determine a parameterization $\boldsymbol{\theta}_k$ using measurement data $\boldsymbol{\eta}^{\text{meas}}$ in the form of a probability distribution $p(\boldsymbol{\theta}_k|\boldsymbol{\eta}^{\text{meas}}, M_k)$. Before this step one may, for some classes of models, have a prior parameter distribution $p(\boldsymbol{\theta}_k|M_k)$ available. For empirical models, it is uncommon to have significant prior knowledge about model parameters, so that the parameter distribution $p(\boldsymbol{\theta}_k|M_k)$ is taken as constant for each value of the parameter vector (within a range of possible values). The parametrization {calibration} of the model using measurement data $\boldsymbol{\eta}^{\text{meas}}$ leads to a posterior (updated) parameter distribution $p(\boldsymbol{\theta}_k|\boldsymbol{\eta}^{\text{meas}}, M_k)$ from Bayes' theorem

$$p(\boldsymbol{\theta}_k|\boldsymbol{\eta}^{\text{meas}}, M_k) = \frac{p(\boldsymbol{\eta}^{\text{meas}}|M_k, \boldsymbol{\theta}_k)\, p(\boldsymbol{\theta}_k|M_k)}{p(\boldsymbol{\eta}^{\text{meas}}|M_k)} \tag{3.19}$$

with $p(\boldsymbol{\eta}^{\text{meas}}|M_k, \boldsymbol{\theta}_k)$ representing the likelihood of the parameter set $\boldsymbol{\theta}_k$ to have generated the measured viscosity data. We use the expression 'likelihood', as opposed to 'probability density', when model outcomes are evaluated in the light of experimental data.

In eq (3.19), $p(\boldsymbol{\eta}^{\text{meas}}|M_k)$ merely acts as a normalizing constant for the posterior of the parameters $p(\boldsymbol{\theta}_k|\boldsymbol{\eta}^{\text{meas}}, M_k)$. For Monte Carlo techniques, the normalizing constant is not required to produce a sample of the posterior parameter distribution. In the case of model comparison, however, $p(\boldsymbol{\eta}^{\text{meas}}|M_k)$ is the targeted quantity and is referred to as Bayesian model evidence (BME). The BME values need to be evaluated with numerical or approximative techniques (explained in more detail in the following section).

The posterior parameter distribution of each model $p(\boldsymbol{\theta}_k|\boldsymbol{\eta}^{\text{meas}}, M_k)$ can now be used to determine the predictive distribution $p\left(\boldsymbol{\eta}_k^{\text{pred}}|\boldsymbol{\eta}^{\text{meas}}, M_k\right)$ by running the respective model with a sample from the posterior parameter distribution.

If predictions shall be made beyond a training data set, a predictive distribution can also be obtained for viscosity values that have not (yet) been used

for Bayesian updating. In this case, $\boldsymbol{\eta}^{\text{meas}}$ and $\boldsymbol{\eta}_k^{\text{pred}}$ would not refer to the same data points. We will differentiate between different data sets used for training and evaluation of the models in our case study (see implementation details).

Bayesian Model Selection

Here, we briefly present the statistical framework of the BMS analysis. We refer to Draper [125] and Hoeting et al. [126] for a more comprehensive introduction to BMS and BMA. In these approaches, all probabilities and statistics are conditional on the considered set of models. While the choice of the model set is inherently subjective, the rigorous statistical procedure to obtain posterior model weights allows for an objective model comparison within the considered set.

The prior model probability is denoted as $P(M_k)$ and reflects a prior preference of one model over another. This preference before updating any model with measurement data, is subjective. One can for example prefer a certain model over another because it is known to obey physical limits. In our study, we have no preference for one or the other model and assign equal values for the prior model probability $P(M_k)$ that necessarily sum up to one. In BMS, the prior model probability is updated to a posterior probability $P(M_k|\boldsymbol{\eta}^{\text{meas}})$ of model M_k. The posterior probability is meaningful because it gives the probability that model M_k is the best model from the set of considered models in light of the data set $\boldsymbol{\eta}^{\text{meas}}$, given by

$$P(M_k|\boldsymbol{\eta}^{\text{meas}}) = \frac{p(\boldsymbol{\eta}^{\text{meas}}|M_k)\,P(M_k)}{\sum_{i=1}^{N_m} p(\boldsymbol{\eta}^{\text{meas}}|M_i)\,P(M_i)} \tag{3.20}$$

This step is analogous to the derivation of the posterior parameter distribution, eq (3.19). The posterior probabilities can then be used as weights for model ranking and model selection. Note that $P(\cdot)$ refers to a discrete probability distribution for a discrete set of model alternatives, whereas $p(\cdot)$ refers to the probability density function of a continuous variable such as model parameters.

BMS is able to identify an optimal compromise between bias (related to model performance) and variance (related to model complexity, i.e. number of adjustable parameters) because it implicitly follows Occam's razor [164,165]. We will now briefly explain the mechanism in BMS that allows for the balancing of model complexity with model performance. Model complexity is encoded in

the flexibility of model predictions, which results from the interplay between the model structure (including the number of parameters) and the allowed flexibility in model parameters (described as prior parameter distributions; for empirical models often assumed constant within bounds). The parameter uncertainty is propagated through the model to the predictions, resulting in a predictive distribution. The predictions are then evaluated based on their mismatch with the measurement data. The more accurate and precise the predictive distribution, the higher is the BMS score $P\left(M_k|\boldsymbol{\eta}^{\text{meas}}\right)$ of the model. A large variance or a high bias in predictions, however, will be penalized by BMS. Through this mechanism, the model which offers the best trade-off solution will rank first in the model comparison.

According to the section 'Bayesian updating', *BME* quantifies the likelihood of model M_k with parameter vector $\boldsymbol{\theta}_k$ to exactly reproduce the measured data, integrated {marginalized} over the model's parameter space Ω_k, as

$$BME = p\left(\boldsymbol{\eta}^{\text{meas}}|M_k\right) = \int_{\Omega_k} p\left(\boldsymbol{\eta}^{\text{meas}}|M_k, \boldsymbol{\theta}_k\right) p\left(\boldsymbol{\theta}_k|M_k\right) d\boldsymbol{\theta}_k \tag{3.21}$$

As in eq. 3.19, $p\left(\boldsymbol{\theta}_k|M_k\right)$ denotes the prior distribution of the parameters $\boldsymbol{\theta}_k$ for model M_k. Evaluating the integral in eq. 3.21 numerically[166] typically requires a high computational effort. Mathematical approximations by so-called information criteria have therefore gained popularity instead. The most frequently used ones within the context of BMS are the Kashyap information criterion (KIC)[167], the Bayesian information criterion (BIC)[168], and the Akaike information criterion (AIC)[169]. These information criteria are computationally efficient, but are prone to yield biased model ranking results as noticed in a variety of studies[170–173]. Schöniger et al.[174] have benchmarked the information criteria against reference solutions for *BME* under both synthetic and real-world conditions. Results have clearly demonstrated that information criteria are poor approximations of *BME* in the case of nonlinear models (i.e. nonlinear in terms of their parameters, not necessarily in the model's input-to-output relation) and produce misleading model ranking results. To obtain reliable model weights, there is still no alternative to brute-force numerical evaluation. In this study we determine *BME* via brute-force Monte Carlo (MC) integration of eq. 3.21.

To interpret the significance of model weights in favor of a specific model, we use the Bayes factor *BF*[166]. It is given by the ratio of *BME* values for

the respective best model M_k and its competitor M_l (e.g. the model with the second highest BME score), as

$$BF\left(M_k, M_l\right) = \frac{p\left(\boldsymbol{\eta}^{\text{meas}}|M_k\right)}{p\left(\boldsymbol{\eta}^{\text{meas}}|M_l\right)} \tag{3.22}$$

Jeffreys[175] provided a rule of thumb for classifying Bayes factor values according to their significance: a Bayes factor of 1 to 3 is "not worth more than a bare mention", a factor of 3 to 10 can be understood as "substantial" evidence in favor of model M_k, a factor between 10 and 100 represents "strong" evidence and a factor of more than 100 can be interpreted as "decisive". These thresholds can be used as guidance when choosing a model from a set.

3.3 Application of Bayesian Model Selection to Thermodynamic Models

In this chapter, we explain how the BMS procedure is applied to select a model among several candidate models for the case of viscosity calculation based on group contributions. First, we discuss how experimental errors (in a statistical sense) impact model building and model selection. Then, we describe how the BMS routine is implemented numerically in our specific study. Finally, we introduce two cases that differ in two aspects: (1) in the experimental viscosity data used for conditioning the models (referred to as training data set, $\boldsymbol{\eta}^{\text{meas}}_{\text{train}}$), and (2) in the data that shall be predicted (referred to as evaluation data set, $\boldsymbol{\eta}^{\text{meas}}_{\text{eval}}$). Results of our case study will be discussed in the following chapter.

3.3.1 Impact of Experimental Errors

Experimental errors influence the proposed models at three stages (see fig. 3.2). Experimental errors in liquid densities $\rho_{liq,i}$ and vapor pressures $p_{sat,i}$ (Part A in fig. 3.2) lead to uncertainty in the PC-SAFT group parameters m_α, σ_α and ε_α, which subsequently lead to uncertainty in the calculated residual entropy s_{res} and Chapman-Enskog viscosity η_{CE}. Further, experimental errors in the measured viscosity η^{meas} lead to uncertainty in the viscosity model parameters A_α, B_α, C_α and D (Part B in fig. 3.2) and therefore to uncertain model predictions $\boldsymbol{\eta}^{\text{pred}}_{\text{train}}$ and $\boldsymbol{\eta}^{\text{pred}}_{\text{eval}}$, respectively. To finally obtain the model ranking, the distribution of predicted viscosities at the conditions of the evaluation

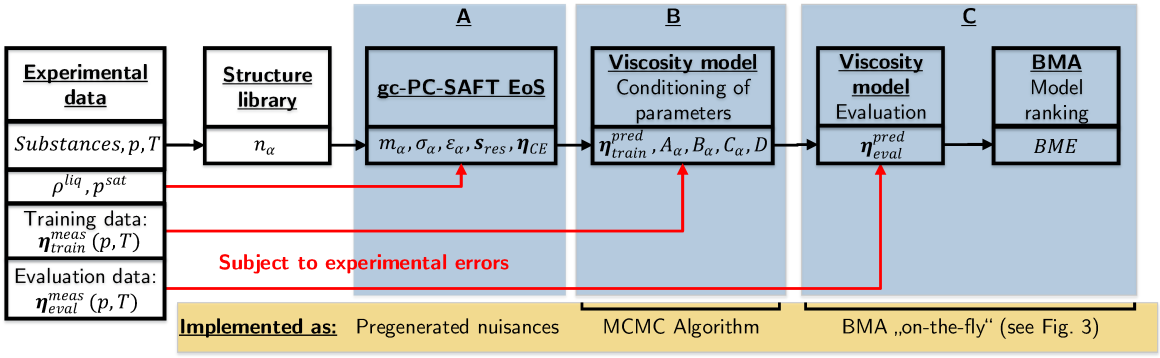


Figure 3.2: Schematic illustration of the steps involved in model building, model evaluation and model ranking based on experimental data. Red lines highlight the influence of experimental errors.

data set $\boldsymbol{\eta}_{\text{eval}}^{\text{pred}}$ is assessed in the light of the experimental data at the evaluation conditions $\eta_{\text{eval}}^{\text{meas}}$, which also are subject to experimental errors (Part C in fig. 3.2).

The Bayesian framework natively accounts for the various experimental errors in model building and in ranking candidate models. Details on the numerical implementation and the assumptions about measurement noise are provided in the next section.

3.3.2 Implementation of the Model Selection Framework

Motivation for the Proposed Two-Step Procedure

A key ingredient of the Bayesian approach to model parametrization and ranking lies in the definition of prior knowledge about parameter distributions. Especially model ranking via BMS relies heavily on the assumption that the chosen parameter priors reflect the current state of knowledge reasonably well; these prior distributions implicitly enter the definition of model complexity as they declare parameters to be more or less constrained. Only if the prior is a reasonable representation of current knowledge, a fair comparison is warranted. Because we are concerned with purely empirical, data-driven models, the definition of valid parameter ranges *before* any data have been taken into account seems arbitrary and hard to justify. Only together, the combination of model equations and training data make up a ready-to-apply data-driven model. We therefore consider the more natural situation where we have already gathered a set of training data that can be used for Bayesian updating

(conditioning) according to eq. 3.19 in a first step, and perform model ranking based on the state of knowledge *after* step 1 in a second step according to eq. 3.20. We summarize this approach as *two-step BMS*. The situation where a previous posterior is used as current prior is commonly known as sequential Bayesian updating[176]. In this study, the second step is a validation of the trained model from step one.

Step 1: Model Conditioning with Markov Chain Monte Carlo Sampling

We denote the measured and the predicted data set for step 1 as $\boldsymbol{\eta}_{\text{train}}^{\text{meas}}$ and $\boldsymbol{\eta}_{\text{train}}^{\text{pred}}$, respectively. Conceptually, this training step {model calibration} corresponds to Part B in fig. 3.2. For implementation, we use an Markov Chain Monte Carlo (MCMC) algorithm to produce posterior (updated) parameter samples, i.e. samples from the posterior parameter distribution $p\left(\boldsymbol{\theta}_k|\boldsymbol{\eta}_{\text{train}}^{\text{meas}}, M_k\right)$. We choose the MCMC approach for its efficiency when confronted with uninformative parameter priors. We speak of uninformative parameters when no significant prior knowledge is available, i.e. when a random parameter guess is as good as any other guess. In our case, we define all prior parameter distributions as uniform and only prescribe generous lower and upper bounds (lb and ub, respectively). Hence, for each element i of the vector $\boldsymbol{\theta_k}$ we use

$$p\left(\theta_{i,k}|M_k\right) = \begin{cases} \frac{1}{ub-lb} & \text{for } lb < \theta_{i,k} < ub \\ 0 & \text{for } \theta_{i,k} < lb \;\text{ or }\; \theta_{i,k} > ub \end{cases} \tag{3.23}$$

These bounds were deliberately set wide to ensure they do not limit parameter conditioning. These uninformative priors reflect our reluctance to define parameter distributions in a data-driven model before training the model with a training data set. Hence, we start from vague prior knowledge and let the training data speak for themselves in form of the posterior distribution from the current updating step.

The algorithm of the MCMC routine is schematically illustrated in fig. 3.3 (Part B, cf. also Part B in fig. 3.2). Starting with a randomly drawn sample from the prescribed parameter prior, a single viscosity group parameter is randomly chosen and randomly perturbed (parameter move $\theta^{i-1} \rightarrow \theta^{\text{trial}}$) with a maximum movement step size Δ_θ. We now have a set of trial parameters θ^{trial}. Then, the likelihood $L_{\text{train}}^{\text{trial}}$ of the trial parameters θ^{trial} for the given

training data set is calculated. For brevity, we use the shorthand notation $L_{\text{train}}^{\text{trial}}$ for $p\left(\boldsymbol{\eta}_{\text{train}}^{\text{meas}}|M_k, \boldsymbol{\theta}_k^{\text{trial}}\right)$. The trial parameters are accepted as new parameters, according to the probability $p_{\text{trial}}^{\text{accept}}$, defined as

$$p_{\text{trial}}^{\text{accept}} = \min\{1, \frac{L_{\text{train}}^{\text{trial}}}{L_{\text{train}}^{i-1}}\} \tag{3.24}$$

The acceptance is evaluated by comparing the $p_{\text{trial}}^{\text{accept}}$ to a random number $\mathcal{R}$ uniformly drawn between zero and unity, so that the trial parameters are accepted as new parameters if $\mathcal{R} < p_{\text{trial}}^{\text{accept}}$. Then, $\theta^i = \theta^{\text{trial}}$ and $L_{\text{train}}^i = L_{\text{train}}^{\text{trial}}$. Otherwise, the trial parameters θ^{trial} and $L_{\text{train}}^{\text{trial}}$ are discarded and the old condition is confirmed as the new condition, $\theta^i = \theta^{i-1}$ and $L_{\text{train}}^i = L_{\text{train}}^{i-1}$. Subsequently, the next MCMC step starts again with a random perturbation of one of the viscosity group parameters. MCMC steps are repeated until a final convergence criterion is met (see below).

Because we use uniform priors $p\left(\boldsymbol{\theta}_k|M_k\right)$ that do not affect this acceptance ratio (unless a move exceeds the bounds), Bayes' updating rule as given by eq. 3.19 is satisfied. Note that the normalizing constant in eq. 3.19 (BME), which is still unknown at this point, drops out when calculating the ratio $L_{\text{train}}^{\text{trial}}/L_{\text{train}}^{i-1}$.

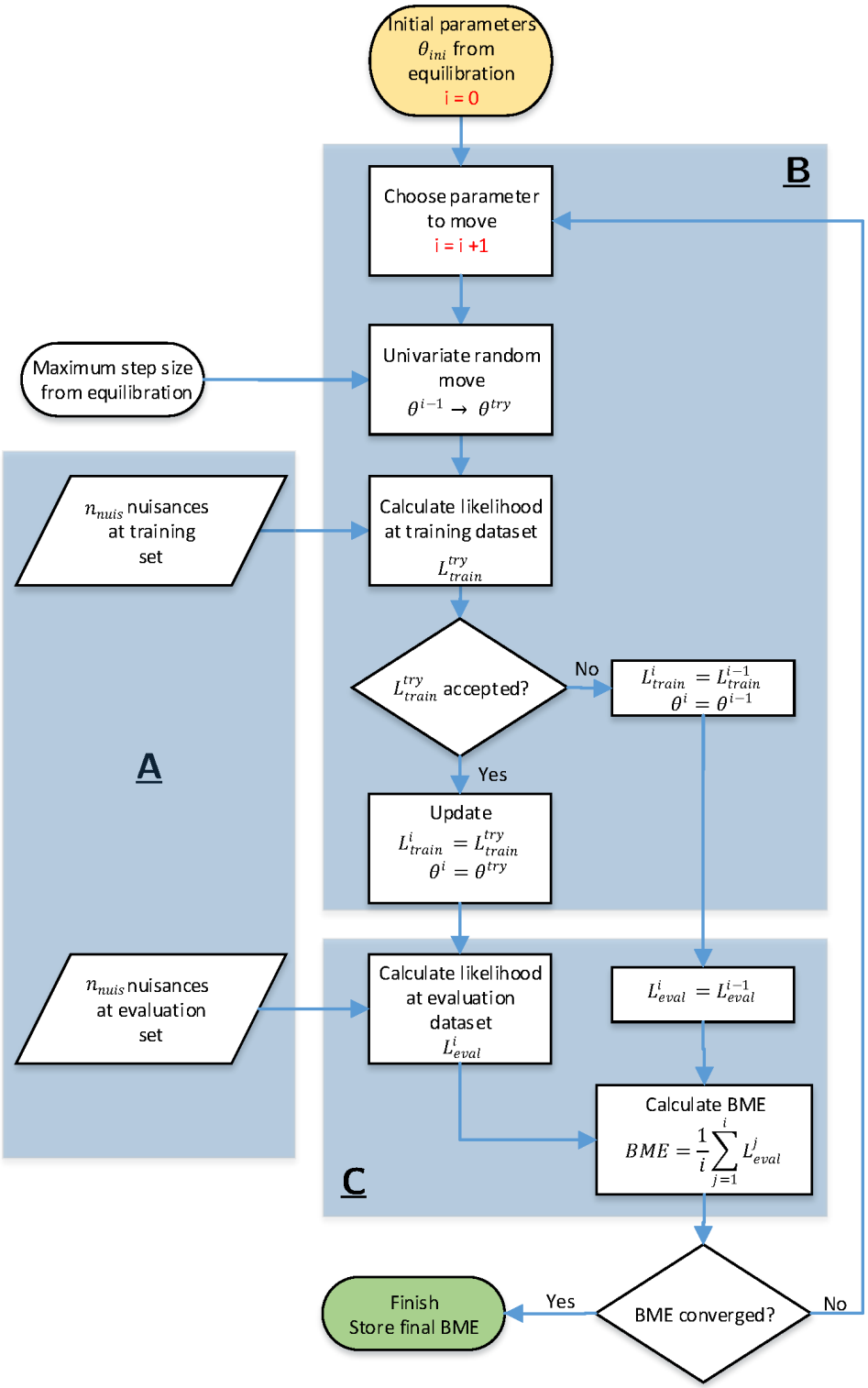


Figure 3.3: Algorithmic flow chart to illustrate the implementation of the Bayesian framework for model parametrization (Part B) based on a training data set and model ranking (Part C) based on an evaluation data set. Experimental errors are treated as nuisances in the case of liquid densities and vapor pressures (Part A) or are addressed by the formulation of the likelihood function in the case of viscosities (Parts B and C).

The maximum step width Δ_θ we use for perturbation was obtained during an equilibration period carried out for each model. In the equilibration period, the maximum step width Δ_θ was adjusted such that the rejection rate ($1 - p_{\text{trial}}^{\text{accept}}$, eq. 3.24) approached a predefined value. Only when the rejection rate maintained stable for 50,000 MCMC cycles without further adjustment of Δ_θ, we assumed the equilibration to be finished. We then started with the production of the samples from the posterior parameter distribution $p\left(\boldsymbol{\theta}_k | \boldsymbol{\eta}_{\text{train}}^{\text{meas}}, M_k\right)$. We confirm that the resulting sample is representative and independent of the rejection rate through running and comparing calculations with different target rejection rates of 40 %, 50 % and 60 %. We demonstrate insensitivity to the rejection rate in the supporting information. Our presented results are based on the 60 % rejection rate.

Accounting for Experimental Errors

Experimental errors enter the MCMC algorithm so far only at one stage, namely the evaluation of the likelihood. We define the likelihood function $p\left(\boldsymbol{\eta}_{\text{train}}^{\text{meas}} | M_k, \boldsymbol{\theta}_k^i\right)$ in eq. 3.19 as the distribution of experimental errors $p\left(\boldsymbol{e}_\eta\right)$ and assume a normal distribution centered on zero with covariance matrix $\boldsymbol{R}$, as

$$\begin{aligned} L_{\text{train}}^{\text{trial}} &= p\left(\boldsymbol{\eta}_{\text{train}}^{\text{meas}} | M_k, \boldsymbol{\theta}_k^{\text{trial}}\right) \\ &= 2\pi^{-N_s/2} \left|\boldsymbol{R}\right|^{-1/2} \\ &\exp\left[-\frac{1}{2}\left(\boldsymbol{\eta}_{\text{train}}^{\text{meas}} - \boldsymbol{\eta}_{\text{train}}^{\text{pred,trial}}\right)^T \boldsymbol{R}^{-1} \left(\boldsymbol{\eta}_{\text{train}}^{\text{meas}} - \boldsymbol{\eta}_{\text{train}}^{\text{pred,trial}}\right)\right] \end{aligned} \tag{3.25}$$

$\boldsymbol{R}$ is a diagonal matrix with length $N_s = N_{s,\text{train}}$ equal to the size of the training data set $\boldsymbol{\eta}_{\text{train}}^{\text{meas}}$. This likelihood definition assumes that experimental errors are uncorrelated, but do not necessarily have constant variance. In our case, we assume that the variance is proportional to the actually measured value, i.e. we assume multiplicative (relative) errors. We investigate the influence of the assumed variance in experimental errors on the outcome of model selection by repeating the entire procedure of model parametrization and ranking for five different relative error levels, as described in the next section. For more detail on the appropriate choice of likelihood (or objective) function we refer to Beck and Woodbury[177].

However, as pointed out in the previous section, experimental errors not

only influence our level of trust in the measured data as encoded in the likelihood function, but they also introduce uncertainty into the PC-SAFT group parameters and the resulting values of residual entropy s_{res} and Chapman-Enskog viscosity η_{CE} (cf. fig. 3.2, Part A). Since we are not interested in obtaining probability distributions of s_{res} or η_{CE}, but only in the final viscosity model parameters A_α, B_α, C_α and D (Part B in fig. 3.2), we adopt a numerical trick to deal with this "internal uncertainty": we treat the uncertain PC-SAFT group parameters as so-called *nuisances*[178] in the MCMC scheme.

To this end, we determine the joint distribution of m_α, σ_α and ε_α by running PC-SAFT based on measured liquid densities ρ^{liq} and vapor pressures p^{sat} that have been randomly perturbed with synthetic experimental errors. This type of perturbation analysis is known as parametric bootstrap[179]. We assume normally distributed relative errors of 2.6 % in p^{sat} and 1% in ρ^{liq}, respectively, according to the DIPPR database[180]. Then, we shift the obtained distribution to match its expected value with the PC-SAFT group parameters published by Sauer et al.[118]. More information about the uncertainties in the PC-SAFT group parameters can be found in the supporting information. Finally, we generated n_{nuis} realizations of m_α, σ_α and ε_α by drawing from this constructed distribution. Each generated nuisance results in a distinct residual entropy s_{res} and Chapman-Enskog viscosity η_{CE}. Then, in the actual MCMC routine, the likelihood of a parameter set $\boldsymbol{\theta}_k^{\text{trial}}$ is evaluated as a function of s_{res} and η_{CE}. All n_{nuis} likelihood values for this specific viscosity model parameter set $\boldsymbol{\theta}_k^{\text{trial}}$ are averaged to obtain one representative likelihood value $L_{\text{train}}^{\text{trial}}$, i.e., we average over the nuisance parameters m_α, σ_α and ε_α. This is a common procedure in MCMC algorithms to handle uncertainties that are not of primary interest to the modeler, yet are of relevance for the final outcome and hence need to be accounted for within the MCMC scheme. The number of pregenerated nuisances n_{nuis} was chosen such that likelihood values were guaranteed to converge when averaged over the ensemble of nuisances. In our case, we generated $n_{\text{nuis}} = 15,000$ realizations.

Step 2: Model Ranking with Monte Carlo Integration

In step 2, we use a second data set for evaluation which will be denoted by $\boldsymbol{\eta}_{\text{eval}}^{\text{meas}}$ (measured data) and $\boldsymbol{\eta}_{\text{eval}}^{\text{pred}}$ (predicted data), respectively. The predictive distribution of the evaluation data, after conditioning on the training data set,

is denoted as $p\left(\boldsymbol{\eta}_{\text{eval}}^{\text{pred}}|\boldsymbol{\eta}_{\text{train}}^{\text{meas}}, M_k\right)$.

The key quantity to obtain model weights is *BME* as given by the integral in eq. 3.21. Brute-force MC integration has been proven to yield consistent results for evaluating this type of equation[174]. Therefore, to implement BMS, we change from the efficient MCMC approach (step 1) to a brute-force MC scheme (step 2, part C in fig. 3.2). Brute-force MC integration determines *BME* as the average likelihood over the whole parameter space of a model M_k, represented by an ensemble of N_{MC} realizations drawn from the parameter distribution, as

$$BME \approx \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} L_{\text{eval}}^{i} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} p\left(\boldsymbol{\eta}_{\text{eval}}^{\text{meas}}|\boldsymbol{\eta}_{\text{train}}^{\text{meas}}, M_k, \boldsymbol{\theta}_k^i\right) \tag{3.26}$$

As opposed to the formulation in eq. 3.21, we do not draw from the prior parameter distribution $p\left(\boldsymbol{\theta}_k|M_k\right)$, but from the posterior parameter distribution $p\left(\boldsymbol{\theta}_k|\boldsymbol{\eta}_{\text{train}}^{\text{meas}}, M_k\right)$ obtained from MCMC (step 1), acknowledging that with respect to the *evaluation* data set, the MCMC parameter sample reflects our current (prior) state of knowledge. The likelihood L_{eval}^{i} of predictions generated by a specific parameter set $\boldsymbol{\theta}_k^i$ is now based on the agreement between predictions and measurements of the evaluation data set (as opposed to using the training data set in step 1, eq. 3.25). Again, the nuisances need to be taken into account (i.e. the uncertainty in m_α, $\sigma_\alpha$and ε_α, part A in fig. 3.2).

We ensure that N_{MC} is chosen sufficiently large by monitoring the convergence of the *BME* estimate. Finally, we compute the model weights according to eq. 3.20.

Two-Step BMS: Intertwining Steps 1 and 2

Conceptually, the two steps undertaken here (step 1: model conditioning, step 2: model ranking) are subsequent steps that are implemented with two distinct numerical schemes (step 1: MCMC, step 2: MC integration). For the sake of efficiency, we have combined them in one global cycle (see fig. 3.3). For each parameter set accepted in step 1, the likelihood of the evaluation data set is determined and the *BME* estimate is updated. This allows us to define the convergence of *BME* as termination criterion for the entire algorithm.

We are not aware of such a combination-type of implementation for BMS in the literature and highly recommend our two-step BMS approach, based on

experiences from this study, for choosing between purely data-driven models. The specific challenges of developing empirical models are well addressed by our proposed approach, because (1) the need to define informative priors is obviated by the use of MCMC for pre-conditioning, (2) experimental errors can be accounted for in several technically sound ways, tailored to the specific interest of the modeler, and (3) BMS is conducted in a numerical framework that has been proven to yield accurate model ranking results.

3.3.3 Investigated Cases

In this work, we apply the BMS framework for two cases. Both cases are problems that, in a similar form, appear in industrial or academic applications. In the first case, we investigate the prediction (extrapolation) of high-pressure data (evaluation data) from low-pressure data (training data) of pure substances (fig. 3.4a). In the second case, we assess the prediction (extrapolation) of viscosities of long n-alkanes (evaluation data) from experimental data of shorter n-alkanes (training data) (fig. 3.4b). In both cases we chose a small number of data points to demonstrate the usefulness of BMS even when only few data is available and to keep the computational effort reasonable. It is the modeler's responsibility to choose a data set that is representative of the task the competing models are designed for. All data is taken from the Dortmund Datenbank[153].

For case 1 (pressure extrapolation), only data of n-pentane is used. We choose n-pentane because extensive data over a wide range of pressure is available. The training data set {calibration data} is chosen within a moderate pressure range of 0.1 MPa to 5 MPa, which corresponds to a range that is reasonably well covered by experimental data for many substances. The evaluation data points lay in the range from 24 MPa to 790 MPa, corresponding to a pressure range that requires a much higher experimental effort and is therefore sparsely available in literature.

In case 2 (carbon chain length extrapolation), the evaluation set {validation data} consists only of substances that are not part of the training set. This case mimics the often encountered practical situation where properties of a substance (in absence of experimental data) shall be estimated based on models of other, chemically similar substances with better experimental coverage. No restrictions regarding the pressure or temperature were made when choosing

the specific data points of case 2.

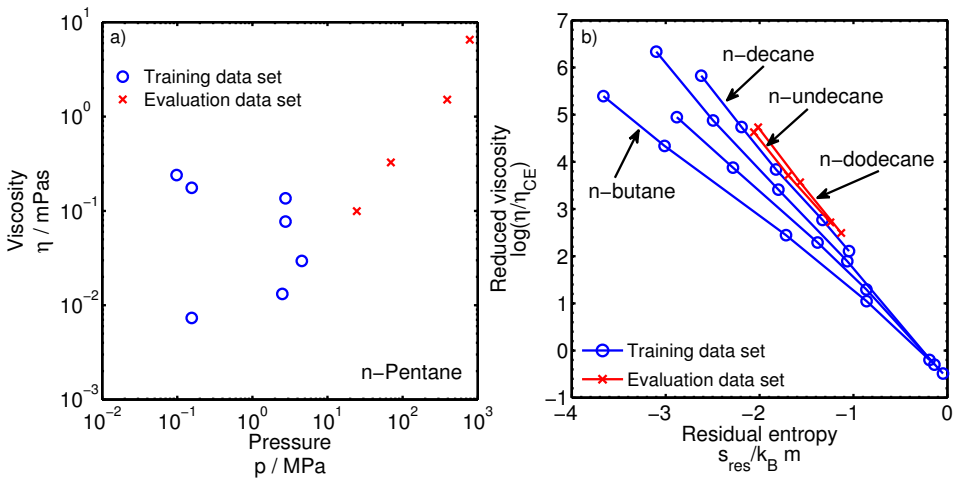


Figure 3.4: Training and evaluation data set of the two investigated cases. a) Case 1 (Pure substance pressure extrapolation), b) case 2 (group contribution-based carbon chain length extrapolation).

To demonstrate how assumptions about the experimental error in the viscosity influence the model ranking, we investigate different assumptions about experimental errors in viscosity data: we use five different values for the assumed standard deviation of experimental errors (see $\boldsymbol{R}$ in eq. 3.25) ranging from 3% to 10% of the measured value. While state of the art measurement techniques at moderate conditions can be seen at the low end of the investigated range, higher uncertainties are to be expected when relying on experimental data from diverse sources including older data and data obtained from various measurement techniques. For the sake of illustration, we here chose rather straight forward assumptions (i.e. normal distribution of errors of various magnitude) about the experimental errors. Though, BMS is able to handle arbitrary types or distributions of experimental errors.

3.4 Results and Discussion

3.4.1 Measures for Model Performance

For assessing the models, we use performance measures. One such indicator, the 'measurement bias' (*MBIAS*), captures the ability of a model to reproduce measurement data. It is defined as the deviation of a model's predictive

mean from the actually measured data value η_j^{meas}, normalized by the standard deviation of experimental error. *MBIAS* is given by

$$MBIAS_j = \frac{1}{\sqrt{\boldsymbol{R}_{jj}}} \left(\tilde{\eta}_j^{\text{pred}} - \eta_j^{\text{meas}} \right) \tag{3.27}$$

with $\tilde{\eta}_j^{\text{pred}}$ denoting the posterior mean of the predictive distribution of a model M_k. By normalization with respect to the experimental error standard deviation, we can interpret and compare results for different data points more easily across different scenarios of error and across the many magnitudes of viscosity.

To indicate how uncertain a model's prediction is (i.e., how large the spread in the posterior ensemble is), we provide Bayesian credible intervals. A 95 % credible interval, e.g., tells the modeler the range of viscosity values that, according to the model, will occur with a probability of 95 %. We determine Bayesian credible intervals of *MBIAS* to judge them in relation to the assumed standard deviation in experimental errors.

To evaluate the trade-off between bias and variance of the various candidate models, we further use statistics that average performance (bias) and ensemble spread (variance) over all data points N_s. We define the weighted mean squared error *WMSE* as

$$WMSE = \frac{1}{N_s} \sum_{j=1}^{N_s} \frac{1}{\boldsymbol{R}_{jj}} \left(\tilde{\eta}_j^{\text{pred}} - \eta_j^{\text{meas}} \right)^2, \tag{3.28}$$

and the weighted mean variance *WMVAR* as

$$WMVAR = \frac{1}{N_s} \sum_{j=1}^{N_s} \frac{1}{\boldsymbol{R}_{jj}} \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \left(\eta_{i,j}^{\text{pred}} - \tilde{\eta}_j^{\text{pred}} \right)^2 \tag{3.29}$$

with N_s for the number of data point in the training or in the evaluation data set, respectively. Further possible performance measures can be found in the supporting information.

3.4.2 Case 1: Pressure Extrapolation

First, we will present the results of model evaluation with respect to case 1 (pressure extrapolation, see fig. 3.4a).

Model Conditioning Based on Training Data set (Step 1)

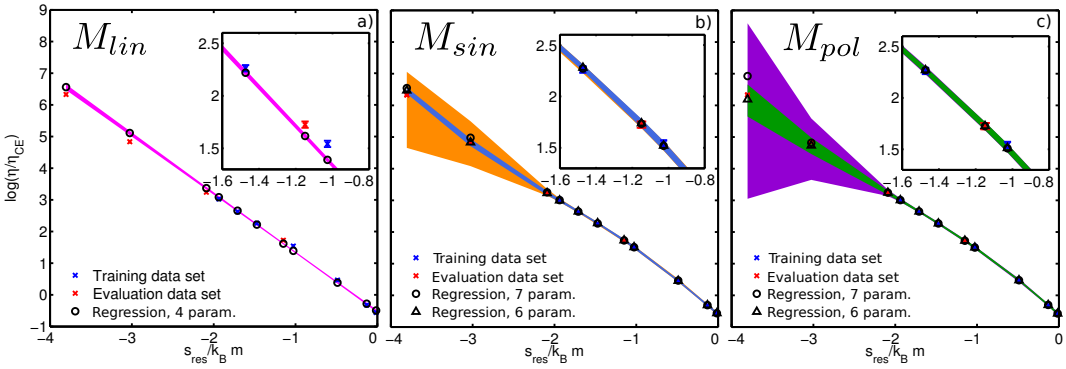


Figure 3.5: Training data set, evaluation data set and model predictions (sample mean at each condition is indicated by black symbols) with Bayesian 95 % credible intervals (colored bands) of all five competing models after conditioning on training data of case 1 (pressure extrapolation) with an assumed relative experimental error of $\boldsymbol{e}_\eta = 3\%$. a) Linear model, b) sinusoidal models (6 parameter version in blue, 7 parameter version in orange) and c) polynomial models (6 parameter version in green, 7 parameter version in purple).

The performance of all five competing models in reproducing the training data ($s_{\text{res}}/(k_B\ m) > -2$) and predicting the evaluation data set ($s_{\text{res}}/(k_B\ m) < -1.2$) after conditioning can be seen in fig. 3.5. In the range of residual entropies covered by the training data (blue crosses), the models show narrow confidence intervals (colored bands). Though, the insets show that only the non-linear models accurately represent the training data. That behaviour is shown in more detail in fig. 3.6a. While the posterior mean prediction of both the polynomial and the sinusoidal models (in both parameter variants using six or seven adjustable parameters) always lies within one standard deviation of experimental errors, the simpler linear model M_{lin} over- or underestimates the measurement values by more than one standard deviation for all data points. Clearly, the linear model with four adjustable group-contribution parameters is not flexible enough to fit the given data sufficiently well. On the other hand, the increased flexibility of the polynomial and sinusoidal models leads to a higher predictive uncertainty, which results in larger credible intervals. These intervals do not seem to be too wide but rather adequate. In summary, in terms of reproducing the measurement data of the training data set, the models with more parameters are clearly to be favored here because they produce accurate and, in comparison to the experimental uncertainty, relatively precise

predictions.

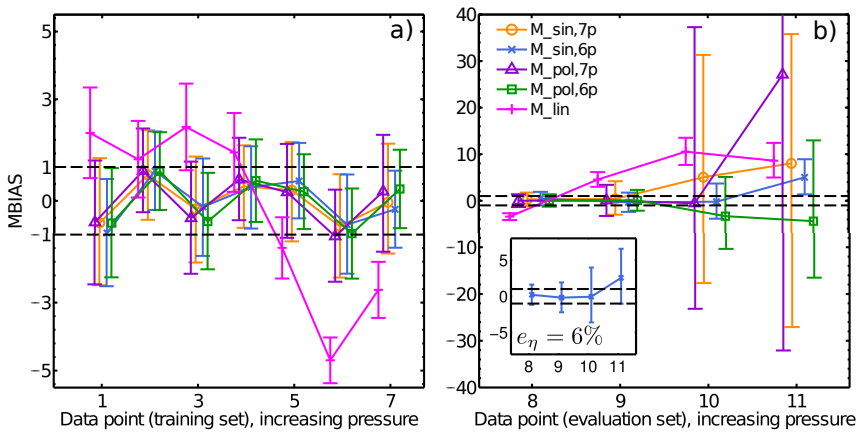


Figure 3.6: Measurement bias with Bayesian 95 % credible intervals of all five competing models after conditioning on training data of case 1 (pressure extrapolation) with an assumed relative experimental error of $\boldsymbol{e}_\eta = 3\%$. a) Predictions of training data set, b) predictions of evaluation data set. The dashed lines mark one experimental error standard deviation. Perfect agreement between predictions and measurements corresponds to zero measurement bias. The inset shows the MBIAS of $M_{sin,6p}$ at a higher relative experimental error of $\boldsymbol{e}_\eta = 6\%$.

Fig. 3.5 ($s_{\text{res}}/(k_B\ m) < -1.2$) and fig. 3.6b show the performance of the competing models in predicting the evaluation data set. As expected, performance in this validation step (extrapolation mode) is worse than in the training period so that this diagram is shown on a different axis scale, as compared to fig. 3.6a.

Again, the simplest model M_{lin} with only four adjustable parameters produces the largest deviations from the actually measured data and the narrowest credible intervals. This combination is undesirable, because we would obtain very confident (highly certain), but wrong predictions from this model. The nonlinear models with six adjustable parameters $M_{\text{pol,6p}}$ and $M_{\text{sin,6p}}$ produce small deviations (measurement bias) in combination with more realistic uncertainty bounds. Overall, the smallest deviations in the predictions are obtained with $M_{\text{sin,6p}}$ along with clearly smaller confidence intervals than $M_{\text{pol,6p}}$. However, data point 11 falls outside the credible interval of $M_{\text{sin,6p}}$. In contrast, the 95 % credible interval of predictions by $M_{\text{pol,6p}}$ contains the actually measured value in all cases. Hence, the polynomial variant should be preferred in terms of reliable predictions.

The two most flexible models with seven adjustable parameters, $M_{\text{pol,7p}}$

and $M_{\mathrm{sin,7p}}$, cannot generally improve on predictive skill as compared to their six-parameter variants. On the other hand, both models produce much wider credible intervals especially at the higher end of the pressure range (see fig. 3.5). The upper limit of the credible interval produced by $M_{\mathrm{pol,7p}}$ at data point 11 is even larger than the axis limit chosen in fig. 3.6b. These wide uncertainty bounds are highly undesired, because predictions by these models provide very little information about the actual data value to be expected.

The effect that highly flexible models perform worse than simpler models in extrapolation is of course expected and is known as the bias-variance dilemma (see comments in the introduction). We investigate this dilemma in more detail by separating the two components bias and variance as proposed e.g. by Bardow[181].

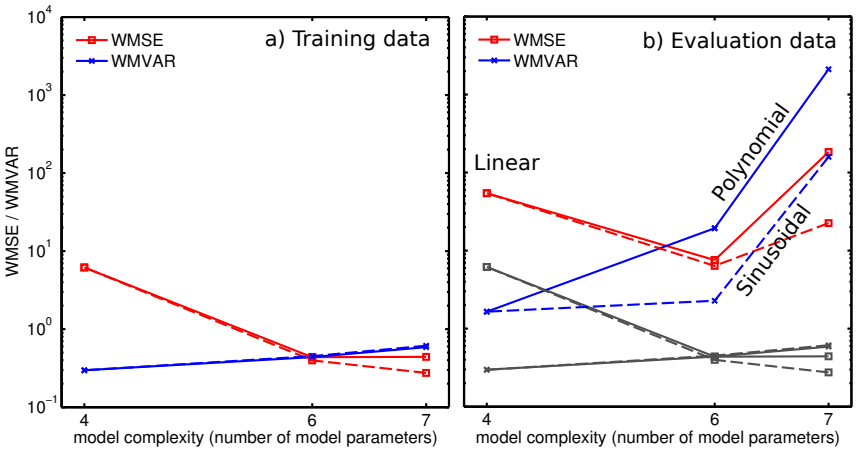


Figure 3.7: Bias-variance behavior of all five competing models after conditioning on training data of case 1 (pressure extrapolation) with an assumed relative experimental error of $\boldsymbol{e}_\eta = 3\%$. Solid lines represent the polynomial models and dashed lines represent the sinusoidal models. Averaged over the a) training data set and b) evaluation data set with grey lines to compare the results in the evaluation data set versus the results in the training data set.

The bias and variance measures are averaged over all data points of the training data set (data points 1 to 7) and shown in fig. 3.7a. The red line corresponds to the weighted mean squared error *WMSE* that remains when reproducing the training data set, and the blue lines show the predictive variance. As intuitively expected, the bias shrinks with increasing model complexity (i.e., increasing number of parameters), while the variance grows.

Specifically, the increase in variance between the four-parameter-model M_{lin} and the six-parameter-models $M_{\mathrm{pol,6p}}$ and $M_{\mathrm{sin,6p}}$ is relatively small (even

though we are looking at a logarithmic scale in fig. 3.7a), while the reduction in bias is significant. Hence, choosing a six-parameter-model seems legitimate and advisable. Note that both, the *WMSE* and the *WMVAR*, act on the same scale (squared model predictions), such that their absolute values can be directly compared and the trade-off can be literally performed without considering any weighting factors due to differences in units. However, choosing between the six-parameter-models and the seven-parameter-models is more ambiguous, because the increase in variance is accompanied with only small reductions in predictive bias.

It seems difficult to make an informed and objective decision when only considering the training data set. However, here, we are in the lucky position to have an evaluation data set at hand which allows us to assess the models' skill in prediction beyond training conditions. The resulting error (*WMSE*) and variance (*WMVAR*) in predicting the evaluation data set is shown in fig. 3.7b. Regarding the bias part (*WMSE*), it turns out that the two models with intermediate complexity ($M_{\mathrm{pol,6p}}$ or $M_{\mathrm{sin,6p}}$) produce by far the lowest error in extrapolation mode. Choosing the more complex (7 parameter) models for their low error in the training period would have been misleading, because these models produce even larger ($M_{\mathrm{pol,7p}}$) or nearly equally large ($M_{\mathrm{sin,7p}}$) errors in extrapolation as compared to the linear model. However, the large overall error of $M_{\mathrm{pol,7p}}$ mainly stems from drastically overestimating the viscosity at data point 11, while the other data points in the evaluation set are nicely met. In this case, the most complex model $M_{\mathrm{pol,7p}}$ has over-fitted the noise in the training data, which bears the risk of such "performance outliers" beyond training conditions. The linear model, on the other hand, under-fits the data due to its lack of flexibility, which triggers a generally inferior performance. Looking at the variance part (*WMVAR*) in fig. 3.7b suggests to choose the sinusoidal model with 6 parameters, $M_{\mathrm{sin,6p}}$, due to its lowest error combined with the low variance in extrapolation.
Overall, our interpretation of fig. 3.6 slightly favors $M_{\mathrm{pol,6p}}$ due to its reliable predictions while fig. 3.7 suggests that $M_{\mathrm{sin,6p}}$ offers the best compromise between bias and variance. At this point, a modeler's choice from these two models (i.e. $M_{\mathrm{sin,6p}}$ or $M_{\mathrm{pol,6p}}$) seems ambiguous and difficult to defend. In the following chapter we look at this dilemma through the eyes of BMS.

Bayesian Model Ranking Based on Evaluation Data set (Step 2)

In contrast to the subjective choice of a model based on fig. 3.6 and fig. 3.7, we now use BMS for an objective choice. After preconditioning the five models by parametrization on the training data set, we evaluate and compare them based on the evaluation data set (see data sets in fig. 3.4a).

We have evaluated *BME* for each of the competing models under all five scenarios of experimental error $\boldsymbol{e}_\eta$, ranging from 3 % to 10 %. Fig. 3.8a shows the resulting values of *BME* as a function of relative experimental error. In the case of the simplest linear model M_{lin}, an increase in experimental error leads to an increase in *BME*. This seems plausible, because the larger the assumed experimental error, the lower the penalty for deviations from the actually measured values. In the case of the other four modeling approaches, however, we see a different trend: *BME* values are almost stable or slightly decreasing with increasing experimental error. This behavior can be explained by the fact that these more flexible models benefit from more accurate data with lower experimental errors, since we have used consistent experimental errors for both conditioning and model evaluation. The conditioning process is more efficient and better predictions are obtained, which in turn lead to higher *BME* values. The linear model, in contrast, cannot take advantage of more accurate training data because of a lack of flexibility. It thus can only benefit from a milder judgment of model validity as incorporated by the likelihood function and the definition of the error covariance matrix $\boldsymbol{R}$ (eq. 3.25).

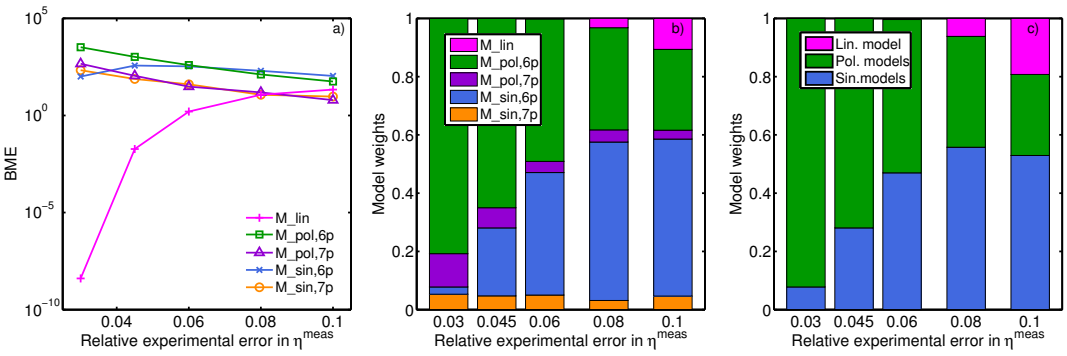


Figure 3.8: Model ranking results based on evaluation data of case 1 (pressure extrapolation) as a function of assumed relative experimental error. a) *BME* values of the five individual models, b) model weights for the five individual models, c) aggregated model weights for the different model types.

Individual *BME* values for the competing models are translated to model

weights according to eq. 3.20. Here, due to a lack of any preference, we choose uniform (i.e., equal) prior model weights. Fig. 3.8b shows the resulting model weights as a function of relative experimental errors. As expected, model choice is most clear in the case of lowest experimental errors, and becomes more and more ambiguous (less decisive) with increasing experimental errors, because the differences in *BME* values decrease (see fig. 3.8b). Even the simplest linear model obtains a non-negligible weight in the case of the highest amount of experimental error assumed here. This trend reflects the principle of parsimony encoded in BMS: if the available data are not informative enough (because the data set is too small or measurement noise is too dominant) to justify a specific level of model complexity, a simpler model will be favored instead. This behavior of BMS has been investigated and confirmed by Schöniger et al. [182].

In our case study setup, the polynomial modeling approach with six adjustable parameters $M_{\mathrm{pol,6p}}$ clearly wins the model competition for assumed errors of 3 % to 4.5 %. For larger errors, the sinusoidal approach with the same level of complexity takes over. Thus, interestingly, model choice not only becomes more ambiguous with increasing experimental error, but even model preference is changed. Since the quality of the data has such a major influence on model ranking, realistic assumptions are crucial to make informed decisions.

Recalling our analysis of fig. 3.6 and fig. 3.7, we showed the difficulties of identifying a clearly best model. BMS gives us a rigorous, objective measure. Due to the fact that $M_{\mathrm{sin,6p}}$ completely misses data point 11 in fig. 3.6b, the model is strongly penalized by BMS. As comparison, the inset in fig. 3.6b shows the MBIAS values of $M_{\mathrm{sin,6p}}$ at the evaluation data set when assuming an experimental error in the viscosity of $\boldsymbol{e}_\eta = 6\%$. The MBIAS values remain nearly unchanged, though, the confidence intervals now enclose the experimental values for all data points of the evaluation data set. Hence, for $M_{\mathrm{sin,6p}}$, the BME increases with increasing assumed experimental errors. Contrarily, for $M_{\mathrm{pol,6p}}$, the increasing assumed experimental error overly widens the confidence intervals which is also penalized by BMS and leads to the decrease of BME with increasing experimental error.

Finally, a modeler might be interested in the question which type of modeling approach (i.e., linear, polynomial, or sinusoidal) performs best, independently from the number of adjustable parameters. To this end, we have now re-weighted the *BME* values with prior model weights that correspond to one third for each type (i.e. 33.3 % for the linear approach, 16.7 % for each of

the two polynomial approaches $M_{\mathrm{pol,6p}}$ and $M_{\mathrm{pol,7p}}$, and analogously 16.7 % for each of the two sinusoidal approaches $M_{\mathrm{sin,6p}}$ and $M_{\mathrm{sin,7p}}$). The resulting aggregated model-type weights are displayed in fig. 3.8c. Similarly to the ranking of the individual models, the polynomial approach clearly wins for errors smaller than 6 %. With a Bayes factor BF of 12, there is strong evidence (according to the rule of thumb provided by Jeffreys [175]) in favor of the polynomial approach assuming the lowest experimental error considered here (3 %). For higher experimental errors, differences in model weights should not be interpreted as significant due to BF values lower than 3. If experimental errors are assumed to be higher than 6 %, model preference changes and the sinusoidal approach ranks first (but with still insignificant differences in model weights). A final, general decision between the two nonlinear modeling approaches is therefore not possible with respect to pressure extrapolation, while the two nonlinear approaches should clearly be preferred over the poorly performing linear model with at least strong evidence in all experimental error scenarios considered here.

3.4.3 Case 2: Carbon Chain Length Extrapolation

Following the same route of discussion in case 2 as for case 1 (pressure extrapolation), we will first present the results of conditioning on the training data set (see fig. 3.4b), with a focus on the bias-variance behavior of the individual models when confronted with carbon chain length extrapolation. Then, we will present model ranking results based on the evaluation data set.

Model Conditioning Based on Training Data set (Step 1)

The performance of all five competing models after conditioning can be seen in fig. 3.9. For this case, we show both the reproduction of the training data and the prediction of the evaluation data in one graph, because the training data cover a much more comprehensive range of experimental conditions (i.e., temperature, pressure and substance). Therefore, the differences in performance within the training period and beyond are not that drastic and we can show the performance measures on the same scale.

Because the training data set is more comprehensive and requires adequate model performance over a wide range of conditions, the overall performance is worse than in the previous case. In the current case, none of the proposed

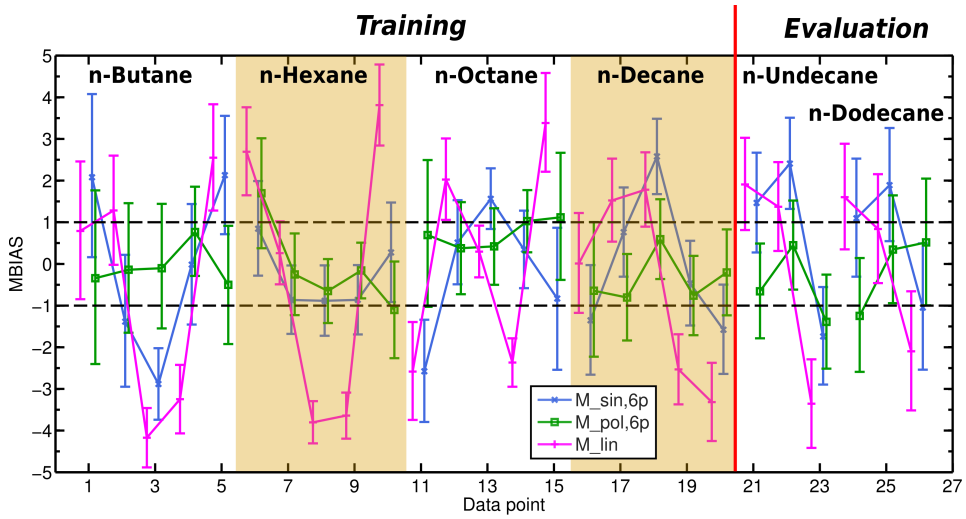


Figure 3.9: Measurement bias with Bayesian 95 % credible intervals of all five competing models after conditioning on training data of case 2 (carbon chain length extrapolation) with an assumed relative experimental error of $\boldsymbol{e}_\eta = 3\%$. Predictions of training data set correspond to data points 1 to 20, predictions of evaluation data set to data points 21 to 26. The dashed lines mark one experimental error standard deviation. Perfect agreement between predictions and measurement data corresponds to zero measurement bias.

modeling approaches yields a posterior mean prediction that always lies within the band of plus/minus one standard deviation of experimental errors. Still, the linear model performs significantly worse than the nonlinear approaches. Also, credible intervals produced by the linear model are again too narrow, while the intervals of the nonlinear models seem more realistic (i.e., they are mostly wide enough to enclose the actually measured value). [Remarkably, when taking into account the credible intervals, all posterior mean predictions of model $M_{\mathrm{pol,6p}}$ lie within the band of plus/minus one standard deviation of experimental errors.]

For better readability, we do not show the measurement bias produced by the nonlinear approaches with seven adjustable parameters ($M_{\mathrm{pol,7p}}$ and $M_{\mathrm{sin,7p}}$), but refer to fig. 3.10 which features the weighted mean squared error *WMSE* (red lines) and weighted mean variance *WMVAR* (blue lines) as performance measures for all five competing models in both the training data set and evaluation data set.

Regarding the results in the training data set, the linear model shows the lowest variance but also the highest bias. For the polynomial model, when moving from 6 to 7 parameters, a slight increase in variance can be traded for

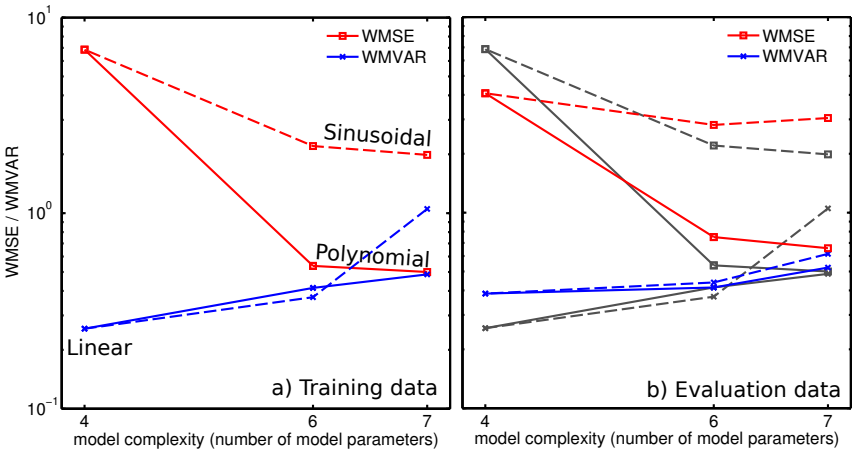


Figure 3.10: Bias-variance behavior of all five competing models after conditioning on training data of case 2 (carbon chain length extrapolation) with an assumed relative experimental error of $\boldsymbol{e}_\eta = 3\%$. Solid lines represent the polynomial models and dashed lines represent the sinusoidal models. Averaged over the a) training data set and b) evaluation data set with grey lines to compare the results in the evaluation data set versus the results in the training data set.

a slight reduction in bias. For the sinusoidal model a similarly small reduction in bias is accompanied by a much larger increase in variance. Hence, it seems advisable to favor the less complex variant of the sinusoidal model, $M_{\text{sin,6p}}$ over its more complex variant. Choosing between the two variants of the polynomial model seems much more difficult. Since both polynomial variants produce a much lower error in the training period than the sinusoidal models while showing similar predictive variances, the polynomial approach seems much more suitable for carbon chain length extrapolation.

To verify this conclusion, we again use the evaluation data set and determine the error and variance when predicting its data points (fig. 3.10b). The six-parameter-variant of the sinusoidal model produces a slightly lower error than its seven-parameter counterpart. The small increase of variance further supports the choice of $M_{\text{sin,6p}}$ over $M_{\text{sin,7p}}$. In the case of the polynomial approach, the more complex variant shows a slightly lower bias (i.e. *WMSE*) but comes along with a slight increase in variance. Hence, neither fig. 3.10a nor fig. 3.10b imply a clear favorite between the two versions of the polynomial model.

These findings again show that we are dealing with a very different situation of data availability: In case 2, the training data set is comprehensive

enough to constrain even the most flexible models considered here, such that adverse effects of too much flexibility on the performance when predicting the evaluation data set do not occur or only to a very small extent. This also means that, with respect to the bias-variance trade-off, we should focus on the bias part. Here, model bias is much more relevant than model variance when looking at the absolute values in fig. 3.10 as compared to fig. 3.7. When confronted with data for various carbon chain lengths, the polynomial model structure obviously is much better able to fit the given data, and the allowed degrees of freedom are used to fit the data even better without yet over-fitting them.

Model Comparison Based on Evaluation Data set (Step 2)

Model ranking results based on the evaluation data set are shown as a function of relative experimental error in fig. 3.11. Since model bias seems to be the decisive factor here, it is not surprising that most *BME* values decrease with increasing levels of measurement noise (fig. 3.11a), because model conditioning becomes less and less effective. Only for the linear model and the sinusoidal model with six adjustable parameters, exceptions from this trend occur. These models cannot benefit from a more informative training data set as much as they are penalized for deviations in the evaluation period.

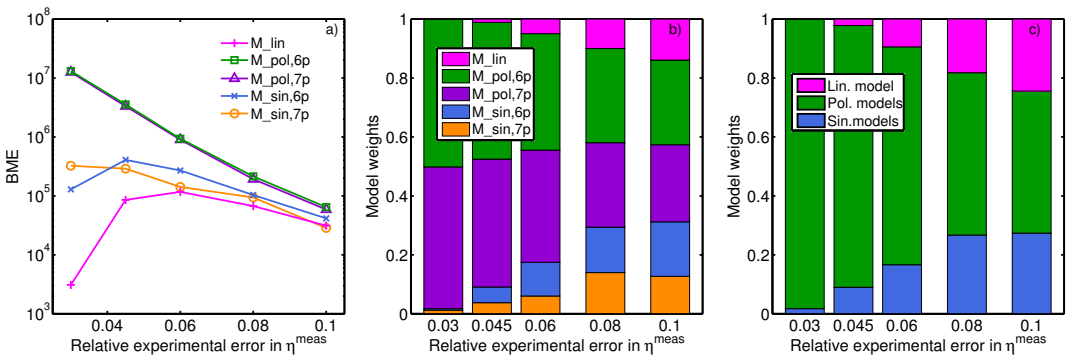


Figure 3.11: Model ranking results based on evaluation data of case 2 (carbon chain length extrapolation) as a function of assumed relative experimental error. a) *BME* values of the five individual models, b) model weights for the five individual models, c) aggregated model weights for the different model types.

Like in case 1 (pressure extrapolation), the decisiveness in model choice suffers when experimental errors increase. For the lowest experimental error

assumed here, BMS ranks the two polynomial models as almost equally likely, while all other models receive negligible posterior model weights (fig. 3.11b). The larger the assumed experimental error, the more uniform the outcome of the model weights.

When looking at aggregated probabilities for model types, the decision in favor of the polynomial approach is unambiguous for the complete range of experimental errors considered here (fig. 3.11c). Assuming experimental errors of 3 % or 4.5 %, Bayes factor values BF of 60 or 8 are achieved, respectively; this corresponds to strong or at least substantial evidence in favor of the polynomial approach when applying the rule of thumb by Jeffreys [175].

Overall, the polynomial approach proves to be a robust and skillful model to predict viscosities under arbitrary conditions. The fact that both variants of the polynomial model, $M_{\mathrm{pol,6p}}$ and $M_{\mathrm{pol,7p}}$, perform equally well indicates two things: on the one hand, the chosen value to fix the additional parameter D in the less flexible version is almost perfect and cannot be outperformed by any randomly drawn parameter set. On the other hand, the data set used for conditioning is comprehensive enough to efficiently constrain the parameter distribution in the case of the more flexible variant such that it is not overly penalized by BMS for its larger complexity.

3.5 Summary and Conclusions

In this work we present an application of Bayesian model selection to thermodynamic models on the example of calculating viscosities via Rosenfeld's entropy scaling approach. We have compared five regression approaches that vary in type (linear, sinusoidal and polynomial) and complexity (four, six and seven adjustable parameters). In our analysis, we have put a special focus on the treatment of experimental errors since they play a central role in data-driven model building.

We have chosen a Bayesian approach to model parametrization and model ranking, because it can elegantly account for experimental errors, including any type of assumption about their statistics. As a result of Bayesian updating, we obtain distributions of parameters and predictions which reflect our current state of knowledge in light of the available data and their attached experimental errors. We obtain predictions with credible intervals instead of deterministic values. This provides a more sound basis for model choice, because uncertainty

intervals might change a purely maximum-performance based preference order. For that reason, Bayesian concepts seem well-suited to address model choice in the field of thermodynamics and has motivated us to transfer the methodology to thermodynamic model selection.

Technically, we have distinguished between model training (Bayesian updating) in a first step and model evaluation (Bayesian model selection) in a second step. After preconditioning the five models by conditioning on a training data set (Bayesian updating), we have evaluated and compared them based on an evaluation data set in a Bayesian model selection framework (BMS). This two-step BMS procedure ensures that (1) the empirical models are compared in a competitive state where they have already been conditioned on a reasonable amount of data, and (2) that BMS can be carried out numerically independent from model conditioning, which allows us to choose the most suitable technique (MC integration) for *BME* evaluation while taking advantage of the efficient MCMC algorithm for model conditioning.

Following this procedure, we have implemented two representative application cases of the entropy scaling approach: a pressure extrapolation (case 1) of pure substances and a carbon chain length extrapolation of n-alkanes using a group contribution approach (case 2). In both cases, we have investigated and discussed the bias-variance behavior of the competing models and strategies for finding a reasonable trade-off in great detail. Beyond the merits of such an analysis, BMS yields objective, statistically rigorous model weights. The underlying quantity *BME* takes into account the complete predictive distribution of a model, not only averaged measures of bias and variance. The resulting model weights can be interpreted with regard to their significance which provides a solid basis for model selection.

Further, we have investigated the influence of experimental errors on the model decision process by implementing five different levels of relative errors. We have found that model choice not only becomes more ambiguous with increasing experimental error, but even model preference can change. In general, any kind of assumption about experimental errors could be incorporated, e.g. data-point specific confidence intervals. The flexibility of BMS in handling arbitrary types of experimental errors is a great advantage when confronted with thermodynamic model selection, since experimental data usually originate from very different sources and techniques and have been obtained under potentially very different conditions.

We have demonstrated the usefulness of applying BMS to thermodynamic model selection. Especially in case 1 (pressure extrapolation), conventional model evaluation did not provide an optimal model choice or even would have been misleading. By applying BMS, the superiority of the six-parameter polynomial model at low assumed experimental errors could be discovered. In case 2, BMS revealed the benefits of the polynomial models for carbon chain length extrapolation. Even though we could not find a clear favorite, the sound knowledge that either of the polynomial models is a valid choice is of great use. Therefore, aiming at a most reliable model for universal application, we find the recently introduced six parameter polynomial model $M_{pol,6p}$ to be the most adequate one. Due to the rigorous nature of BMS, these findings could be easily compared to any other type of competing model or extended to additional sets of training and evaluation data with respective assumptions about experimental error.

4. Pure Substance and Mixture Viscosities based on Entropy-Scaling and an Analytic Equation of State

This chapter is a literal quote of the publication: O. Lötgering-Lin, M. Fischer, M. Hopp, J. Gross; Ind. Eng. Chem. Res., 2018, 57 (11), pp 4095-4114.

Additions implemented in this chapter compared to the original journal-publication are indicated by square brackets.

Specific contributions of the authors of the original publication are as follows: Madlen Hopp conducted an early feasibility study on the application of the Entropy Scaling approach for mixtures under my supervision as part of her Master thesis.
Matthias Fischer carried out the molecular simulations applied in this work and took part in the discussion of the results of the molecular simulations that were used in the development of the mixture model.
I, under the supervision of Joachim Gross, defined the model for the calculation of mixture viscosities and conducted the calculations of mixture viscosities using the PCP-SAFT EoS as well as the comprehensive study of 566 mixtures considered in this chapter.

This study proposes a simple model for viscosities, based on entropy scaling, for real substances and mixtures. The residual entropy is calculated with the perturbed chain polar statistical associating fluid theory (PCP-SAFT). The model requires two or three pure component parameters, noting, however, that an entirely predictive group-contribution approach as proposed in our previous work [Loetgering-Lin O., Gross J., Ind. Eng. Chem. Res. 2015, 54, 7942-7952][, also see chapter 2,] gives also very good results. Overall 140

real substances are considered with relative mean deviations from experimental data of about 5% (without excluding 'outliers'). We performed molecular simulations for mixtures of simple model fluid in order to determine a suitable mixture model. A completely predictive approach for viscosities of real mixtures is thereby obtained. The model is evaluated for 566 mixtures with about 34500 experimental data points of various complexity (i.e. nearly ideal systems as well as highly asymmetric mixtures). Mixtures of non-polar substances and mixtures with at least one polar, but non-hydrogen-bonding component are predicted very accurately with relative mean deviations of in average 6.2% (173 mixtures considered) and 5.3% (126 mixtures considered), respectively. Limitations of the model are found for mixtures with hydrogen-bonding (associating) components such as amines and alcohols, where deviations are systematically higher. Lastly, we present results of mixture viscosities using the purely predictive group-contribution framework and find similar results for the predictive approach.

4.1 Introduction

In economic and conceptual optimization of industrial processes, transport properties are often of critical importance, but at the same time these properties are scarcely available as experimental data, especially for mixtures. Moreover, the available experimental data often covers a rather limited range of pressure, temperature, and composition. Different approaches to predict transport properties are currently used and further developed, e.g., corresponding states principles[14,183,184], Friction Theory[41,43], Free Volume Theory[38–40,147], Thermodynamic Scaling[63,65,185], Isomorph Theory[54,56,57] or, based on the kinetic theory, the Vesovic-Wakeham approaches[19,28–31]. Specifically, entropy scaling and Isomorph Theory (which incorporates both 'entropy scaling' and 'thermodynamic scaling' in a more general framework) provide the potential of covering a very wide range of pressure and temperature in a simple and reliable manner.

The entropy scaling approach goes back to Rosenfeld[1,67] who showed that for simple, spherical monatomic model liquids dimensionless forms of transport properties such as viscosity (η^*), thermal conductivity, or self-diffusion can be described as a function of the residual entropy only (i.e., $\eta^*(s_{res})$, with $s_{res}(T,p)$ instead of $\eta(T,p)$), with one universal set of scaling parameters. This entropy scaling was rediscovered in a more microscopic framework by

Dzugutov[68]. Goel et al.[47] investigated this concept for the viscosity of chains of Lennard-Jones spheres (LJ-chain) of varying chain length. They found that the dimensionless viscosity of each LJ-chain still collapses onto a single curve, solely dependent on the residual entropy. However, these LJ-chains of different chain length can no longer be described with one universal parameter set but need to be parametrized individually. Interestingly, the component-specific parameters correlate well with the chain length. Studies by Chopra et al.[45] compare the entropy scaling approach for various isomers.

Subsequent investigations of model fluids and of real substances confirmed the monovariable, but fluid-specific relation between dimensionless transport properties and the residual entropy. Examples are dumbbell-shaped particles[46], Gaussian-core fluids[69,70], n-alkanes[71,72,84], CO_2[74] and N_2[75], refrigerants[76], and water[48,77–81]. Specifically, the works on Gaussian-core fluids and water reveal limitations of the entropy scaling approach when dealing with fluids that show anomalous behavior in transport properties (i.e., increased particle mobility upon isothermal compression).

For gases, as pointed out by Rosenfeld[67] and others[51,71,82], Rosenfeld's definition of the dimensionless viscosity leads to a collapse onto a single curve $\eta^*(s_{res})$ but shows a steep increase in the low density regime. This steep and diverging increase calls for a complex ansatz function and obscures the calculation of transport properties in the low density limit. To alleviate this problem, different approaches for obtaining a dimensionless form of the transport properties were proposed. Some authors suggested so-called generalized Rosenfeld transport coefficients[82] in which the second Virial coefficient is applied. Others apply the dilute gas Chapman-Enskog formulation[2,149] in two different ways: either to calculate a residual viscosity (i.e., the difference of the actual viscosity and its dilute gas value)[71,85] or by division of the actual viscosity by the Chapman-Enskog value[48,50,51,73].

Studies on the application of the entropy scaling approach to mixtures are scarce. Delage-Santacreu et al.[85] studied mixtures of Mie n-6 fluids within the framework of Isomorph Theory[54,56,57] (which is closely related to entropy scaling) and formulated empirical mixing rules that allow the prediction of mixture viscosities from parameters obtained for pure fluids. Novak[96] obtained promising results for viscosities of natural gas mixtures based on the entropy scaling approach in a predictive segment based framework.

In this study, we propose a simple model for viscosities of mixtures based

on entropy scaling. The model is comprehensively compared to experimental data of real substances. As a preparatory step, we develop pure substance parameters for a wide range of substances. Molecular simulations of model fluids are conducted for developing a mixture model. For mixtures or real substances, we assess the proposed approach as a predictive model (i.e., without adjustable mixture parameters). Strengths and limitations of the model are identified. As a last step, we evaluate the model as an entirely predictive group-contribution model, similar to our earlier work[48] [, see chapter 2], but now extended to mixtures.

4.2 Theoretical Background

4.2.1 Entropy Scaling of Pure Substances

In the original article, Rosenfeld limited consideration to simple, spherically symmetric fluids and disclosed that there exists, to excellent approximation, a monovariable dependence of a dimensionless form of the viscosity, $\eta^+ (s_{res})$, on the molar residual entropy only[1,67]. The molar residual entropy is the molar entropy minus the molar ideal gas entropy, according to $s_{res} (\rho, T) = s (\rho, T) - s^{\mathrm{ig}} (\rho, T)$, with ρ and T as the molar density and temperature, respectively.

Further, Rosenfeld showed that for liquids the logarithmic dimensionless viscosity η^+ can be well described by a linear ansatz

$$\ln \eta^+ = \ln \left(\frac{\eta}{\eta_R} \right) = A^R + B^R \frac{s_{res}}{N_A k_B} \tag{4.1}$$

where η_R is the reference viscosity to obtain a dimensionless form of the viscosity, initially introduced by Rosenfeld as

$$\eta_R = \rho^{\frac{2}{3}} \left(N_A{}^{\frac{1}{3}} M k_B T \right)^{\frac{1}{2}} \tag{4.2}$$

In eq. 4.1, A^R and B^R are the corresponding scaling parameters, N_A denotes the Avogadro constant, k_B is Boltzmann's constant, and M is the molar mass.

Interestingly, for monatomic liquids, Rosenfeld's viscosity scaling parameters A^R and B^R are approximately universal; i.e., A^R and B^R have approximately the same value for various spherically symmetric species. Investigations of nonspherical model fluids and real molecules, however, revealed that

the scaling parameters are component-specific. Correlations for the scaling parameters A^R and B^R based on their chain length[47] or "number of segments"[73] were proposed.

As reported by Rosenfeld and others[51,67,71,82], a pitfall of Rosenfeld's original formulation is a steep increase in the dimensionless viscosity η^+ when approaching small absolute s_{res}-values, i.e., for the low density region. This steep and diverging increase is not captured by the simple linear ansatz given in eq. 4.1 so that the original approach was limited to dense phases.

Several approaches on how to deal with the increase in the dimensionless viscosity at low densities were published. One promising approach[50,51] applies the Chapman-Enskog viscosity $\eta_{\mathrm{CE},i}$[149] of substance i to formulate a dimensionless form of the viscosity $\eta_i^* = \dfrac{\eta_i}{\eta_{\mathrm{CE},i}}$. The applied Chapman-Enskog viscosity $\eta_{\mathrm{CE},i}$ is given as

$$\eta_{\mathrm{CE},i} = \frac{5}{16} \frac{\sqrt{M_i k_B T / (N_A \pi)}}{{\sigma_i}^2 \, \Omega_i^{(2,2)*}} \tag{4.3}$$

where the collision integral $\Omega_i^{(2,2)*}$ is calculated using the empirical correlation by Neufeld et al.[151] We note that this form of the Chapman-Enskog viscosity $\eta_{\mathrm{CE},i}$ (eq. 4.3) differs from the segment-based formulation given in our previous publication[48][, presented in chapter 2,] by a substance-specific factor. This factor, within the argument of the logarithm in eq. 4.1, however, only increments the parameter A^R on the right hand side of the equation. We provide more detail on the rational of this form in the appendix 4.A.

The PCP-SAFT equation of state[105] considers molecules as chains of tangentially bonded segments. The (noninteger) number of segments per chain, m_i is thereby a pure component parameter. We define the dimensionless residual entropy of a pure substance i in the form

$$s_i^* = \frac{s_{res,i}}{N_A \; k_B \; m_i} \tag{4.4}$$

We find it advantageous[48][, as shown in chapter 2] to divide the residual entropy $s_{res,i}$ by the segment number m_i because the dimensionless residual entropy s_i^* then spans similar values for very different substances. Our earlier study showed a monovariable behavior $\eta^*(s^*)$, to very good approximation, for many different substances[48][, see chapter 2]. Using the Chapman-Enskog reference, we observed for n-alkanes a nearly linear relation $\eta^*(s^*)$ for the entire entropy range. The highest deviations were seen when approaching the ideal

gas limit.

For substances with hydrogen-bonds, such as alcohols or water, however, deviations from the nearly linear entropy scaling (eq. 4.1) can be observed. We therefore applied a third order polynomial ansatz function[48,186][, as described and evaluated in chapter 2 and chapter 3], as

$$\ln \eta_i^* = A_i + B_i s_i^* + C_i s_i^{*2} + D_i s_i^{*3} \tag{4.5}$$

where A_i to D_i are adjustable pure substance viscosity parameters.

4.2.2 Entropy Scaling of Mixtures

The model for predicting the viscosity of pure fluids (based on entropy scaling) proposed by some of the authors[48][, see chapter 2,] is here extended to mixtures. For guiding the model development, we performed a comprehensive study on simple model fluids using molecular simulations. Based on this study, for mixtures of K substances, we propose the following form

$$\ln \eta_{\text{mix}}^* = \sum_{i=1}^{K} x_i A_i + \sum_{i=1}^{K} \frac{x_i m_i}{\overline{m}} B_i s^* + \sum_{i=1}^{K} \frac{x_i m_i}{\overline{m}} C_i s^{*2} + \sum_{i=1}^{K} \frac{x_i m_i}{\overline{m}} D_i s^{*3} \tag{4.6}$$

with

$$s^* = \frac{s_{res}\left(\rho, T, \boldsymbol{x}\right)}{N_A \; k_B \; \overline{m}} \tag{4.7}$$

Here, x_i is the mole fraction of component i and $\overline{m} = \sum_{i=1}^{K} x_i m_i$ is the averaged segment number of the mixture. The bold $\boldsymbol{x}$ represents the vector of mole fractions of all components in the mixture. The ratio $x_i m_i / \overline{m}$ in eq. 4.6 is the fraction of segment number of species i among all segment numbers in the mixture. The residual entropy of the mixture is defined as

$$\begin{aligned} s_{res}\left(\rho, T, \boldsymbol{x}\right) &= s\left(\rho, T, \boldsymbol{x}\right) - s^{\text{ig}}\left(\rho, T, \boldsymbol{x}\right) \\ &= s\left(\rho, T, \boldsymbol{x}\right) - \left(\sum_{i=1}^{N} x_i s_i^{\text{ig}}\left(\rho, T\right) - N_A k_B \sum_{i=1}^{N} x_i \ln x_i \right) \end{aligned} \tag{4.8}$$

The parameters A_i to D_i in eq. 4.6 are the pure substance viscosity parameters introduced in eq. 4.5. Hence, our formulation of the mixture viscosity does not introduce any new adjustable parameters, in accordance to our objective

of defining a purely predictive model for mixture viscosities.

As Chapman-Enskog viscosity for mixtures, $\eta_{\mathrm{CE,mix}}$, we use the approximation of Wilke[15,187], given by

$$\eta_{\mathrm{CE,mix}} = \sum_{i=1}^{N} \frac{x_i \eta_{\mathrm{CE},i}}{\sum\limits_{j=1}^{N} x_j \phi_{ij}} \tag{4.9}$$

where

$$\phi_{ij} = \frac{\left(1 + \left(\eta_{\mathrm{CE},i}/\eta_{\mathrm{CE},j}\right)^{1/2} \left(M_j/M_i\right)^{1/4}\right)^2}{\left(8\left(1 + M_i/M_j\right)\right)^{1/2}}. \tag{4.10}$$

4.2.3 Calculation of s_{res} Using the Perturbed-Chain Polar Statistical Associating Fluid Theory

Based on the statistical associating fluid theory (SAFT) developed by Chapman et al.[188] on the basis of Wertheim's perturbation theory for short-ranged highly directional interactions[97–100], a variety of SAFT-type equations of state (EoS) evolved. One of them is the perturbed chain SAFT EoS (PC-SAFT) by Gross and Sadowski[105]. Further development led to additional terms to also describe polar substances (PCP-SAFT)[104,110,111]. The residual specific Helmholtz energy a^{res} is then expressed as

$$a^{\mathrm{res}}(\rho, T, \boldsymbol{x}) = a^{\mathrm{hc}} + a^{\mathrm{disp}} + a^{\mathrm{assoc}} + a^{\mathrm{polar}} \tag{4.11}$$

where a^{hc} describes the hard-chain contribution. The further terms a^{disp}, a^{assoc} and a^{polar} express the dispersive, associating and polar contributions to the intermolecular potential, respectively. Within the SAFT framework, molecules are described by at least three pure component parameters: the segment number m_i, the segment size parameter σ_{ii} and the dispersion energy parameter ε_{ii}. For polar substances the dipole moment μ_i and, in case of quadrupolar substances, the quadrupole moment Q_i are additionally used. Hydrogen bonding interactions are described with the association term, which introduces the association energy parameter ε^{AB} and the association volume κ^{AB} as pure component parameters. All parameters used in this study were taken from literature[13,104,105,110,111,189]. The PCP-SAFT parameters and references used in this work are listed in the Supporting Information.

From the residual Helmholtz energy given in eq. 4.11, any other residual

static thermodynamic property can be derived. The residual entropy s_{res} is calculated according to

$$s_{res}\left(\rho, T, \boldsymbol{x}\right) = -\left(\frac{\partial a_{res}}{\partial T}\right)_{\rho,\boldsymbol{x}} \tag{4.12}$$

The PCP-SAFT model can be applied with a binary interaction parameter. Throughout this study, we target at predicting mixture viscosities and we set the binary interaction parameter to zero, $k_{ij} = 0$, for all pairs of substances, unless explicitly stated otherwise. For more information regarding the PCP-SAFT EoS and the theoretical framework, we refer to the original publications[104,105,110–112].

4.2.4 Procedure for Adjusting Pure Component Parameters

The pure component viscosity parameters A_i to D_i are adjusted to experimental pure substance viscosities. Experimental viscosities were taken from the Dortmund Datenbank[153]. Due to the declining accuracy of the PCP-SAFT equation of state in the vicinity of the critical point, data points within the range of conditions

$$(0.5\ p_{crit} < p < 1.5\ p_{crit}) \wedge (0.8\ T_{crit} < T < 1.2\ T_{crit}) \tag{4.13}$$

were omitted during parameter regression. For all calculations and for the reported values of deviations between model and experimental data reported in this work, however, all data points (also those close to the critical point) are included.

For many substances, there exist few or no experimental data of gas viscosities in the literature. The A_i parameter then degenerates (i.e., it is highly correlated to the remaining parameters B_i to D_i), which usually leads to unphysical predictions for extrapolations beyond training conditions. Therefore, we use A_i parameters obtained from the group contribution approach of our earlier study[48][, see chapter 2,] whenever possible. We note that the value of the A_i parameter used in this study is obtained as $A_i = A_i^{gc} + \ln(\sqrt{1/m_i})$ from the A_i^{gc} parameter of the previously published group contribution approach[48] due to the different representation of the Chapman-Enskog viscosity $\eta_{\mathrm{CE},i}$, as

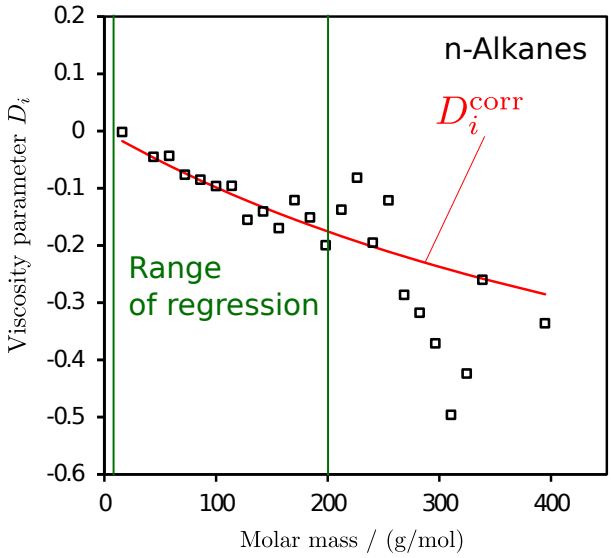


Figure 4.1: Correlation of the viscosity parameter D_i with the molar mass for n-alkanes. Squares are the individually adjusted values of the D parameter with the red line as the proposed regression used to obtain D_i^{corr} as defined in eq. 4.14. The red line results from regression to the range indicated by the green lines.

discussed in the appendix.

For substances, where the experimental data only covers a narrow range of residual entropies, the C_i and D_i parameters, describing the non-linearity in the Entropy-Scaling, are also highly correlated and not uniquely identifiable. [That means, different combinations of values for C_i and D_i parameters lead to approximately the same accuracy in describing experimental viscosities.] Therefore, we applied a regression approach for the D_i parameter. Fig. 4.1 shows the individually adjusted D_i parameters versus the molar mass. We found, that the adjusted D_i parameter scales with the molar mass, which confirms previous findings[48][, shown in chapter 2]. We approximate the D_i parameter with a linear regression approach, given by

$$D_i^{\mathrm{corr}} = (-1.25594 - 888.1232/(M_i/(g/mol)))^{-1} \tag{4.14}$$

where the coefficients were regressed to the individually adjusted D_i values within the range indicated by the green boundaries in fig. 4.1. We chose a linear approach in reciprocal variables because of the non-diverging behaviour of eq. 4.14 for molecules of high molecular mass, like polymers.

Thus, if a substance under investigation can be assembled from published

Table 4.1: Parameters of the LJ Potential of the Investigated Model Mixtures

Mixture	σ_{22}/σ_{11}	$\varepsilon_{22}/\varepsilon_{11}$	ε_{12}
I	1	0.7	$\sqrt{\varepsilon_{11}\varepsilon_{22}}$
II	0.8	1	$\sqrt{\varepsilon_{11}\varepsilon_{22}}$
III	1	1	$0.75\sqrt{\varepsilon_{11}\varepsilon_{22}}$

functional groups[48][, developed in chapter 2], two parameters are adjusted (B_i, C_i), otherwise three (A_i, B_i, C_i). These viscosity parameters of each substance were optimized by minimizing the squared relative deviations of calculated to experimental viscosities for all data points of the considered substance using a Levenberg-Marquardt algorithm[154]. Viscosity parameters of all investigated substances are listed in the Supporting Information.

4.2.5 Molecular Simulation Details

In order to systematically assess the entropy scaling of mixtures, we investigated simple model fluids and mixtures using molecular simulations. The particles interact through the Lennard-Jones (LJ) pair potential

$$\phi\left(r_{ij}\right) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{4.15}$$

depending on the distance r_{ij} between two particles i and j, the energy parameter ε_{ij}, and the size parameter σ_{ij}. Cross-wise parameters are calculated according to the Lorentz and Berthelot combining rules, as $\sigma_{ij} = 0.5(\sigma_{ii} + \sigma_{jj})$ and $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$, respectively.

We simulated three binary model mixtures. Mixture (I) is a size-asymmetric mixture, where $\sigma_{22} = 0.7 \cdot \sigma_{11}$, while the ε_{ii} parameters are the same for both components (i.e., $\varepsilon_{11} = \varepsilon_{22}$). The components of mixture (II) vary in their energy parameter ($\varepsilon_{11} \neq \varepsilon_{22}$), while $\sigma_{22} = \sigma_{11}$. Mixture (III) is a mixture of two identical Lennard-Jones fluids, where the cross-energy parameter is lower than the energy of the two constituents (and the cross-energy parameter thus deviates from the Berthelot combining rule), with $\varepsilon_{12} = 0.75\varepsilon_{11} = 0.75\varepsilon_{22}$. An overview of the parameters of the mixtures is given in tab. 4.1.

Molecular Dynamic (MD) simulations were used to calculate the viscosity along isochors or isobars, respectively. We initialize the problem using a preceding Monte Carlo (MC) simulation for the equilibration of a system of 600

LJ particles in a cubic box. The equilibration was carried out in two different ways.

In order to calculate the viscosity along values of constant volume, the MC run was carried out in a NVT ensemble (i.e., constant volume V, temperature T, and number of particles of each species of the mixture N_i). The energy is sampled, and a configuration k with potential energy $E_k \approx \langle E \rangle$ serves as a starting point for the subsequent MD simulation for determining the viscosity. Velocities according to a Maxwelll distribution were allocated to the Lennard-Jones particles. The entire procedure has the advantage that a weakly coupled (Berendsen) thermostat is sufficient for preserving the temperature during the MD simulation.

For determining viscosities along isobars, we perform preceding MC simulations in the NpT ensemble, where the pressure p is kept constant and where the volume and the energy are sampled. A configuration k is selected as a starting point for the MD simulation, where the volume is $V_k \approx \langle V \rangle$ and the energy is $E_k \approx \langle E \rangle$. With this procedure, we avoid barostats and perform MD simulations in the NVT ensemble with weakly coupled thermostats for simulation results along defined pressure p.

The viscosity is determined from the Green-Kubo formalism[150,190–193] which uses the autocorrelation function of the shear-stress-tensor which is sampled in every time step during the MD simulation. The system is integrated for the reduced time $t^* = t\sqrt{\frac{\varepsilon_{11}}{M_1 \sigma_{11}^2}} = 6956$ with a reduced time step of $\Delta t^* = 6.956 \cdot 10^{-4}$ using a velocity Verlet algorithm. Dimensionless quantities are defined with parameters of the first component for all mixtures (σ_{11}, ε_{11}, M_1). The molecular mass M_i of all species is equal and defined as unity. We verified good agreement of our results with literature values for viscosities of LJ fluids at various state points[194,195]

The usual dimensionless variables for the LJ-fluid are here applied. We calculated the viscosities along three isochors at $\rho^* = \rho\sigma_{11}^3 = 0.7, 0.8, 0.9$ at four temperatures $T^* = T/(\varepsilon_{11}/k_B) = 0.8, 1.0, 1.2, 1.5$. Beside the pure components, we considered three different compositions ($x_1 = 0.25, 0.5, 0.75$). Additionally, we carried out simulations at a pressure $p^* = p\sigma_{11}^3/\varepsilon_{11} = 0.15$ in a temperature regime from $T^* = 0.8$ to $T^* = 1.2$ for various compositions. The resulting dimensionless viscosities ($\eta^* = \eta\frac{\sigma_{11}^2}{\sqrt{\varepsilon_{11} M_1}}$) are also reduced with the parameters of the first component of the considered mixture.

The PCP-SAFT equation of state is not strictly a model for LJ fluids.

The equation of state, however, gives a reasonable description of LJ fluids and mixtures[196]. We also obtained the residual entropy from MC simulation results of this study using a histogram reweighting technique, and we used a highly accurate equation of state for LJ fluids developed by Thol et al.[197] for calculating the residual entropy, as shown in the Supporting Information. The conclusions drawn from the molecular simulation study do not depend on the choice between the PC-SAFT model or a more accurate LJ-equation of state.

4.3 Results: Pure substances

We consider pure substances first and mixtures thereafter. The development of a viscosity-model for mixtures was guided by results from molecular simulations. The study of model fluids through molecular simulation is described first. For mixtures of real substances, we present a comprehensive study of 566 binary mixtures of real substances with about 34500 experimental data points from the literature. For structuring the presentation of results, we further distinguish three classes of real substances: (1) nonpolar, (2) polar but nonassociating, and (3) self-associating (hydrogen-bonding) components.

As a measure for discussing deviations of the proposed model to experimental data, we use the Average Absolute Deviation

$$AAD\% = (100/N_{data}) \sum_{i}^{N_{data}} ((|\eta_{i,calc} - \eta_{i,exp}|)/\eta_{i,exp}) \tag{4.16}$$

where N_{data} is the number of data points, as well as the relative deviation $\delta_i\%$, defined as

$$\delta_i\% = 100 \cdot (\eta_{i,calc} - \eta_{i,exp})/\eta_{i,exp}. \tag{4.17}$$

All parameters of the viscosity model are listed in the appendix. Deviations of the model to experimental data, $AAD\%$, for all substances are given in the Supporting Information.

4.3.1 Lennard-Jones model fluid

In a first step, we evaluate our approach with simulated viscosities of pure Lennard-Jones (LJ) fluids. The dimensionless simulated viscosities η^* versus the residual entropy s^*, which was obtained with the PCP-SAFT EoS, can be

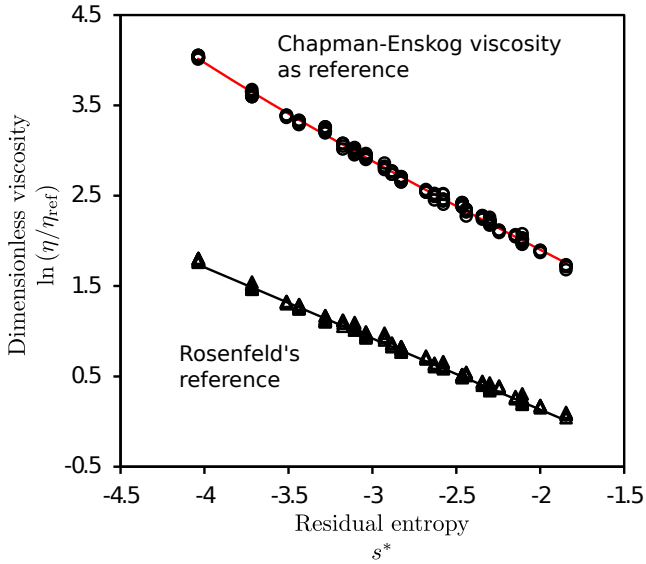


Figure 4.2: Dimensionless simulated viscosities of a Lennard-Jones fluid. Triangles represent the dimensionless viscosity using Rosenfeld's reference viscosity (eq. 4.2). The black line is the result of the linear correlation (eq. 4.1), see the Supporting Information. Black circles are the dimensionless simulated viscosities using the Chapman-Enskog viscosity η_{CE} (eq. 4.3) as the reference viscosity. The red line is the result of our correlation.

seen in Fig 4.2. Using either Rosenfeld's reference (eq. 4.2) or the Chapman-Enskog viscosity (eq. 4.3), all simulated data points collapse onto a single curve. A similar diagram, where results of a highly accurate equation of state for the Lennard-Jones fluid is used for calculating the residual entropy is shown in the Supporting Information. It is thereby concluded that the correlations mildly change when an equation of state for LJ-fluids is used for s^* instead of the PCP-SAFT model, but the general entropy scaling behavior and all conclusions of our analysis are unaltered.

4.3.2 Nonpolar substances

As members to the class of nonpolar substances, we investigate n-alkanes, branched alkanes, cycloalkanes, alkenes, and aromatics. Fig. 4.3 (a) presents deviations (as $AAD\%$) of predicted viscosities from experimental data of n-alkanes ranging from methane to n-hexatriacontane (C36). We show deviations in two different ways. The crosses connected with a solid line are the $AAD\%$ values. As a measure for the distribution of deviations, we additionally present *box-and-whisker*-marks of the $AAD\%$. In that representation, 50% of all data

points fall within the box; hence, the bottom and top edge of the box are the first and third quartiles (i.e., 25 % and 75 % of the data points), respectively. The range from the top of the box to the upper end of the whisker represents the $AAD\%$-range where the 25% of the data points with the highest $AAD\%$-values lie. The red band inside the box gives the median (i.e., 50 % of data points have an $AAD\%$ value lower than this value and vice versa). The dots are outliers, where we use the common definition of a maximal whisker length of 1.5 times the box length. That means that outliers have a distance from the mean of at least 2.7 standard deviations. Outliers are not taken into account in the calculation of the percentiles. We would like to note that we did not discriminate the experimental data taken from the Dortmund Datenbank but used all the data available. We found that, as expected, data points in the vicinity of the critical point show higher $AAD\%$ and, hence, might occur as outliers. The second common characteristic of nearly all remaining outliers is their origin from rather old publications (i.e., 1990 and older).

Fig. 4.3 (a) shows that an entropy scaling approach can lead to remarkably low errors. Considering all investigated n-alkanes (from C1 to C36) the averaged $AAD\%$ is 3.91%. For all but two n-alkanes, the complete box is below 10 % (and for most n-alkanes even below 5%), which means, taking into account all conditions with experimental viscosities available, 75 % of the model's predictions deviate less than 10 % (or 5%, respectively) from the experimental data. We emphasize once again that all data have gone into the calculation of the $AAD\%$ values, also those data points that are labeled as outliers according to the above statistical analysis.

Figures 4.3 (b) to 4.3 e) present results of other members of the group of nonpolar substances, i.e., alkenes, branched and cyclic alkanes, and rather nonpolar aromatics. Overall, the observations made for n-alkanes also apply for these substances: we find surprisingly low deviations of calculated viscosities from experimental data, considering the fact that data from wide ranges of conditions entered our study. Three exceptions with significantly higher $AAD\%$, namely AAD-values above 10%, are neopentane, squalane and propylbenzene.

In order to visualize data points that are perceived as outliers, we give a detailed analysis for the dimensionless viscosity η^* of n-butane in fig. 4.4. The inset gives (non-logarithmic) relative deviations $\delta_i\%$ of all data points i for varying s^*. Blue circles are the data points that are marked as outliers in the *Box-and-whisker*-representation of fig. 4.3 (a). In the inset, the red line

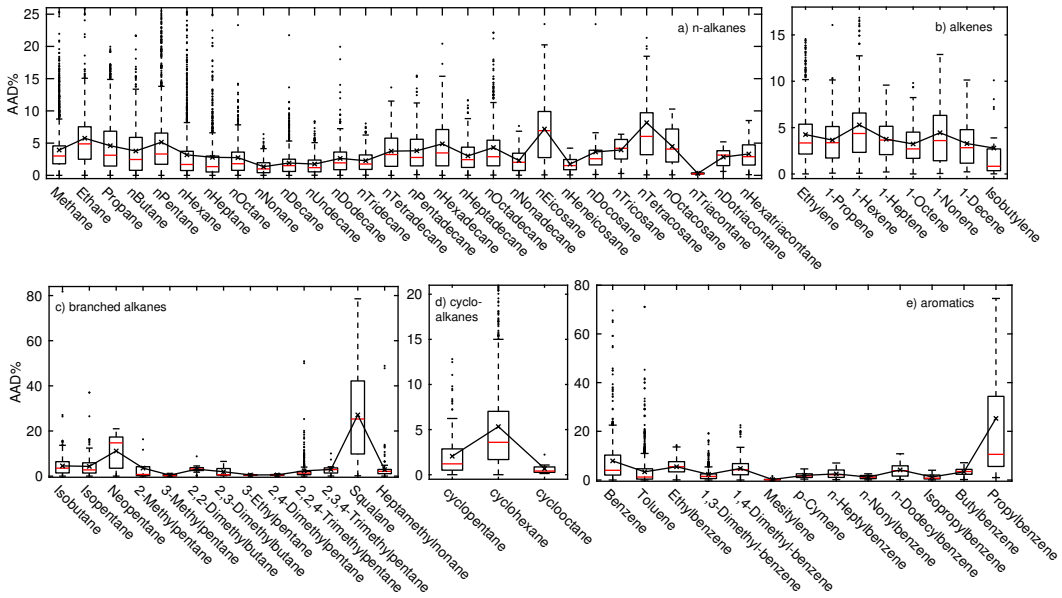


Figure 4.3: Deviations of proposed model from experimental data for pure non-polar, nonassociating substances. Crosses, connected with a solid line, are the $AAD\%$ of each substance. *Box-and-whisker*-representation of the distribution of deviations. Dots are outliers.

marks the residual entropy at critical conditions, $s^{*,crit} \approx -0.45$. Overall, the experimental viscosities of n-butane all collapse onto a single curve. In the vicinity of the critical point deviations increase. Also, when approaching the ideal gas limit, accuracy of the correlation declines. Whereby, for dilute gases, also experimental uncertainties are known to often increase.

4.3.3 Polar, Nonassociating Substances

As members in the class of polar, nonassociating species, we study aldehydes, ketones, esters, and ethers, as well as hydrofluoro- and chlorofluorocarbons (referred to as hfc+cfc). The results are summarized in fig. 4.5. The averaged $AAD\%$ values of each chemical family are 8.82% for aldehydes, 5.66% for ketones, 3.26% for esters, 10.12% for ethers, and 5.13% for hydrofluoro- and chlorofluorocarbons. The boxes for nearly all substances are completely below 10 % which means that the deviations to 75% of the available data points of each substance are below 10 %. Exceptions with significantly higher $AAD\%$, namely, AAD-% values above 10%, are ethanal, di-n-pentyl-ether, diethyl-ether, and R23.

In Figures 4.6 and 4.7, we show the detailed results of one rather good and

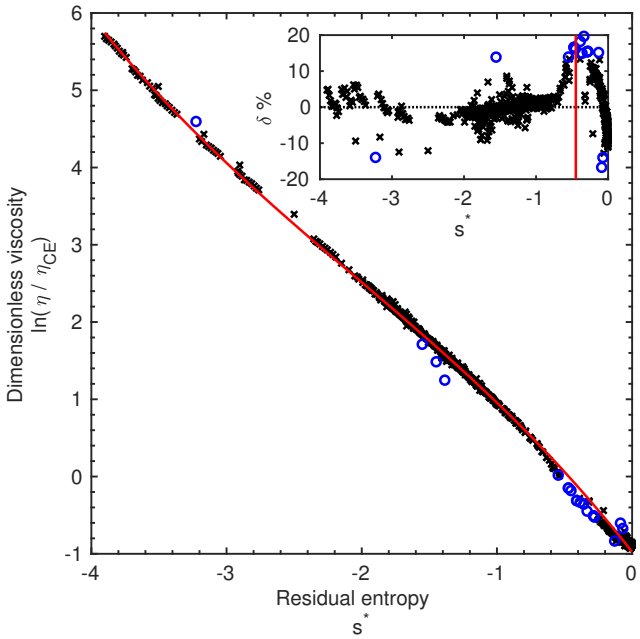


Figure 4.4: Dimensionless viscosity of n-butane versus the residual entropy s^*. Crosses are experimental data. The red line is the result of the correlation with parameters as given in the Supporting Information. The inset shows the relative deviations $\delta\%$ of the calculated viscosities. The vertical red line in the inset is the calculated residual entropy at the critical point $s^{*,crit}$. Blue circles are the data points that are shown as outliers in the *box-and-whisker*-representation of fig. 4.3 (a). Experimental data taken from DDBST[153].

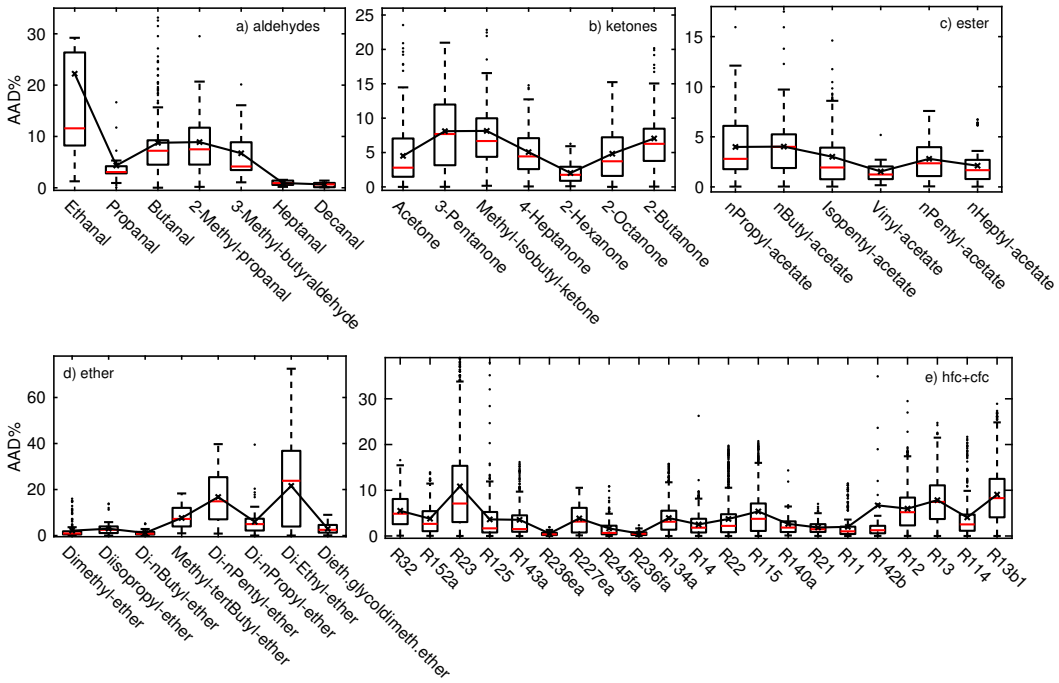


Figure 4.5: Deviations of proposed model from experimental data for pure polar, nonassociating substances. Crosses, connected with a solid line, are the *AAD%* of each substance. *Box-and-whisker*-representation of the distribution of deviations. Dots are outliers.

one poorly described substance, pentafluoroethane (R125) and diethyl ether, respectively. For R125 (fig. 4.6), the data points collapse onto a single curve that can be described with the given parameters rather well. Only in the gaseous phase, the data points are scattering. For diethyl ether (fig. 4.7), the data points still collapse reasonably well onto a single curve. However, the data points close the critical residual entropy $s^{*,crit} \approx -0.45$ are not on a low-curvature line with the gas-phase data points at $s^{*,crit} \to 0$. The full set of data around the critical residual entropy $s^{*,crit} \approx -0.45$ were reported in one publication in the year 1935. Although we as authors from a nonexperimental study have to be very reserved with such speculations, we see the possibility that this data set is not very accurate. Nonetheless, the errors we report, of course include all data.

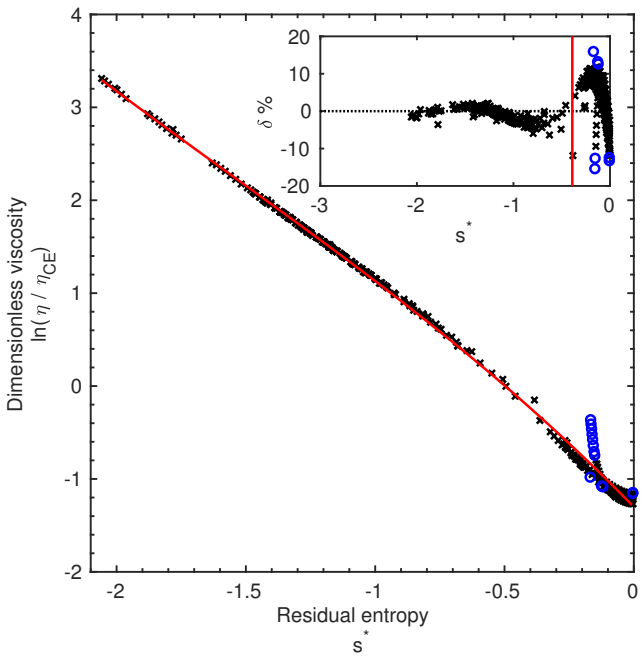


Figure 4.6: Dimensionless viscosity of pentafluoroethane (R125) as an example of a substance with average deviations of the proposed model (red line, with parameters given in the Supporting Information) from experimental data (crosses). The inset shows the relative deviations $\delta\%$ of calculated viscosities. The vertical red line in the inset is the calculated residual entropy at the critical point $s^{*,crit}$. Blue circles are outliers. Experimental data taken from DDBST[153].

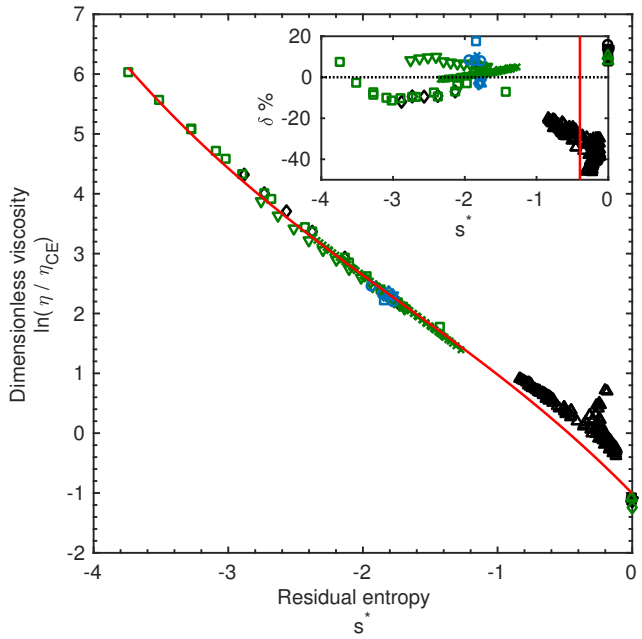


Figure 4.7: Dimensionless viscosity of diethyl-ether as an example of a substance with large deviations of the proposed model (red line, with parameters given in the Supporting Information) from experimental data (crosses). The inset shows the relative deviation $\delta\%$ of calculated viscosities. The vertical red line in the inset is the calculated residual entropy at the critical point $s^{*,crit}$. Experimental data taken from DDBST[153]. Different colors and symbols indicate data sets from different publications.

4.3.4 Associating (Hydrogen-Bonding) Substances

As polar and self-associating substances we investigated amines and alcohols. Deviations of the viscosity-model to experimental data are presented in fig. 4.8. The averaged $AAD\%$ values are 4.69% for amines and 6.38% for alcohols. The complete boxes as well as the $AAD\%$ values for most substances are below 10%. Exceptions with significantly higher $AAD\%$, namely AAD-values above 10%, are dimethylamine, methanol, 1-decanol, 1-undecanol, 1-dodecanol, and $1,2$-ethandiol.

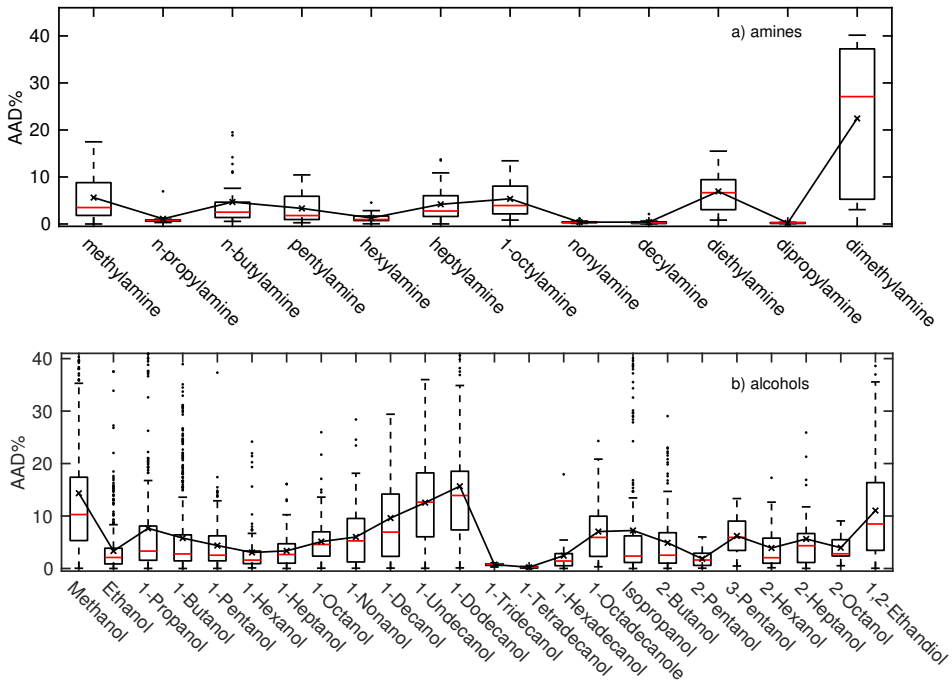


Figure 4.8: Deviations of proposed model from experimental data for pure polar, associating substances. Crosses, connected with a solid line, are the $AAD\%$ of each substance. *Box-and-whisker*-representation of the distribution of deviations. Dots are outliers.

In fig. 4.9, we show the detailed results of 1-nonanol. For this substance, we obtain an AAD-value that is close to the overall average of AAD-values in this class of substances. Experimental data covers a fairly wide range of residual entropy. The data points collapse well onto a single curve which is described very satisfyingly by our ansatz and parameters provided. The diagram shows that for alcohols the curvature gets more pronounced as compared to nonpolar substances and necessitates the proposed non-linear approach. The curvature was found to be most pronounced for alcohols in comparison to other chemical

families here investigated.

Fig. 4.10 gives the results of methylamine. We select methylamine because experimental data for the vapor phase is available. Data points in the liquid phase (i.e., high absolute residual entropies s^*) collapse onto one single curve, though, not as well as seen in previous examples. In the gaseous phase (i.e., low absolute residual entropies s^*) we observe an error of 10 - 15 %. Please note, the ideal gas limit is purely predictive in this case because the A_i parameter of the viscosity model is calculated from the group-contribution approach[48][, described in chapter 2].

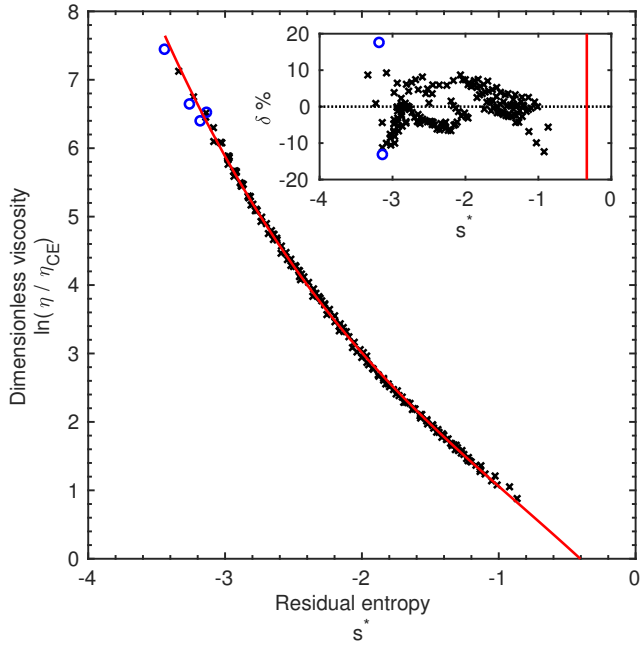


Figure 4.9: Dimensionless viscosity of 1-nonanol versus the residual entropy s^*. Crosses are experimental data. The red line is the result of the correlation with parameters as given in the Supporting Information. The inset shows the relative deviation $\delta\%$ of the calculated viscosities. The vertical red line in the inset is the calculated residual entropy at the critical point $s^{*,crit}$. Blue circles are outliers. Experimental data taken from DDBST[153].

Overall, we find that, with a few exceptions, entropy scaling can be used to give an accurate description of pure substance liquid viscosities with only two to three adjustable parameters per substance. Due to the simple structure of the polynomial correlation function, these parameters are reasonably well-behaved for extrapolations beyond training conditions (see also [chapter 2] Lötgering-Lin et al.[186]). For gas viscosities the accuracy can decrease, whereby we here decided to define the A_i parameter from a group-contribution

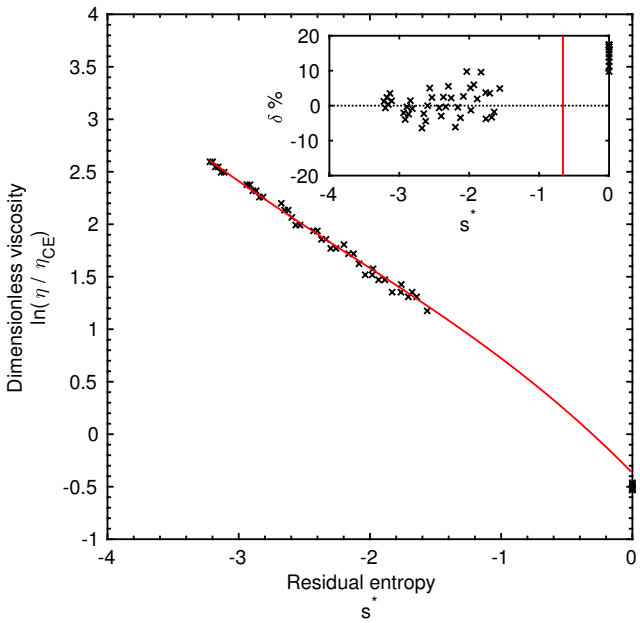


Figure 4.10: Dimensionless viscosity of methylamine versus the residual entropy s^*. Crosses are experimental data. The red line is the result of the correlation with parameters as given in the Supporting Information. The inset shows the relative deviation $\delta\%$ of the calculated viscosities. The vertical red line in the inset is the calculated residual entropy at the critical point $s^{*,crit}$. Experimental data taken from DDBST[153].

approach[48][, described in chapter 2,] and noting that experimental uncertainties often increase for gaseous conditions.

4.4 Results: Mixtures

In this section, we give detailed results for a selection of representative mixtures. Tables with $AAD\%$ values of all investigated mixtures are provided in the Supporting Information. First, we regard mixtures of model fluids as characterized by molecular simulations. For presenting the results for real mixtures, we find it advantageous to also distinguish between (1) mixtures of nonpolar species, (2) mixtures with at least one polar (nonassociating) substance, and (3) systems with at least one associating (hydrogen-bonding) component.

4.4.1 Mixtures of Lennard-Jones Model Fluids

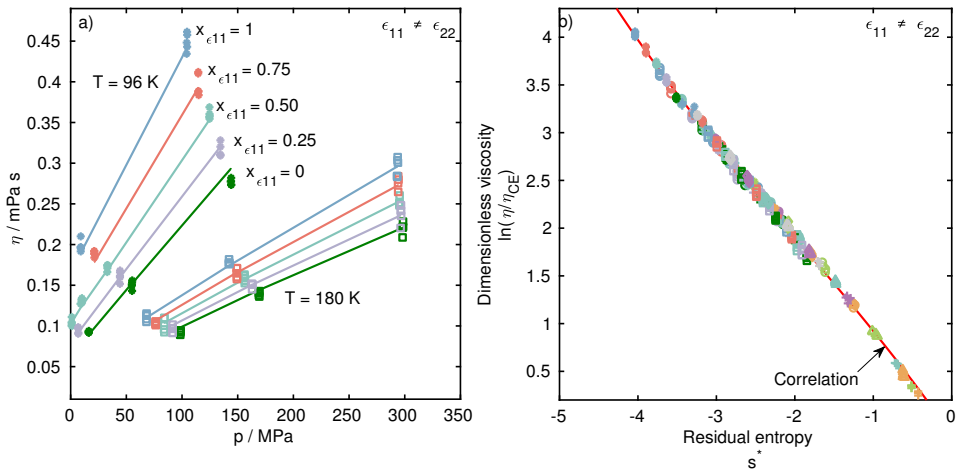


Figure 4.11: Viscosities of LJ mixture I ($\varepsilon_{11} \neq . \varepsilon_{22}$) at two temperatures and various compositions (in different colors).(a) Comparison of model predictions (lines) to results from molecular simulations (symbols).(b) Dimensionless mixture viscosities as a function of the residual entropy. Comparison of model (red line) to results from molecular simulation (colored symbols).

This section presents results of three different binary mixtures of Lennard-Jones fluids, as summarized in tab. 4.1. Fig. 4.11 provides mixture viscosities of model mixture I: two Lennard Jones fluids with same size parameter $\sigma_{11} = \sigma_{22}$, but with different energetic parameter $\varepsilon_{11} \neq . \varepsilon_{22}$. Mixture viscosities at different compositions and conditions (i.e., temperature T, density ρ, and

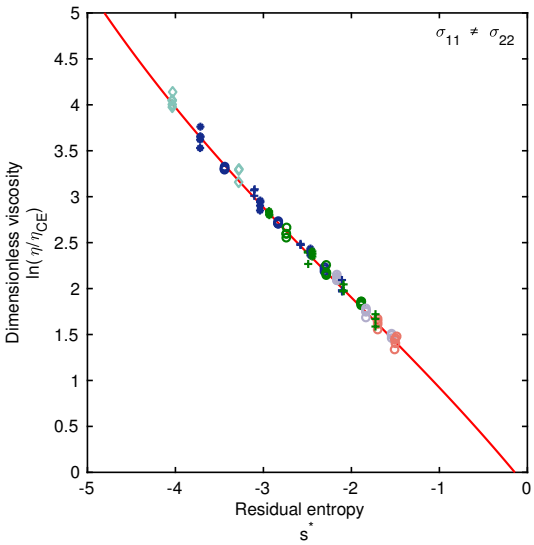


Figure 4.12: Simulated viscosities of mixture II ($\sigma_{11} \neq . \sigma_{22}$). Comparison of model (line) to results from molecular simulations (symbols). Different symbol types are for different temperatures; different colors indicate different compositions.

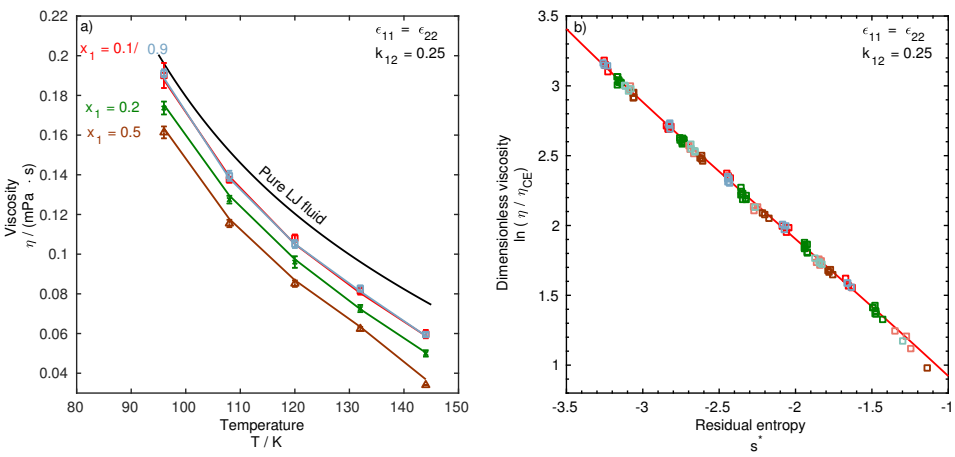


Figure 4.13: Simulated isobaric viscosities of mixture III (deviation from Berthelot equation, $\varepsilon_{12} = 0.75\varepsilon_{11} = 0.75\varepsilon_{22}$) at two temperatures and various compositions. (a) Comparison of model predictions (lines) to results from molecular simulations (symbols).(b) Dimensionless mixture viscosities as a function of the residual entropy. Comparison of model (red line) to results from molecular simulation (colored symbols).

pressure p) all collapse well onto a single curve, as shown in fig. 4.11b. For this mixture, entropy scaling implies that any mixture behaves like a single pseudopure substance. The red line is the result of our mixture model. Because the viscosity parameters for monatomic Lennard Jones fluids are universal, i.e., in this mixture, the two fluids with different energetic parameters ε_{11} and ε_{22} are described with the same set of A_i to D_i viscosity parameters, the mixture viscosity correlation only gives one single curve that represents the simulated dimensionless viscosities η^* well. Consequently, $\eta(p, T)$ is described just as well, as shown in 4.11 (a).

In model mixture II the two Lennard Jones fluids have the same energy parameter $\varepsilon_{11} = \varepsilon_{22}$, without correction to the Berthelot approximation, but now with different size parameters $\sigma_{11} \neq . \sigma_{22}$. Results of our molecular simulations are analyzed in fig. 4.12. All mixture viscosities again collapse to a single curve that is well described by the model.

In model mixture III we look at a mixture that consists of two identical Lennard Jones fluids ($\sigma_{11} = \sigma_{22}$, $\varepsilon_{11} = \varepsilon_{22}$) but with strong deviation from the Berthelot equation, $\varepsilon_{12} = 0.75\varepsilon_{11}$. As shown in fig. 4.13a, this system exhibits an interesting behaviour of the viscosity η: Starting from the viscosity of pure LJ-fluid 1, the mixture viscosities decrease with increasing composition of substance 2, with a lowest viscosity value for the equimolar mixture. The mixture is symmetric, so that for increasing mole fractions of pure LJ-fluid 2 beyond the equimolar mixture, the viscosity of course increases again. The proposed model for mixture viscosities is in very good agreement with the data from molecular simulations.

4.4.2 Mixtures of Nonpolar Substances

We here give detailed results for mixtures with both components being neither polar nor self-associating. Fig. 4.14 presents the *box-and-whisker*-representation of all binary mixtures of nonpolar species with available experimental data that contain n-octane. All mixtures are predicted with an average deviation of 3.89%. We observe no dependence of the quality of the model on the shape-asymmetry of the binary mixture (i.e., for mixtures of long chains with short chains). The high deviation in the mixture containing squalane was anticipated because squalane was already weakly represented as a pure substance (fig. 4.3 (c)). The results shown in fig. 4.14 are representative for the class of mixtures

of nonpolar substances in general; only 35 of 173 mixtures of nonpolar species are predicted with $AAD\%$s of above 10%. These mixtures can be grouped as follows:

- Mixtures containing squalane, which itself is badly represented as a pure substance.
- Some methane or ethane mixtures, especially those with high asymmetry in shape or chain length.
- Mixtures containing aromatics are not consistently well described, i.e., some are well described, but some show $AAD\%$-values between 10% and 20%. We did not identify a pattern for the varying deviations.
- Mixtures containing ethylene show deviations greater than 50%, even though the pure substance is adequately described with an $AAD\%$ of 4.25% averaged over 527 data points.

For contextualization of those mixtures that give higher deviations, we would like to remind, that our work did not consider binary interaction parameters k_{ij}. Some improvement could be expected for mixtures with ethylene or mixtures with methane, for example, where k_{ij}-parameters typically improve the representation of phase equilibrium properties. An example on the effect of applying an adjusted binary interaction parameter k_{ij} is given further below.

Within the group of nonpolar mixtures, we present more detailed results for four different types of mixtures. The simplest cases are mixtures of n-alkanes of comparable chain length. Fig. 4.15 shows the viscosity of the binary mixture n-heptane + n-nonane at two different mole fractions as a function of the temperature. As the overall $AAD\%$ of 2.48 % suggests, all data points are well described for the full range of pressure (up to 717 bar), temperature and concentration covered by the experimental data points. Compared to the model mixtures (Figures 4.11 to 4.13), when dealing with mixtures of real (i.e., non-monatomic) substances, dimensionless mixture viscosities η^* no longer collapse onto one single universal curve but fan out between the pure substance viscosities. This behavior is well represented by the model.

As representative examples for mixtures with substantial asymmetry in molecular size or shape, fig. 4.16, fig. 4.17 and fig. 4.18 show the results of

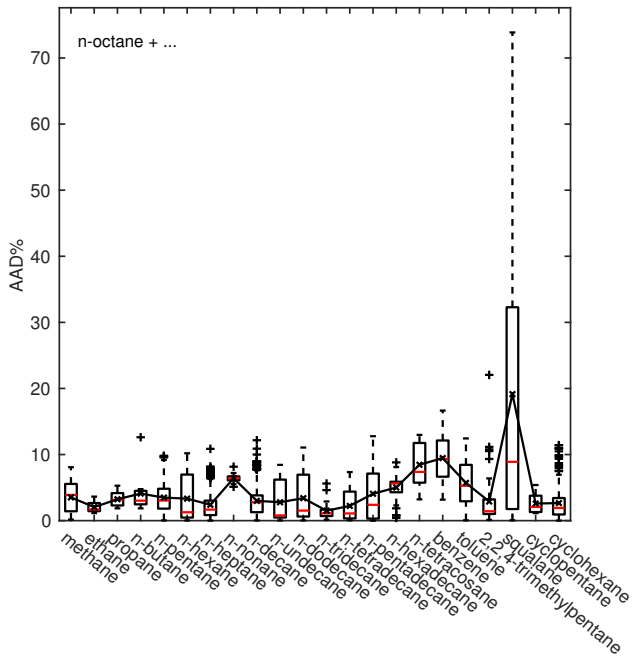


Figure 4.14: *Box-and-whisker*-representation of the investigated mixtures of nonpolar substances that contain n-octane. Crosses, connected with a solid line, are the $AAD\%$ of each substance. Dots are outliers.

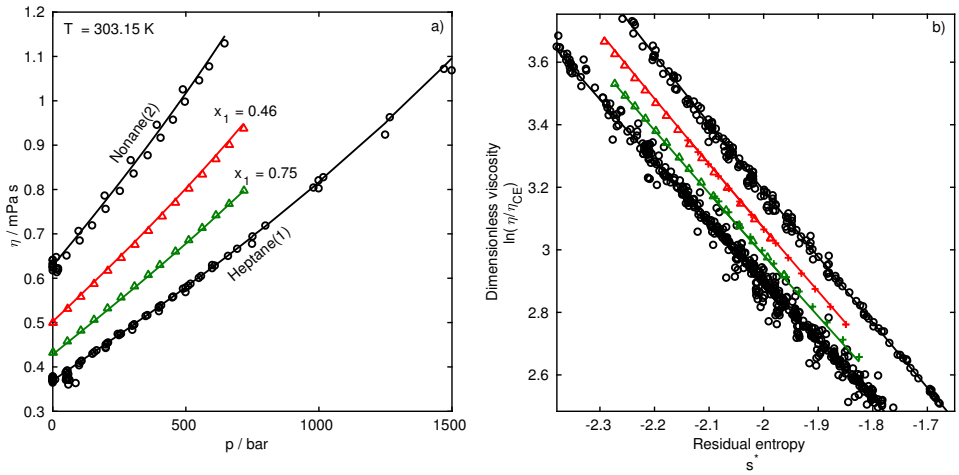


Figure 4.15: Viscosities of n-heptane + n-nonane. In both graphs, black symbols and lines present the pure substance viscosities, colored symbols and lines mixture viscosities. Symbols are experimental data[198–200]; lines are results of the proposed model.(a) shows viscosities $\eta(p, T)$ at one temperature $T = 303.15K$ at two different concentrations.(b) presents the entropy scaled viscosity $\eta^*(s_{res})$ at various temperatures, pressures, and compositions.

n-octane + n-tetradecane as a mixture with asymmetry in chain length, n-dodecane + cyclohexane as an example of a system with asymmetry in shape, and n-tridecane + 2, 2, 4, 4, 6, 8, 8-heptamethylnonane as a mixture of a linear molecule and a highly branched linear molecule. Very good agreement between the model and experimental data is achieved. We note that for n-dodecane + cyclohexane and n-tridecane + 2, 2, 4, 4, 6, 8, 8-heptamethylnonane experimental data covers high pressures of up to 1800 bar. Even at these high pressures, the model gives accurate predictions of mixture viscosities. Results of further mixtures of nonpolar substances can be found in the Supporting Information.

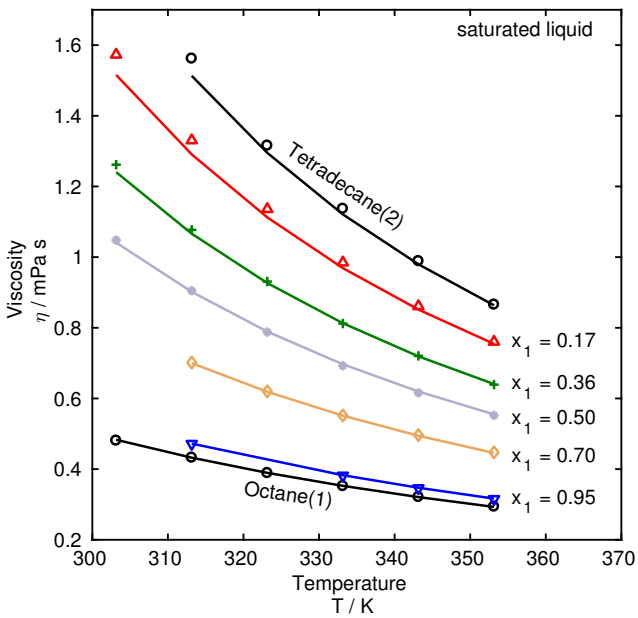


Figure 4.16: Saturated liquid viscosities of n-octane + n-tetradecane. Comparison of proposed model (lines) with experimental data (symbols) [198,201,202]. Black symbols for viscosity of pure substances; colored symbols for mixtures.

4.4.3 Mixtures involving polar, nonassociating substances

The second class of mixtures investigated are systems with at least one polar but nonself-associating (hydrogen-bonding) component. In this group we consider esters, ethers, ketones, hydrofluorocarbons (hfc), and chlorofluorohydrocarbons (cfc). We first regard mixtures with two polar substances. Mixtures with one polar and one nonpolar species are more asymmetric in their interactions and are considered in the same subsection, but further below. Across all mixtures, the proposed model leads to an average deviation of 6.27% to

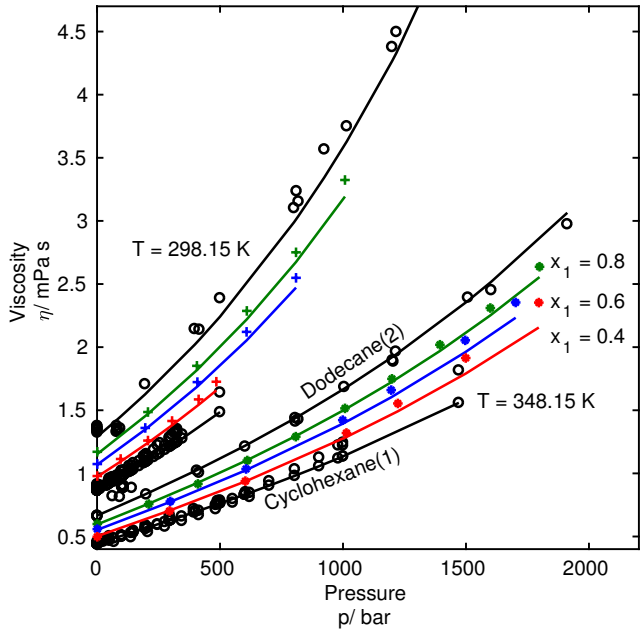


Figure 4.17: Viscosities of cyclohexane + n-dodecane at two temperatures. Comparison of proposed model (lines) with experimental data (symbols)[203,204]. Black symbols for viscosity of pure substances; colored symbols for mixtures.

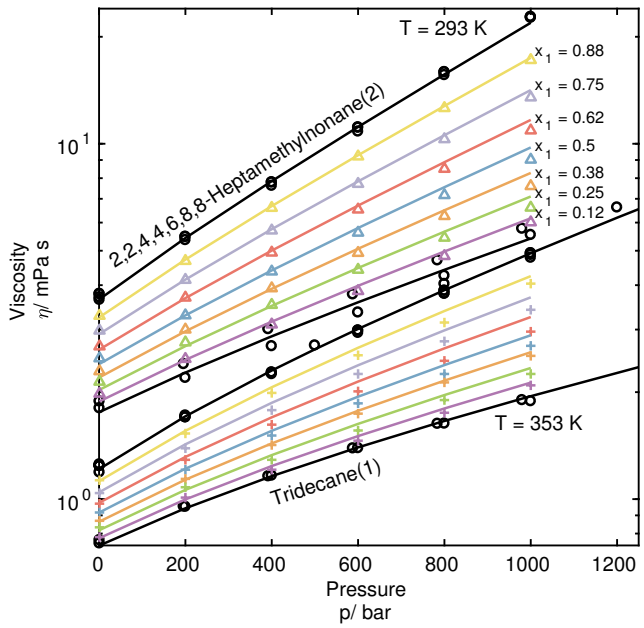


Figure 4.18: Viscosities of n-tridecane + $2,2,4,4,6,8,8$-heptamethylnonane at two temperatures. Comparison of proposed model (lines) with experimental data (symbols)[205]. Black symbols for viscosity of pure substances; colored symbols for mixtures.

all experimental data, whereby only 4 out of 26 mixtures are predicted with $AAD\%$ values above 10% (Supporting Information).

Fig. 4.19 shows the viscosities of the system 1,1,1-trifluoroethane (R143a) in mixture with pentafluoroethane (R125) at molar fraction $x_{R143a} = 0.59$ as a function of pressure for several temperatures. Overall, mixture viscosities are accurately predicted with an $AAD\%$ of 5.7%. The main graph shows the liquid viscosities which are described well by the model. The inset shows gas viscosities. Mixture gas viscosities appear to be represented less accurately, although the deviation is still modest, with $AAD_{\text{gas}}\%$ of 5.14%. We emphasize, however, that for viscosities of this gas mixture the dominantly important A_i parameter stems from the group contribution approach[48][, see chapter 2], so that the gas phase values can be considered as predicted.

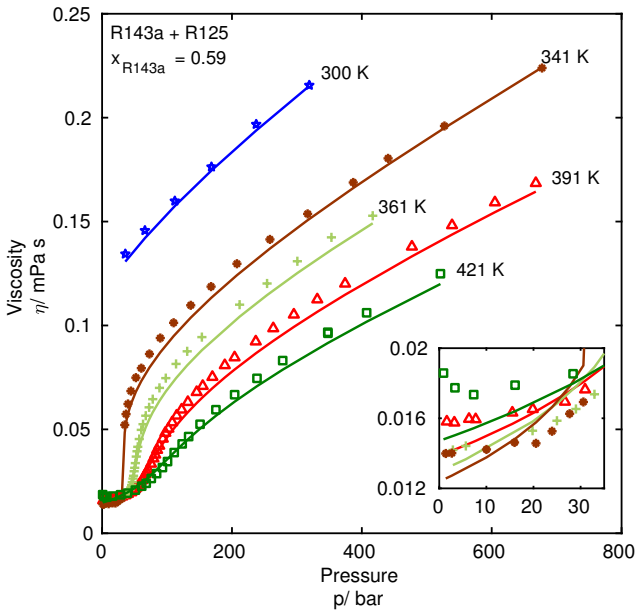


Figure 4.19: Isopleth viscosities of R143a + R125 at $x_{R143a} = 0.59$ and five different temperatures. Symbols represent experimental data[206]; lines are model predictions. The insert emphasizes the pressure range of gas viscosities.

The following group of mixtures consisting of one nonpolar and one polar substance is interesting because it introduces asymmetry in intermolecular interactions. A *box-and-whisker*-representation of all systems containing *n*-octane and a polar (nonassociating) component is given in fig. 4.20. For all of these mixtures, the average deviation ($AAD\%$) is 3.44%. This subset is representative for the entire group of all considered mixtures involving one nonpolar and one polar species: only 5 out of 100 mixtures within the class of

'nonpolar + polar' mixtures are predicted with an $AAD\%$ higher than 10%.

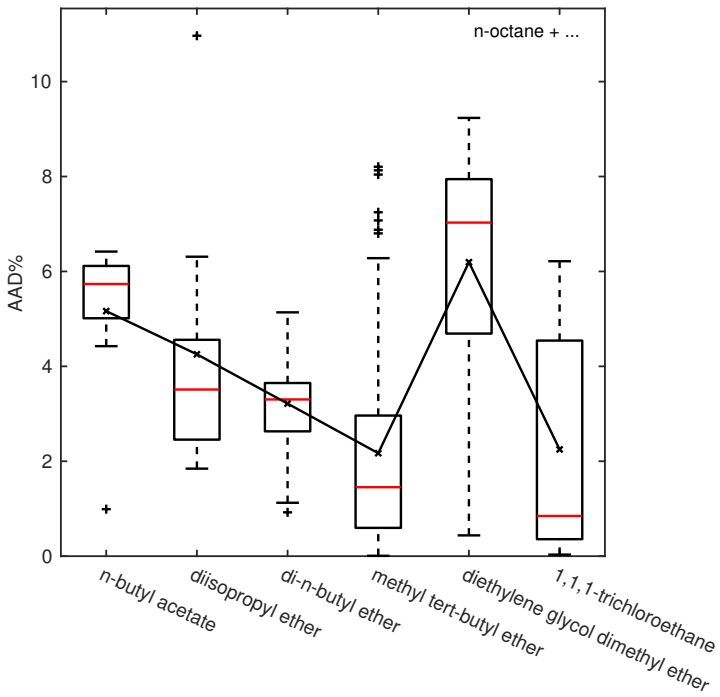


Figure 4.20: Binary systems of n-octane in mixture with a polar substance. Crosses, connected with a solid line, are the $AAD\%$ of each substance. *Box-and-whisker*-representation of the distribution of deviations. Dots are outliers.

Prominent representatives of this group [(i.e. class of 'nonpolar + polar' mixtures)] are mixtures of various alkanes with methyl-tert-butylether (MTBE). In fig. 4.21 we show results of n-nonane with MTBE. The full range of compositions and temperatures are accurately predicted with an $AAD\%$ of 1.19%. This result is especially remarkable because the pure substance viscosity parameters of MTBE were adjusted to very few data points covering only a narrow range of residual entropy, as can be seen in fig. 4.21b.

The mixture difluoromethane (R32) + propane is interesting in two aspects. First, it is a rather nonideal mixture, which exhibits an azeotropic vapor-liquid equilibrium behavior. Without the use of a binary interaction parameter k_{ij}, the VLE is described rather poorly from PCP-SAFT, as fig. 4.22a shows (dashed lines). We use this mixture to investigate how sensitive the predicted mixture viscosities are regarding the value of k_{ij}. Second, the liquid viscosity in this mixture can be well below viscosities of both pure substances at the same temperature, as shown in fig. 4.22b. Simply estimating the mixture viscosity by interpolating pure substance viscosities would lead to very

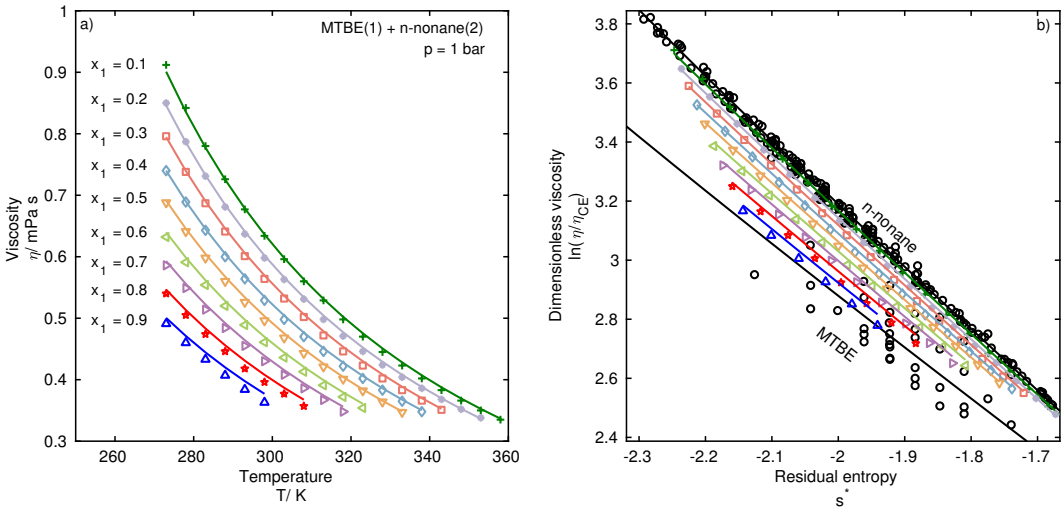


Figure 4.21: Viscosities of MTBE + n-nonane at $p = 1$ bar. Comparison of model predictions (lines) with experimental data (symbols)[207]. Black symbols present the pure substance viscosities; colored symbols are for mixture viscosities. (a) Mixture viscosities as $\eta(p, T)$. (b) Entropy scaled viscosity $\eta^*(s_{res})$ at different compositions.

inaccurate results. The results of the model in fig. 4.22b confirm that (similar to model mixture III) mixture viscosities of such highly nonideal systems can be accurately described, here with an $AAD\%$ of 3.21%. The k_{ij} value of 0.095 was adjusted to VLE data of the mixture at various temperatures. In fig. 4.22b we also include results for $k_{ij} = 0$. Absolute values of mixture viscosities are for that case overestimated with an $AAD\%$ of 11.24%, although the temperature dependence is qualitatively well described. Viscosities lower than both pure substances are also predicted with $k_{ij} = 0$ but to a low degree. We conclude that the model with $k_{ij} = 0$ is promising for screening-type studies, where one seeks estimates for a wide range of substances and mixtures. The quality of the predictions significantly improve, however, when a binary interaction parameter k_{ij} is available, for example, by adjusting to phase equilibrium data.

In order to assess the proposed viscosity model for mixtures in the vicinity of the critical point and at dense supercritical conditions, we regard carbon dioxide (CO_2) in mixture with methane. This mixture is of significant industrial interest. We modeled CO_2 as a quadrupolar substance, with PCP-SAFT parameters taken from the literature[111] and listed in the Supporting Information. Experimental and predicted mixture viscosities of this system are shown in fig. 4.23. The mixture is very well represented with an $AAD\%$ of 5.24%.

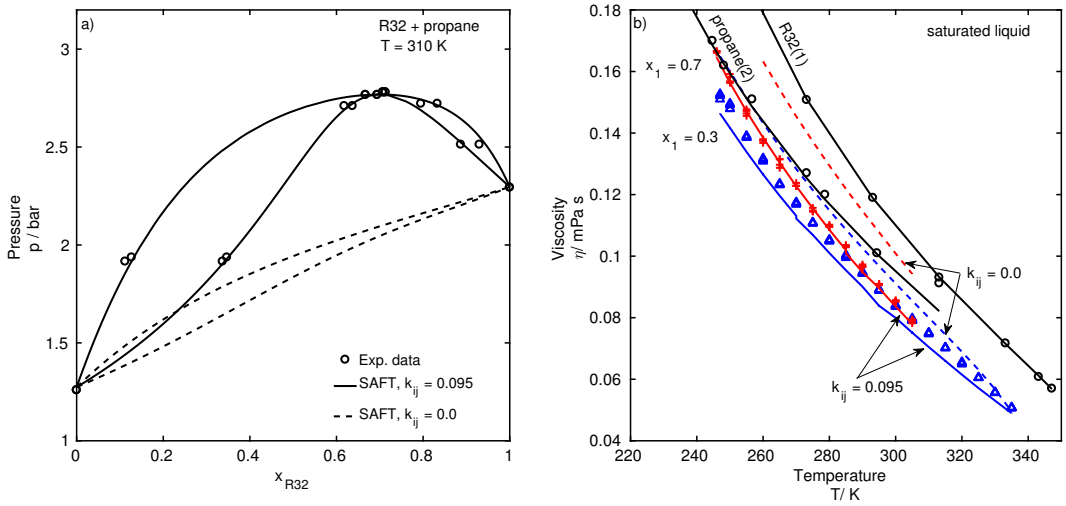


Figure 4.22: Physical properties of mixture difluoromethane (R32) + propane.(a) Vapor-liquid equilibrium at $T = 310$ K calculated with the PCP-SAFT EoS with $k_{ij} = 0.095$ (solid lines) and $k_{ij} = 0$ (dashed lines) in comparison with experimental data (symbols)[208].(b) Mixture viscosities at saturation pressure along two isopleths, also calculated with $k_{ij} = 0.095$ (lines) and with $k_{ij} = 0$ (dashed lines) in comparison to experimental data[209] (symbols).

We consider this a good result because the viscosity has a very steep increase with pressure in this region of phase space, where small errors in predicted pressure lead to large deviations in viscosity. Also the other mixtures containing CO_2 that were considered are well described with an average $AAD\%$ of 5.9%. Results of further mixtures involving polar, nonassociating substances can be found in the Supporting Information.

4.4.4 Limitations of the Proposed Model: Mixtures with Associating (Hydrogen-Bonding) Substances

The last class of mixtures we study are mixtures with at least one self-associating component (i.e., here, alcohols and amines). To avoid asymmetries in the molecular interaction in a first step, we regard mixtures of two self-associating substances. Because one can speak of some degree of symmetry in intermolecular interactions, these mixtures are less demanding for thermodynamic models, compared to mixtures with one self-associating and one nonassociating substance. In fig. 4.24, we present systems of 1-pentanol in mixture with other alcohols. We first observe that for mixtures containing methanol or ethanol the model shows high deviations from experimental data. This finding is more

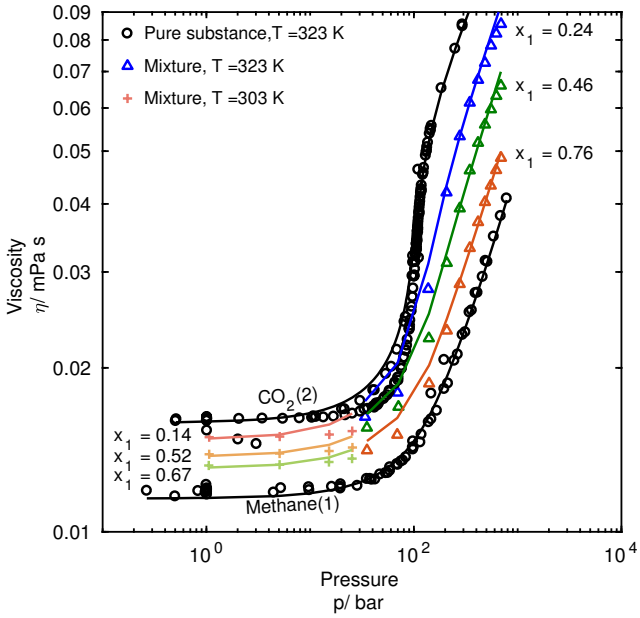


Figure 4.23: Viscosities of methane + CO_2 at two temperatures and different compositions at conditions close to the critical point. Comparison of model predictions (lines) to experimental data[210,211] (symbols). Black symbols represent the pure substances; colored symbols are for mixture data.

general: Almost all mixtures containing either methanol or ethanol are not well represented. For both, pure methanol and pure ethanol, extensive data is available which is well represented by the proposed model. Other mixtures, that do not contain methanol or ethanol, are rather well described, as the inset in fig. 4.24 confirms. Omitting the mixtures containing methanol or ethanol, the average $AAD\%$ value is 5.12%. This number is representative for all alcohol-alcohol mixtures that do not contain methanol or ethanol.

The results of mixtures containing amines fall in two categories. For mixtures of n-propylamine + alcohols (again with the exception of methanol and ethanol), the model is in good agreement with experimental data. The mixtures of n-butylamine + alcohols, however, are very poorly predicted by the model.

Next, we regard mixtures with one nonpolar and one polar and self-associating component. We first analyze mixtures containing n-octane and an alcohol, fig. 4.25. Compared to n-octane mixtures considered in previous sections, the n-octane mixtures containing an alcohol are less accurately represented. Apart from mixtures containing either methanol or ethanol, several other mixtures lead to $AAD\%$ values above 10% or even 20%. Regarding the complexity of

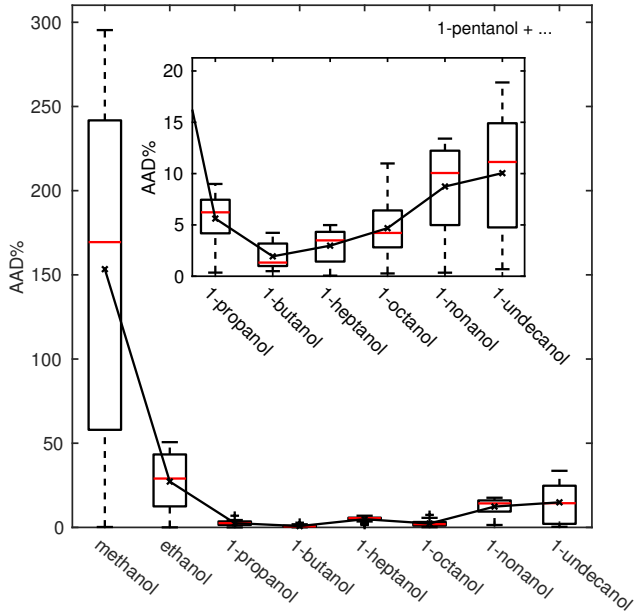


Figure 4.24: Mixtures 'associating + associating' that contain 1-pentanol. Crosses, connected with a solid line, are the *AAD*% of each substance. *Box-and-whisker* representation of the distribution of deviations. Dots are outliers.

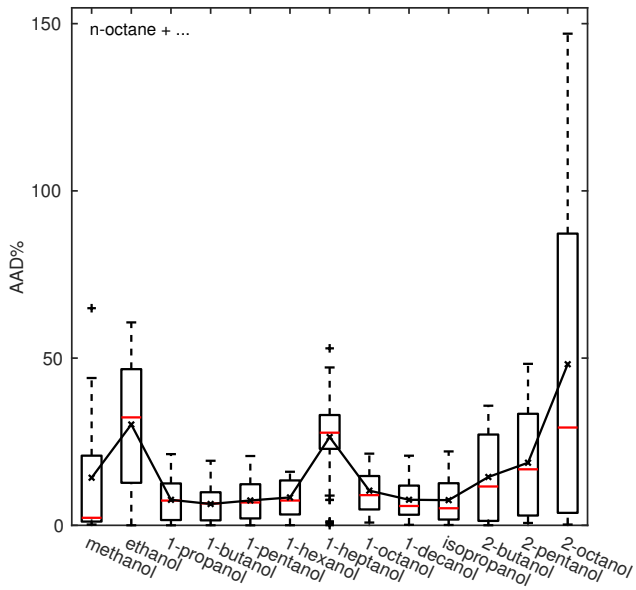


Figure 4.25: Mixtures 'nonpolar + associating' that contain *n*-octane. Crosses, connected with a solid line, are the *AAD*% of each substance. *Box-and-whisker*-representation of the distribution of deviations. Dots are outliers.

these mixtures and the fact that, due to our objective of predicting mixtures, we did not adjust any k_{ij} values to these mixtures, the results are on the one hand still satisfying. On the other hand, it is important to clarify whether the deviations are due to the approximation $k_{ij} = 0$, i.e., if improved viscosity predictions can be obtained when a k_{ij}-value is adjusted to, say, vapor-liquid equilibria. We adjusted k_{ij} values to VLE data of selected mixtures and compared $AAD\%$ values for viscosities. Of course, no viscosity information entered the regression of k_{ij}-values. The results for this study are given in tab. 4.2. For mixtures containing either methanol or ethanol, we find that using an adjusted k_{ij} value does not improve the viscosity-results of model predictions significantly. Methanol + n-pentane as well as ethanol + n-hexane remain at high $AAD\%$ values. The results for ethanol + n-pentane, which is satisfyingly represented already without k_{ij} (fig. 4.26), are not altered significantly. For mixtures containing longer alcohols, however, the improvement of the model performance scales with the magnitude of the adjusted k_{ij} value: The results of the system 1-hexanol + n-hexane, which requires a rather moderate k_{ij}-value to represent the VLE appropriately, are only slightly improved by considering an adjusted k_{ij}-value. The systems n-octane + 1-propanol (fig. 4.27) and n-octane + 1-butanol require larger k_{ij}-values (0.03 and 0.044, respectively) and viscosity deviations of our model are reduced significantly for these mixtures when using the adjusted k_{ij}-values.

The observations made for n-octane + alcohol mixtures are representative for other mixtures of alcohols with n-alkanes, cycloalkanes and alkenes: The viscosity of most mixtures are accurately predicted by the proposed model, although some systems show significantly higher deviations than found for systems without associating species. No systematic or clear reason for these exceptions was found. As a tendency, we see that mixtures containing a secondary alcohol in mixture with a nonpolar substance are overall less accurately predicted by the viscosity model than mixtures containing a primary alcohol.

Mixtures of aromatics and alcohols as well as mixtures of one polar, nonselfassociating component with a (self-)associating substance are weakly predicted by the proposed approach, with $AAD\%$ values that exceed 20%, as seen in Fig. 4.28. All results of mixtures involving associating substances can be found in the Supporting Information. The development of the viscosity model proposed in this work was guided by a molecular simulation study. This study revealed a set of suitable mixing rules for the viscosity model (eq. 4.6). The

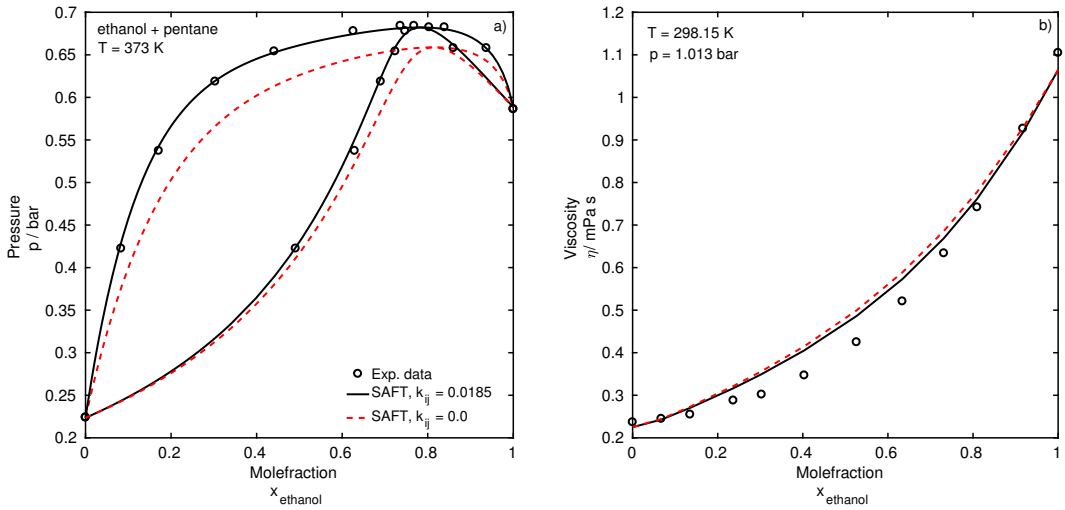


Figure 4.26: Physical properties of mixture ethanol + n-pentane.(a) Vapor-liquid equilibrium at $T = 373$ K from PCP-SAFT EoS with $k_{ij} = 0.0185$ (solid lines) and $k_{ij} = 0$ (dashed lines) in comparison to experimental data[212] (symbols).(b) Predicted mixture viscosities along isothermal and isobaric conditions, also for $k_{ij} = 0.0185$ (solid lines) and $k_{ij} = 0$ (dashed lines) in comparison to experimental data[213] (symbols).

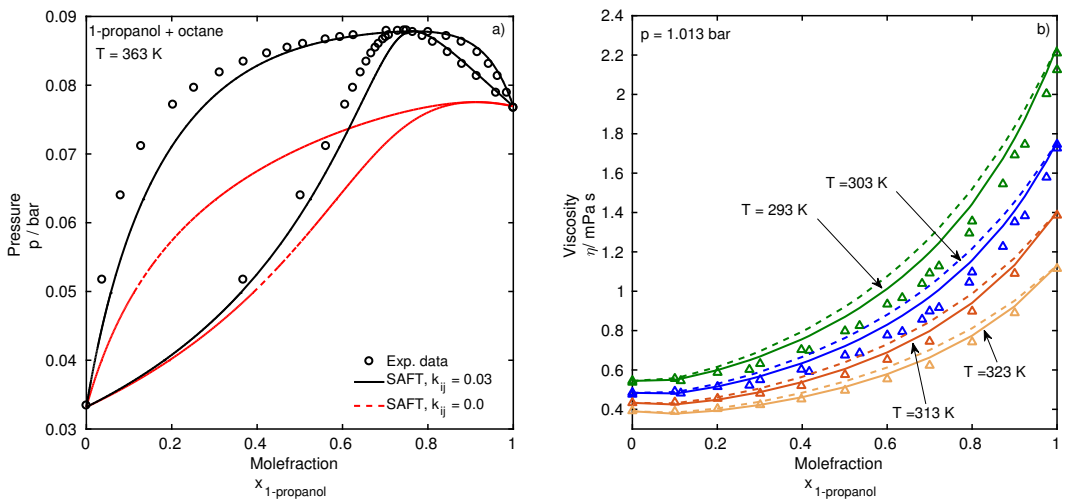


Figure 4.27: Physical properties of mixture 1-propanol + n-octane.(a) Vapor-liquid equilibrium at $T = 363$ K from PCP-SAFT EoS with $k_{ij} = 0.03$ (solid lines) and $k_{ij} = 0$ (dashed lines) in comparison to experimental data[214] (symbols).(b) Predicted mixture viscosities along several isotherms and pressure $p = 1.013$bar, also with $k_{ij} = 0.03$ (solid lines) and $k_{ij} = 0$ (dashed lines) in comparison to experimental data[213,215,216] (symbols).

Table 4.2: Influence of k_{ij} values on the prediction of mixture viscosities of mixtures of type 'nonpolar + associating'.

	k_{ij}	$AAD\%$	k_{ij}	$AAD\%$
methanol + n-pentane	0.0	32.84%	0.012	32.44
ethanol + n-pentane	0.0	9.10%	0.018	7.63
ethanol + n-hexane	0.0	21.17%	0.019	20.12%
1-hexanol + n-hexane	0.0	6.30%	0.012	5.52%
1-propanol + n-octane	0.0	7.63%	0.030	4.11%
n-octane + 1-butanol	0.0	6.38%	0.044	1.87%

molecular simulations were performed for model fluids and mixtures of model fluids interacting with Lennard-Jones potentials. The results of this section suggest that hydrogen-bonding interactions need to be considered in a reliable viscosity model. A possible approach is to consider the bond-fraction (or monomer fraction) of the association term in the PC-SAFT model in the viscosity model. This might be a subject of a subsequent study.

4.4.5 Prediction of Mixture Viscosities: Summary of Results

A comprehensive study of 566 binary mixtures of real substances with about 34500 experimental data points from literature was conducted. Out of these 566 mixtures, 173 mixtures are mixtures of nonpolar substances and 126 are mixtures including polar but nonself-associating substances. The remaining 267 mixtures consist of at least one component that is self-associating.

Fig. 4.28 provides the results of all investigated systems, with colors indicating the average $AAD\%$. On the main diagonal are the pure substances that were discussed in the previous section. The off-diagonal elements are for binary mixtures. The diagram in summary indicates very good agreement for nonassociating components, but higher deviations for mixtures with associating (hydrogen-bonding) substances.

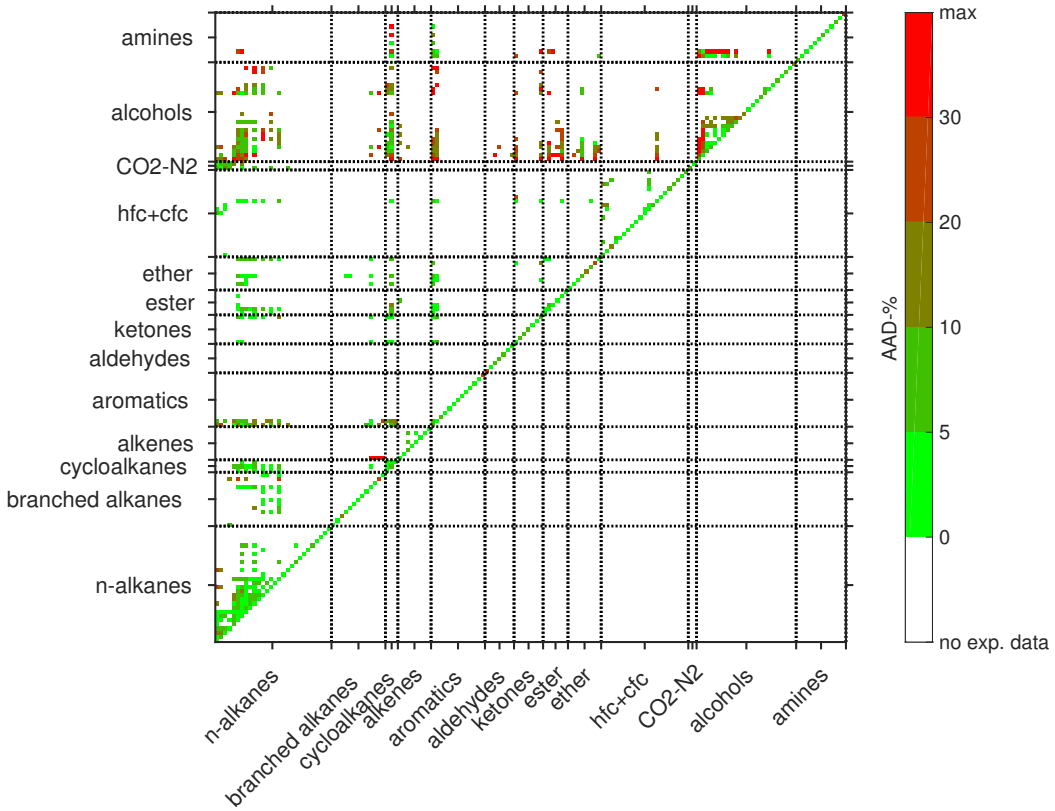


Figure 4.28: Summary all substances and mixtures investigated, according to chemical families. The main diagonal are the pure substances. Off-diagonal entries are 566 binary mixtures with about 34500 experimental data points. Color coding gives the deviations (as a range of $AAD\%$-values) of the model from experimental data.

4.4.6 Prediction of Mixture Viscosities Based on the Group Contribution Approach

The previously described results are based on pure substance viscosity parameters determined by individual adjustment to pure substance viscosity data. For substances with scarce or missing experimental viscosity data, group contribution (*gc*) approaches are very powerful. In *gc*-approaches, molecules are decomposed into functional groups α, e.g., CH_3 or CH_2 in the case of *n*-alkanes. Based on these functional groups, and the corresponding functional group parameters, the desired molecule can be constructed and pure substance properties can be predicted. To calculate the residual entropy s_{res} based on group contributions, we use the so-called homosegmented PC-SAFT model[118] (*gc*-PC-SAFT). For the viscosity, we consider our previously published model[48] [presented in chapter 2].

To assess predictions of mixture viscosities with a *gc*-approach, we apply the *gc*-method to mixtures of *n*-alkane+*n*-alkane-, *n*-alkane+aromatic- and *n*-alkane+cycloalkane-mixtures. Figures 4.29 - 4.31 show the AAD% values of various mixtures of each class of mixtures. Every symbol represents the AAD% averaged over all data points of one mixture. Symbols close to the black bisecting line resemble equal results with either of the two modeling approaches (*gc*-based or non-*gc*-based). Points below the bisecting line represent mixtures that are more accurately predicted with the *gc*-based approach, and vice versa.

Overall the mixture viscosities can be predicted surprisingly well with the *gc*-approach. To make this point clear, the *gc*-results shown in fig. 4.29 are based on one single set of six group contribution viscosity parameters (and six *gc*-PC-SAFT parameters) for all *n*-alkanes present in the mixtures. The subsequent results shown in fig. 4.30 and fig. 4.31 confirm, that also for mixtures of non-linear molecules the *gc* approach leads to good predictions of mixture viscosities. This is a valuable result, for example for problems of solvent and process design[12,13,217].

4.5 Conclusions

In this work, we propose an approach for predicting viscosities of pure substances and mixtures based on entropy scaling. The proposed model only needs two adjustable parameters if group contribution parameters are avail-

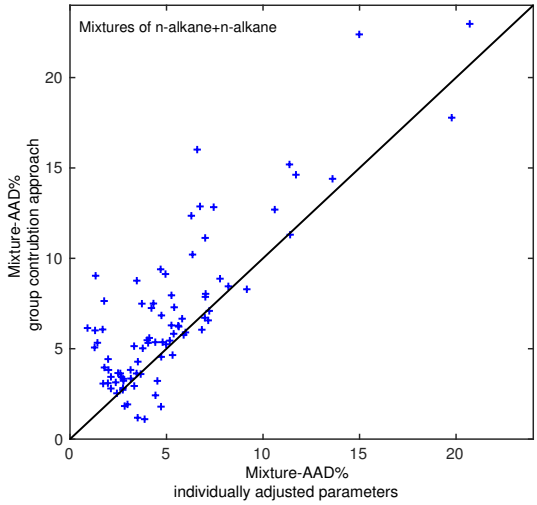


Figure 4.29: Comparison of deviations (i.e., $AAD\%$) of predicted viscosities of mixtures obtained with either a fully group contribution based approach (y-axis) or with individually adjusted pure substance parameters (x-axis) for mixtures of n-alkanes + n-alkanes.

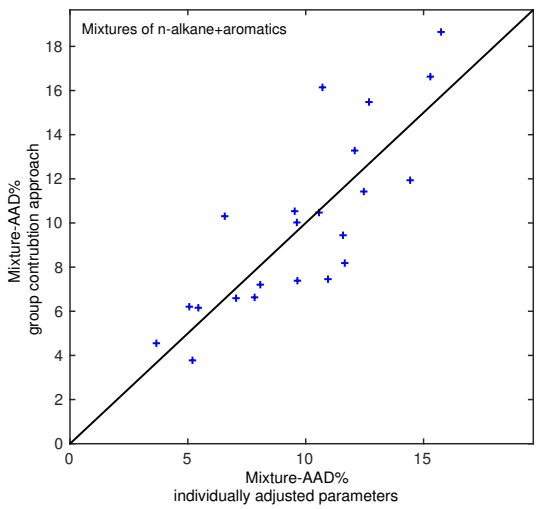


Figure 4.30: Comparison of deviations (i.e., $AAD\%$) of predicted viscosities of mixtures obtained with either a fully group contribution based approach (y-axis) or with individually adjusted pure substance parameters (x-axis) for mixtures of n-alkanes + aromatics.

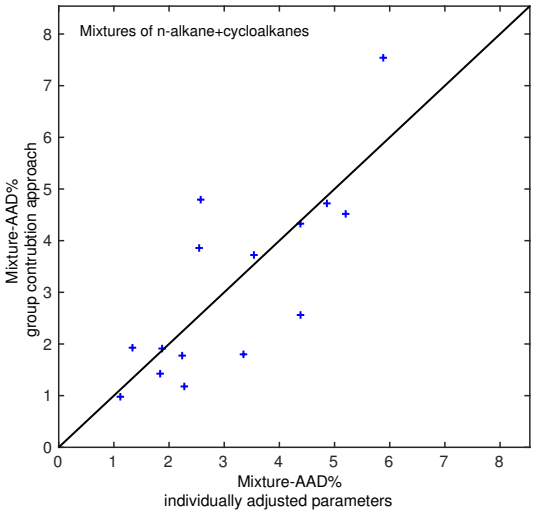


Figure 4.31: Comparison of deviations (i.e., $AAD\%$) of predicted viscosities of mixtures obtained with either a fully group contribution based approach (y-axis) or with individually adjusted pure substance parameters (x-axis) for mixtures of n-alkanes + cycloalkanes.

able (see Lötgering-Lin and Gross [48] [or chapter 2]), otherwise three adjustable parameters. We adjusted parameters for 140 pure substances. Pure substance viscosities can be described over wide ranges of pressure and temperature.

Molecular simulations were performed for mixtures of Lennard-Jones fluids in order to guide in the development of a mixture model for the viscosity. The proposed viscosity model does not introduce any further adjustable parameters for mixtures so that mixture viscosities are calculated in a predictive manner based on the pure substance viscosity parameters. Viscosities of mixtures that do not contain hydrogen-bonding (associating) substances can be accurately predicted. Overall, 299 mixtures without associating components were investigated with average relative deviations (i.e., $AAD\%$) of predicted viscosities from experimental data of 6.0%. Only 44 of the 299 mixtures showed average deviations greater than 10%. Even rather non-ideal mixtures (for example mixtures of a polar with a nonpolar species, or strongly size-asymmetric mixtures) can be predicted with very satisfying accuracy. Our study was performed assuming the binary-interaction parameter $k_{ij} = 0$, but our work also shows that a binary interaction parameter k_{ij} independently adjusted to vapor-liquid equilibria improves the prediction of mixture viscosities.

For mixtures that contain associating components, the viscosity model more often leads to deviations (AAD) that exceed 10 %. Especially mixtures that

include methanol or ethanol show high deviations. Introducing an adjustable binary viscosity parameter that should lead to better correlation results for mixtures with associating substances was not covered by the objective of this work, namely to predict mixture viscosities. Instead, we speculate that the degree of hydrogen bonding (as available from the association term of the PCP-SAFT model) should enter the entropy scaling model for an improved prediction of viscosities in associating mixtures.

[The mixing rules of the proposed model are not based on the molar mass but on the chain length of the constituents. Hence, mixtures of components of similar shape but very different molar mass are not represented accurately. To describe such asymmetric mixtures, literature suggests the implementation of more complex, molar mass weighted mixing rules[43,184,218].]

Appendix

4.A Modifications to the Viscosity Parameters

We here add the superscript 'gc' to the viscosity parameter A_i to indicate it originates from a combination of group contribution parameters as given in our previous publication[48][, presented in chapter 2 of this work] (i.e., A_i^{gc}). In this work, we do not use a Chapman-Enskog formulation that is connected to the segment number m. Hence, to obtain the viscosity parameter A_i as used in eqs 4.5 and 4.6 from A_i^{gc}, a slight modification has to be introduced. The group-contribution form of the Chapman-Enskog viscosity applied was defined as[48]

$$\eta_{\mathrm{CE},i}^{\mathrm{gc}} = \frac{5}{16} \frac{\sqrt{M_i k_B T / (m_i N_A \pi)}}{{\sigma_i}^2 \, \Omega_i^{(2,2)*}} = \eta_{\mathrm{CE},i} \sqrt{1/m_i} \tag{4.18}$$

Applying this to eq. 4.5 yields

$$\ln \eta_i^* = A_i^{\mathrm{gc}} + \ln(\sqrt{1/m_i}) + B_i s^* + C_i {s^*}^2 + D_i {s^*}^3 \tag{4.19}$$

Comparing eq. 4.19 with eq. 4.5 then gives

$$A_i = A_i^{\mathrm{gc}} + \ln\left(\sqrt{1/m_i}\right) \tag{4.20}$$

4.B Reformulated, EoS-independent formulation of the viscosity parameters

When applying an equation of state that does not provide molecular parameters, correlations provided by Chung et al.[36] can be applied, given as

$$\frac{\tilde{\varepsilon}_i}{k_B} = \frac{T_{c,i}}{1.2593} \tag{4.21}$$

$$\tilde{\sigma}_i = 0.809 V_{c,i}^{\frac{1}{3}} \tag{4.22}$$

where the tilde indicates the use of Chung's approximation for the molecular parameters.

The Chapman-Enskog viscosity obtained by using the parameters $\tilde{\sigma}_i$ and $\tilde{\varepsilon}_i$, referred to as $\tilde{\eta}_{\mathrm{CE},i}$, is connected to the Chapman-Enskog viscosity obtained with molecular parameters from the PCP-SAFT EoS by

$$\eta_{\mathrm{CE},i} = \tilde{\eta}_{\mathrm{CE},i} \frac{\tilde{\sigma}_i^2 \ \tilde{\Omega}_i^{(2,2)*}}{\sigma_i^2 \ \Omega_i^{(2,2)*}} \tag{4.23}$$

where variables without the tilde are determined with the molecular parameters as used in the PCP-SAFT EoS.

The viscosity parameter $\tilde{A}_i$ that is based on Chung's approximation for molecular parameters can therefore be deduced from A_i by

$$\tilde{A}_i = A_i + \ln\left(\frac{\tilde{\sigma}_i^2 \ \tilde{\Omega}_i^{(2,2)*}}{\sigma_i^2 \ \Omega_i^{(2,2)*}}\right) \tag{4.24}$$

The pure substance viscosity parameters that include the segment number m_i can be written as

$$\tilde{B}_i = \frac{B_i}{m_i} \tag{4.25}$$

$$\tilde{C}_i = \frac{C_i}{m_i^2} \tag{4.26}$$

$$\tilde{D}_i = \frac{D_i}{m_i^3} \tag{4.27}$$

Then, the EoS-independent viscosity model reads as

$$\ln \frac{\eta_i}{\tilde{\eta}_{\mathrm{CE},i}} = \tilde{A}_i + \tilde{B}_i \left(\frac{s_{res}}{N_A\ k_B} \right) + \tilde{C}_i \left(\frac{s_{res}}{N_A\ k_B} \right)^2 + \tilde{D}_i \left(\frac{s_{res}}{N_A\ k_B} \right)^3 \tag{4.28}$$

4.C Correction of Previously Published Alkene-Parameters of the Group Contribution Viscosity Model

In our previous work[48][, presented in chapter 2], we unintentionally only used 1-alkenes to adjust the viscosity group parameters for the functional groups $=CH_2$ and =CH–. Alkenes other than 1-alkenes were not part of the training set. When using these parameters for alkenes with a double bond inside the molecule, nonphysical viscosity values are obtained. We now readjusted the viscosity group parameters for alkenes with a double bond at different positions. The new parameters are given in tab. 4.3. These parameters can be applied to alkenes independent of the position of the double bond. To clarify the chemical structure of the named groups, please see fig. 4.32.

Table 4.3: New viscosity group parameters for alkenes. $D = 0.01245 = const.$ replacing values given in ref.[48] [and chapter 2].

functional group α	$A_\alpha \times 10^3$	$B_\alpha \times 10$	$C_\alpha \times 10^2$
$=CH_2$	-9.8955	-1.4164	-8.6680
=CH–	1.2549	-1.5930	-7.7575
=C<	-10.5416	-0.7296	24.3935

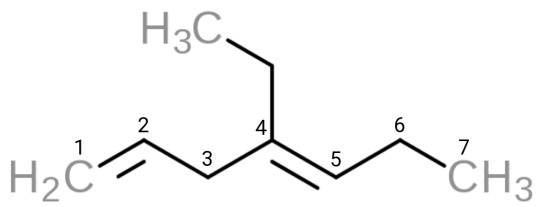


Figure 4.32: Example molecule to clarify the structure of the given groups. The functional group $=CH_2$ represents atom number 1, group =CH– represents atoms number 2 and 5, and group =C< represents atom number 4.

5. Prediction of visosities of mixtures with more than two components

5.1 Introduction

In the preceding chapter, the viscosity model was applied to pure substances and binary mixtures. All equations (most importantly eq. 4.6 and eq. 4.9) can also be applied to mixtures consisting of more than 2 constituents. In this chapter, the viscosity model is applied to 17 ternary mixtures, 21 quaternary mixtures and 3 quinary mixtures.

No additional equations have to be introduced, so that this chapter merely presents results for predicted viscosities of multicomponent mixtures. All calculations are carried out without binary interaction parameters (i.e. $k_{ij} = 0$). As before, no further parameters were adjusted. Hence, the calculated viscosities are purely predictive, based on the pure substance parameters presented in chapter 4.

5.2 Results

5.2.1 Ternary Mixtures

All data of ternary viscosities were obtained from the Dortmund Datenbank (DDB)[153]. Out of the 17 ternary mixtures, 12 mixtures are mixtures of n-alkanes. Out of these 12 n-alkane mixtures, six contain methane. Overall, the ternary mixtures of n-alkanes are well represented, with an average $AAD\%$ value of 5.20% (see tab. 5.1). The available data consists of vapor

as well as liquid viscosities which are both equally well predicted. As observed for binary mixtures of n-alkanes, accuracy of the viscosity model declines with increasing difference of the chainlength of the components of the mixture. Fig. 5.1 compares the results of the prediction for the ternary mixture eicosane+tetracosane+heptane to experimental data as an example of a well represented asymmetric mixture of n-alkanes.

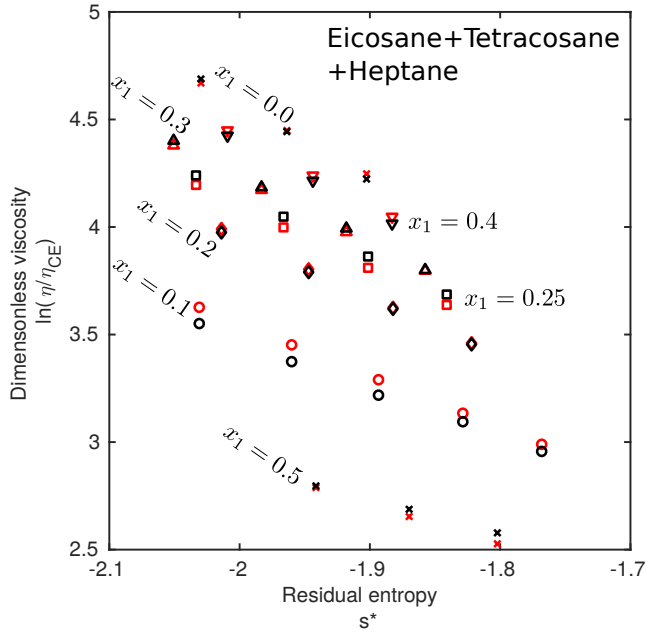


Figure 5.1: Experimental[219] (black symbols) and calculated (red symbols) viscosities of the ternary mixture eicosane+tetracosane+heptane. Each symbol represents a constant molar concentration of eicosane.

Data of five ternary mixtures containing a polar substance were available in the DDB. Based on that data, also ternary mixtures containing polar components can be accurately predicted, although, overall, slightly worse than for mixtures without polar components. Especially the mixtures containing benzene show significantly higher $AAD\%$ values (see tab. 5.1).

5.2.2 Quaternary Mixtures

Out of the 21 quaternary mixtures, 11 mixtures consist of n-alkanes only. The liquid viscosity of these 11 mixtures are accurately represented with an average $AAD\%$ of 6.22 % (see table 5.2). Similar to ternary mixtures, also quaternary mixtures that contain polar components (here, mildly polar substances, namely

Table 5.1: Deviations of predicted mixture viscosities to experimental data as AAD% of ternary mixtures.

Component1	Component2	Component3	N_{dat}	$AAD\%$	$N_{dat,liq}$	$AAD\%_{liq}$	$N_{dat,vap}$	$AAD\%_{vap}$
Methane	Ethane	Propane	3	6.06	0	-	3	6.06
Methane	Ethane	Butane	3	6.61	0	-	3	6.61
Methane	Propane	Butane	3	7.04	0	-	3	7.04
Methane	Decane	Tetradecane	24	6.19	24	6.19	0	-
Methane	Decane	Hexadecane	24	16.08	24	16.08	0	-
Methane	Decane	Octadecane	24	9.40	24	9.40	0	-
Ethane	Propane	Butane	3	5.29	0	-	3	5.29
Pentane	Octane	Decane	530	4.58	530	4.58	0	-
Hexane	Octane	Hexadecane	10	6.39	10	6.39	0	-
Heptane	Octane	Nonane	5	6.70	5	6.70	0	-
Eicosane	Tetracosane	Heptane	26	3.13	26	3.13	0	-
Decane	Eicosane	Tetracosane	18	3.40	18	3.40	0	-
Hexane	Cyclohexane	Benzene	19	19.62	19	19.62	0	-
Benzene	Cyclohexane	Tetradecane	256	10.32	256	10.32	0	-
Toluene	Cyclohexane	Pentane	33	8.66	33	8.66	0	-
Toluene	Cyclohexane	Hexane	30	7.79	30	7.79	0	-
Chlorodifluoromethane[R22]	Pentafluoroethane[R125]	Propane	14	6.93	0	-	14	6.93

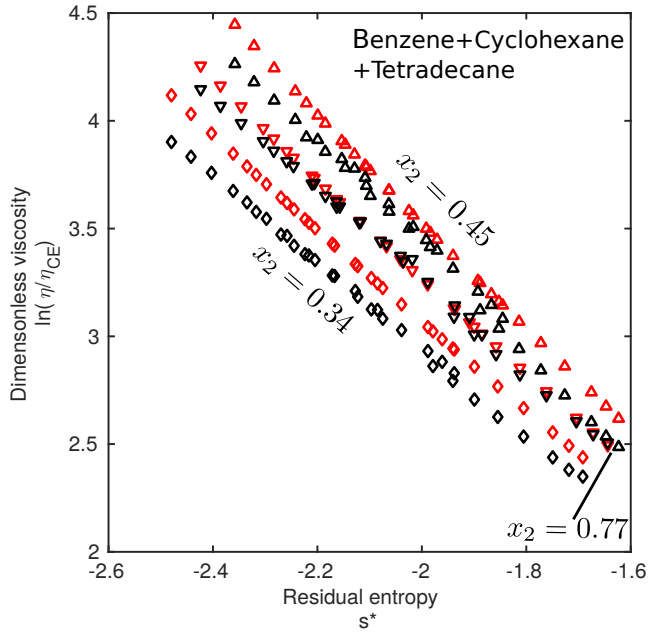


Figure 5.2: Experimental[220] (black symbols) and calculated (red symbols) viscosities of the ternary mixture benzene+cyclohexane+tetradecane. Each symbol represents a constant molar concentration of cyclohexane.

benzene, toluene and ethylbenzene) are predicted less accurately than mixtures of non-polar substances with an average $AAD\%$ value of 13.56 %.

5.2.3 Quinary Mixtures

Only one of the three quinary mixtures with experimental available in the DDB consists of n-alkanes only. The other two mixtures both contain mildly polar substances, i.e. toluene and ethylbenzene. Viscosities of all three quinary mixtures are equally well predicted by the model with an average $AAD\%$ of 10.62 %; also see tab. 5.3.

Table 5.2: Deviations of predicted liquid mixture viscosities to experimental data as AAD% of quaternary mixtures.

Component1	Component2	Component3	Component4	N_{dat}	$AAD\%$
Methane	Ethane	Propane	Butane	3	6.32
Hexane	Decane	Dodecane	Hexadecane	16	6.40
Hexane	Octane	Decane	Tetracosane	15	11.57
Hexane	Octane	Dodecane	Hexadecane	5	9.02
Heptane	Decane	Tridecane	Hexadecane	10	10.14
Heptane	Octane	Dodecane	Hexadecane	10	7.76
Heptane	Nonane	Dodecane	Hexadecane	18	5.56
Heptane	Octane	Undecane	Tridecane	28	11.55
Octane	Undecane	Tridecane	Pentadecane	10	10.34
Octane	Decane	Undecane	Pentadecane	10	8.03
Decane	Dodecane	Tetradecane	Hexadecane	90	2.35
Decane	Hexane	Cyclohexane	Benzene	23	14.61
Benzene	Ethylbenzene	Heptane	Cyclooctane	30	13.53
Hexane	Ethylbenzene	Cyclohexane	Toluene	24	16.59
Benzene	Toluene	Ethylbenzene	Heptane	20	10.22
Octane	Ethylbenzene	Cyclohexane	Toluene	24	17.19
Octane	Hexane	Ethylbenzene	Toluene	24	9.64
Toluene	Ethylbenzene	Heptane	Cyclooctane	10	13.37
Benzene	Toluene	Heptane	Cyclooctane	20	14.65
Octane	Hexane	Cyclohexane	Toluene	24	11.03
Pentane	Toluene	Heptane	Cyclohexane	15	14.67

Table 5.3: Deviations of predicted liquid mixture viscosities to experimental data as AAD% of quinary mixtures.

Component1	Component2	Component3	Component4	Component5	N_{dat}	AAD
Heptane	Octane	Undecane	Tridecane	Pentadecane	7	10.05
Octane	Hexane	Ethylbenzene	Cyclohexane	Toluene	22	10.52
Benzene	Toluene	Ethylbenzene	Heptane	Cyclooctane	20	10.93

6. An optimal experimental design study for predicting the viscosity of pure substances using a group contribution model

Work presented in this chapter has been conducted with the support of Dr. Andreas Geiges from the research group of Prof. Dr.-Ing. Wolfgang Nowak, Institute for Modelling Hydraulic and Environmental Systems (LS^3)/SimTech, University of Stuttgart, Germany.

*Specific contributions to the work presented in this chapter are as follows: Andreas Geiges provided the code for the Preposterior Data Impact Assessor (PreDIA) * and advised on implementing it into the existing thermodynamic framework. Both Wolfgang Nowak and Andreas Geiges advised as experts on Bayesian Statistics, especially in the analysis of the calculated results.*
I, under the supervision of Joachim Gross, developed the code embedding PreDIA in the thermodynamic framework, defined the investigated example applications, conducted the calculations and evaluated the results presented in this chapter.

Recent developments in the calculation of transport properties allow the prediction of increasingly complex pure substances and mixtures. The aim of this work is to estimate the expected model uncertainty in the light of possible new information (i.e. experiments). The Preposterior Data Impact Assessor (PreDIA)[135] approach is applied for optimal experimental design (OED) in combination with the prediction of viscosities using Entropy Scaling. The

*The code can be freely downloaded from www.github.com/predia/source

goal is to select optimal experimental conditions such that the expected model uncertainty is minimized.
In this chapter, „design“ refers to the conditions an experiment is carried out at, i.e. selected substance(s), temperature and pressure. The viscosity model proposed in chapter 2 and chapter 4 is used to describe the dependence of the viscosity on pressure and temperature. From chapter 3, the procedure to obtain model parameter distributions (i.e. the MCMC step) while taking experimental and model errors into account, is adopted.

6.1 Introduction

The preceding chapters introduced the background and results of the proposed viscosity model. A statistical framework for the model assessment was described and applied. Because the introduced viscosity model is data driven (i.e. semi-empirical), experimental data is needed to adjust the viscosity parameters. When approaching more complex fluids, one often lacks experimental data (especially at high pressures). The higher the desired experimental pressures, the higher the effort and technical requirements. Some questions posed to an experimentalist for planning experiments particularly meaningful for building and parameterizing reliable viscosity models are:

1. For a limited pressure range technically accessible, at which experimental conditions should measurements be carried out to parameterize the viscosity model for predicting viscosities at high pressure?

2. Which substances are best suited to parameterize chemical groups, that are transferable to other substances?

3. When parameterizing group-contribution parameters, how many measurements for which substances are optimal for parameterizing a certain functional group or specific „target substance“ for a defined overall experimental effort (i.e. number of experiments)?

For addressing these questions for arbitrary models (i.e. also non-linear models) in one single statistical framework for optimal experimental design, the already well-investigated Preposterior Data Impact Assessor (PreDIA) approach was applied[135].

Any optimal experimental design (OED) approach requires expressing the relevant goal as an objective function. For the acquisition of new experimental data, this objective function will be the expected improvement of the model quality or, likewise, reduction of model uncertainty, due to new data. A Bayesian framework was chosen for the uncertainty analysis, because different information sources can be accounted for (including expert knowledge, assumptions and prior distributions). Within this framework, the improvement of the model is evaluated by comparing the „prior uncertainty“ (before further experiments) in a particular model output with the expected „posterior uncertainty“ (the expected uncertainty after additional measurements at the

proposed conditions were taken into account). A rigorous and linearization-free measure of data impact within a Monte-Carlo framework is used to capture any degree of dependence. Relying on a Monte-Carlo framework allows for a stochastic uncertainty analysis of deterministic models like the proposed model for predicting viscosities. The prior uncertainty in the model parameters is approximated by an ensemble of parameter realizations. Applying the model on the individual realizations allows to transfer the input uncertainty through the model towards the model output uncertainty.

PreDIA applies Bayesian inference by weighted representations of posterior model states, repeatedly applied for potential future measurement values. Reweighting of the model prior allows for effective analysis of expected data impact, independent of the actual future measurement values, only dependent on the proposed experimental design, even for non-linear correlations.

This chapter applies concepts of the preceding chapters and adds another layer of consideration by introducing the optimal experimental design framework (PreDIA) that is based on Bayesian statistics. The reader is challenged by the combination of three approaches:

1. The (group contribution) viscosity model based on the Entropy Scaling approach (Chapter 2 and Chapter 4)

2. The definition of the prior distribution of viscosity parameters applying the Monte Carlo approach (Chapter 3)

3. Non-linear (Bayesian) statistical analysis to estimate the expected impact of future experiments (PreDIA)

6.2 Theoretical Background

The Entropy Scaling approach is used to describe the dependence of the viscosity on pressure and temperature. Both, the pure substance approach as well as the group contribution approach are applied within the PreDIA framework. For more information about the theoretical background of these methods, see chapter 4.2.1 and chapter 2. To express our prior knowledge [†] (i.e. model parameter distributions before additional measurements are considered), the Monte Carlo algorithm introduced in chapter 3.2.3 and chapter 3.3.2 was used. In this chapter, focus will be on the PreDIA method and its application in the optimal experimental design for parametrizing the viscosity model.

6.2.1 Preposterior Data Impact Assessor (PreDIA)

Stochastic optimal experimental design

Data impact is generally the effect of data on a model. This is often referred to as information content of measurement data. PreDIA evaluates this data impact in a Bayesian methodology, meaning, Bayesian Inference is used to incorporate the new data into the model. Bayesian Inference preserves the effect of non-linear models that every future data value (also repetitive measurements at the same conditions, as also the outcome of repeated measurements will vary within the experimental accuracy) differently affects the model and thus leads to a different data impact. Hence, to obtain a reasonable measure for the data impact, all possible measurement outcomes have to be evaluated. PreDIA is designed to perform this task in an efficient way. The goal is to evaluate many different possible posterior states and derive a so-called preposterior state (i.e. a prediction of the posterior state) and the corresponding model uncertainty. Despite the fact that each posterior state is different and depends on the measurement outcome, the expected model uncertainty only depends on the chosen experimental design.

Non-linear Bayesian data impact

As a matter of principle, any sufficiently accurate conditioning method or Bayesian inference scheme can be employed in nonlinear data impact analy-

[†] The expression „prior knowledge“ will be explained more thoroughly later in this chapter

sis. One flexible approach is the Bootstrap filter[221], a nonlinear and accurate conditioning method based on weighting of distributions. The Bootstrap filter reweights a given prior set of realizations based on their goodness of fit to data, defined through the Bayesian likelihood L of realization θ_i to match the measurement values η_0. **Realization** θ_i here means, a set of uncertain **model parameters** of which the uncertainty, in some sense, is to be minimized in this optimal experimental design study.

Gaussian error models are commonly used to simulate noisy data through $\eta_i(d) = f(\theta_i, d) + \varepsilon_i$, where ε_i is a sample drawn from a normal (Gaussian) distribution $\mathcal{N}(0, R_\varepsilon)$ with covariance matrix R_ε and d as an, for now arbitrary, vector containing the design parameters to be investigated. In order to compute the likelihood of a realization θ_i and a given data set η_0, the Bootstrap filter would use the residual vector $\Delta_i = \eta_0 - f_y(\theta_i, d)$ and, for n_η measurement values, evaluate the Likelihood L as

$$L(\theta_i|\eta_0) = \frac{1}{\left((2\pi)^{n_\eta} \det R_\varepsilon\right)^{1/2}} \exp\left[-\frac{1}{2}\left(\Delta_i\right)^T R_\varepsilon^{-1}\left(\Delta_i\right)\right]. \tag{6.1}$$

Normalizing the likelihoods according to Bayes theorem yields the weight vector used in the Bootstrap filter as

$$w_i = \frac{L(\theta_i|\eta_0)}{\sum_{j=1}^{n_\theta} L(\theta_j|\eta_0)} \tag{6.2}$$

which represents one posterior state, based on one particular measurement data set η_0.

By using the re-weighting of realizations to express the posterior distribution given one particular hypothetical measurement data value, the same prior can be used to generate different possible posterior states by using different weights that represent different measurement outcomes. So, the inference of many statistically stable posterior states requires only one sufficiently large prior as a source for fitting realizations. For this reason, repetition for (1) many possible sets of data values per given design and (2) many possible designs in the preposterior analysis is relatively cheap and favors the use of the Bootstrap filter. Averaging over many synthetic posterior states led to the term *preposterior analysis*[222,223] and is the core of nonlinear data impact analysis.

6.2.2 Application of PreDIA in the experimental design of viscosity measurements

The implementation of PreDIA in the planning of viscosity measurements (here, with the specific goal to minimize model uncertainty) basically consists of four parts: Part I (see fig. 6.1) is the definition of the example application, defined by ranges of pressure p, temperature T and substance(s) ζ. Part II introduces experimental and model uncertainties and deals with the computation of the prior model parameter distributions. Part III defines the objective function to be minimized by the optimizer and defines the goal of the optimal experimental design. Part IV in fig. 6.1 illustrates the PreDIA algorithm, forming the core of the optimal experimental design framework.

The following sections will go into more detail about parts I, III and IV. Part II (Uncertainties and prior model parameter distributions) is covered in detail in chapter 3 [‡], only the basic principles will be repeated in the following sections.

6.2.3 Data and model uncertainty and its representation in the prior model parameters

Just as in chapter 3, also in this study of optimal experimental design, we are faced with experimental errors that influence the model (i.e. its parametrization). Besides experimental error, models are always approximate and hence additionally contribute to model uncertainty. In this study, experimental errors in the viscosity, liquid densities and vapor pressures are taken into account. Experimental uncertainties in the viscosity directly influence the viscosity parameters (as captured in vector θ, see chapters 2 and 4) and its distribution $p(\theta)$. The residual entropy s_{res} and Chapman-Enskog viscosity η_{CE} are calculated using the introduced homosegmented *gc* PCP-SAFT EoS and are, hence, subject to experimental errors in the liquid density ρ^{liq} and vapor pressure p^{sat}.

As illustrated in the previous section, (Bayesian) statistical tools like PreDIA work by reweighting prior model parameter distributions in the light of new information. Those parameter distributions have to reflect the mentioned model uncertainties as well as experimental errors. Since we work with a (semi-)

[‡]For data and model uncertainties, see chapters 3.2.3, 3.3.1 and fig. 3.2. For details on the MCMC algorithm, see 3.3.2.

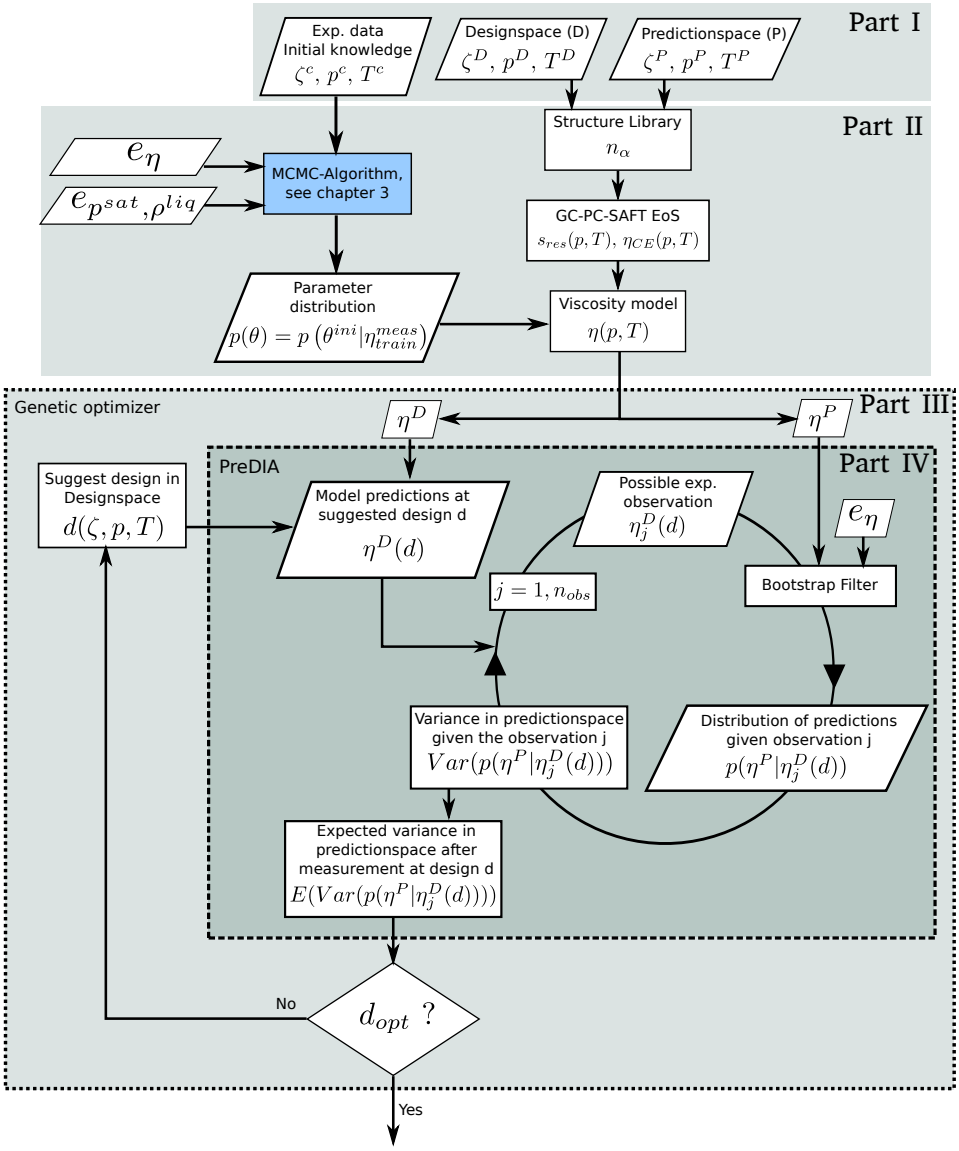


Figure 6.1: Illustration of the implementation of PreDIA and the thermodynamic model for the calculation of the viscosity based on group contributions. The MCMC-Algorithm is explained in more detail in Lötgering-Lin et al.[186] or chapter 3.

empirical model, defining a prior parameter distribution without taking experimental data into account is not appealing. Instead, the Markov Chain Monte Carlo (MCMC) algorithm explained in chapter 3[186] is applied. The so-obtained parameter distributions were earlier (chapter 3) regarded as posterior in the context of the MCMC routine, but these parameter distributions will here serve as the prior for PreDIA.

6.2.4 Combination of PreDIA and the Entropy Scaling approach

Given a sample of N_{MC} realizations of viscosity parameter sets θ, randomly drawn from the prior parameter distribution $p(\theta)$ (i.e. $p(\theta^{ini}|\eta^{meas}_{\text{train}})$, the result from the MCMC step, the distribution of predicted viscosities at any given state point (i.e. substance ζ, pressure p and temperature T, combined in d) $p(\eta(d))$ can be calculated. The Designspace D represents the range of pressures p^D, temperatures T^D and substances ζ^D accessible with a given experimental setup. Furthermore, Predictionspace P is analogously defined by pressure p^P, temperature T^P and substances ζ^P, that is, the target substance, for which the uncertainty of model predictions $\eta^P = \eta(\zeta^P, p^P, T^P)$ (task-driven experimental design[224]) is to be minimized. The aim of this work is to find the design d_{opt}, such that the model uncertainty in the Predictionspace P is minimized.

The objective function is formulated as model uncertainty, i.e.

$$\phi(p(\eta^P|\eta^D(d))) = E\left(\frac{\sqrt{Var(p(\eta^P|\eta^D(d)))}}{E(p(\eta^P))}\right) \tag{6.3}$$

where $E(...)$ denotes the expected value and the term $Var(p(\eta^P|\eta^D(d)))$ describes the model uncertainty in the Predictionspace after taking into account the viscosities at the proposed design d. While at this point, the prior model uncertainty is known, the model uncertainty after taking into account the viscosities at the proposed design is not known yet.

If measurements at the design d would be carried out (for discussion, the „not yet existent“ possible measurements are referred to as η_0), the real model uncertainty in the Predictionspace could be approached in two different ways:

1. Repetition of the complete MCMC procedure to obtain a new parameter distribution $p(\theta|\eta_0)$, which then could be used to calculate the variance in the Predictionspace. This path would result in the most accurate description of the model uncertainty but also comes with the highest computational effort. In this case, we would speak of the *posterior* model uncertainty.

2. Recomputation of the weights (i.e. probability) w_i of each of the N_{MC} parameter sets θ_i drawn from the prior parameter distribution $p(\theta)$ in the light of the new information $\eta_0(d)$ by applying the Bootstrap Filter

as explained before (eq. 6.1 and eq. 6.2), as

$$w_i = \frac{L(\theta_i|\eta_0(d))}{\sum_{j=1}^{N_{MC}} L(\theta_j|\eta_0(d))} \tag{6.4}$$

These weights can then be used to estimate the variance of the model in the Predictionspace given the knowledge of that specific measurement value(s) η_0, $Var(p(\eta^P|\eta_0(d)))$. As this model uncertainty stems from reweighting of a prior distribution based on the new measurement, we here speak of the *weighted posterior* model uncertainty. If the used prior is representative, the weighted posterior model uncertainty should be a good approximation of the posterior model uncertainty described above.

Obviously, the goal here is to approximate the above mentioned *(weighted) posterior* model uncertainty after potential measurements at design d without in fact knowing the measurement outcome. The expected model uncertainty (and hence an indicator for the impact of measurements at design d) is called the *preposterior* model uncertainty. It is obtained by generating samples for different possible measurement outcomes η_0 and weighting them through eq. 6.4.

In non-linear data impact analysis, the specific measurement value $\eta_0(d)$ is not known but a distribution of predicted viscosities at a fixed design d, $p(\eta^D(d))$. To obtain the *preposterior* model uncertainty in the Predictionspace, we draw two random samples: One sample represents a set of N_{obs} potential measurement outcomes η_0 (observations), drawn from $p(\eta^D(d))$ [§]. The second is a sample of N_{MC} realizations (i.e. parameter sets) drawn from the prior parameter distribution $p(\theta)$ that will be used as a representative sample of the models predictions to determine the model uncertainty in the Predictionspace. The Bootstrap filter is applied to obtain the weights of each of the N_{MC} realizations, successively assuming each of the N_{obs} potential measurement outcomes. That means, each of the N_{MC} parameter sets can be assigned with a weight w_i, assuming any one of the N_{obs} observations to be the (potentially) real outcome of a measurement η_0 in eq. 6.4. Hence, a $N_{MC} \times N_{obs}$ weight matrix W is

[§]A design is not linked to a specific measurement value but to many different potential measurement outcomes. The probability of each potential measurement outcome is given by $p(\eta^D(d))$, determined by evaluating the model in the Designspace with the prior parameter distribution $p(\theta)$ (i.e. model is assumed to be the best current estimator).

obtained. The weight matrix W is used to calculate the expected conditional prediction variance as

$$Var(p(\eta^P|\eta^D(d))) \approx \frac{1}{N_{obs}} \sum_{j=1}^{N_{obs}} \frac{\nu_{1,j}}{\nu_{1,j}^2 - \nu_{2,j}} \left\{ \sum_{i=1}^{N_{MC}} (\eta_i^P)^2 W_{ij} - \left(\sum_{i=1}^{N_{MC}} \eta_i^P W_{ij} \right)^2 \right\} \tag{6.5}$$

with $\nu_{1,j} = \sum_{i=1}^{N_{MC}} w_{ij}$ and $\nu_{2,j} = \sum_{i=1}^{N_{MC}} w_{ij}^2$

This expected conditional prediction variance in the Predictionspace (pre-posterior) can now be used in eq. 6.3 to estimate the model uncertainty after a (hypothetical) measurement at design d was conducted. The goal of the optimization scheme then is to find the optimal design d_{opt}, such that the objective function is minimized, as

$$d_{opt} = \underset{d \in D}{arg\ min}\{\phi(d)\} \tag{6.6}$$

6.2.5 Investigated example applications

Before PreDIA can be applied, three ingredients always have to be provided. First, the training data points $\eta_{\text{train}}^{meas}$ that are used in the MCMC scheme to obtain a reasonable prior knowledge (training of the model) have to be selected. Second, the Designspace D in which the experiments are allowed to be carried out has to be defined. Third, the Predictionspace P that is defined by range of conditions in which the objective function is evaluated has to be chosen. For each ingredient the substance(s) ζ, pressures p and temperatures T must be defined.

Within this study, two example applications (ExA) were considered. The first example application, ExA^{press}, assumes having measured three viscosities at low pressures (0.1 MPa) of one pure component, n-octane, shown in fig. 6.2 (black diamonds). This defines our initial knowledge that will be used in the MCMC scheme to generate the prior model parameter distributions. This is the only step where actual experimental data is needed during the optimal experimental design procedure. Starting from this point, it is desired to optimize the conditions (pressure and temperature within Designspace D, see fig. 6.2) at which additional measurements would be most meaningful such that the expected uncertainty of the predictive model within a specific (in this case nearly equally wide) range of pressure and temperature (Predictionspace P,

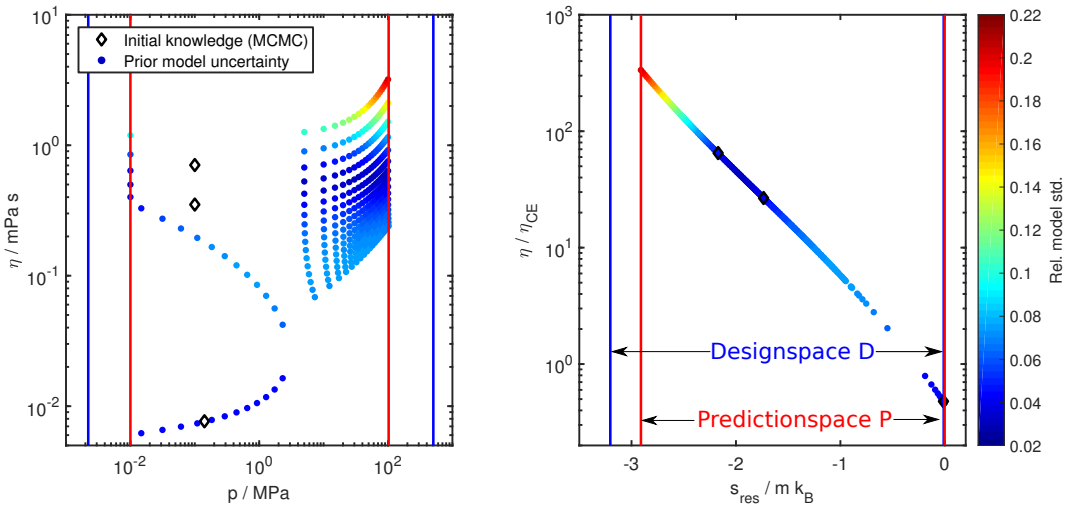


Figure 6.2: Initial experimental datapoints of n-octane used in the MCMC step of example application ExA^{press} (diamonds). Range of conditions in Design- (blue line) and Predictionspace (red line). Model uncertainty, based on the prior knowledge obtained from the MCMC step, is represented by the colored circles as relative standard deviation.

see fig. 6.2) is minimized. To be able to verify the results, pressure and temperature pairs in the Designspace D are chosen on the basis of conditions of experimental data from the Dortmund Datenbank[153]. The Predictionspace P is chosen as an equidistant, linear grid in the temperature and pressure. This linear grid ensures that the optimization result is not biased by accumulation of data points in a certain state point region. In this example application, for both, the MCMC step as well as the Bootstrap filter, an experimental error in the viscosity of $e_\eta = 4.5\%$ is assumed.

In the second example application, ExA^{gc}, the aim is to reduce the model uncertainty of a target substance by using experimental data of other substances of similar structure (i.e. chemical families with shared functional groups). This example application analyses the transferability of gc parameters and addresses the question which measurements are most valuable for ensuring sensitivity on the target substance. As a target substance, defining the Predictionspace P, a long, strongly branched alkane is chosen: $2,2,4,4,6,8,8$-heptamethylnonane. As Designspace D and initial datapoints for the MCMC step, n-alkanes and branched alkanes with only few „sidegroups“ were selected, as can be seen in fig. 6.3 and table 6.1. Higher expected errors of $e_\eta = 10\%$ were assumed as the accuracy of the model in predicting viscosities of branched alkanes signif-

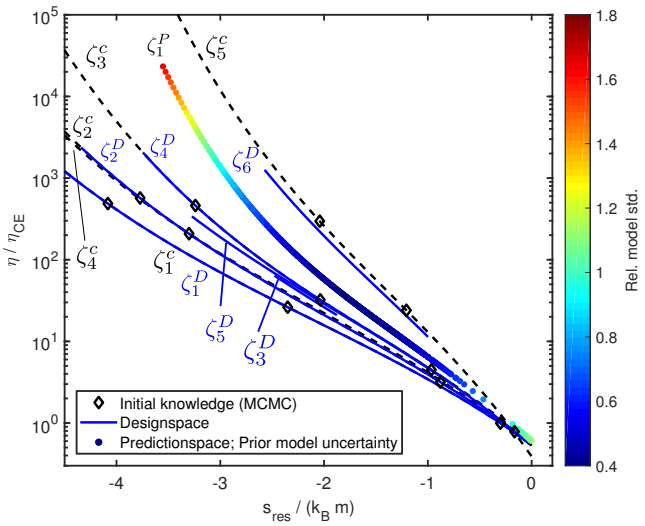


Figure 6.3: Initial experimental datapoints used in the MCMC step of example application ExA^{gc} (diamonds), see table 6.1, together with black dashed lines representing the $\eta(s_{\text{res}})$ relation of the corresponding substance. Range of conditions in Design- (blue lines) and model uncertainty in the Predictionspace, based on the prior knowledge obtained from the MCMC step (colored circles) as relative standard deviation.

icantly increases as compared to linear n-alkanes[48](see chapter 2) and these, less accurate predicted viscosities serve as potential measurement outcomes in the PreDIA approach.

In both example applications, shown in fig. 6.2 and fig. 6.3, the coloured circles represent the grid points in Predictionspace P. The color indicates the model's prior uncertainty, given as relative standard deviation of model predictions, analogous to eq. 6.3 given as

$$\phi(p(\eta^P)) = E\left(\frac{\sqrt{Var(p(\eta^P))}}{E(p(\eta^P))}\right) \tag{6.7}$$

$$= \frac{1}{N_P}\sum_{i=1}^{N_P}\frac{\sqrt{Var(p(\eta_i{}^P))}}{\frac{1}{N_{MC}}\sum_{j=1}^{N_{MC}}\eta_{ij}^P} \tag{6.8}$$

In both example applications, the peaks of model uncertainty occur in the liquid region at high densities.

As optimizer a genetic algorithm (GA)[225] available in MatLAB R2014b

Table 6.1: Substances and range of conditions used in ExA^{gc} in the MCMC step (initial knowledge), Designspace D and Predictionspace P as shown in fig. 6.3.

	Substance	Abbreviation	p-range / MPa		T-range / K	
			p_{min}	p_{max}	T_{min}	T_{max}
Initial knowledge	Isobutane	ζ_1^c	3.19	31.45	140.00	498.15
	Isopentane	ζ_2^c	4.90	1176.80	348.15	453.15
	2,2,4-trimethylpentane	ζ_3^c	3.92	501.40	323.15	523.15
	n-pentane	ζ_4^c	2.52	490.33	293.15	453.15
	n-octadecane	ζ_5^c	2.55	40.83	323.15	473.15
Designspace	Isobutane	ζ_1^D	0.02	800.00	120.00	548.15
	Isopentane	ζ_2^D	0.02	1176.80	253.15	523.15
	2,3-dimethylbutane	ζ_3^D	0.10	147.10	313.15	313.15
	2,2,4-trimethylpentane	ζ_4^D	0.10	503.10	197.93	548.15
	n-hexane	ζ_5^D	0.06	500.00	178.15	623.15
	n-hexadecane	ζ_6^D	0.10	425.10	318.15	531.95
Predictionspace	2,2,4,4,6,8,8-heptamethylnonane	ζ_1^P	0.01	101.74	240.00	740.00

was applied. A genetic algorithm is chosen in this case due to its simple integration when dealing with mixed-integer problems. The GA optimizer was applied with default numerical parameters. It is noteworthy, however, that the results generated by heuristic methods, such as GA, can strongly depend on the chosen optimizer parameters.

6.3 Results & Discussion

6.3.1 Single substance optimization

First, results of the ExA^{press} example application will be discussed, where one single pure substance serves as Design- and Predictionspace, see fig. 6.2. Results shown for this example application are based on $N_{MC} = 50000$ realizations and $N_{obs} = 2000$ observations (i.e. hypothetical measurement outcomes). For more information on convergence behaviour and sample-sizes, also see SI D.1 and SI D.3.1.

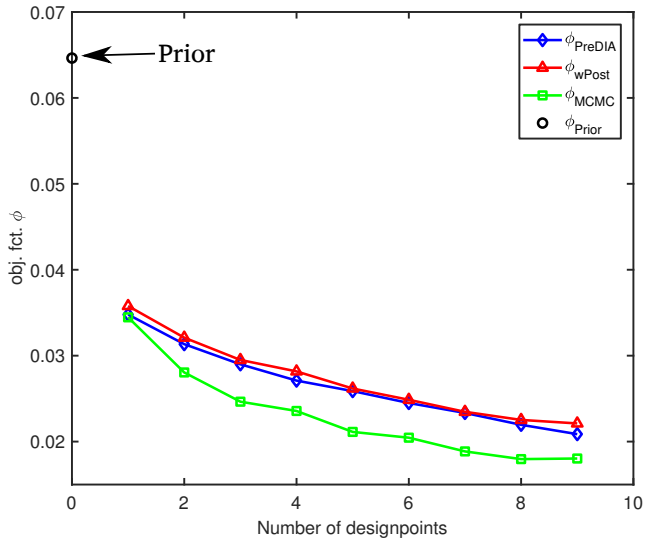


Figure 6.4: ExA^{press}: Objective function ϕ as defined in eq. 6.3 obtained at different number of simultaneously optimized design points. The black circle represents the prior model uncertainty. Blue diamonds represent the estimated objective function ϕ_{PreDIA} according to eq. 6.3. Red triangles are obtained by reweighting (see chapter 6.2.4) the prior parameter distribution with experimental data at the suggested design points, giving the actual weighted posterior objective function ϕ_{wPost}. Green squares are the posterior result from the MCMC algorithm using the extended training data set $\eta^{meas}_{\text{train}+d_{opt}}$ (see chapter 6.2.4).

Fig. 6.4 shows the expected value of the objective function $\phi_{PreDIA}(d)$ (see eq. 6.3) of calculations with increasing number of design points applying the PreDIA approach. As reference, the black circle represents the prior model uncertainty. As expected, the objective function decreases with increasing number of design-points; further datapoints (better knowledge) decrease the model variance. The reduction of expected model uncertainty with each addi-

tional data point declines. The red triangles and green squares in fig. 6.4 are only accessible because many experimental values for n-octane exist that can be used for evaluation of PreDIA. Not applying these data points in the first place only serves the purpose of mimicking a practical application of PreDIA. Both, the red triangles as well as the blue squares, will be discussed more thoroughly in later paragraphs of this chapter.

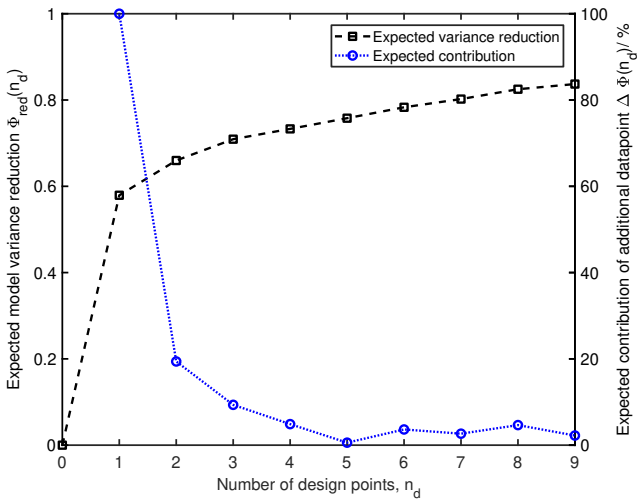


Figure 6.5: Expected reduction of relative model variance (black circles) in the Predictionspace and the expected contribution of each additional design point to the overall variance-reduction.

To illustrate the expected benefit of additional measurements, fig. 6.5 shows the expected reduction of relative variance in the Predictionspace with increasing number of hypothetically added experimental data points. The expected reduction of relative model variance, $\Phi_{red}(n_d)$, is formulated as an average over the gridpoints in the Predictionspace as

$$\Phi_{red}(n_d) = E\left(\frac{Var(p(\eta^P)) - Var(p(\eta^P|\eta^D(d(n_d))))}{Var(p(\eta^P))}\right) \tag{6.9}$$

$$= \frac{1}{N_P}\sum_{i=1}^{N_P}\frac{Var(p(\eta_i{}^P)) - Var(p(\eta_i{}^P|\eta^D(d(n_d))))}{Var(p(\eta_i{}^P))} \tag{6.10}$$

where N_P is the number of gridpoints in the Predictionspace P. Adding a single additional data point decreases the model variance drastically by approximately 60%. The further reduction of the model variance by adding more data points decreases until the reduction of model variance approximately scales lin-

early with the number of additional data points. For nine added data points the variance is expected to be lowered by 80%.

The expected contribution (blue circles, right axis) of each additional data point illustrates that behaviour more clearly, given by

$$\Delta\Phi(n_d) = E\left(\frac{Var(p(\eta^P|\eta^D(d(n_d-1)))) - Var(p(\eta^P|\eta^D(d(n_d))))}{Var(p(\eta^P)) - Var(p(\eta^P|\eta^D(d(n_d))))}\right) \quad (6.11)$$

$$= \frac{1}{N_P}\sum_{i=1}^{N_P}\frac{Var(p({\eta_i}^P|\eta^D(d(n_d-1)))) - Var(p({\eta_i}^P|\eta^D(d(n_d))))}{Var(p({\eta_i}^P)) - Var(p({\eta_i}^P|\eta^D(d(n_d))))} \quad (6.12)$$

Fig. 6.5 shows that after three additional measurements the contribution of further data points is nearly constant, corresponding to the linear increase in the expected reduction of model variance. This means, already after six experimental viscosities (three experimental viscosities used as „initial knowledge“, three additional experimental viscosities to be measured), the viscosity model is well parametrized. I continue by considering the results for three additional, simultaneously optimized design points in more detail in the remainder of the chapter.

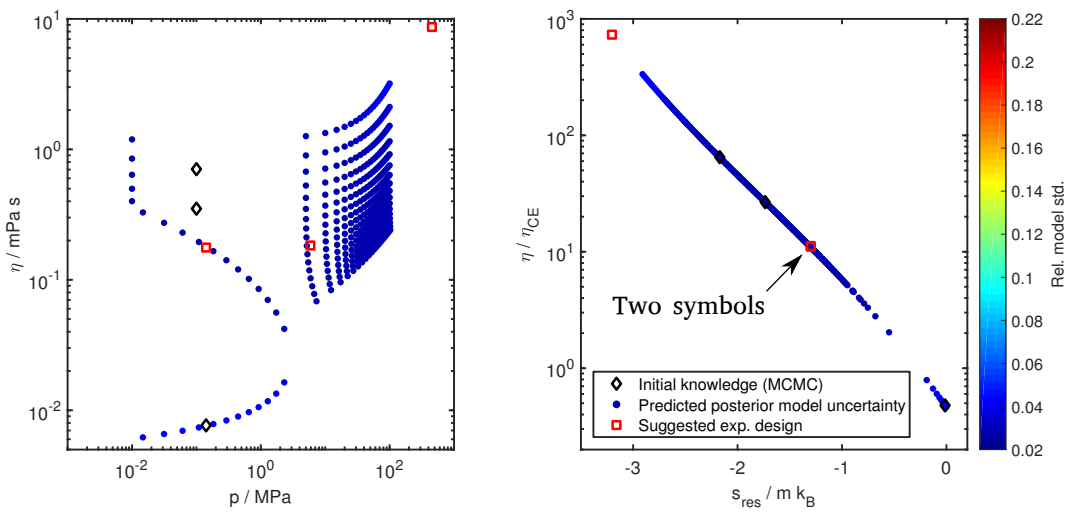


Figure 6.6: Coloured dots represent the expected, preposterior model uncertainty given by eq. 6.3 with three additional optimally chosen design points (red squares). Black diamonds represent the initial knowledge used in the MCMC step.

In fig. 6.6 the red squares present the next three design points in example application ExA^{press} that are expected to be optimal for reducing model un-

certainty. The black diamonds are the initial experimental data points used in the MCMC step as already illustrated in fig. 6.2. Carrying out three additional measurements at conditions marked by the red squares is expected to result in a model uncertainty as indicated by the coloured dots. Comparing fig. 6.2 and fig. 6.6 the decrease of model uncertainty is most evident at high pressures. But, simultaneously, also regions in the Predictionspace of already low model uncertainty are expected to be further improved, resulting in overall lower model uncertainty. Interestingly and non-intuitively, the two suggested experimental conditions at medium temperature are chosen at pressures that end up at nearly the same residual entropy.

To evaluate the accuracy of the predicted reduction of model variance, the posterior model objective function ϕ_{MCMC} is presented in fig. 6.4 as green squares. ϕ_{MCMC} (green squares) was obtained by running the MCMC step with an updated training data set $\eta^{meas}_{\text{train}+d_{opt}}$ according to the proposed optimal experimental design. Hence, $\eta^{meas}_{\text{train}+d_{opt}}$ consists of the initially used experimental data $\eta^{meas}_{\text{train}}$ and actual experimental values at the proposed (optimal) experimental conditions $\eta^{meas}(d_{opt})$. The MCMC step then provides a new prior parameter distribution $p(\theta^{d_{opt}}) = p(\theta^{ini}|\eta^{meas}_{\text{train}+d_{opt}})$. By applying $p(\theta^{d_{opt}})$ to the model and evaluation in the Predictionspace, the objective function ϕ_{MCMC} (green squares) can be directly calculated. Hence, ϕ_{MCMC} (green squares) is a measure for the „real“ posterior model uncertainty and thus serves as an evaluation criteria for the PreDIA approach. We note that PreDIA is avoiding the re-run of the computationally demanding MCMC step and tries to estimate $p(\theta^{ini}|\eta^{meas}_{\text{train}+d_{opt}})$ predictively.

The predicted (preposterior) objective function ϕ_{PreDIA} (blue diamonds) can now be directly compared with the MCMC (posterior) outcome ϕ_{MCMC} (green squares). PreDIA accurately predicts the trend (especially for the first additional measurement). Though there is an offset between the predicted objective function ϕ_{PreDIA} (blue diamonds) and the „real“ objective function values ϕ_{MCMC} (green squares), which increases with increasing number of design-points. One has to bear in mind the difference of knowledge underlying the results illustrated by the blue diamonds (ϕ_{PreDIA}) and the green squares (ϕ_{MCMC}). ϕ_{PreDIA} (blue diamonds) was calculated using the prior parameter distribution $p(\theta)$ that has been conditioned only on the initial knowledge and was averaged over N_{obs} possible, hypothetical measurement outcomes. In contrast to that, for ϕ_{MCMC} (green squares) the experimental values of the initial

knowledge and at the suggested, optimal experimental conditions were used.

As a better measure for the accuracy of PreDIA, fig. 6.4 also shows the weighted posterior objective function ϕ_{wPost} as red triangles. To obtain ϕ_{wPost} (red triangles), again PreDIA was applied, though, the Bootstrap filter does not average over the N_{obs} hypothetically possible measurement values but conditions (re-weights) the prior parameter distribution $p(\theta)$ on the same knowledge as in ϕ_{MCMC} (i.e. actual experimental values at the suggested optimal conditions). Hence, basically, eq. 6.5 is applied with only one observation, which is the real experimental value.

In this example application (ExA^{press}), the additional knowledge used in PreDIA to obtain ϕ_{wPost} (red triangles) does not change the objective function significantly compared to ϕ_{PreDIA} (blue diamonds). This means, even with knowledge of the experimental data within the PreDIA framework, the „real“ model uncertainty, calculated by re-running the MCMC procedure, is not represented more accurately. Obviously, PreDIA can not be expected to give a better estimation of the posterior state when averaging over hypothetical measurement outcomes (ϕ_{PreDIA}, blue diamonds) than when indeed using the experimental data (ϕ_{wPost}, red triangles). Hence, in this example application, PreDIA gives a satisfyingly accurate, purely predictive estimation of the impact of experimental data on model uncertainty.

Overall, the results illustrated by the three symbol types in fig. 6.4, blue diamonds, red triangles and green squares, can be summarized as follows: The PreDIA result (blue diamonds) is based only on three experimental data points and is based on two assumptions: First, that the prior parameter distribution $p(\theta)$ is also representative when additional (hypothetical) experiments are considered. Second, that PreDIA correctly predicts the posterior parameter distribution based on hypothetical measurements. The red triangles use actual experimental data to assess the second point, namely the predictive power of PreDIA to predict the posterior distribution. The good agreement of blue to red symbols confirms that the posterior is indeed well predicted. The assumption that the prior, i.e. the parameter distribution $p(\theta)$ obtained for three experimental data points, is transferable to cases, where additional measurements are available is assessed by comparing red to green symbols. The comparison reveals that, although quantitative discrepancies result from assuming $p(\theta)$ as unchanged, the qualitative results are very well captured.

6.3.2 Optimized parameterization of group contributions

This second example application, ExA^{gc}, is significantly more demanding. The number of adjustable parameters is twice the number of adjustable parameters in the previous example application. More parameters require more realizations for obtaining a representative distribution of viscosity predictions. For the description of the prior state, $N_{MC} = 100000$ realizations were used in this example application. The distributions of the group contribution parameters of the n-alkanes and branched alkanes are, in this chapter, obtained via the MCMC approach and, hence, adjusted simultaneously, not in successive order [¶]. Therefore, the parameter distributions are convoluted and not independent of each other. Additional datapoints in, e.g., a branched alkane, shift the weighting of the objective function towards the branched alkanes and, hence, lead to a better representation of the branched alkanes, but might lead to a worse representation of the n-alkanes.

Fig. 6.7 presents the objective function $\phi(d)$ for the second example application ExA^{gc} with d representing a proposed design (defined by pressures, temperatures and substance(s)). Blue diamonds represent the expected, preposterior reduction of model variance. As a reference, the prior model variance (i.e. without additional measurements) is shown as a black circle. For this example application, a maximum of 4 optimal conditions for measurements (designs) were considered, because predictions are less reliable than in the first example application ExA^{press} due to the increased problem complexity. As expected, the preposterior objective function (and hence the expected model uncertainty) declines with increasing number of additional datapoints. The expected impact is highest for the first additional datapoint and declines with further additional datapoints. This behavior can also be seen in fig. 6.8. Surprisingly, the suggested design consisting of four experimental conditions results in an increased expected preposterior model variance as opposed to three additional experiments. This is non-physical and shows the limits of the proposed approach. The additional design-points were not obtained sequentially but simultaneously, which means, the 4 optimized design points can be completely different from the proposed 3 optimal design-points. As

[¶]For parametrizing the group contribution model for viscosity in chapter 2, the model parameters for n-alkanes and branched alkanes were adjusted successively, i.e. the n-alkane viscosity parameters were fixed during the adjustment of the group contribution viscosity parameters of the branched alkanes.

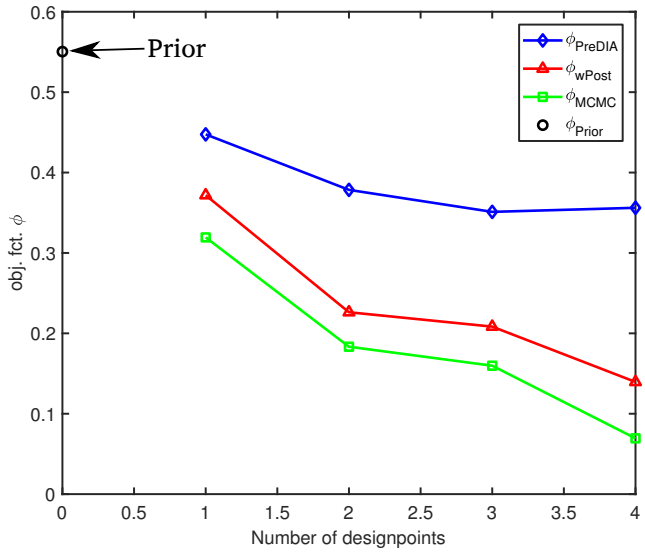


Figure 6.7: Objective function ϕ as defined in eq. 6.3 obtained at different number of simultaneously optimized design points in the example application ExA^{gc}. The black circle represents the prior model uncertainty. Blue diamonds represent the expected objective function. Red triangles are obtained by reweighting the prior parameter distribution with actual experimental data at the suggested design points. Green squares are the result from the MCMC algorithm using the extended training data set $\eta^{meas}_{\text{train}+d_{opt}}$.

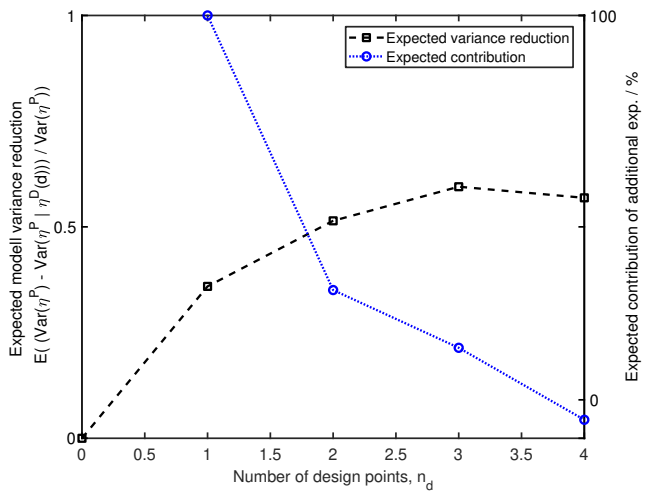


Figure 6.8: Expected reduction of relative model variance (black circles) and the expected contribution of each additional design point to the overall reduction of model variance.

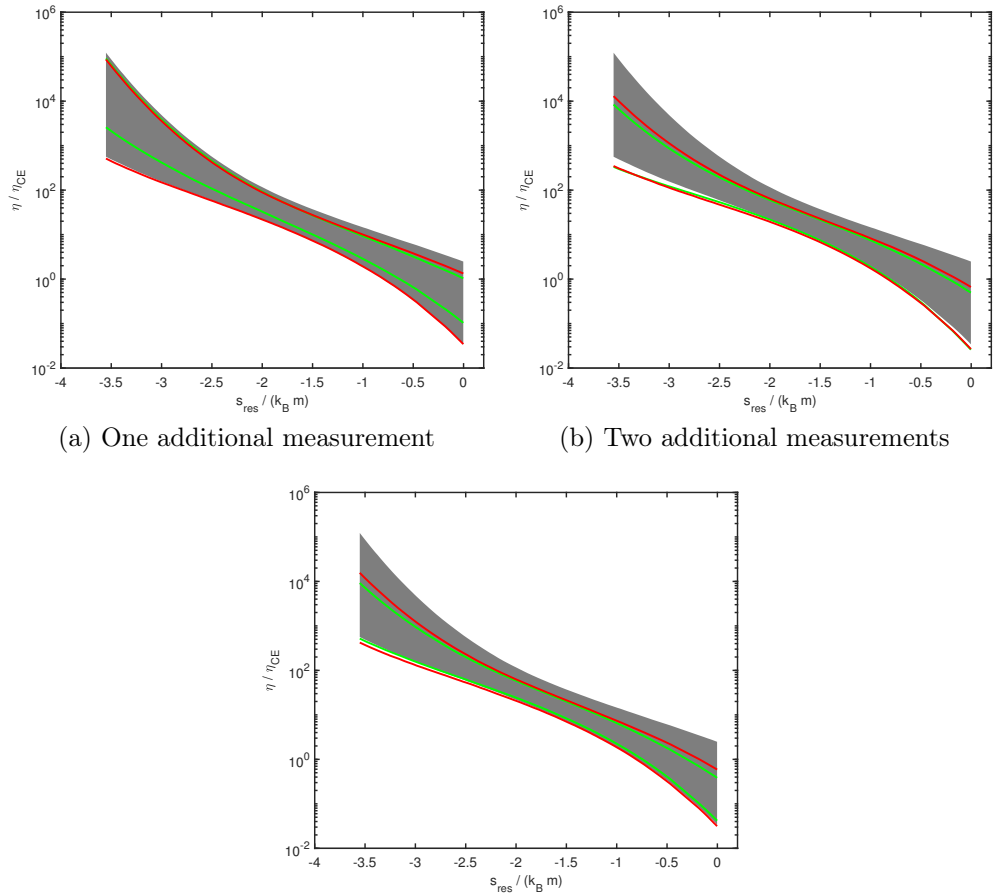
mentioned before, in this study a heuristic optimizer is applied which is not guaranteed to always result in the global optimum. The less strongly reduced variance for 4 design-points, compared to 3 design-points can either be caused

by a local optimum (design-points were not obtained sequentially) or due to predictions being less and less reliable with increasing number of simultaneously optimized design-points. An evaluation of sample-sizes for this example application, showing that sample-sizes are chosen sufficiently large, can be found in the SI D.3.2 The results suggest a recalibration of the model (repetition of the MCMC step, part II in fig. 6.1) after the first three design-points and then re-investigate if and where additional measurements should be conducted. Hence, within this study, the results of three additional, optimally chosen design-points are investigated further.

Fig. 6.7 also shows the objective functions ϕ_{wPost} (red triangles) and ϕ_{MCMC} (green squares). The diagram allows to evaluate the quality of PreDIA's predictions for the model variance. For both, ϕ_{wPost} (red triangles) and ϕ_{PreDIA} (blue diamonds), the PreDIA approach (i.e. reweighting of a prior parameter distribution based on new knowledge) is applied. For ϕ_{PreDIA} (blue diamonds) the expected impact of a measurement is obtained by averaging over hypothetical measurement outcomes, whereas for ϕ_{wPost} (red triangles) the prior parameter distribution is reweighted based on the actual measurement outcome (i.e. evaluating the PreDIA approach after the measurement has been conducted). Therefore, ϕ_{wPost} (red triangles) serves as a reference about how predictive PreDIA could perform if it had knowledge of the measurement outcome. In contrast, ϕ_{PreDIA} (blue diamonds) is a complete prediction of the impact of measurements. In this example application, ExA^{gc}, ϕ_{PreDIA} (blue diamonds) and ϕ_{wPost} (red triangles) are not in close agreement. Qualitatively, the trend is predicted correctly by PreDIA, but especially the impact of two measurements is underestimated significantly. The non-physical increase of model uncertainty for four additional measurements compared to three measurements is not confirmed by ϕ_{wPost} (red triangles) or ϕ_{MCMC} (green squares). That supports the above described interpretation of the trend of ϕ_{PreDIA} (blue diamonds), resulting in the suggestion to recalibrate (generate a new prior) after three measurements. Another possible explanation is a local optimum.

Overall, additional datapoints beyond the first additional measurement shift the viscosity in the Predictionspace to lower values with more narrow confidence intervals, as can be seen in fig. 6.9 when comparing the prior distributions (grey area) with the posterior (green curves) and weighted posterior distributions (red curves) [‖]. Both effects, the decrease of absolute viscosity values and

[‖] Curves represent the upper and lower bounds of the 95% confidence intervals



(a) One additional measurement

(b) Two additional measurements

(c) Three additional measurements

Figure 6.9: 95% confidence intervals in the Predictionspace of example application ExA^{gc}. Grey area represents the prior distribution, green curves the posterior obtained by another MCMC step and red curves the weighted posterior distribution.

the narrowing of the confidence intervals, result in lower values of the objective function as defined in eq. 6.3. Especially for two additional measurements (fig. 6.9b), the lower bounds of both the posterior (green curves) and weighted posterior distributions (red curves) are below (i.e. outside) the prior distribution (grey area). PreDIA, in simplified terms, mainly reweighs the prior distribution. The fact that PreDIA can hardly shift the expected viscosity distributions to viscosities that significantly lower than the prior distributions might also explain why the results of PreDIA, ϕ_{PreDIA} (blue diamonds in fig. 6.7), diverge from ϕ_{MCMC} and ϕ_{wPost} (green squares and red triangles in fig. 6.7, respectively).

The following fig. 6.10 and table 6.2 present the proposed next three condi-

tions to carry out measurements. It turns out that up to the optimization of three additional design-points, the only substance to carry out measurements for is 2,2,4-trimethylpentane. PreDIA identified 2,2,4-trimethylpentane as an important substance, because it contains three of the four functional groups the target substance (i.e., Predictionspace) consists of: It is the only substance in the Designspace that contains a quaternary carbon. It also contains a tertiary and two primary carbons. The results from PreDIA show that the applied prior (i.e. the initial parameter distribution based on the defined initial knowledge, see 6.1) provides enough information regarding the secondary carbon. With four optimized experimental conditions, also other substances, namely isobutane and n-hexane, are suggested by PreDIA.

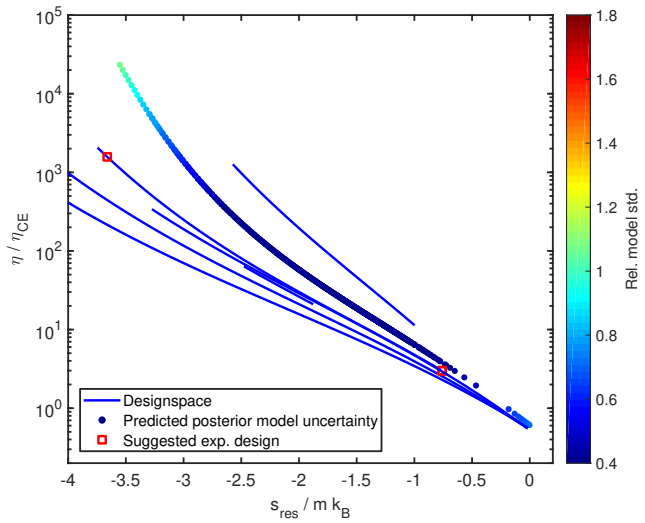


Figure 6.10: Coloured dots represent the expected, preposterior model uncertainty given by eq. 6.3 with three additional „optimally chosen“ design points (red squares). Black diamonds represent the initial knowledge used in the MCMC step.

6.4 Summary and Conclusion

This chapter investigates the PreDIA approach for optimizing thermodynamic properties. The Monte-Carlo approach from chapter 3 is applied to determine prior parameter distributions whereby the initial knowledge is defined by a set of a priori known experimental values from literature.

The goal of this study was to determine optimal experimental conditions in a defined Designspace for decreasing model uncertainty in a defined Predic-

Table 6.2: PreDIA-Optimal conditions for the next up to 4 experiments in the example application ExA^{gc}.

Substance	Temperature T / K	Pressure p / MPa	Residual entropy s_{res} / -
One additional measurement			
2,2,4-trimethylpentane	544.25	2.50	-0.76
Two additional measurement			
2,2,4-trimethylpentane	298.14	499.40	-3.75
2,2,4-trimethylpentane	544.25	2.50	-0.76
Three additional measurement			
2,2,4-trimethylpentane	197.93	70.10	-3.66
2,2,4-trimethylpentane	544.25	2.50	-0.76
2,2,4-trimethylpentane	544.25	2.50	-0.76
Four additional measurement			
Isobutane	468.15	0.10	-0.00
2,2,4-trimethylpentane	298.14	499.40	-3.75
2,2,4-trimethylpentane	543.15	2.46	-0.77
n-hexane	323.00	44.66	-2.00

tionspace. Two example applications of different complexity were investigated. In the first example application, ExA^{press}, measurements at higher pressures were planned based on viscosities at ambient pressure in the same substance, here, n-octane. Up to nine additional measurements were optimized. Multiple experimental conditions were not optimized sequentially but simultaneously. It was shown that the PreDIA approach is predictive beyond the initial knowledge and is able to accurately estimate the impact of additional measurements on model uncertainty. The accuracy was evaluated by comparing the reduction of predicted model variance (preposterior) with posterior parameter distributions obtained by reweighting (weighted posterior). Because comprehensive experimental data actually exists for n-octane, it is possible to rather rigorously assess the predictions from PreDIA. That was done by repeating the MCMC procedure (posterior) based on results of the measurements at the optimized conditions. The comparison of preposterior results and weighted posterior results showed that PreDIA predicts the impact of measurements generally well. A more accurate description of data impact by the preposterior as compared to the weighted posterior can not be expected. However, comparing the weighted posterior and the posterior obtained by a MCMC procedure revealed an offset. This indicates that certain effects are not captured by the predictive PreDIA method. A possible approach to alleviate this issue could be to determine a „broader“ prior viscosity parameter distribution by allowing higher uncertainties in the MCMC step than in the actual optimal experimental design procedure.

In the second example application, the presented group contribution viscos-

ity approach (chapter 2) was applied to plan experiments that reduce the model variance in a target substance (2, 2, 4, 4, 6, 8, 8-heptamethylnonane) by carrying out measurements in other substances sharing functional groups, namely primary, secondary, tertiary and quaternary carbons embedded in n-alkanes and branched alkanes. Up to four optimal experimental conditions were determined. In this example application, PreDIA was not able to predict the data impact with high quantitative agreement, especially with increasing number of planned experiments. Although, the trend was captured well for up to three additional measurements. A comparison of weighted posterior and of posterior results showed that the offset observed in the first example application can also be seen in this example application, again, indicating a prior parameter distribution that might not fully cover all effects of additional measurements. For instance, it was shown that the reduction of the objective function by PreDIA is mainly obtained by reducing the model uncertainty, while in fact measurements in this case also lead to a shift of the confidence intervals to lower viscosities, which also leads to lower values of the objective function.

Further studies could include

- an optimization study that is not limited by a predefined set of substances as Designspace but uses a relaxed space of molecular parameters. Composing the optimal substance for carrying out the measurements is then part of the optimization problem (as similarly done by e.g. Bardow et al.[226]).

- the determination of optimal pure substance measurements to decrease model uncertainty regarding the prediction of mixture viscosities.

- the determination of optimal measurement conditions with the goal to discriminate different property models. As shown in chapter 3, models can be ranked to decide which model is best suited for different applications. PreDIA could be used to determine experimental conditions at which differences in the models are expected to be more significant. Another interesting decision making process is the molecular design of e.g. solvents in process optimization[13,140,141,217]. Experiments could be planned to maximize differences in the objective function upon which the most suited molecule is identified.

7. Conclusion

This work proposes a new model to correlate and a new model to predict the viscosity of pure substances and mixtures via the entropy scaling approach. The model needs three parameters per substance if each substance is parametrized individually. By applying a group contribution approach to predict pure substance parameters from functional groups, that number is reduced to three adjustable parameters per functional group. Overall 140 substances were investigated. Applying a third-order polynomial ansatz function for describing $\ln(\eta^*)$ as a function of residual entropy s_{res}, Entropy Scaling was shown to be a reliable and accurate framework for the description of pure substance viscosities with an average relative deviation of 5.21% using individually adjusted parameters. Applying the group contribution approach, relative deviations are only slightly higher for most chemical families. Only 1-alcohols show a significantly higher relative deviation of 10.98% applying the group contribution approach as compared to 6.81% with individually adjusted parameters.

For non-associating and weakly self-associating substances (e.g., amines) the dimensionless viscosity $\ln(\eta^*)$ shows a nearly linear relation to the residual entropy s_{res}. For strongly self-associating substances (e.g., alcohols), the second- and third-order terms of the polynomial become more important, i.e. the relation between $\ln(\eta^*)$ and s_{res} shows a more pronounced curvature and explains why deviations using the group contribution approach for 1-alcohols are higher. Further systematic studies of associating substances, possibly taking the degree of dimerization (accessible from SAFT-type models) into account, might be a further step for improving the description of associating substances.

Beyond accuracy as a sole measure for the quality of the model, Bayesian Model Selection (BMS) provided information about the robustness of the proposed model regarding extrapolations in pressure and predictions for substances excluded from the training set. Bayesian Model Selection was shown

to be a helpful tool, especially for developing or selecting empirical models. Besides the examples shown in chapter 3, BMS is broadly applicable to assess and compare (physical property) models. It does not imply any assumption about model error distributions and is applicable to linear as well as non-linear models. The MCMC sampling applied in this work for generating prior knowledge, i.e. parameter distributions based on the training data set, leads to an increased computational effort. The resulting parameter distributions, however, are valuable for further statistical analysis, e.g. for optimal experimental design studies or data analysis.

The optimal experimental design study presented as part of this thesis is one example of how these parameter distributions can subsequently be used. In two examples, the application of the Preposterior Data Impact Assessor (PreDIA), an approach based on Bayesian Statistics for obtaining optimal experimental conditions, for thermodynamic modeling is presented and evaluated. Besides the direct and obvious benefit of optimal conditions, in this case to reduce model uncertainty, the results also support the robustness and charming simplicity of the proposed model: Both, the model with parameters individually adjusted for each substance and the group contribution model require only few data points to obtain reliable predictions of pure substance viscosities.

An extension of the proposed model for mixtures was developed. The extension does not introduce any further adjustable parameters but is purely predictive. Predictions for 34500 data points of 566 mixtures showed that the mixture model is able to accurately predict viscosities of mixtures of non-polar as well as polar substances. Similar to the results for pure substances, also mixtures containing strongly associating substances show significantly higher than average deviations. Applying a binary interaction parameter (adjusted to VLE data) improved the results for some of the mixtures containing self-associating substances, but only had a minor effect on the mixtures containing ethanol or methanol. Taking the degree of dimerization into account might be beneficial in the description of mixtures containing associating components. Because mixture viscosities are predicted based on the combination of pure substance parameters, an intriguing asset is the capability to use pure substance parameters that were obtained from the group contribution approach.

A. Supporting information to chapter 2: A Group Contribution Method for Viscosities Based on Entropy Scaling Using the Perturbed-Chain Polar Statistical Associating Fluid Theory

A.1 Investigation of supercooled liquids

Intrigued by an inspiring question by an anonymous reviewer, we investigated how our correlation performs for supercooled liquids. Only for few substances there was sufficient experimental data at supercooled conditions available: few n-alkanes, 1-propanol and water. The experimental reduced viscosities are presented in fig. A.1. Interestingly, the reduced viscosites of the five investigated substances still show a monovariable dependence on the residual entropy s_{res}. Though, for the self-associating substances (i.e. water and 1-propanol), viscosities at supercooled conditions show a steep increase which was not captured by our correlation. For the investigated n-alkanes, we did not find that steep increase of the viscosity at supercooled conditions: experimental data align well along our correlation result.

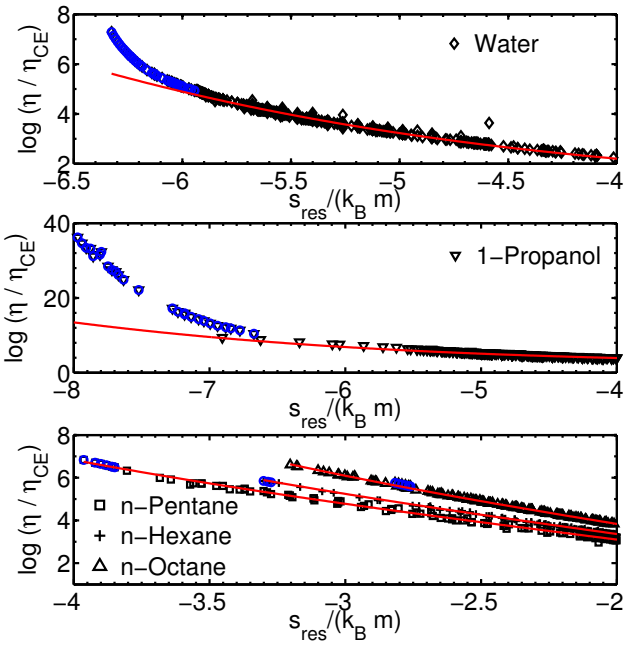


Figure A.1: Experimental viscosities of selected substances (symbols) together with the calculated viscosities (red lines) using the parameters given in Table 1 of the publication. Additionally to the data used in the publication (black symbols), we added data at supercooled conditions (blue symbols) of water[227–230], 1-propanol[231,232] and three n-alkanes[233].

A.2 References of experimental data

For the full list of the investigated substances and the corresponding references, please see the original Supporting Information, available free of charge on the ACS Publications website at DOI: 10.1021/acs.iecr.5b01698.

B. Supporting information to chapter 3: Bayesian Model Selection helps to Choose Objectively between Thermodynamic Models: A Demonstration of Selecting a Viscosity-Model based on Entropy Scaling

B.1 Uncertainties and Correlation in PC-SAFT parameters

To account for experimental errors in liquid densities ρ^{liq} and vapor pressures p^{sat}, we determined the joint distributions of the PC-SAFT group parameters m_α, σ_α and ε_α. To this end, we adjusted m_α, σ_α and ε_α to experimental data of n-butane up to n-hexatriacontane that was randomly perturbed with normally distributed synthetic experimental errors. We assumed relative errors of 2.6 % in p^{sat} and 1 % in ρ^{liq}. These resulting distributions are given as colored dots in Figure B.1 for the functional group CH_2 and in Figure B.2 for CH_3.

As expected, we found a strong correlation between the PC-SAFT group parameters. From these distributions, we calculated the correlation matrix and shifted the parameter distributions such that the mean values match the parameters proposed by Sauer et al. [118]. The correlation matrices together with

Table B.1: Statistics of all PC-SAFT group parameters used together with the correlation matrix.

Parameter	Mean value	95% CI (%)	Correlation matrix		
			m_{CH2}	σ_{CH2}	ε_{CH2}/k_B
m_{CH2}	0.4561	0.78	1	-0.7022	-0.9782
σ_{CH2}	3.89	0.37		1	0.6149
ε_{CH2}/k_B	239.01	0.34			1
Parameter	Mean value	95% CI (%)	Correlation matrix		
			m_{CH3}	σ_{CH3}	ε_{CH3}/k_B
m_{CH3}	0.6120	2.28	1	-0.7423	-0.9757
σ_{CH3}	3.7202	1.26		1	0.6476
ε_{CH3}/k_B	229.90	1.35			1

the 95 % confidence intervals (in percent of the mean value) of each PC-SAFT group parameter are given in Table B.1.

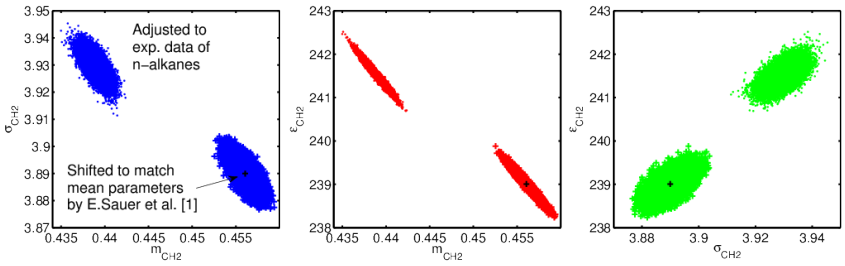


Figure B.1: Uncertainties and correlation in PC-SAFT $CH2$-parameters. Dots represent the parameters adjusted to experimental data that has been randomly perturbed with normally distributed synthetic errors. Colored crosses are the distributed parameters after shifting the mean (black cross) by applying the correlation matrix given in Table B.1.

B.2 Convergence behaviour of *BME* at different rejection rates

Figure B.3 and Figure B.4 show the evolution of *BME* over the MCMC steps carried out at different rejection rates. In both cases, the *BME* values converge well and hence demonstrate insensitivity against the chosen rejection rates. Final ensemble sizes (size of the posterior sample obtained from MCMC) ranged between 400,000 in the case of the six-parameter polynomial model

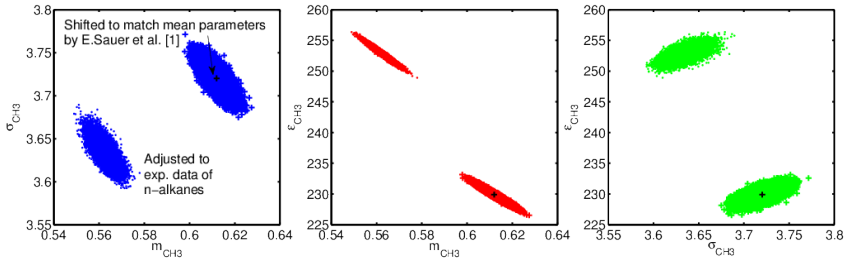


Figure B.2: Uncertainties and correlation in PC-SAFT $CH3$-parameters. Dots represent the parameters adjusted to experimental data that has been randomly perturbed with normally distributed synthetic errors. Colored crosses are the distributed parameters after shifting the mean (black cross) by applying the correlation matrix given in Table B.1.

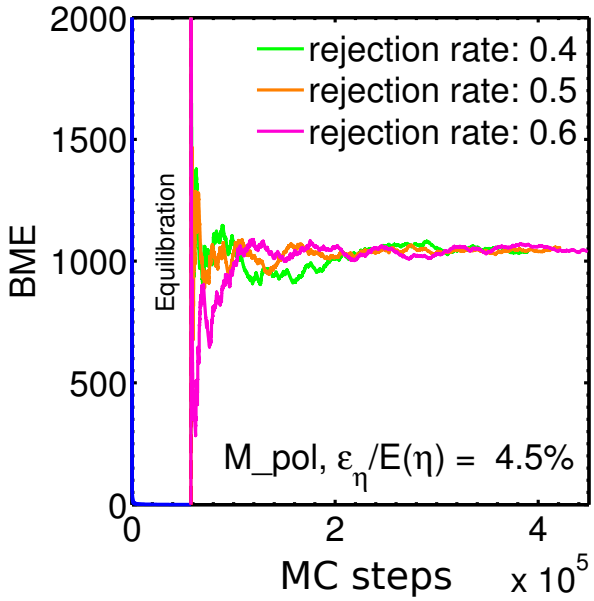


Figure B.3: BME as a function of MC steps in case 1 (pressure extrapolation) with an assumed relative experimental error of $e_\eta = 4.5\%$ at three different rejection rates.

$M_{\text{pol,6p}}$ and 2,000,000 in the case of the seven-parameter sinusoidal model $M_{\text{sin,7p}}$. The larger sample size required for model $M_{\text{sin,7p}}$ reflects its unfavorable convergence behaviour for obtaining a stable BME value.

B.3 Further interpretation of Bias and Variance

For a more detailed interpretation of the results, we introduce two additional performance measures; the weighted squared error WSE and the weighted

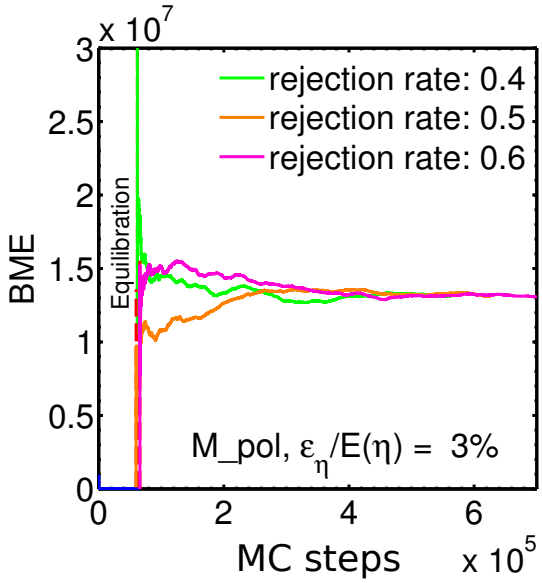


Figure B.4: *BME* as a function of MC steps in case 2 (carbon chain length extrapolation) with an assumed relative experimental error of $e_\eta = 3\%$ at three different rejection rates.

variance *WVAR*. While *MBIAS* contains information about over- or underestimation of a measured value via its sign, we use the weighted squared error (*WSE*) as a second performance measure on an absolute, positive scale. *WSE* is defined as *MBIAS* squared, as

$$WSE_j = MBIAS_j^2 = \frac{1}{R_{jj}} \left(\tilde{\eta}_j^{\text{pred}} - \eta_j^{\text{meas}} \right)^2 \tag{B.1}$$

As a measure for the ensemble spread, we use the weighted variance (*WVAR*) defined as the variance of model predictions at a data point j normalized by the experimental error variance, as

$$WVAR_j = \frac{1}{R_{jj}} \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \left(\eta_{i,j}^{\text{pred}} - \tilde{\eta}_j^{\text{pred}} \right)^2 \tag{B.2}$$

B.3.1 Case 1: Pressure Extrapolation

In Figure B.5a, the weighted squared error *WSE* illustrates the predictive bias of each model in both, reproducing the training data set (data points 1 to 7) and in predicting the evaluation data set (data points 8 to 11). It becomes obvious that the linear model produces a much higher bias than the

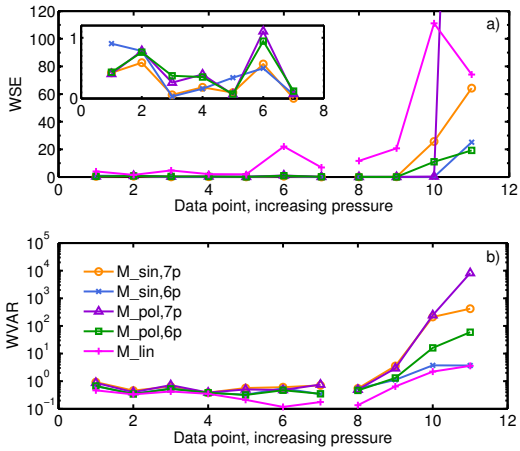


Figure B.5: Bias-variance behavior of all five competing models after conditioning on training data of case 1 (pressure extrapolation) with an assumed relative experimental error of $e_\eta = 3\%$. a) Weighted squared error (WSE) and b) weighted variance (WVAR) when predicting the training set (data points 1 to 7) and the evaluation set (data points 8 to 11).

other models in the training period, and that only the nonlinear modeling approaches with an intermediate flexibility (six adjustable parameters, $M_{\mathrm{pol,6p}}$ and $M_{\mathrm{sin,6p}}$) produce acceptable errors in the evaluation period. Note that the *WSE* of $M_{\mathrm{pol,7p}}$ at data point 10 even lies outside the shown axis range at a value of 734.

Figure B.5b illustrates the predictive uncertainty in both the training and the evaluation period as weighted variance *WVAR*. As discussed with respect to the width of credible intervals, the simplest linear model yields the lowest variance in predictions, while the nonlinear model variants with the highest flexibility (seven adjustable parameters, $M_{\mathrm{pol,7p}}$ and $M_{\mathrm{sin,7p}}$) produce the largest variance especially at the high end of the pressure range.

C. Supporting information to chapter 4: Pure Substance and Mixture Viscosities based on Entropy-Scaling and an Analytic Equation of State

C.1 Correlation of the Lennard-Jones fluid

In the results section we showed for pure Lennard Jones (LJ) fluids that viscosity data follow the entropy scaling approach. Data for viscosity, when made dimensionless with an appropriate expression, collapse onto a single curve with residual entropy. In the main text, we obtained the residual entropy s_{res} from the PCP-SAFT equation of state. Because the PC-SAFT equation of state is not strictly a model for LJ fluids, we will here consider the entropy scaling behavior of LJ fluids for values of residual entropy s_{res} taken from a highly accurate LJ-model proposed by Thol et al.[197].

Furthermore, we obtain s_{res} from Monte Carlo simulations using the histogram reweighting technique and we thereby confirm that the the LJ-model proposed by Thol et al.[197] is giving reliable results for entropy. A comparison of the PC-SAFT model and of the LJ-model proposed by Thol et al.[197] with the results of molecular simulations of our work is shown in Fig C.1. As anticipated, some deviation of residual entropies calculated from the PC-SAFT model from the simulated results are observed, whereas the LJ equation of state is in excellent agreement with the simulation data.

Different values of residual entropies lead to a different behavior of the en-

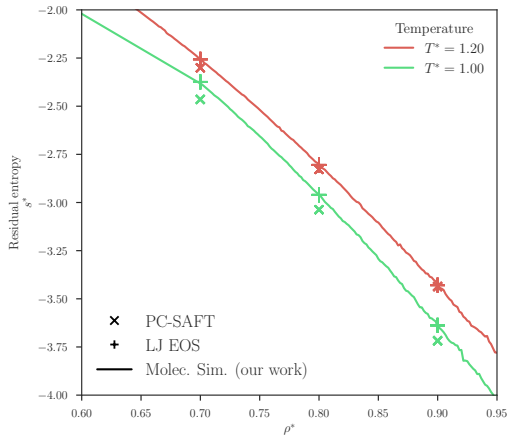


Figure C.1: Dimensionless residual entropy with dimensionless density for two temperatures. Comparison of results from PC-SAFT equation of state and from an accurate LJ equation of state[197] to results from molecular simulations (using histogram reweighting for simulations in the grand canonical ensemble).

tropy scaled dimensionless viscosity. Fig C.2 confirms that the general entropy scaling behavior is observed for the LJ-fluid with s_{res} as calculated from an accurate LJ-equation of state. With slightly different parameters, however, the entropy scaling behavior is also observed when s_{res} is calculated from the PC-SAFT model.

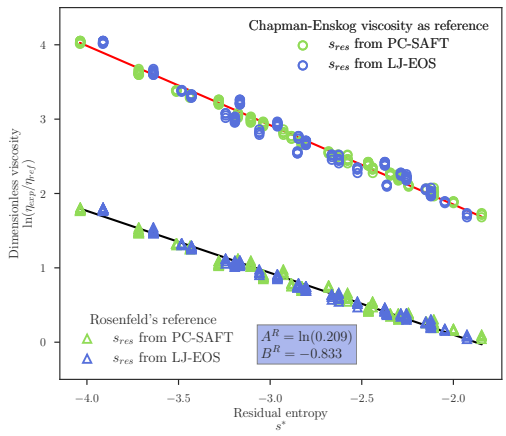


Figure C.2: Dimensionless simulated viscosities of a Lennard-Jones fluid. The residual entropy is thereby calculated from the PC-SAFT equation of state (green symbols) and from the accurate LJ-model of Thol et al.[197] (blue symbols). 'Triangles' represent the dimensionless viscosity using Rosenfeld's reference viscosity. The black line is the result of the linear correlation. 'Circles' are the dimensionless simulated viscosities using the Chapman-Enskog viscosity η_{CE} as reference viscosity.

By adjusting A^R and B^R of eq. (1) (of the main text) to the viscosities shown in Fig C.2, using the LJ equation of state for the entropy, we obtain $A^R = \ln(0.209)$ and $B^R = -0.833$. These values are in reasonable agreement with Rosenfelds originally published scaling parameters[1,67] ($A_0^R = \ln(0.2)$ and $B_0^R = -0.8$).

C.2 Application of viscosity model to Lennard-Jones chains

To assess the proposed expression for the low-density limit, where the A-parameter of our viscosity model[48] determines the viscosity, we compare results for the low density viscosity of fully flexible Lennard Jones chains (of various chain length) with viscosities determined from molecular simulations by Delage-Santacreu et al.[85]. Viscosity parameter A_{LJ} is based on the expression given in Table C.1, applied to chain fluids as described in Appendix A of the main text (eq 20). Figure C.3 confirms that the proposed model is in good agreement to the simulated viscosities.

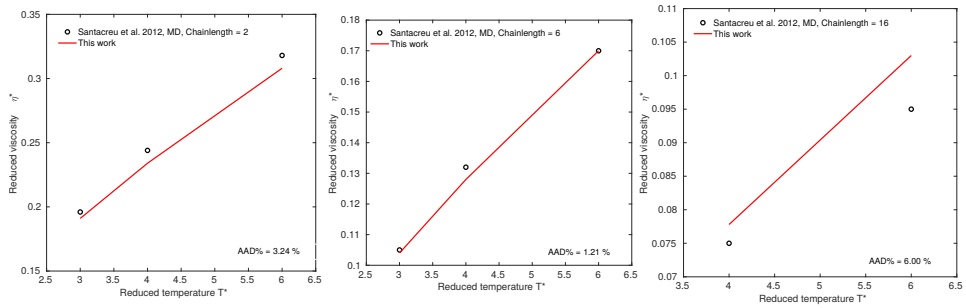


Figure C.3: Viscosities of Lennard-Chains of various chain length. Black circles are MD results by Santacreu et al.[234], the red lines are the results of the proposed model. The applied viscosity parameter A_{LJ} corresponds to the parameter given in Table C.1, adjusted for higher chain lengths as given in the Appendix A of the main text (eq 20).

C.3 Viscosity parameters

Table C.1: Pure substance viscosity parameters of the monatomic, spherical Lennard-Jones model fluid.

Substance	A_i	B_i	C_i	D_i	
Lennard-Jones	-0.1659	-1.1768	-0.1068	-0.0178	*

Table C.2: Pure substance viscosity parameters of n-alkanes. The asterix indicates that the A_i parameter was individually adjusted. Otherwise, the A_i parameter was combined from functional group parameters.

Substance	A_i	B_i	C_i	D_i	
Methane	-0.0595	-0.8908	-0.0348	-0.0177	*
Ethane	-0.5741	-1.5923	-0.1959	-0.0325	*
Propane	-0.8013	-1.9972	-0.2907	-0.0467	*
nButane	-0.9763	-2.2413	-0.3690	-0.0605	
nPentane	-1.0978	-2.4577	-0.4397	-0.0737	
nHexane	-1.2035	-2.5958	-0.4816	-0.0865	
nHeptane	-1.2979	-2.6936	-0.4951	-0.0988	
nOctane	-1.3837	-2.8569	-0.5461	-0.1107	
nNonane	-1.4629	-3.0058	-0.5842	-0.1222	
nDecane	-1.5367	-3.1490	-0.6154	-0.1334	
nUndecane	-1.6062	-3.2485	-0.6500	-0.1441	
nDodecane	-1.6719	-3.3902	-0.6956	-0.1546	
nTridecane	-1.7345	-3.4695	-0.7043	-0.1647	
nTetradecane	-1.7945	-3.5557	-0.7385	-0.1744	
nPentadecane	-1.8521	-3.7647	-0.8216	-0.1839	
nHexadecane	-1.9077	-3.8125	-0.8290	-0.1931	
nHeptadecane	-1.9615	-3.8823	-0.8386	-0.2021	
nOctadecane	-2.0137	-4.0711	-0.9298	-0.2107	
nNonadecane	-2.0644	-4.0653	-0.8991	-0.2191	
nEicosane	-2.1139	-4.1759	-0.9573	-0.2273	
nHeneicosane	-2.1622	-4.2184	-0.9136	-0.2353	
nDocosane	-2.2094	-4.3373	-0.9708	-0.2430	
nTricosane	-2.2556	-4.2312	-0.8951	-0.2505	
nTetracosane	-2.3010	-4.6041	-1.0303	-0.2578	
nOctacosane	-2.4748	-5.1281	-1.2775	-0.2852	
nTriacontane	-2.5579	-4.7635	-1.0517	-0.2979	
nDotriacontane	-2.6388	-5.1000	-1.2007	-0.3100	
nHexatriacontane	-2.7953	-5.6550	-1.4638	-0.3325	

Table C.3: Pure substance viscosity parameters of branched alkanes.

Substance	A_i	B_i	C_i	D_i
Isobutane(2-Methylpropane)	-0.9521	-2.2586	-0.3438	-0.0605
Isopentane(2-Methylbutane)	-1.0805	-2.4514	-0.4390	-0.0737
Neopentane(2	-0.7986	-2.0899	-0.2188	-0.0737
2-Methylpentane	-1.1910	-2.5695	-0.4813	-0.0865
3-Methylpentane	-1.1910	-2.5416	-0.4690	-0.0865
2,2-Dimethylbutane	-0.9375	-2.1668	-0.2861	-0.0865
2,3-Dimethylbutane	-1.1770	-2.4259	-0.3992	-0.0865
3-Ethylpentane	-1.2889	-2.6369	-0.5094	-0.0988
2,4-Dimethylpentane	-1.2789	-2.7334	-0.5402	-0.0988
2,2,4-Trimethylpentane	-1.1464	-2.7647	-0.5615	-0.1107
2,3,4-Trimethylpentane	-1.3623	-2.9320	-0.5877	-0.1107
Squalane	-2.6069	-5.3734	-1.3691	-0.2979
2,2,4,4,6,8,8-Heptamethylnonane	-1.3808	-3.2389	-0.6518	-0.1931

Table C.4: Pure substance viscosity parameters of cycloalkanes. The asterix indicates that the A_i parameter was individually adjusted.

Substance	A_i	B_i	C_i	D_i	
cyclopentane	-0.8485	-2.3053	-0.3801	-0.0718	*
cyclohexane	-0.9982	-2.3823	-0.3169	-0.0847	*
cyclooctane	-1.3132	-2.4521	-0.3023	-0.1090	*

Table C.5: Pure substance viscosity parameters of alkenes. The asterix indicates that the A_i parameter was individually adjusted. Otherwise, the A_i parameter was combined from functional group parameters.

Substance	A_i	B_i	C_i	D_i	
Ethylene	-0.4941	-1.4818	-0.1407	-0.0304	*
1-Propene	-0.7728	-1.9599	-0.2842	-0.0447	
1-Hexene	-1.1640	-2.4505	-0.4432	-0.0847	
1-Heptene	-1.2614	-2.6535	-0.5149	-0.0971	
1-Octene	-1.3495	-2.8423	-0.5594	-0.1090	
1-Nonene	-1.4305	-2.7664	-0.5102	-0.1206	
1-Decene	-1.5058	-3.2891	-0.7212	-0.1318	
Isobutylene(2-Methylpropene)	-0.9838	-2.1027	-0.3034	-0.0585	*

Table C.6: Pure substance viscosity parameters of aromatics.

Substance	A_i	B_i	C_i	D_i
Benzene	-1.0929	-2.2871	-0.4125	-0.0792
Toluene	-1.1874	-2.4505	-0.4581	-0.0918
Ethylbenzene	-1.2856	-2.6929	-0.5605	-0.1039
1,3-Dimethyl-benzene	-1.2749	-2.5804	-0.5289	-0.1039
1,4-Dimethyl-benzene	-1.2749	-2.6129	-0.5418	-0.1039
Mesitylene	-1.3566	-2.6547	-0.5489	-0.1157
p-Cymene	-1.4434	-2.8975	-0.6410	-0.1270
n-Heptylbenzene	-1.6693	-3.4376	-0.8015	-0.1589
n-Nonylbenzene	-1.7939	-3.5061	-0.7608	-0.1785
n-Dodecylbenzene	-1.9630	-3.7720	-0.8318	-0.2058
Isopropylbenzene	-1.3671	-2.8169	-0.6018	-0.1157
Butylbenzene	-1.4557	-2.8655	-0.5777	-0.1270
Propylbenzene	-1.3743	-2.1888	-0.3417	-0.1157

Table C.7: Pure substance viscosity parameters of CO_2 and N_2. The asterix indicates that the A_i parameter was individually adjusted.

Substance	A_i	B_i	C_i	D_i	
CO2	-0.5138	-1.3768	-0.2074	-0.0467	*
N2	-0.1964	-0.9461	-0.0310	-0.0303	*

Table C.8: Pure substance viscosity parameters of aldehydes.

Substance	A_i	B_i	C_i	D_i
Ethanal	-0.3574	-0.7933	0.1404	-0.0467
Propanal	-0.4766	-1.3477	-0.0555	-0.0604
Butanal	-0.5809	-2.0176	-0.3375	-0.0737
2-Methyl-propanal	-0.5688	-2.2652	-0.4374	-0.0737
3-Methyl-butyraldehyde	-0.6655	-1.8956	-0.2519	-0.0865
Heptanal	-0.8376	-2.8879	-0.6488	-0.1107
Decanal	-1.0452	-2.9497	-0.6641	-0.1441

Table C.9: Pure substance viscosity parameters of ketones.

Substance	A_i	B_i	C_i	D_i
Acetone	-1.0551	-2.2200	-0.3547	-0.0604
3-Pentanone	-1.2629	-2.5335	-0.4528	-0.0865
Methyl-Isobutyl-ketone	-1.3442	-3.0229	-0.6378	-0.0988
4-Heptanone	-1.4324	-2.8019	-0.5333	-0.1107
2-Hexanone	-1.3512	-2.5662	-0.4335	-0.0988
2-Octanone	-1.5078	-2.8420	-0.5162	-0.1222
2-Butanone	-1.1653	-2.4287	-0.4373	-0.0737

Table C.10: Pure substance viscosity parameters of esters.

Substance	A_i	B_i	C_i	D_i
nPropyl-acetate	-1.2543	-2.2953	-0.2853	-0.1005
nButyl-acetate	-1.3379	-2.7683	-0.5104	-0.1123
Isopentyl-acetate	-1.4118	-2.9353	-0.5861	-0.1238
Vinyl-acetate(Ethenyl-acetate)	-1.1242	-2.4374	-0.4095	-0.0864
nPentyl-acetate	-1.4154	-2.7149	-0.4738	-0.1238
nHeptyl-acetate	-1.5560	-3.0556	-0.5314	-0.1456

Table C.11: Pure substance viscosity parameters of ethers. The asterix indicates that the A_i parameter was individually adjusted. Otherwise, the A_i parameter was combined from functional group parameters.

Substance	A_i	B_i	C_i	D_i	
Dimethyl-ether	-0.8518	-1.8921	-0.2578	-0.0487	
Diisopropyl-ether	-1.3532	-2.8157	-0.5234	-0.1005	*
Di-nButyl-ether	-1.3556	-2.8110	-0.4985	-0.1238	
Methyl-tertButyl-ether	-0.9246	-2.4010	-0.4260	-0.0883	
Di-nPentyl-ether	-1.4947	-3.6239	-0.8585	-0.1456	
Di-nPropyl-ether	-1.1967	-2.5254	-0.4272	-0.1005	
Di-Ethyl-ether	-1.0070	-2.2992	-0.3894	-0.0755	
Diethyleneglycoldimethylether	-1.2095	-2.6243	-0.4929	-0.1270	

Table C.12: Pure substance viscosity parameters of hydro(chloro)fluorocarbons. The asterix indicates that the A_i parameter was individually adjusted.

Substance	A_i	B_i	C_i	D_i	
R32	-1.0833	-1.9968	-0.2503	-0.0546	*
R152a	-1.1237	-2.1489	-0.2769	-0.0680	*
R23	-1.0556	-2.0562	-0.2825	-0.0717	*
R125	-1.3005	-2.8541	-0.5396	-0.1155	*
R143a	-1.1297	-2.2024	-0.3203	-0.0846	*
R236ea	-0.9770	-2.4251	-0.2831	-0.1409	*
R227ea	-1.3718	-3.0118	-0.5540	-0.1543	*
R245fa	-1.4285	-2.9879	-0.4947	-0.1269	*
R236fa	-1.1680	-2.6971	-0.4772	-0.1409	*
R134a	-1.2914	-2.6782	-0.4505	-0.1004	*
R143a	-0.8530	-1.4570	0.3634	-0.0881	*
R227ea	-1.0079	-2.2718	-0.4248	-0.0868	*
R115	-1.1464	-2.7197	-0.5569	-0.1427	*
R140a	-1.0325	-2.3470	-0.4703	-0.1264	*
R21	-0.9850	-2.3249	-0.4743	-0.1012	*
R115	-0.9935	-2.5208	-0.6068	-0.1295	*
R142b	-0.9930	0.1100	7.0428	-0.0991	*
R125	-0.9671	-2.5578	-0.6434	-0.1163	*
R134a	-0.9757	-2.4426	-0.5724	-0.1025	*
R114	-1.1623	-2.9065	-0.6941	-0.1550	*
R13b1	-0.9604	-2.6583	-0.7764	-0.1385	*

Table C.13: Pure substance viscosity parameters of alcohols.

Substance	A_i	B_i	C_i	D_i	
Methanol	-0.57700	-0.44059	0.03545	-0.01064	*
Ethanol	-1.08153	-1.19980	-0.13292	-0.04870	*
1-Propanol	-0.87560	-1.47721	-0.08931	-0.06237	
1-Butanol	-1.00650	-1.88906	-0.12523	-0.07554	
1-Pentanol	-1.11880	-2.22062	-0.25493	-0.08825	
1-Hexanol	-1.21800	-2.06340	-0.21639	-0.10052	
1-Heptanol	-1.30750	-2.24414	-0.26729	-0.11237	
1-Octanol	-1.38960	-2.58867	-0.28276	-0.12383	
1-Nonanol	-1.46570	-2.71167	-0.49511	-0.13491	
1-Decanol	-1.53710	-2.97397	-0.50795	-0.14562	
1-Undecanol	-1.60440	-3.44231	-0.55890	-0.15600	
1-Dodecanol	-1.6684	-3.9471	-0.9305	-0.1661	
1-Tridecanol	-1.7295	-2.8003	-0.5484	-0.1758	
1-Tetradecanol	-1.7882	-2.8133	-0.5546	-0.1852	
1-Hexadecanol	-1.8993	-3.0227	-0.5234	-0.2033	
1-Octadecanole	-2.0036	-3.7363	-0.8864	-0.2203	
Isopropanol	-0.8470	-1.0735	0.2102	-0.0624	
2-Butanol	-0.9864	-1.3818	0.2030	-0.0755	
2-Pentanol	-1.1043	-1.0389	0.4671	-0.0883	
3-Pentanol	-1.1043	-0.9823	0.4199	-0.0883	
2-Hexanol	-1.2075	-1.7111	0.0319	-0.1005	
2-Heptanol	-1.3001	-1.7492	0.0215	-0.1124	
2-Octanol	-1.3845	-1.1782	0.3771	-0.1238	
1,2-Ethandiol	-0.7689	-0.9258	-0.1133	-0.0642	

Table C.14: Pure substance viscosity parameters of amines.

Substance	A_i	B_i	C_i	D_i
methylamine	-0.3690	-1.2753	-0.2168	-0.0335
n-propylamine	-0.7383	-1.7723	-0.0548	-0.0614
n-butylamine	-0.8691	-2.3216	-0.5911	-0.0746
pentylamine	-0.9812	-2.6533	-0.7457	-0.0874
hexylamine	-1.0803	-2.3348	-0.3037	-0.0997
heptylamine	-1.1698	-3.4917	-1.0288	-0.1116
1-octylamine	-1.2518	-2.9995	-0.9478	-0.1230
nonylamine	-1.3279	-3.1517	-1.0221	-0.1341
decylamine	-1.3992	-3.2065	-1.0390	-0.1449
diethylamine	-0.6824	-2.1290	-0.5989	-0.0746
dipropylamine	-0.8987	-2.1083	-0.2763	-0.0997
dimethylamine	-0.3807	2.1585	0.9391	-0.0477

C.4 Tables with $AAD\%$ of all investigated substances and mixtures

C.4.1 Pure substances

Table C.15: Deviations of calculated viscosities and simulated viscosities for Lennard-Jones fluids.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
Lennard-Jones	205	2.72	6	2.46

Table C.16: Deviations of calculated and experimental pure substance viscosities for n-alkanes.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
n-alkanes	17115	3.91		
Methane	2439	3.93	195	3.00
Ethane	2476	5.76	32	4.88
Propane	1663	4.54	44	3.14
nButane	727	3.78	20	2.46
nPentane	2000	5.15	118	3.31
nHexane	1425	3.18	119	1.69
nHeptane	1220	2.80	96	1.36
nOctane	1054	2.75	41	1.80
nNonane	264	1.35	10	1.00
nDecane	993	1.92	53	1.51
nUndecane	427	1.75	21	1.20
nDodecane	366	2.63	16	1.94
nTridecane	168	2.27	10	1.76
nTetradecane	155	3.76	0	3.23
nPentadecane	261	3.81	7	2.78
nHexadecane	252	4.89	4	3.48
nHeptadecane	317	3.01	8	2.44
nOctadecane	434	4.32	37	2.89
nNonadecane	273	2.30	1	2.03
nEicosane	66	7.19	0	6.93
nHeneicosane	30	1.75	0	1.52
nDocosane	24	3.63	1	2.57
nTricosane	11	3.94	0	4.20
nTetracosane	22	8.17	3	6.04
nOctacosane	14	4.50	1	4.08
nTriacontane	3	0.29	0	0.23
nDotriacontane	12	2.84	0	3.14
nHexatriacontane	19	3.32	1	2.90

Table C.17: Deviations of calculated and experimental pure substance viscosities for branched alkanes.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
branched_alkanes	2456	9.55		
Isobutane(2-Methylpropane)	705	4.53	10	3.53
Isopentane(2-Methylbutane)	332	4.32	10	2.76
Neopentane(2	154	11.15	1	14.75
2-Methylpentane	9	3.57	2	0.73
3-Methylpentane	7	0.42	0	0.29
2,2-Dimethylbutane	19	3.09	4	3.17
2,3-Dimethylbutane	7	1.96	0	0.61
3-Ethylpentane	6	0.53	1	0.50
2,4-Dimethylpentane	9	0.56	1	0.52
2,2,4-Trimethylpentane	442	2.36	52	1.33
2,3,4-Trimethylpentane	16	2.96	2	2.91
Squalane	567	27.14	2	25.31
2,2,4,4,6,8,8-Heptamethylnonane	183	2.89	12	2.11

Table C.18: Deviations of calculated and experimental pure substance viscosities for cycloalkanes.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
cycloalkanes	1359	4.50		
cyclopentane	334	2.05	13	1.21
cyclohexane	1017	5.33	53	3.59
cyclooctane	8	0.68	1	0.41

Table C.19: Deviations of calculated and experimental pure substance viscosities for aromatics.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
aromatics	5295	4.81		
Benzene	1307	7.85	57	3.99
Toluene	2139	3.36	129	1.23
Ethylbenzene	410	5.46	2	5.45
1,3-Dimethyl-benzene	623	2.33	61	1.48
1,4-Dimethyl-benzene	595	4.76	9	4.23
Mesitylene	32	0.22	0	0.16
p-Cymene	15	1.91	1	1.68
n-Heptylbenzene	42	2.51	0	2.20
n-Nonylbenzene	15	1.18	1	1.16
n-Dodecylbenzene	30	4.15	0	3.80
Isopropylbenzene	14	1.25	0	0.77
Butylbenzene	30	3.37	1	3.50
Propylbenzene	43	25.29	5	10.54

Table C.20: Deviations of calculated and experimental pure substance viscosities for alkenes.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
alkenes	1546	4.18		
Ethylene	527	4.25	41	3.34
1-Propene	279	3.65	4	3.37
1-Hexene	306	5.29	22	4.36
1-Heptene	163	3.75	1	3.66
1-Octene	106	3.23	3	2.70
1-Nonene	87	4.45	0	3.60
1-Decene	53	3.28	0	2.83
Isobutylene(2-Methylpropene)	25	2.79	5	0.83

Table C.21: Deviations of calculated and experimental pure substance viscosities for carbondioxide and nitrogen.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
CO2	3173	7.85	51	5.01
N2	2689	4.40	266	3.17

Table C.22: Deviations of calculated and experimental pure substance viscosities for aldehydes.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
aldehydes	370	8.82		
Ethanal	21	22.22	4	11.58
Propanal	17	4.34	5	3.07
Butanal	233	8.77	30	7.22
2-Methyl-propanal	61	8.90	2	7.50
3-Methyl-butyraldehyde	18	6.73	1	4.15
Heptanal	10	0.93	0	0.91
Decanal	10	0.65	0	0.66

Table C.23: Deviations of calculated and experimental pure substance viscosities for ketones.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
ketones	885	5.66		
Acetone	235	4.51	8	2.80
3-Pentanone	151	8.11	1	7.69
Methyl-Isobutyl-ketone	100	8.14	7	6.67
4-Heptanone	117	5.06	3	4.44
2-Hexanone	87	2.02	0	1.76
2-Octanone	105	4.83	0	3.73
2-Butanone	90	7.05	7	6.27

Table C.24: Deviations of calculated and experimental pure substance viscosities for esters.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
ester	584	3.26		
nPropyl-acetate	118	3.99	0	2.80
nButyl-acetate	153	4.02	5	4.00
Isopentyl-acetate	106	3.01	10	1.94
Vinyl-acetate(Ethenyl-acetate)	15	1.54	1	1.24
nPentyl-acetate	102	2.80	0	2.36
nHeptyl-acetate	90	2.11	10	1.67

Table C.25: Deviations of calculated and experimental pure substance viscosities for ethers.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
ether	824	10.12		
Dimethyl-ether	166	2.14	23	0.93
Diisopropyl-ether	93	2.97	4	2.50
Di-nButyl-ether	76	1.16	3	0.80
Methyl-tertButyl-ether	37	8.39	0	7.64
Di-nPentyl-ether	16	16.75	0	14.88
Di-nPropyl-ether	78	6.00	5	4.97
Di-Ethyl-ether	296	21.58	0	23.81
Diethyleneglycoldimethylether	62	2.93	1	2.48

Table C.26: Deviations of calculated and experimental pure substance viscosities for hfc+cfc.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
hfc+cfc	7833	5.13		
R32	227	6.01	1	4.76
R152a	269	3.78	9	2.65
R23	743	10.86	40	7.10
R125	469	3.63	13	1.67
R143a	269	3.54	29	1.51
R236ea	73	0.54	2	0.36
R227ea	35	3.85	1	3.15
R245fa	154	1.66	13	0.67
R236fa	191	0.54	2	0.47
R134a	833	3.93	33	3.12
R143a	609	2.53	17	1.82
R227ea	869	3.73	77	2.21
R115	389	5.37	32	3.76
R140a	47	2.65	3	1.84
R21	174	1.86	5	1.48
R115	309	2.00	39	1.02
R142b	96	6.69	11	1.26
R125	562	5.98	17	5.20
R134a	663	7.83	8	7.45
R114	311	4.07	33	2.54
R13b1	541	9.06	16	8.29

Table C.27: Deviations of calculated and experimental pure substance viscosities for amines.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
amines	310	4.69		
methylamine	51	5.62	0	3.50
n-propylamine	16	1.10	0	0.77
n-butylamine	31	4.68	6	2.51
pentylamine	24	3.32	1	1.80
hexylamine	24	1.28	0	0.95
heptylamine	24	4.19	2	2.76
1-octylamine	31	5.35	1	3.93
nonylamine	14	0.39	0	0.33
decylamine	14	0.39	1	0.23
diethylamine	66	6.93	0	6.66
dipropylamine	8	0.22	0	0.23
dimethylamine	7	22.45	1	27.10

Table C.28: Deviations of calculated and experimental pure substance viscosities for alcohols.

Component 1	N_{dat}	AAD %	$N_{outliers}$	Median
alcohols	5361	6.38		
Methanol	1049	5.33	79	3.27
Ethanol	691	3.40	45	2.14
1-Propanol	698	7.66	58	3.32
1-Butanol	605	5.79	68	2.78
1-Pentanol	160	4.38	9	2.53
1-Hexanol	144	3.05	12	1.60
1-Heptanol	82	3.36	4	2.66
1-Octanol	220	5.13	5	4.58
1-Nonanol	170	6.00	3	5.26
1-Decanol	186	9.61	1	6.93
1-Undecanol	201	12.55	0	12.64
1-Dodecanol	189	15.67	16	13.91
1-Tridecanol	5	0.74	1	0.83
1-Tetradecanol	8	0.24	0	0.27
1-Hexadecanol	20	2.45	1	1.42
1-Octadecanole	27	7.03	0	5.92
Isopropanol	379	7.24	40	2.37
2-Butanol	160	4.90	13	2.53
2-Pentanol	49	1.85	1	1.62
3-Pentanol	18	6.21	0	5.93
2-Hexanol	49	3.91	0	2.06
2-Heptanol	38	5.63	4	4.34
2-Octanol	21	3.96	1	2.80
1,2-Ethandiol	192	11.04	5	8.49

C.4.2 Mixtures

Table C.29: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of n-alkanes + n-alkanes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
methane	ethane	344	4.57	9	4.01
	propane	722	11.31	73	3.92
	n-butane	181	6.12	4	5.74
	n-pentane	8	3.77	0	2.00
	n-hexane	11	1.88	0	2.05
	n-heptane	90	4.84	2	3.86
	n-octane	42	3.52	0	3.89
	n-decane	195	13.66	0	12.91
	n-tetradecane	24	20.43	0	19.17
	n-octadecane	24	20.55	0	20.97
ethane	methane	344	4.57	9	4.01
	propane	23	11.62	2	6.14
	n-butane	15	5.96	0	5.95
	n-hexane	13	0.68	1	0.53
	n-heptane	92	7.68	1	6.57
	n-octane	13	2.05	0	1.68
	n-tetradecane	21	10.99	0	11.34
	n-octadecane	22	13.80	3	16.34
propane	methane	722	11.31	73	3.92
	ethane	23	11.62	2	6.14
	n-butane	19	6.22	0	6.20
	n-heptane	71	8.76	0	7.84
	n-octane	11	3.23	0	3.16
n-butane	methane	181	6.12	4	5.74
	ethane	15	5.96	0	5.95
	propane	19	6.22	0	6.20
	n-hexane	9	1.94	1	0.63
	n-octane	10	4.09	1	3.02
n-pentane	methane	8	3.77	0	2.00
	n-hexane	11	7.25	0	7.47
	n-heptane	57	6.44	2	2.62
	n-octane	306	3.49	4	3.05
	n-nonane	11	9.29	0	10.10
	n-decane	414	3.24	27	2.35
	n-undecane	11	11.12	0	11.05
	n-dodecane	11	11.59	0	11.67
	n-hexadecane	11	9.75	0	10.73
n-hexane	methane	11	1.88	0	2.05
	ethane	13	0.68	1	0.53
	n-butane	9	1.94	1	0.63
	n-pentane	11	7.25	0	7.47
	n-heptane	134	3.01	0	1.39
	n-octane	83	3.33	0	1.27
	n-nonane	20	7.10	0	7.44
	n-decane	62	8.40	0	9.41
	n-undecane	11	8.63	0	9.50
	n-dodecane	34	5.27	0	3.74
	n-tridecane	27	5.83	4	4.11
	n-tetradecane	41	3.78	0	1.42
	n-hexadecane	240	7.33	9	6.43
n-heptane	methane	90	4.84	2	3.86
	ethane	92	7.68	1	6.57
	propane	71	8.76	0	7.84
	n-pentane	57	6.44	2	2.62
	n-hexane	134	3.01	0	1.39

Table C.29: (Continuation)

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
	n-octane	418	2.40	39	1.67
	n-nonane	93	2.48	2	0.96
	n-decane	362	2.67	41	1.61
	n-undecane	79	2.53	10	1.65
	n-dodecane	31	8.96	0	9.18
	n-tridecane	22	18.10	0	17.37
	n-tetradecane	17	6.47	0	7.86
	n-hexadecane	25	6.87	0	7.05
	n-octadecane	3	3.20	0	2.83
	n-eicosane	32	6.68	0	7.61
	n-docasane	25	5.53	0	5.78
	n-tetracosane	19	6.38	0	5.45
n-octane	methane	42	3.52	0	3.89
	ethane	13	2.05	0	1.68
	propane	11	3.23	0	3.16
	n-butane	10	4.09	1	3.02
	n-pentane	306	3.49	4	3.05
	n-hexane	83	3.33	0	1.27
	n-heptane	418	2.40	39	1.67
	n-nonane	20	6.42	2	6.43
	n-decane	481	3.00	26	2.78
	n-undecane	33	2.79	0	0.79
	n-dodecane	52	3.40	0	1.52
	n-tridecane	22	1.50	2	1.11
	n-tetradecane	67	2.23	0	1.09
	n-pentadecane	22	4.06	0	2.39
	n-hexadecane	35	5.01	5	5.62
	n-tetracosane	19	8.46	0	7.37
n-nonane	n-pentane	11	9.29	0	10.10
	n-hexane	20	7.10	0	7.44
	n-heptane	93	2.48	2	0.96
	n-octane	20	6.42	2	6.43
	n-decane	20	6.87	0	6.77
	n-undecane	11	6.05	0	5.85
	n-dodecane	14	6.01	1	5.99
	n-tetradecane	3	3.73	0	3.62
	n-hexadecane	14	3.49	0	4.17
n-decane	methane	195	13.66	0	12.91
	n-pentane	414	3.24	27	2.35
	n-hexane	62	8.40	0	9.41
	n-heptane	362	2.67	41	1.61
	n-octane	481	3.00	26	2.78
	n-nonane	20	6.87	0	6.77
	n-undecane	11	6.50	0	6.29
	n-dodecane	14	6.19	0	6.17
	n-tridecane	22	2.21	0	1.79
	n-tetradecane	3	4.30	0	3.83
	n-pentadecane	22	5.63	0	5.58
	n-hexadecane	68	2.55	1	1.96
	n-octadecane	5	1.65	0	1.49
	n-eicosane	24	6.56	0	5.87
	n-docasane	20	4.42	0	4.61
	n-tetracosane	16	1.34	0	1.05
n-undecane	n-pentane	11	11.12	0	11.05
	n-hexane	11	8.63	0	9.50
	n-heptane	79	2.53	10	1.65
	n-octane	33	2.79	0	0.79
	n-nonane	11	6.05	0	5.85
	n-decane	11	6.50	0	6.29
	n-dodecane	11	5.59	1	5.71
	n-tridecane	22	3.14	0	3.00

Table C.29: (Continuation)

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
	n-pentadecane	22	6.64	0	6.44
	n-hexadecane	11	3.05	0	2.38
n-dodecane	n-pentane	11	11.59	0	11.67
	n-hexane	34	5.27	0	3.74
	n-heptane	31	8.96	0	9.18
	n-octane	52	3.40	0	1.52
	n-nonane	14	6.01	1	5.99
	n-decane	14	6.19	0	6.17
	n-undecane	11	5.59	1	5.71
	n-tetradecane	3	2.86	0	2.50
	n-hexadecane	14	2.53	0	2.42
	n-heptadecane	63	2.35	3	1.60
	n-tricosane	32	3.71	2	2.99
n-tridecane	n-hexane	27	5.83	4	4.11
	n-heptane	22	18.10	0	17.37
	n-octane	22	1.50	2	1.11
	n-decane	22	2.21	0	1.79
	n-undecane	22	3.14	0	3.00
	n-pentadecane	22	8.53	0	8.49
n-tetradecane	methane	24	20.43	0	19.17
	ethane	21	10.99	0	11.34
	n-hexane	41	3.78	0	1.42
	n-heptane	17	6.47	0	7.86
	n-octane	67	2.23	0	1.09
	n-nonane	3	3.73	0	3.62
	n-decane	3	4.30	0	3.83
	n-dodecane	3	2.86	0	2.50
	n-hexadecane	25	2.42	0	1.38
n-pentadecane	n-octane	22	4.06	0	2.39
	n-decane	22	5.63	0	5.58
	n-undecane	22	6.64	0	6.44
	n-tridecane	22	8.53	0	8.49
n-hexadecane	n-pentane	11	9.75	0	10.73
	n-hexane	240	7.33	9	6.43
	n-heptane	25	6.87	0	7.05
	n-octane	35	5.01	5	5.62
	n-nonane	14	3.49	0	4.17
	n-decane	68	2.55	1	1.96
	n-undecane	11	3.05	0	2.38
	n-dodecane	14	2.53	0	2.42
	n-tetradecane	25	2.42	0	1.38
	n-eicosane	35	9.88	0	9.54
n-heptadecane	n-dodecane	63	2.35	3	1.60
n-octadecane	methane	24	20.55	0	20.97
	ethane	22	13.80	3	16.34
	n-heptane	3	3.20	0	2.83
	n-decane	5	1.65	0	1.49
n-eicosane	n-heptane	32	6.68	0	7.61
	n-decane	24	6.56	0	5.87
	n-hexadecane	35	9.88	0	9.54
	n-tetracosane	3	1.63	0	1.94
n-docasane	n-heptane	25	5.53	0	5.78
	n-decane	20	4.42	0	4.61
n-tricosane	n-dodecane	32	3.71	2	2.99
n-tetracosane	n-heptane	19	6.38	0	5.45
	n-octane	19	8.46	0	7.37
	n-decane	16	1.34	0	1.05
	n-eicosane	3	1.63	0	1.94

Table C.30: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of n-alkanes + branched alkanes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-butane	isobutane	25	5.44	0	5.52
	squalane	99	12.39	14	6.91
n-hexane	2,2,4-trimethylpentane	19	1.91	0	1.55
	squalane	118	20.11	1	11.74
n-heptane	2,2,4-trimethylpentane	117	1.20	5	0.98
n-octane	2,2,4-trimethylpentane	50	2.89	5	1.42
	squalane	80	19.14	0	8.90
n-nonane	2,2,4-trimethylpentane	12	2.44	0	2.27
n-decane	3-methylpentane	5	10.95	0	11.54
	2,2,4-trimethylpentane	16	5.09	0	5.76
n-dodecane	2-methylpentane	11	7.09	0	7.61
	2,2-dimethylbutane	11	3.84	1	4.14
	2,3-dimethylbutane	11	4.36	0	4.76
	2,4-dimethylpentane	3	4.07	0	4.21
	2,2,4-trimethylpentane	43	4.00	2	2.34
	2,2,4,4,6,8,8-heptamethylnonane	90	1.62	3	1.62
n-tridecane	2,2,4,4,6,8,8-heptamethylnonane	378	2.78	15	2.16
n-tetradecane	2-methylpentane	11	4.52	0	5.27
	2,3-dimethylbutane	11	2.87	0	2.63
	2,2,4-trimethylpentane	20	2.61	0	2.97
n-hexadecane	2-methylpentane	16	7.06	0	8.56
	3-methylpentane	16	7.30	0	8.35
	2,2-dimethylbutane	11	2.28	0	1.98
	2,3-dimethylbutane	11	3.45	0	3.74
	2,4-dimethylpentane	3	4.08	0	3.74
	2,2,4-trimethylpentane	21	3.43	1	2.45
	squalane	2	29.35	0	29.35
	2,2,4,4,6,8,8-heptamethylnonane	45	4.73	0	3.70

Table C.31: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of n-alkanes + cycloalkanes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-pentane	cyclohexane	32	5.26	1	4.86
n-hexane	cyclopentane	13	2.25	0	2.43
	cyclohexane	263	5.08	24	3.61
n-heptane	cyclohexane	127	5.23	15	3.11
	cyclooctane	68	3.39	0	3.48
n-octane	cyclopentane	13	2.59	0	2.07
	cyclohexane	190	2.66	16	1.92
n-nonane	cyclohexane	46	2.91	0	3.09
n-decane	cyclopentane	8	4.21	0	3.30
	cyclohexane	81	5.94	0	3.83
	cyclooctane	22	5.60	1	5.55
n-dodecane	cyclopentane	11	4.53	0	4.32
	cyclohexane	105	2.23	3	1.95
n-tetradecane	cyclopentane	11	2.31	0	2.05
	cyclohexane	170	3.44	6	2.63
n-hexadecane	cyclopentane	13	2.34	0	2.51
	cyclohexane	409	5.18	1	4.87

Table C.32: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of n-alkanes + aromatics.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
methane	benzene	13	14.27	0	14.51
	toluene	55	8.40	2	5.19
ethane	benzene	5	26.06	0	20.12
propane	toluene	46	6.27	4	4.28
n-pentane	benzene	13	3.32	1	2.83
	toluene	12	11.60	0	12.12
n-hexane	benzene	262	10.06	13	3.81
	toluene	217	5.17	3	3.91
n-heptane	benzene	393	4.13	19	2.62
	toluene	98	6.65	0	6.80
n-octane	benzene	48	9.47	0	9.33
	toluene	62	5.71	0	5.28
n-nonane	benzene	7	8.03	0	8.29
n-decane	benzene	46	13.87	0	13.61
	toluene	5	17.34	0	17.46
n-dodecane	benzene	37	18.63	0	18.71
n-tridecane	toluene	8	11.39	1	7.37
n-tetradecane	benzene	285	8.43	20	7.33
	toluene	38	11.43	0	12.69
n-hexadecane	benzene	65	19.46	1	15.09
	toluene	83	6.91	0	6.71
n-octadecane	benzene	3	19.50	0	20.10

Table C.33: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of branched alkanes + branched alkanes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
isobutane	squalane	80	2851.39	16	358.51
squalane	isobutane	80	2851.39	16	358.51

Table C.34: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of cycloalkanes + cycloalkanes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
cyclopentane	cyclohexane	9	2.26	0	2.12
	cyclooctane	10	2.43	1	2.70
cyclohexane	cyclopentane	9	2.26	0	2.12
	cyclooctane	33	4.30	0	3.95
cyclooctane	cyclopentane	10	2.43	1	2.70
	cyclohexane	33	4.30	0	3.95

Table C.35: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of cycloalkanes + branched alkanes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
cyclohexane	2,2,4-trimethylpentane	134	3.20	0	2.08

Table C.36: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of cycloalkanes + aromatics.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
cyclopentane	toluene	157	8.69	1	7.90
cyclohexane	benzene	654	8.49	22	6.46
	toluene	246	10.10	0	10.06
cyclooctane	benzene	31	15.42	0	16.01
	toluene	33	13.51	0	14.56

Table C.37: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of alkenes + alkenes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
1-hexene	1-octene	84	6.14	0	5.54
	1-decene	84	5.71	0	5.62
1-octene	1-hexene	84	6.14	0	5.54
	1-decene	30	2.08	3	1.43
1-decene	1-hexene	84	5.71	0	5.62
	1-octene	30	2.08	3	1.43

Table C.38: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of alkenes + branched alkanes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
ethylene	2,2,4-trimethylpentane	73	51.94	0	54.50
	2,3,4-trimethylpentane	60	50.61	0	51.64
	squalane	118	167.64	10	83.72
	2,2,4,4,6,8,8-heptamethylnonane	85	61.16	3	53.81

Table C.39: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of aromatics + aromatics.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
benzene	toluene	480	5.65	6	4.16
toluene	benzene	480	5.65	6	4.16

Table C.40: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of aromatics + branched alkanes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
benzene	2,4-dimethylpentane	3	7.19	0	7.21
	2,2,4-trimethylpentane	11	4.67	0	5.61

Table C.40: (Continuation)

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
	squalane	15	19.37	0	7.26
toluene	2,2,4-trimethylpentane	24	4.63	0	5.20
	2,2,4,4,6,8,8-heptamethylnonane	90	20.86	1	19.36

Table C.41: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of ester + ester.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-propyl acetate	n-butyl acetate	17	4.99	1	4.78
n-butyl acetate	n-propyl acetate	17	4.99	1	4.78
	isopentyl acetate	18	5.88	0	5.80
isopentyl acetate	n-butyl acetate	18	5.88	0	5.80

Table C.42: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of ester + ether.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-propyl acetate	diethylene glycol dimethyl ether	60	3.50	0	4.04
n-butyl acetate	diethylene glycol dimethyl ether	62	2.30	0	1.97

Table C.43: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of ester + hfc/cfc.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-pentyl acetate	1,1,1-trichloroethane	11	4.48	0	4.86

Table C.44: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of ether + ketones.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
di-n-butyl ether	methyl ethyl ketone	33	2.82	0	2.86
diethyl ether	acetone	12	3.57	1	2.13
	methyl ethyl ketone	7	12.08	0	11.64

Table C.45: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of hfc/cfc + hfc/cfc.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
difluoromethane	pentafluoroethane	605	8.63	14	6.04
	1,1,1,2-tetrafluoroethane	328	6.59	0	5.71
	chloropentafluoroethane	4	11.53	0	11.10
	dichlorodifluoromethane	8	15.37	0	10.86
1,1-difluoroethane	chlorodifluoromethane	162	2.22	0	2.01
	chloropentafluoroethane	8	10.44	0	9.85
	dichlorodifluoromethane	19	7.92	0	7.33
trifluoromethane	chlorotrifluoromethane	6	6.44	0	6.26
pentafluoroethane	difluoromethane	605	8.63	14	6.04
	1,1,1-trifluoroethane	186	5.70	1	5.40
1,1,1-trifluoroethane	pentafluoroethane	186	5.70	1	5.40
1,1,1,2-tetrafluoroethane	difluoromethane	328	6.59	0	5.71
carbon tetrafluoride	chlorodifluoromethane	14	4.21	0	4.39
chlorodifluoromethane	1,1-difluoroethane	162	2.22	0	2.01
	carbon tetrafluoride	14	4.21	0	4.39
	chloropentafluoroethane	10	4.81	0	3.51
	1-chloro-1,1-difluoroethane	236	3.51	11	2.63
	dichlorodifluoromethane	219	7.24	19	5.83
	chlorotrifluoromethane	11	6.55	0	6.20
	bromotrifluoromethane	14	6.43	0	7.09
chloropentafluoroethane	difluoromethane	4	11.53	0	11.10
	1,1-difluoroethane	8	10.44	0	9.85
	chlorodifluoromethane	10	4.81	0	3.51
1-chloro-1,1-difluoroethane	chlorodifluoromethane	236	3.51	11	2.63
dichlorodifluoromethane	difluoromethane	8	15.37	0	10.86
	1,1-difluoroethane	19	7.92	0	7.33
	chlorodifluoromethane	219	7.24	19	5.83
chlorotrifluoromethane	trifluoromethane	6	6.44	0	6.26
	chlorodifluoromethane	11	6.55	0	6.20
bromotrifluoromethane	chlorodifluoromethane	14	6.43	0	7.09

Table C.46: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of hfc/cfc + ketones.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
1,1,1-trichloroethane	acetone	16	2.29	0	2.41
	methyl ethyl ketone	11	4.46	0	4.98
dichlorofluoromethane	acetone	14	21.03	0	21.36

Table C.47: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of n-alkanes + ester.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-hexane	n-butyl acetate	33	4.51	0	3.95
	isopentyl acetate	24	4.43	0	4.33
	n-pentyl acetate	11	4.32	0	4.63
n-heptane	n-propyl acetate	7	1.68	1	1.21
	n-butyl acetate	10	6.44	0	6.83
n-octane	n-butyl acetate	10	5.16	1	5.73
n-nonane	n-butyl acetate	11	4.98	0	5.07
n-decane	n-propyl acetate	13	2.68	0	2.44
	n-butyl acetate	7	6.15	0	6.34
n-dodecane	n-butyl acetate	10	5.80	0	6.20
n-tetradecane	n-propyl acetate	13	1.41	2	0.65
	n-butyl acetate	11	1.78	0	1.54
n-hexadecane	n-butyl acetate	11	1.39	0	1.28

Table C.48: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of n-alkanes + ether.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-hexane	diisopropyl ether	10	6.74	1	6.56
	methyl tert-butyl ether	10	5.14	0	5.40
	diethylene glycol dimethyl ether	55	7.14	0	7.81
n-heptane	diisopropyl ether	10	5.23	4	4.85
	methyl tert-butyl ether	78	1.58	4	0.95
	diethylene glycol dimethyl ether	55	6.14	0	6.62
n-octane	diisopropyl ether	10	4.25	1	3.51
	di-n-butyl ether	67	3.22	1	3.30
	methyl tert-butyl ether	106	2.17	7	1.45
	diethylene glycol dimethyl ether	55	6.19	0	7.03
n-nonane	methyl tert-butyl ether	130	1.19	2	1.20
	diethylene glycol dimethyl ether	33	8.23	0	8.95
n-decane	methyl tert-butyl ether	145	2.10	9	1.83
	diethylene glycol dimethyl ether	33	9.80	0	10.93
n-dodecane	diethylene glycol dimethyl ether	33	7.79	0	8.29
n-tetradecane	diethylene glycol dimethyl ether	33	5.65	0	5.62
n-hexadecane	diethylene glycol dimethyl ether	33	4.01	0	3.07

Table C.49: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of n-alkanes + hfc/cfc.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
methane	carbon tetrafluoride	144	9.97	1	9.92
	chlorodifluoromethane	68	2.95	4	2.39
ethane	carbon tetrafluoride	19	3.36	0	3.31
propane	difluoromethane	128	3.21	0	3.43
	chlorodifluoromethane	21	4.11	0	3.83
	chloropentafluoroethane	21	2.21	0	1.91
n-hexane	1,1,1-trichloroethane	24	4.54	0	3.67
n-heptane	1,1,1-trichloroethane	23	1.85	0	1.86
n-octane	1,1,1-trichloroethane	23	2.25	0	0.85
n-decane	1,1,1-trichloroethane	10	2.09	0	2.29
n-dodecane	1,1,1-trichloroethane	9	3.50	0	3.65
n-hexadecane	1,1,1-trichloroethane	6	7.70	0	7.84

Table C.50: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of n-alkanes + ketones.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-hexane	acetone	11	4.70	1	4.38
	methyl ethyl ketone	11	7.18	0	6.97
n-heptane	acetone	17	1.67	0	1.44
	methyl ethyl ketone	19	4.36	0	6.07
n-nonane	methyl ethyl ketone	9	1.15	1	0.92
n-decane	methyl ethyl ketone	9	1.61	1	1.61
n-dodecane	methyl ethyl ketone	9	4.27	0	4.36
n-hexadecane	methyl ethyl ketone	5	11.44	0	12.08

Table C.51: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of CO2-N2 + n-alkanes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
co2	methane	184	5.24	0	4.49
	ethane	20	4.93	0	5.31
	propane	23	10.40	2	6.38
	n-butane	22	7.27	0	7.85
	n-heptane	63	4.32	6	3.70
	n-decane	18	9.74	5	11.33
	n-tetradecane	12	9.17	0	9.55
	n-octadecane	12	7.00	0	5.06
nitrogen	methane	20	3.26	0	3.34
	ethane	20	6.38	0	6.84
	propane	20	7.30	0	7.94
	n-butane	19	7.98	0	8.27
	n-pentane	16	14.13	0	14.48
	n-heptane	2	14.27	0	14.27

Table C.52: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of ether + branched alkanes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
diisopropyl ether	2,2,4-trimethylpentane	73	3.13	1	3.20
methyl tert-butyl ether	2-methylpentane	11	2.48	0	2.68
	3-methylpentane	11	2.02	0	2.29
	2,2,4-trimethylpentane	20	2.79	0	1.65
diethylene glycol dimethyl ether	2,2,4-trimethylpentane	55	4.34	0	4.79

Table C.53: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of ketones + branched alkanes.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
acetone	2,2,4-trimethylpentane	74	2.97	0	3.15
methyl ethyl ketone	2,2,4-trimethylpentane	63	1.78	0	1.85

Table C.54: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of cycloalkanes + ester.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
cyclohexane	n-propyl acetate	11	6.10	0	6.75
	n-butyl acetate	26	7.52	0	7.28
	isopentyl acetate	26	12.92	1	13.78

Table C.55: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of cycloalkanes + ether.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
cyclohexane	diisopropyl ether	9	5.99	0	5.28
	methyl tert-butyl ether	9	6.05	1	5.88
	diethylene glycol dimethyl ether	33	9.11	0	9.40

Table C.56: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of cycloalkanes + hfc/cfc.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
cyclohexane	1,1,1-trichloroethane	19	3.00	0	3.13

Table C.57: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of cycloalkanes + ketones.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
cyclopentane	acetone	39	3.55	0	3.23
	methyl ethyl ketone	39	1.66	1	1.59
cyclohexane	acetone	69	4.78	0	3.45
	methyl ethyl ketone	90	3.27	9	1.65

Table C.58: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of alkenes + ester.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
ethylene	vinyl acetate	75	7.18	2	6.64

Table C.59: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of aromatics + ester.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
benzene	n-propyl acetate	20	3.39	0	3.38
	n-butyl acetate	34	1.88	1	1.72
	isopentyl acetate	24	3.92	1	3.00
toluene	n-butyl acetate	24	2.23	0	2.12
	isopentyl acetate	24	3.47	1	3.13

Table C.60: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of aromatics + ether.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
benzene	diisopropyl ether	10	9.91	1	10.41
	di-n-butyl ether	26	2.29	0	2.38
	methyl tert-butyl ether	54	9.18	0	9.87
	diethyl ether	26	8.50	2	5.03
toluene	di-n-butyl ether	26	4.11	0	4.28
	methyl tert-butyl ether	13	4.45	2	4.73

Table C.61: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of aromatics + hfc/cfc.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
benzene	1,1,1-trichloroethane	35	5.59	0	3.68
toluene	1,1,1-trichloroethane	35	3.95	1	3.50

Table C.62: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of aromatics + ketones.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
benzene	acetone	90	3.16	6	2.90
	methyl ethyl ketone	173	4.58	6	3.80
toluene	acetone	114	8.28	0	8.11
	methyl ethyl ketone	95	7.02	0	7.74

Table C.63: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of amines + alcohols.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-propylamine	methanol	156	72.14	0	64.41
	ethanol	177	16.96	2	19.59
	1-propanol	166	2.32	1	1.89
	1-butanol	177	3.45	0	2.57
	1-pentanol	22	4.83	0	5.47
	1-heptanol	22	4.30	0	4.09
	1-octanol	22	5.34	0	3.90
	1-decanol	22	16.13	2	17.08
	2-butanol	11	2.35	0	2.61
n-butylamine	methanol	154	13.63	0	13.42
	ethanol	165	4.43	7	3.94
	1-propanol	132	36.66	0	37.97
	1-butanol	195	63.52	0	72.40
	1-pentanol	55	59.61	0	60.56
	1-hexanol	55	73.29	0	75.73
	1-heptanol	55	89.21	0	97.38
	1-octanol	55	120.65	0	114.40
	1-decanol	22	83.19	0	66.68
	2-butanol	20	84.66	0	18.09

Table C.64: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of alcohols + alcohols.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
methanol	ethanol	139	20.88	0	22.58
	1-propanol	36	74.86	0	72.88
	1-butanol	11	137.85	0	135.47
	1-pentanol	14	153.38	0	169.44
	1-hexanol	11	204.03	0	185.73
	isopropanol	51	224.13	0	192.57
	2-butanol	11	363.99	0	272.56
ethanol	methanol	139	20.88	0	22.58
	1-propanol	88	12.77	0	13.51
	1-butanol	143	27.41	0	29.87
	1-pentanol	143	27.23	0	29.02
	1-hexanol	20	32.09	0	33.94
	1-heptanol	55	36.26	0	36.55
	1-octanol	65	57.41	0	61.13
	1-nonanol	55	18.46	0	18.15
	1-decanol	55	29.44	0	28.16
	isopropanol	79	36.59	0	35.09
	2-butanol	44	55.49	0	50.24
1-propanol	methanol	36	74.86	0	72.88
	ethanol	88	12.77	0	13.51
	1-butanol	42	3.62	0	4.22
	1-pentanol	54	2.55	1	2.54
	1-heptanol	22	2.03	4	1.75
	1-nonanol	22	11.23	0	12.37
	1-undecanol	22	12.56	0	7.58
	isopropanol	61	7.80	1	8.26
	2-butanol	14	1179.41	2	1492.59
1-butanol	methanol	11	137.85	0	135.47
	ethanol	143	27.41	0	29.87
	1-propanol	42	3.62	0	4.22
	1-pentanol	32	0.78	1	0.63

Table C.64: (Continuation)

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
	1-nonanol	32	11.61	0	13.53
	1-decanol	265	12.06	6	11.39
	isopropanol	10	4.25	0	4.35
	2-butanol	117	7.63	12	7.77
1-pentanol	methanol	14	153.38	0	169.44
	ethanol	143	27.23	0	29.02
	1-propanol	54	2.55	1	2.54
	1-butanol	32	0.78	1	0.63
	1-heptanol	22	4.88	4	5.52
	1-octanol	32	2.29	1	1.94
	1-nonanol	22	12.36	0	14.26
	1-undecanol	22	14.89	0	14.37
1-hexanol	methanol	11	204.03	0	185.73
	ethanol	20	32.09	0	33.94
1-heptanol	ethanol	55	36.26	0	36.55
	1-propanol	22	2.03	4	1.75
	1-pentanol	22	4.88	4	5.52
	1-octanol	32	2.40	1	2.49
	1-nonanol	22	7.48	0	7.15
	1-undecanol	22	18.06	0	18.21
1-octanol	ethanol	65	57.41	0	61.13
	1-pentanol	32	2.29	1	1.94
	1-heptanol	32	2.40	1	2.49
	1-nonanol	77	5.93	0	5.33
	1-decanol	77	13.41	0	12.67
	1-undecanol	55	15.71	0	15.87
1-nonanol	ethanol	55	18.46	0	18.15
	1-propanol	22	11.23	0	12.37
	1-butanol	32	11.61	0	13.53
	1-pentanol	22	12.36	0	14.26
	1-heptanol	22	7.48	0	7.15
	1-octanol	77	5.93	0	5.33
	1-decanol	93	14.96	0	14.59
	1-undecanol	66	16.88	0	15.57
1-decanol	ethanol	55	29.44	0	28.16
	1-butanol	265	12.06	6	11.39
	1-octanol	77	13.41	0	12.67
	1-nonanol	93	14.96	0	14.59
	1-undecanol	87	21.18	0	21.08
1-undecanol	1-propanol	22	12.56	0	7.58
	1-pentanol	22	14.89	0	14.37
	1-heptanol	22	18.06	0	18.21
	1-octanol	55	15.71	0	15.87
	1-nonanol	66	16.88	0	15.57
	1-decanol	87	21.18	0	21.08
isopropanol	methanol	51	224.13	0	192.57
	ethanol	79	36.59	0	35.09
	1-propanol	61	7.80	1	8.26
	1-butanol	10	4.25	0	4.35
	2-butanol	44	6.22	0	4.78
2-butanol	methanol	11	363.99	0	272.56
	ethanol	44	55.49	0	50.24
	1-propanol	14	1179.41	2	1492.59
	1-butanol	117	7.63	12	7.77
	isopropanol	44	6.22	0	4.78

Table C.65: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of n-alkanes + amines.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-hexane	n-butylamine	21	28.43	0	30.88
n-heptane	n-butylamine	33	30.87	0	28.71

Table C.66: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of cycloalkanes + amines.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
cyclohexane	n-propylamine	22	3.25	2	3.37
	n-butylamine	33	20.42	7	3.73
	n-hexylamine	9	3.03	0	2.80
	n-octylamine	11	70.95	0	75.39
	n-decylamine	7	39.09	0	45.31

Table C.67: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of aromatics + amines.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
benzene	n-propylamine	33	5.78	6	3.22
	n-butylamine	28	7.50	4	4.94
	n-hexylamine	9	5.75	0	5.44
	n-octylamine	9	13.48	0	14.65
	n-decylamine	7	3.29	0	3.01
toluene	n-propylamine	22	6.13	0	4.15
	n-butylamine	11	1.43	0	1.00

Table C.68: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of n-alkanes + alcohols.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
methane	methanol	56	17.78	3	7.79
	1-butanol	12	18.20	0	19.09
	isopropanol	12	18.64	0	20.05
ethane	methanol	54	16.12	8	10.86
	ethanol	9	8.24	0	8.32
	1-butanol	8	27.89	0	28.91
	isopropanol	9	13.05	0	12.04
propane	methanol	48	21.32	0	15.18
n-pentane	methanol	45	32.84	0	34.13
	ethanol	12	9.10	0	7.65
	1-propanol	12	6.15	0	5.96
	1-butanol	20	10.57	0	10.36
	isopropanol	12	30.75	1	12.40
n-hexane	methanol	13	29.94	0	17.16
	ethanol	288	21.17	0	21.59
	1-propanol	155	10.10	0	10.40
	1-butanol	49	9.89	0	10.46

Table C.68: (Continuation)

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
	1-pentanol	92	18.54	0	12.21
	1-hexanol	269	6.30	10	4.73
	1-heptanol	22	10.40	0	8.70
	1-octanol	27	12.36	0	13.34
	1-decanol	49	7.21	1	5.51
	isopropanol	35	9.15	0	6.28
	2-butanol	18	23.43	0	23.74
n-heptane	methanol	14	29.98	0	21.80
	ethanol	447	17.68	0	18.04
	1-propanol	255	8.22	0	7.89
	1-butanol	41	5.84	1	5.16
	1-pentanol	24	8.51	0	8.42
	1-hexanol	24	5.70	0	5.80
	1-heptanol	57	5.78	2	4.50
	1-octanol	28	5.08	0	3.99
	1-decanol	49	7.43	2	7.05
	1-dodecanol	28	17.21	0	11.10
	isopropanol	10	3.28	2	1.80
n-octane	methanol	10	14.24	1	2.25
	ethanol	176	30.15	0	32.28
	1-propanol	232	7.63	0	7.40
	1-butanol	104	6.38	0	6.56
	1-pentanol	116	7.43	0	6.83
	1-hexanol	25	8.35	0	7.42
	1-heptanol	44	26.39	7	27.70
	1-octanol	27	10.37	0	9.03
	1-decanol	31	7.64	0	5.79
	isopropanol	68	7.53	0	5.11
	2-butanol	39	14.45	0	11.61
	2-pentanol	39	18.72	0	16.75
	2-octanol	21	48.16	0	29.24
n-nonane	1-decanol	14	6.70	2	4.23
n-decane	ethanol	103	49.80	0	57.17
	1-propanol	13	3.07	1	3.48
	1-butanol	16	4.87	0	5.05
	1-hexanol	16	5.26	0	4.61
	1-octanol	16	8.27	0	8.51
	1-decanol	29	7.61	0	6.42
	isopropanol	71	5.01	0	4.18
	2-butanol	39	9.51	0	9.26
	2-pentanol	39	13.06	0	11.15
	2-heptanol	21	62.09	0	59.55
	2-octanol	21	46.57	0	31.28
n-dodecane	1-hexanol	21	32.62	0	31.98
	1-heptanol	21	63.34	0	55.66
	1-octanol	21	25.63	0	17.93
	isopropanol	39	6.03	0	5.70
	2-butanol	39	6.66	0	6.38
	2-pentanol	39	8.90	0	5.83
	2-heptanol	21	28.10	0	33.77
	2-octanol	42	21.43	0	25.38
n-tetradecane	1-hexanol	21	17.66	0	14.49
	1-octanol	21	12.66	0	10.93
	1-decanol	21	19.39	0	17.09
	1-dodecanol	21	24.15	0	23.68
	2-octanol	13	7.56	0	4.68
n-hexadecane	1-butanol	108	8.91	0	8.77
	1-pentanol	19	10.77	0	11.12
	1-hexanol	36	14.46	0	16.61
	1-heptanol	33	16.00	0	16.62

Table C.68: (Continuation)

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
	1-octanol	48	6.38	0	5.14
	isopropanol	48	5.21	0	4.67

Table C.69: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of cycloalkanes + alcohols.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
cyclopentane	ethanol	39	12.37	0	12.77
	1-propanol	111	9.15	9	7.68
	1-butanol	74	5.57	0	3.92
	1-pentanol	67	9.67	0	5.31
	1-heptanol	79	4.13	0	3.16
	isopropanol	77	11.37	0	11.20
	2-butanol	74	18.44	0	17.62
	2-pentanol	39	25.36	0	22.51
cyclohexane	methanol	216	83.24	0	61.87
	ethanol	224	28.27	0	26.56
	1-propanol	114	6.76	0	6.22
	1-butanol	174	4.49	0	4.47
	1-pentanol	59	6.09	0	4.36
	1-hexanol	44	9.29	0	9.32
	1-heptanol	11	8.69	0	8.43
	1-octanol	33	10.08	0	10.28
	1-nonanol	11	4.47	0	4.50
	1-decanol	22	5.88	0	6.16
	1-dodecanol	14	22.11	0	24.03
	isopropanol	57	8.38	0	7.42
	2-butanol	81	13.17	0	8.43
	2-pentanol	50	17.74	0	11.03
	2-octanol	12	11.90	0	13.04

Table C.70: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of alkenes + alcohols.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
ethylene	methanol	7	23.84	0	24.48
	1-pentanol	75	12.69	0	11.34
	1-heptanol	75	15.14	1	11.47
	1-nonanol	100	14.90	1	13.07
1-hexene	1-butanol	33	11.08	0	12.68

Table C.71: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of aromatics + alcohols.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
benzene	methanol	222	73.51	8	30.29
	ethanol	117	23.04	1	24.73
	1-propanol	72	14.40	0	14.52

Table C.71: (Continuation)

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
	1-butanol	116	18.52	0	19.40
	1-pentanol	24	22.91	0	24.25
	1-hexanol	55	24.13	0	25.30
	isopropanol	79	298.79	1	25.12
	2-butanol	20	40.55	0	47.08
	2-octanol	32	36.49	0	33.24
toluene	methanol	152	73.57	0	73.15
	ethanol	336	16.94	2	16.05
	1-propanol	246	12.22	0	10.61
	1-butanol	94	16.70	0	16.87
	1-pentanol	112	17.83	0	19.32
	1-hexanol	55	16.74	0	17.87
	1-heptanol	94	15.30	0	12.04
	1-octanol	126	21.27	0	22.28
	2-pentanol	55	33.79	0	34.10
	2-heptanol	42	28.18	0	27.72
	2-octanol	42	37.04	0	39.30

Table C.72: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of ester + amines.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-butyl acetate	n-butylamine	10	28.54	0	28.31
isopentyl acetate	n-butylamine	10	43.99	0	45.21

Table C.73: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of ketones + amines.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
acetone	n-propylamine	11	25.07	0	30.14
methyl ethyl ketone	n-butylamine	10	33.18	0	37.29

Table C.74: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of aldehydes + alcohols.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
butanal	ethanol	12	64.03	3	60.65
2-methylpropanal	1-butanol	73	27.93	8	22.17
decanal	ethanol	33	17.27	0	20.12

Table C.75: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of ester + alcohols.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
n-butyl acetate	methanol	16	86.86	0	75.73
	ethanol	77	25.39	0	24.72
	1-propanol	75	10.05	7	4.98
	1-butanol	69	16.77	9	7.26
	1-pentanol	55	34.35	7	9.72
	isopropanol	11	42.94	0	20.84
isopentyl acetate	ethanol	84	30.38	0	31.22
vinyl acetate	ethanol	11	33.01	0	36.51
	1-butanol	11	24.16	0	27.60
	1-octanol	11	22.42	0	24.54
	1-decanol	11	10.81	0	11.59
n-pentyl acetate	methanol	12	168.80	0	165.37
	ethanol	16	26.23	0	24.61
	1-propanol	16	9.58	0	9.86
	1-butanol	60	12.36	0	11.02
	1-pentanol	60	13.32	1	12.47
	1-hexanol	21	26.50	0	25.99
	1-heptanol	16	22.08	0	22.96
	1-octanol	16	21.86	0	21.52

Table C.76: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of ether + alcohols.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
diisopropyl ether	ethanol	126	27.26	0	29.64
di-n-butyl ether	1-propanol	66	17.17	0	19.21
methyl tert-butyl ether	methanol	36	64.64	0	65.64
	ethanol	147	5.64	0	4.34
	1-propanol	172	8.94	18	4.80
	1-butanol	163	6.13	0	5.94
	1-pentanol	171	5.21	0	4.75
	1-hexanol	148	2.15	14	1.37
	isopropanol	152	7.74	0	7.02
	2-butanol	11	6.09	0	4.41
diethyl ether	methanol	19	77.42	0	88.87
	ethanol	9	27.48	0	36.63
	1-propanol	20	17.25	0	17.71
	1-butanol	18	15.45	0	15.79
	1-pentanol	11	3.04	0	3.00
diethylene glycol dimethyl ether	1-propanol	121	20.08	0	20.82
	1-butanol	33	16.74	0	18.46

Table C.77: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of hfc/cfc + alcohols.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
1,1,1-trichloroethane	methanol	27	129.33	0	106.43
	ethanol	44	27.76	0	28.62
	1-propanol	11	14.06	0	11.30
	1-butanol	11	12.15	0	12.74
	1-hexanol	10	11.94	0	12.84

	1-heptanol	48	11.19	0	12.41
	2-butanol	11	28.97	0	32.09

Table C.78: Deviations of predicted mixture viscosities to experimental data as $AAD\%$ of ketones + alcohols.

Component 1	Component 2	N_{dat}	AAD %	$N_{outliers}$	Median
acetone	methanol	149	36.93	0	35.90
	ethanol	105	18.69	0	19.94
	1-propanol	18	24.77	0	28.62
	1-butanol	64	22.92	0	16.82
	1-hexanol	5	26.91	1	31.13
	isopropanol	18	28.91	0	28.88
methyl ethyl ketone	methanol	38	34.01	0	40.10
	ethanol	63	16.78	0	19.66
	1-butanol	147	17.85	0	16.81
	1-pentanol	36	14.47	0	14.61
	isopropanol	126	18.78	0	16.12
	2-butanol	176	35.10	1	25.54
	2-pentanol	84	19.26	2	16.73
	2-heptanol	48	22.18	0	22.55

C.5 Applied PCP-SAFT parameters

Table C.79: Applied pure component PCP-SAFT parameters

Substance	m	σ/Å	(ε/k)/K	κ^{AB}	(ε^{AB}/k)/K	DM/D	Q/D	Source
Methane	1	3.7039	150.03	0	0	0	0	105
Ethane	1.6069	3.5206	191.42	0	0	0	0	105
Propane	2.002	3.6184	208.11	0	0	0	0	105
Butane	2.3316	3.7086	222.88	0	0	0	0	105
Pentane	2.6896	3.7729	231.2	0	0	0	0	105
Hexane	3.0576	3.7983	236.77	0	0	0	0	105
Heptane	3.4831	3.8049	238.4	0	0	0	0	105
Octane	3.8176	3.8373	242.78	0	0	0	0	105
Nonane	4.2079	3.8448	244.51	0	0	0	0	105
Decane	4.6627	3.8384	243.87	0	0	0	0	105
Undecane	4.9082	3.8893	248.82	0	0	0	0	105
Dodecane	5.306	3.8959	249.21	0	0	0	0	105
Tridecane	5.6877	3.9143	249.78	0	0	0	0	105
Tetradecane	5.9002	3.9396	254.21	0	0	0	0	105
Pentadecane	6.2855	3.9531	254.14	0	0	0	0	105
Hexadecane	6.6485	3.9552	254.7	0	0	0	0	105
Heptadecane	6.9809	3.9675	255.65	0	0	0	0	105
Octadecane	7.3271	3.9668	256.2	0	0	0	0	105
Nonadecane	7.7175	3.9721	256	0	0	0	0	105
Eicosane	7.9849	3.9869	257.75	0	0	0	0	105
N-heneicosane	8.7495	3.9192	252.3877	0	0	0	0	235
N-docosane	8.9886	3.9466	253.9514	0	0	0	0	235
N-tricosane	9.3851	3.9416	254.1439	0	0	0	0	235

Table C.79: Continuation

Substance	m	σ/Å	(ε/k)/K	κ^{AB}	(ε^{AB}/k)/K	DM/D	Q/D	Source
N-tetracosane	9.7959	3.9384	253.5506	0	0	0	0	235
N-octacosane	11.4145	3.9318	252.3643	0	0	0	0	235
N-triacontane	11.4888	4.0142	258.3031	0	0	0	0	235
N-dotriacontane	12.1112	4.0303	258.4649	0	0	0	0	235
N-hexatriacontane	13.0294	4.0906	261.2998	0	0	0	0	235
Acetaldehyde	2.1188	3.246	229.8505	0	0	2.6891	0	235
Propanal	2.6001	3.2872	235.21	0	0	2.72	0	110
Butanal	2.8825	3.4698	247.09	0	0	2.72	0	110
2-methylpropanal	2.6157	3.5699	251.5072	0	0	2.6981	0	235
3-methylbutyraldehyde	3.0004	3.6188	255.2047	0	0	2.6082	0	235
Heptanal	3.8527	3.66	260.6984	0	0	2.5782	0	235
Decanal	5.2885	3.6634	251.9655	0	0	2.5782	0	235
Methylamine	1.6191	3.337	246.3797	0.03	1103.6088	1.3101	0	235
N-propylamine	3.4083	3.1114	217.4541	0.03	86.0572	1.1692	0	235
N-butylamine	1.7814	4.2631	293.8144	0.03	1497.5173	1.391	0	235
N-pentylamine	1.6177	4.656	302.1019	0.03	1923.3059	1.5499	0	235
N-hexylamine	3.9434	3.5463	247.4674	0.03	82.5108	1.5889	0	235
N-heptylamine	1.688	5.0417	328.7725	0.03	2220.2939	1.6608	0	235
N-octylamine	1.9349	4.9721	315.3046	0.03	2476.9879	1.421	0	235
N-nonylamine	2.0093	5.07	314.0984	0.03	2731.5672	0.8994	0	235
N-decylamine	2.4141	4.9185	313.772	0.03	2644.5966	1.409	0	235
Diethylamine	1.3688	4.7262	295.2136	0.03	1547.4422	0.9204	0	235
Di-n-propylamine	3.6653	3.6551	240.4467	0.03	39.2881	1.07	0	235
N-dodecylamine	5.812	3.817	260.8335	0.03	53.2229	0.8814	0	235
Ethylamine	1.6164	3.7728	263.487	0.03	1213.4586	1.2201	0	235
Dimethylamine	2.2493	3.3738	226.4848	0.03	959.5053	1.0313	0	235

Table C.79: Continuation

Substance	m	σ/Å	(ε/k)/K	κ^{AB}	(ε^{AB}/k)/K	DM/D	Q/D	Source
Trioctylamine	11.35	3.7563	227.9647	0	0	0.8004	0	[235]
Ethylene	1.5425	3.4523	179.37	0	0	0	2	[111]
Propylene	1.9597	3.5356	207.19	0	0	0	0	[105]
1-hexene	2.9853	3.7753	236.81	0	0	0	0	[105]
1-heptene	3.2863	3.8138	243.7559	0	0	0.6296	0	[235]
1-octene	3.7424	3.8133	243.02	0	0	0	0	[105]
1-nonene	4.0617	3.8452	247.0709	0	0	0.5996	0	[235]
1-decene	4.4503	3.8544	248.2291	0	0	0.4197	0	[235]
Methanol	1.5073	3.3248	211.5975	0.03	2519.7116	1.6998	0	[235]
Ethanol	2.3723	3.1964	203.8245	0.03	2514.0609	1.6908	0	[235]
1-propanol	3.4604	3.0742	217.3745	0.03	2044.5298	1.6788	0	[235]
1-butanol	4.2768	3.0659	217.7082	0.03	1846.2029	1.6698	0	[235]
1-pentanol	4.305	3.2421	228.6214	0.03	1835.0315	1.7	0	[235]
1-hexanol	4.3974	3.3897	235.5309	0.03	1930.6231	1.6488	0	[235]
1-heptanol	4.7248	3.461	237.6069	0.03	1956.8424	1.7388	0	[235]
1-octanol	5.7204	3.3562	228.8515	0.03	1869.8446	1.6488	0	[235]
1-nonanol	4.749	3.729	253.6291	0.03	2073.3555	1.6099	0	[235]
1-decanol	5.7328	3.596	243.7089	0.03	1967.8312	1.6189	0	[235]
1-undecanol	7.3595	3.38	225.1381	0.03	1888.8681	1.6698	0	[235]
1-dodecanol	6.2353	3.6999	249.0952	0.03	2002.9613	1.6908	0	[235]
1-tridecanol	5.5859	3.9862	265.7028	0.03	2212.4364	1.6488	0	[235]
1-tetradecanol	5.9103	3.9832	266.3354	0.03	2238.1838	1.55	0	[235]
1-hexadecanol	7.0917	3.8845	259.4044	0.03	2140.2545	1.6698	0	[235]
1-octadecanol	7.3343	3.9962	266.5143	0.03	2150.7482	1.6608	0	[235]
Isopropanol	4.0249	2.918	198.5953	0.03	1871.8788	1.6608	0	[235]
2-butanol	4.4207	3.0151	205.1808	0.03	1718.0488	1.6608	0	[235]

Table C.79: Continuation

Substance	m	σ/Å	(ε/k)/K	κ^{AB}	(ε^{AB}/k)/K	DM/D	Q/D	Source
2-pentanol	5.2839	3.0076	200.0793	0.03	1713.2429	1.6668	0	[235]
3-pentanol	4.7983	3.0947	203.9658	0.03	1806.0011	1.6399	0	[235]
2-hexanol	4.7812	3.2816	217.9545	0.03	1793.3615	1.6578	0	[235]
2-heptanol	5.135	3.3462	219.5951	0.03	1878.5922	1.6488	0	[235]
2-octanol	5.7545	3.3545	220.9018	0.03	1721.3869	1.6488	0	[235]
Ethylene glycol	2.6124	3.1536	310.114	0.03	2711.6591	2.4103	0	[235]
Benzene	2.2463	3.7852	296.24	0	0	0	5.5907	[111]
Toluene	2.8149	3.7169	285.69	0	0	0	0	[105]
Ethylbenzene	3.0799	3.7974	287.35	0	0	0	0	[105]
1,3-dimethylbenzene	3.1861	3.7563	283.98	0	0	0	0	[105]
1,4-dimethylbenzene	3.1723	3.7781	283.77	0	0	0	0	[105]
Mesitylene	3.671	3.7284	277.6576	0	0	0	0	[235]
P-cymene	3.6373	3.907	284.1096	0	0	0	0	[235]
N-heptylbenzene	4.5229	3.9994	289.0709	0	0	0	0	[235]
N-nonylbenzene	5.6327	3.9104	275.8787	0	0	0	0	[235]
N-dodecylbenzene	6.8535	3.9041	270.8303	0	0	0	0	[235]
Cumene	3.3779	3.8376	281.5324	0	0	0.3897	0	[235]
Butylbenzene	3.7662	3.8727	283.07	0	0	0	0	[105]
Propylbenzene	3.3438	3.8438	288.13	0	0	0	0	[105]
Isobutane	2.2616	3.7574	216.53	0	0	0	0	[105]
Isopentane	2.562	3.8296	230.75	0	0	0	0	[105]
Neopentane	2.3543	3.955	225.69	0	0	0	0	[105]
2-methylpentane	2.9317	3.8535	235.58	0	0	0	0	[105]
3-methylpentane	2.8852	3.8605	240.48	0	0	0	0	[105]
2,2-dimethylbutane	2.6008	4.0042	243.51	0	0	0	0	[105]
2,3-dimethylbutane	2.6853	3.9545	246.07	0	0	0	0	[105]

Table C.79: Continuation

Substance	m	σ/Å	(ε/k)/K	κ^{AB}	(ε^{AB}/k)/K	DM/D	Q/D	Source
3-ethylpentane	3.1886	3.887	247.0555	0	0	0	0	235
2,4-dimethylpentane	3.1235	3.9346	239.7005	0	0	0	0	235
2,2,4-trimethylpentane	3.1352	4.0918	250.0059	0	0	0	0	235
2,3,4-trimethylpentane	3.198	4.025	258.5959	0	0	0	0	235
Squalane	9.734	4.2525	263.7664	0	0	0	0	235
2,2,4,4,6,8,8-heptamethylnonane	5.017	4.3017	266.5519	0	0	0	0	235
Cyclopentane	2.3655	3.7114	265.83	0	0	0	0	105
Cyclohexane	2.5303	3.8499	278.11	0	0	0	0	105
Cyclooctane	2.9236	3.9831	304.8636	0	0	0	0	235
Propyl ethanoate	3.7658	3.4289	235.42	0	0	1.78	0	110
Butyl ethanoate	3.9629	3.5482	242.27	0	0	1.87	0	110
Isopentyl acetate	4.0348	3.6672	247.4168	0	0	1.7987	0	235
Vinyl acetate	3.3946	3.2692	232.2234	0	0	1.7897	0	235
N-pentyl acetate	4.1066	3.646	249.3581	0	0	1.7208	0	235
N-heptyl acetate	5.3414	3.5818	240.1371	0	0	1.8707	0	235
Dimethyl ether	2.2634	3.2723	210.29	0	0	1.3	0	110
Diisopropyl ether	3.5466	3.6737	216.8508	0	0	1.1302	0	235
Di-n-butyl ether	4.4126	3.6997	233.9837	0	0	1.1692	0	235
Methyl tert-butyl ether	2.9286	3.711	233.9573	0	0	1.361	0	235
Di-n-pentyl ether	4.8197	3.8391	246.5082	0	0	1.1992	0	235
Di-n-propyl ether	3.4881	3.6961	234.0444	0	0	1.2112	0	235
Diethyl ether	2.9726	3.5127	219.53	0	0	1.15	0	110
Diethylene glycol dimethyl ether	4.2282	3.5722	258.4078	0	0	1.97	0	235
Acetone	2.7447	3.2742	232.99	0	0	2.88	0	110

Table C.79: Continuation

Substance	m	σ/Å	(ε/k)/K	κ^{AB}	(ε^{AB}/k)/K	DM/D	Q/D	Source
3-pentanone	3.2786	3.5159	248.69	0	0	2.82	0	[110]
Methyl isobutyl ketone	3.473	3.6296	250.53	0	0	2.6891	0	[235]
4-heptanone	3.7347	3.6958	259.1972	0	0	2.5002	0	[235]
2-hexanone	3.5951	3.5814	254.4017	0	0	2.6801	0	[235]
2-octanone	4.467	3.6222	252.4353	0	0	2.4613	0	[235]
Methyl ethyl ketone	2.9835	3.4239	244.99	0	0	2.78	0	[110]
Acetic acid	0.9907	4.4948	316.3575	0.03	2575.4822	1.7388	0	[235]
Propionic acid	1.7652	3.8743	223.4035	0.03	3336.825	1.7508	0	[235]
N-butyric acid	2.3552	3.774	227.9723	0.03	3392.9358	1.6488	0	[235]
Carbon dioxide	1.5131	3.1869	163.33	0	0	0	4.4	[111]
Nitrogen	1.1504	3.3848	91.4	0	0	0	1.43	[111]
Difluoromethane	2.4382	2.8148	162.6954	0	0	1.9786	0	[235]
1,1-difluoroethane	2.5999	3.1426	179.9699	0	0	2.2619	0	[235]
Trifluoromethane	2.5662	2.8588	140.9286	0	0	1.6488	0	[235]
Pentafluoroethane	3.0838	3.1297	154.4771	0	0	1.5409	0	[235]
1,1,1-trifluoroethane	2.4877	3.2839	161.4797	0	0	2.3204	0	[235]
1,1,1,2,3,3-hexafluoropropane	3.681	3.1918	177.2769	0	0	0.2009	0	[235]
1,1,1,2,3,3,3-heptafluoropropane	3.5407	3.301	160.9843	0	0	1.9396	0	[235]
1,1,1,3,3-pentafluoropropane	3.5888	3.1797	184.7	0	0	1.5559	0	[235]
1,1,1,3,3,3-hexafluoropropane	3.6025	3.2428	170.2242	0	0	1.9816	0	[235]
1,1,1,2-tetrafluoroethane	3.1242	3.0549	165.8855	0	0	2.0581	0	[235]
Carbon tetrafluoride	2.1861	3.1358	122.3822	0	0	0	0	[235]

Table C.79: Continuation

Substance	m	σ/Å	(ε/k)/K	κ^{AB}	(ε^{AB}/k)/K	DM/D	Q/D	Source
Chlorodifluoromethane	2.4408	3.1487	185.7335	0	0	1.4579	0	[235]
Chloropentafluoroethane	2.7628	3.5063	170.7613	0	0	0.5186	0	[235]
1,1,1-trichloroethane	2.4793	3.7527	275.4749	0	0	1.7808	0	[235]
Dichlorofluoromethane	2.4475	3.3353	229.5168	0	0	1.2891	0	[235]
Isobutene	2.2515	3.6636	223.2165	0	0	0.5006	0	[235]
Trichlorofluoromethane	2.2845	3.6905	248.8305	0	0	0.4497	0	[235]
1-chloro-1,1-difluoroethane	2.4533	3.4805	202.0841	0	0	2.1405	0	[235]
Dichlorodifluoromethane	2.2262	3.5608	205.6771	0	0	0.5096	0	[235]
Chlorotrifluoromethane	2.1909	3.3787	162.3126	0	0	0.5096	0	[235]
1,2-dichlorotetrafluoroethane	2.7766	3.6468	202.3572	0	0	0.6685	0	[235]
Bromotrifluoromethane	2.1338	3.5126	185.5387	0	0	0.6505	0	[235]

D. Supporting information to chapter 6: An optimal experimental design study for predicting the viscosity of pure substances using a group contribution model

D.1 Summary of the simulation parameters

For PreDIA the preposterior expectation is based on a certain number of N_{obs} of possible measurement observations and N_{MC} realizations. The sample-sizes depend on the example application and were chosen such that converged calculations are ensured, see tab. D.1.

Table D.1: Number of observations and realizations in both investigated example applications

Example application	N_{obs}	N_{MC}
ExA^{press}	2000	50000
ExA^{gc}	4000	100000

D.2 Expected reduction of model uncertainty versus reduction of model error

In fig. D.1 we present the sum of the squared errors (SSE) of the model as

$$SSE = \sum_{i=1}^{n_{exp}} \left(\eta_i^{calc} - \eta_i^{meas}\right)^2 \tag{D.1}$$

Here, n_{exp} is the number of data points that are available in the Dortmund Datenbank[153]. As compared to all previous measures of accuracy, this value is not a measure for the model variance but the model deviation from experimental data. The PreDIA does not aim at decreasing the deviation between model predictions and experimental data but to decrease the variance in the model's predictions. Nevertheless, if the error of the experimental data is low compared to the model error, decreasing the model variance should also decrease the deviations of the model's predictions compared to experimental data.

The viscosity parameters θ used for each number of additional data points n_{design} were separately adjusted to the initial experimental viscosity values $\eta_{\text{train}}^{meas}$ and the experimental viscosity values at the proposed optimal design point $\eta^{meas}(d_{opt}(n_{design}))$. We find that already after two additional data points, the accuracy of the model compared to experimental data does not improve any further. On the one hand, this indicates the model to be very robust. On the other hand, this was also the result shown in fig. 6.5. Hence, we conclude that for this example application already five experimental data points (of which three were chosen by optimization) suffice to parametrize the viscosity model for the complete, technically accessible range of pressure and temperature.

D.3 Evaluation of convergence and sample-size

To quickly assess if sample-sizes can be the cause for the deviations in the objective functions ϕ_{MCMC} and ϕ_{PreDIA}, calculations were performed with various number of realizations (i.e. N_{MC}) and various number of observations (i.e. number of hypothetical measurement outcomes, N_{obs}).

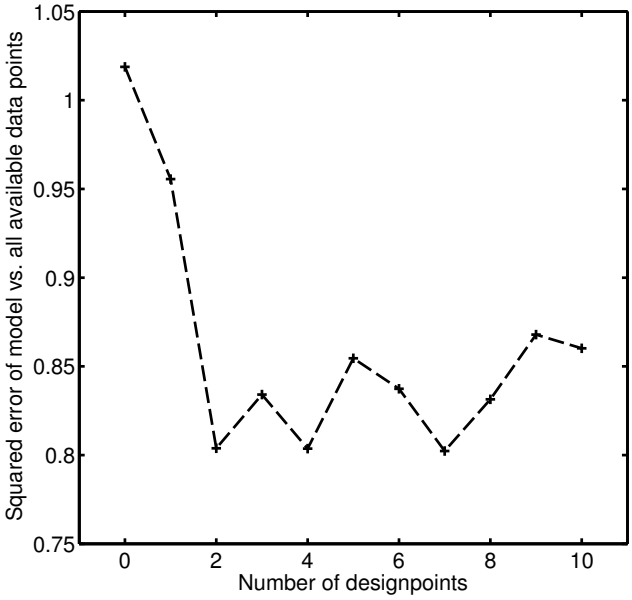


Figure D.1: Squared error of calculated viscosities versus experimental data. The model was parametrized on experimental data at the conditions suggested by the presented PreDIA procedure (ExA^{press}) with increasing number of experimental data points. All available data from the Dortmund Datenbank[153] was used and can also be found in[48].

D.3.1 Example application ExA^{press}

Over the number of optimized design-points, sample-sizes decrease (see fig. D.2). The relative decrease of sample-size is comparable for the different number of used realizations. Interestingly, the resulting objective functions do not depend on the number of realizations (see fig. D.3). This suggests, that the number of realizations is sufficiently large to be representative for the model variance.

Another cause for non-converged results can be an insufficient number of observations (i.e. hypothetical measurement outcomes), N_{obs}. To assess this possible issue, calculations with $N_{obs} = 2000$ and $N_{obs} = 4000$ observations and $N_{MC} = 50000$ realizations were carried out. Sample-sizes over the number of optimized design-points only weakly vary with the number of observations (see fig. D.4). No effect on the objective functions (see fig. D.5) was found.

Overall, the evaluation of sample-sizes for the example application ExA^{press} shows, that the observed deviations in the objective functions (see fig. 6.4) are not caused by insufficiently small sample-sizes or number of observations.

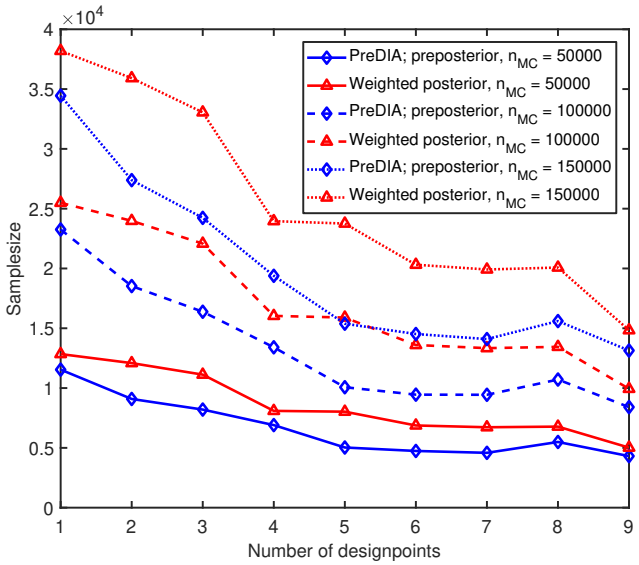


Figure D.2: Sample-sizes over the number of optimized designs with $N_{obs} = 2000$ observations but different number of realizations N_{MC} in example application S_{Press}

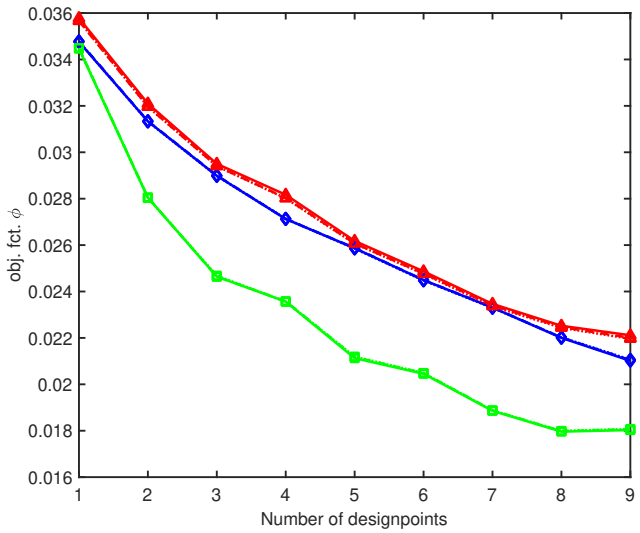


Figure D.3: Objective functions (ϕ_{PreDIA} in blue, ϕ_{wPost} in red and ϕ_{MCMC} in green) over the number of optimized designs with $N_{obs} = 2000$ observations but different number of realizations N_{MC} in example application S_{Press}. Results from different N_{MC} are shown as dotted and dashed lines, though, differences in the results are barely distinguishable.

D.3.2 Example application ExA^{gc}

Over the number of optimized design-points, sample-sizes decrease (see fig. D.6). The relative decrease of sample-size is significantly higher when starting

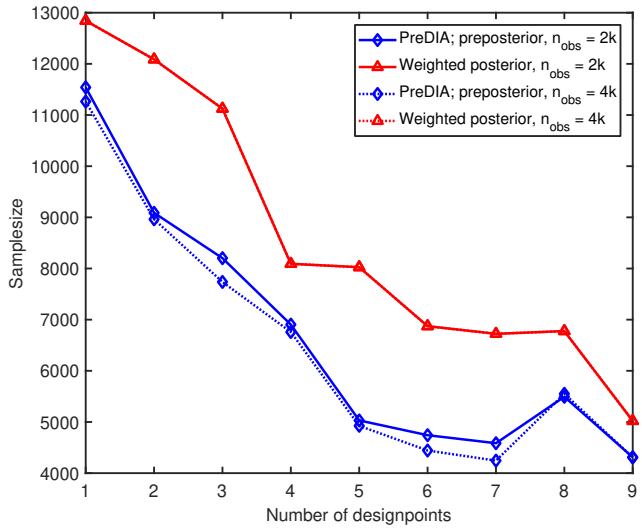


Figure D.4: Sample-sizes over the number of optimized designs with $N_{MC} = 50000$ realizations but different number of observations N_{obs} in example application S_{Press}

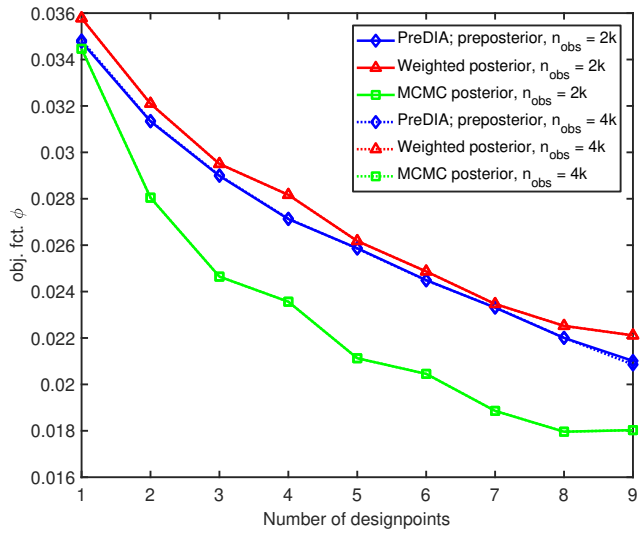


Figure D.5: Objective functions (ϕ_{PreDIA} in blue, ϕ_{wPost} in red and ϕ_{MCMC} in green) over the number of optimized designs with $N_{MC} = 50000$ realizations but different number of observations N_{obs} in example application S_{Press}

with $N_{MC} = 100000$ realizations as compared to $N_{MC} = 10000$ realizations. Interestingly, the resulting objective functions do not change significantly (see fig. D.7). This suggests, that the number of realizations is sufficiently large to be representative for the model variance.

As before (see chapter D.3.1) calculations with varying number of observations, $N_{obs} = 1000$ and $N_{obs} = 4000$, observations and $N_{MC} = 100000$ real-

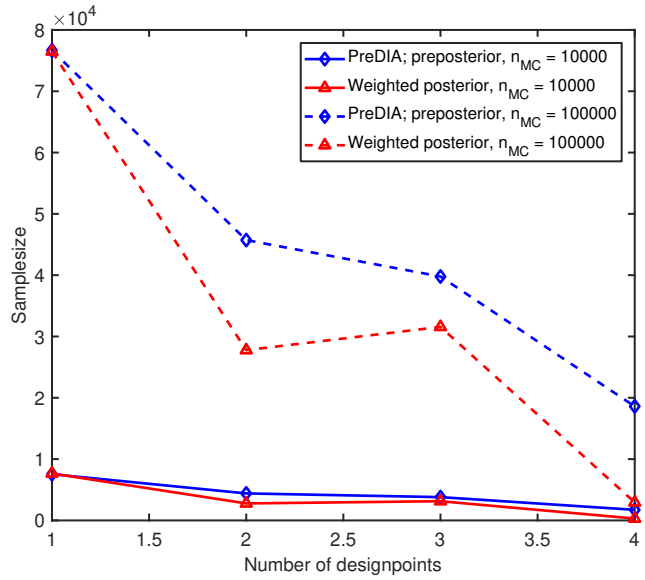


Figure D.6: Sample-sizes over the number of optimized designs with $N_{obs} = 4000$ observations but different number of realizations N_{MC} in example application ExA^{gc}

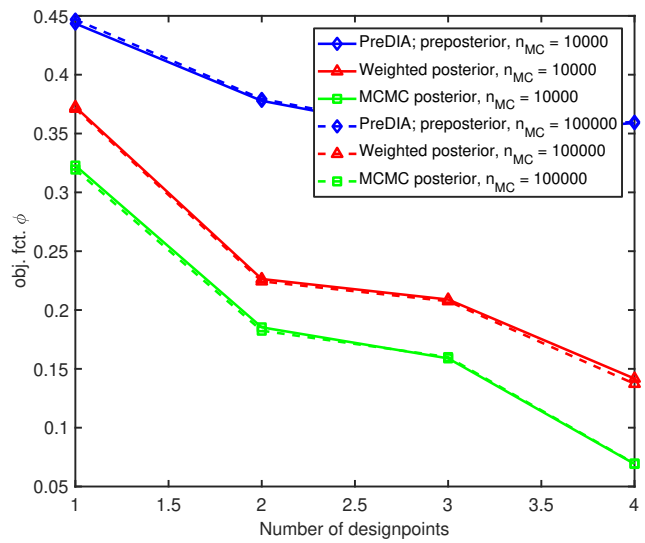


Figure D.7: Objective functions (ϕ_{PreDIA} in blue, ϕ_{wPost} in red and ϕ_{MCMC} in green) over the number of optimized designs with $N_{obs} = 4000$ observations but different number of realizations N_{MC} in example application ExA^{gc}. Results from different N_{MC} are shown as dotted and dashed lines, though, differences in the results are barely distinguishable.

izations were carried out. Sample-sizes over the number of optimized designpoints only weakly vary with the number of observations (see fig. D.8). Again, no effect on the objective functions (see fig. D.9) was found.

Overall, the evaluation of sample-sizes for the example application ExA^{gc} also shows, that the observed deviations in the objective functions (see fig. 6.7) are not caused by insufficiently small sample-sizes or number of observations.

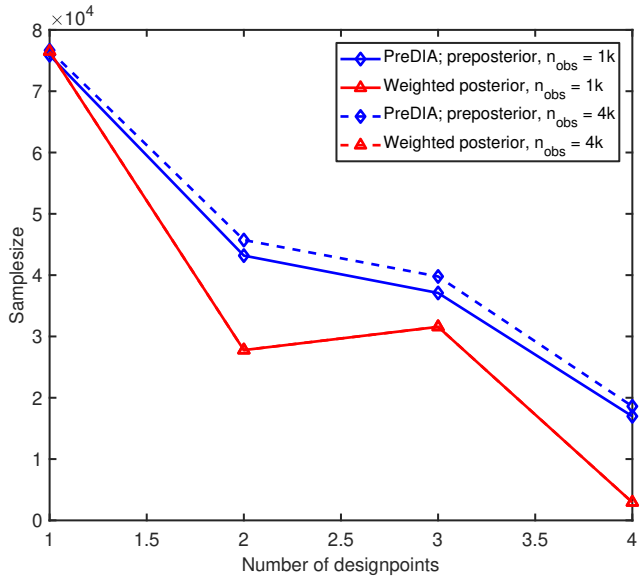


Figure D.8: Sample-sizes over the number of optimized designs with $N_{MC} = 100000$ realizations but different number of observations N_{obs} in example application ExA^{gc}

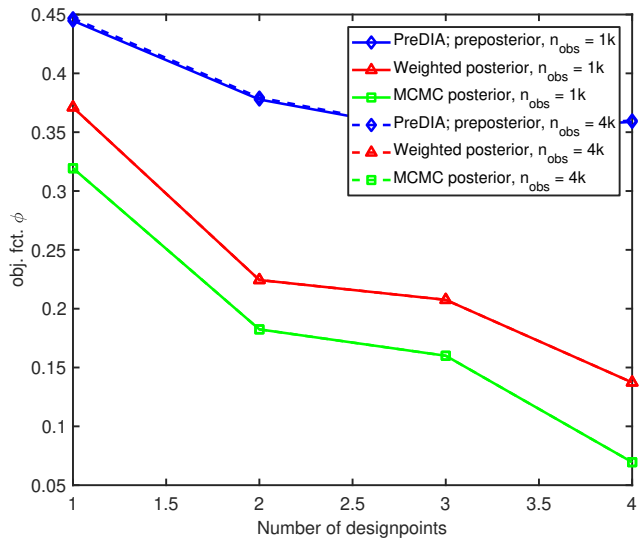


Figure D.9: Objective functions (ϕ_{PreDIA} in blue, ϕ_{wPost} in red and ϕ_{MCMC} in green) over the number of optimized designs with $N_{MC} = 100000$ realizations but different number of observations N_{obs} in example application ExA^{gc}

Bibliography

[1] Y. Rosenfeld, "Relation between the transport coefficients and the internal entropy of simple systems," *Phys. Rev. A*, vol. 15, pp. 2545–2549, 1977.

[2] S. Chapman and T. Cowling, *The Mathematical Theory of Non-uniform Gases.* Cambridge University Press, 1970.

[3] G. K. Batchelor, *An introduction to fluid dynamics.* Cambridge University Press, 1967.

[4] P. Kundu, I. Cohen, and D. Dowling, *Fluid Mechanics.* Elsevier Science, 2012.

[5] N. I. Diamantonis, G. C. Boulougouris, D. M. Tsangaris, M. J. E. Kadi, H. Saadawi, S. Negahban, and I. G. Economou, "Thermodynamic and transport property models for carbon capture and sequestration (CCS) processes with emphasis on CO2 transport," *Chem. Eng. Res. Des.*, vol. 91, no. 10, pp. 1793–1806, 2013.

[6] L.-B. Ouyang, "New correlations for predicting the density and viscosity of supercritical carbon dioxide under conditions expected in carbon capture and sequestration operations," *Open Pet. Eng. J.*, vol. 4, pp. 13–21, 2011.

[7] E. Beckman, "Supercritical and near-critical CO2 in green chemical synthesis and processing," *J. Supercrit. Fluids*, vol. 28, no. 2-3, pp. 121–191, 2004.

[8] A. Vidal, C. Rodriguez, P. Koukouvinis, M. Gavaises, and M. Mchugh, "Modelling of diesel fuel properties through its surrogates using

perturbed-chain, statistical associating fluid theory," *Int. J. Engine Res.*, p. 1468087418801712, 2018.

[9] Z. Franco and Q. Nguyen, "Flow properties of vegetable oil-diesel fuel blends," *Fuel*, vol. 90, no. 2, pp. 838–843, 2011.

[10] N. Yilmaz, "Temperature-dependent viscosity correlations of vegetable oils and biofuel-diesel mixtures," *Biomass Bioenergy*, vol. 35, no. 7, pp. 2936–2938, 2011.

[11] B. Esteban, J.-R. Riba, G. Baquero, A. Rius, and R. Puig, "Temperature dependence of density and viscosity of vegetable oils," *Biomass Bioenergy*, vol. 42, no. 0, pp. 164–171, 2012.

[12] M. Lampe, M. Stavrou, J. Schilling, E. Sauer, J. Gross, and A. Bardow, "Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT," *Comput. Chem. Eng.*, vol. 81, no. Supplement C, pp. 278–287, 2015.

[13] M. Stavrou, M. Lampe, A. Bardow, and J. Gross, "Continuous molecular targeting-computer-aided molecular design (CoMT-CAMD) for simultaneous process and solvent design for CO2 Capture," *Ind. Eng. Chem. Res.*, vol. 53, no. 46, pp. 18 029–18 041, 2014.

[14] G. Galliero, C. Boned, and A. Baylaucq, "Molecular dynamics study of the Lennard-Jones fluid viscosity: Application to real fluids," *Ind. Eng. Chem. Res.*, vol. 44, no. 17, pp. 6963–6972, 2005.

[15] B. Poling, J. Prausnitz, and J. Connell, *The Properties of Gases and Liquids*. McGraw-Hill Professional, 2000.

[16] S. Bock, E. Bich, E. Vogel, A. S. Dickinson, and V. Vesovic, "Calculation of the transport properties of carbon dioxide. I. shear viscosity, viscomagnetic effects, and self-diffusion," *J. Chem. Phys.*, vol. 117, no. 5, pp. 2151–2160, 2002.

[17] R. Hellmann, E. Bich, E. Vogel, A. S. Dickinson, and V. Vesovic, "Calculation of the transport and relaxation properties of methane. I. shear viscosity, viscomagnetic effects, and self-diffusion," *J. Chem. Phys.*, vol. 129, no. 6, p. 064302, 2008.

[18] J. Millat, J. Dymond, and C. De Castro, *Transport Properties of Fluids: Their Correlation, Prediction And Estimation.* Cambridge University Press, 2005, ch. 6, p. 113.

[19] A. S. de Wijn, V. Vesovic, G. Jackson, and J. P. M. Trusler, "A kinetic theory description of the viscosity of dense fluids consisting of chain molecules," *J. Chem. Phys.*, vol. 128, no. 20, p. 204901, 2008.

[20] D. Chandler, "Rough hard sphere theory of the self-diffusion constant for molecular liquids," *J. Chem. Phys.*, vol. 62, no. 4, pp. 1358–1363, 1975.

[21] J. H. Dymond and M. A. Awan, "Correlation of high-pressure diffusion and viscosity coefficients for n-alkanes," *Int. J. Thermophys.*, vol. 10, no. 5, pp. 941–951, 1989.

[22] M. Assael, J. Dymond, M. Papadaki, and P. Patterson, "Correlation and prediction of dense fluid transport coefficients: II. simple molecular fluids," *Fluid Phase Equilib.*, vol. 75, no. 0, pp. 245–255, 1992.

[23] M. Assael, J. Dymond, M. Papadaki, and P. Patterson, "Correlation and prediction of dense fluid transport coefficients. I. n-alkanes," *Int. J. Thermophys.*, vol. 13, no. 2, pp. 269–281, 1992.

[24] M. J. Assael, J. H. Dymond, and P. M. Patterson, "Correlation and prediction of dense fluid transport coefficients. V. aromatic hydrocarbons," *Int. J. Thermophys.*, vol. 13, no. 5, pp. 895–905, 1992.

[25] M. J. Assael, J. H. Dymond, and S. K. Polimatidou, "Correlation and prediction of dense fluid transport coefficients. VI. n-alcohols," *Int. J. Thermophys.*, vol. 15, no. 2, pp. 189–201, 1994.

[26] M. J. Assael, J. H. Dymond, and S. K. Polimatidou, "Correlation and prediction of dense fluid transport coefficients. VII. refrigerants," *Int. J. Thermophys.*, vol. 16, no. 3, pp. 761–772, 1995.

[27] F. Ciotta, J. M. Trusler, and V. Vesovic, "Extended hard-sphere model for the viscosity of dense fluids," *Fluid Phase Equilib.*, vol. 363, pp. 239–247, 2014.

[28] D. D. Royal, V. Vesovic, J. P. M. Trusler, and W. A. Wakeham, "Prediction of the viscosity of dense fluid mixtures," *Mol. Phys.*, vol. 101, no. 3, pp. 339–352, 2003.

[29] D. Royal, V. Vesovic, J. Trusler, and W. Wakeham, "Predicting the viscosity of liquid refrigerant blends: comparison with experimental data," *Int. J. Refrig.*, vol. 28, no. 3, pp. 311–319, 2005.

[30] V. Vesovic and W. Wakeham, "Prediction of the viscosity of fluid mixtures over wide ranges of temperature and pressure," *Chem. Eng. Sci.*, vol. 44, no. 10, pp. 2181–2189, 1989.

[31] A. S. de Wijn, N. Riesco, G. Jackson, J. P. M. Trusler, and V. Vesovic, "Viscosity of liquid mixtures: The vesovic-wakeham method for chain molecules," *J. Chem. Phys.*, vol. 136, no. 7, p. 074514, 2012.

[32] J. Dymond, "The interpretation of transport coefficients on the basis of the van der waals model," *Physica*, vol. 75, no. 1, pp. 100–114, 1974.

[33] M. J. Assael, J. H. Dymond, M. Papadaki, and P. M. Patterson, "Correlation and prediction of dense fluid transport coefficients. III. n-alkane mixtures," *Int. J. Thermophys.*, vol. 13, no. 4, pp. 659–669, 1992.

[34] A. Allal, C. Boned, and A. Baylaucq, "Free-volume viscosity model for fluids in the dense and gaseous states," *Phys. Rev. E*, vol. 64, p. 011203, 2001.

[35] A. Allal, M. Moha-ouchane, and C. Boned, "A new free volume model for dynamic viscosity and density of dense fluids versus pressure and temperature," *Phys. Chem. Liq.*, vol. 39, no. 1, pp. 1–30, 2001.

[36] T.-H. Chung, M. Ajlan, L. Lee, and K. E. Starling, "Generalized multiparameter correlation for nonpolar and polar fluid transport properties," *Ind. Eng. Chem. Res.*, vol. 27, pp. 671–679, 1988.

[37] P. E. Rouse, "A theory of the linear viscoelastic properties of dilute solutions of coiling polymers," *J. Chem. Phys.*, vol. 21, no. 7, pp. 1272–1280, 1953.

[38] F. Llovell, R. M. Marcos, and L. F. Vega, "Free-volume theory coupled with soft-SAFT for viscosity calculations: Comparison with molecular

simulation and experimental data," *J. Phys. Chem. B*, vol. 117, no. 27, pp. 8159–8171, 2013.

[39] F. Llovell, R. M. Marcos, and L. F. Vega, "Transport properties of mixtures by the soft-SAFT + free-volume theory: Application to mixtures of n-alkanes and hydrofluorocarbons," *J. Phys. Chem. B*, vol. 117, no. 17, pp. 5195–5205, 2013.

[40] I. Polishuk and A. Yitzhak, "Modeling viscosities of pure compounds and their binary mixtures using the modified yarranton-satyro correlation and free volume theory coupled with SAFT+Cubic EOS," *Ind. Eng. Chem. Res.*, vol. 53, no. 2, pp. 959–971, 2014.

[41] S. E. Quiñones Cisneros, C. K. Zéberg-Mikkelsen, and E. H. Stenby, "The friction theory (f-theory) for viscosity modeling," *Fluid Phase Equilibr.*, vol. 169, no. 2, pp. 249–276, 2000.

[42] S. E. Quiñones-Cisneros, C. K. Zéberg-Mikkelsen, and E. H. Stenby, "One parameter friction theory models for viscosity," *Fluid Phase Equilib.*, vol. 178, no. 1, pp. 1–16, 2001.

[43] S. E. Quiñones Cisneros, C. K. Zéberg-Mikkelsen, J. Fernández, and J. García, "General friction theory viscosity model for the PC-SAFT equation of state," *AIChE J.*, vol. 52, no. 4, pp. 1600–1610, 2006.

[44] S. E. Quiñones-Cisneros and U. K. Deiters, "Generalization of the friction theory for viscosity modeling," *J. Phys. Chem. B*, vol. 110, no. 25, pp. 12 820–12 834, 2006.

[45] R. Chopra, T. M. Truskett, and J. R. Errington, "On the use of excess entropy scaling to describe single-molecule and collective dynamic properties of hydrocarbon isomer fluids," *J. Phys. Chem. B*, vol. 114, no. 49, pp. 16 487–16 493, 2010.

[46] R. Chopra, T. M. Truskett, and J. R. Errington, "Excess entropy scaling of dynamic quantities for fluids of dumbbell-shaped particles," *J. Chem. Phys.*, vol. 133, no. 10, p. 104506, 2010.

[47] T. Goel, C. N. Patra, T. Mukherjee, and C. Chakravarty, "Excess entropy scaling of transport properties of Lennard-Jones chains," *J. Chem. Phys.*, vol. 129, no. 16, p. 164904, 2008.

[48] O. Lötgering-Lin and J. Gross, "A group contribution method for viscosities based on entropy scaling using the perturbed-chain polar statistical associating fluid theory," *Ind. Eng. Chem. Res.*, vol. 54, no. 32, pp. 7942–7952, 2015.

[49] O. Lötgering-Lin, M. Fischer, M. Hopp, and J. Gross, "Pure substance and mixture viscosities based on entropy-scaling and an analytic equation of state," *Ind. Eng. Chem. Res.*, vol. 57, no. 11, pp. 4095–4114, 2018.

[50] L. T. Novak, "Self-diffusion coefficient and viscosity in fluids," *Int. J. Chem. React. Eng.*, vol. 9, no. A63, 2011.

[51] L. T. Novak, "Fluid viscosity-residual entropy correlation," *Int. J. Chem. React. Eng.*, vol. 9, no. A107, p. A107, 2011.

[52] N. P. Bailey, U. R. Pedersen, N. Gnan, T. B. Schrøder, and J. C. Dyre, "Pressure-energy correlations in liquids. I. results from computer simulations," *J. Chem. Phys.*, vol. 129, no. 18, p. 184507, 2008.

[53] N. P. Bailey, U. R. Pedersen, N. Gnan, T. B. Schrøder, and J. C. Dyre, "Pressure-energy correlations in liquids. II. analysis and consequences," *J. Chem. Phys.*, vol. 129, no. 18, pp. –, 2008.

[54] N. Gnan, T. B. Schrøder, U. R. Pedersen, N. P. Bailey, and J. C. Dyre, "Pressure-energy correlations in liquids. IV. "isomorphs" in liquid phase diagrams," *J. Chem. Phys.*, vol. 131, no. 23, p. 234504, 2009.

[55] T. B. Schrøder, N. P. Bailey, U. R. Pedersen, N. Gnan, and J. C. Dyre, "Pressure-energy correlations in liquids. III. statistical mechanics and thermodynamics of liquids with hidden scale invariance," *J. Chem. Phys.*, vol. 131, no. 23, pp. –, 2009.

[56] T. B. Schrøder, N. Gnan, U. R. Pedersen, N. P. Bailey, and J. C. Dyre, "Pressure-energy correlations in liquids. V. isomorphs in generalized Lennard-Jones systems," *J. Chem. Phys.*, vol. 134, no. 16, p. 164505, 2011.

[57] J. C. Dyre, "Isomorphs, hidden scale invariance, and quasiuniversality," *Phys. Rev. E*, vol. 88, p. 042139, 2013.

[58] J. C. Dyre, "Hidden scale invariance in condensed matter," *J. Phys. Chem. B*, vol. 118, no. 34, pp. 10 007–10 024, 2014.

[59] N. P. Bailey, L. Bohling, A. A. Veldhorst, T. B. Schrøder, and J. C. Dyre, "Statistical mechanics of roskilde liquids: Configurational adiabats, specific heat contours, and density dependence of the scaling exponent," *J. Chem. Phys.*, vol. 139, no. 18, pp. –, 2013.

[60] A. A. Veldhorst, J. C. Dyre, and T. B. Schrøder, "Scaling of the dynamics of flexible Lennard-Jones chains," *J. Chem. Phys.*, vol. 141, no. 5, p. 054904, 2014.

[61] T. S. Ingebrigtsen, J. R. Errington, T. M. Truskett, and J. C. Dyre, "Predicting how nanoconfinement changes the relaxation time of a supercooled liquid," *Phys. Rev. Lett.*, vol. 111, p. 235901, 2013.

[62] D. Coslovich, , and C. M. Roland, "Thermodynamic scaling of diffusion in supercooled Lennard-Jones liquids," *J. Phys. Chem. B*, vol. 112, no. 5, pp. 1329–1332, 2008.

[63] E. R. López, A. S. Pensado, M. J. P. Comuñas, A. A. H. Pádua, J. Fernández, and K. R. Harris, "Density scaling of the transport properties of molecular and ionic liquids," *J. Chem. Phys.*, vol. 134, no. 14, p. 144507, 2011.

[64] L. Bohling, T. S. Ingebrigtsen, A. Grzybowski, M. Paluch, J. C. Dyre, and T. B. Schrøder, "Scaling of viscous dynamics in simple liquids: theory, simulation and experiment," *New J. Phys.*, vol. 14, no. 11, p. 113035, 2012.

[65] C. M. Roland, S. Bair, and R. Casalini, "Thermodynamic scaling of the viscosity of van der waals, h-bonded, and ionic liquids," *J. Chem. Phys.*, vol. 125, no. 12, p. 124508, 2006.

[66] J. C. Dyre, "Perspective: Excess-entropy scaling," *J. Chem. Phys.*, vol. 149, no. 21, p. 210901, 2018.

[67] Y. Rosenfeld, "A quasi-universal scaling law for atomic transport in simple fluids," *J. Phys.: Condens. Matter*, vol. 11, no. 28, pp. 5415–5427, 1999.

[68] M. Dzugutov, "A universal scaling law for atomic diffusion in condensed matter," *Nature*, vol. 381, no. 6578, pp. 137–139, 1996.

[69] P. Mausbach and H.-O. May, "Transport anomalies in the Gaussian core model fluid," *Z. Phys. Chem.*, vol. 223, pp. 1035–1046, 2009.

[70] W. P. Krekelberg, T. Kumar, J. Mittal, J. R. Errington, and T. M. Truskett, "Anomalous structure and dynamics of the Gaussian-core fluid," *Phys. Rev. E*, vol. 79, p. 031203, 2009.

[71] G. Galliero, C. Boned, and J. Fernández, "Scaling of the viscosity of the Lennard-Jones chain fluid model, argon, and some normal alkanes," *J. Chem. Phys.*, vol. 134, no. 6, p. 064505, 2011.

[72] Z. N. Gerek and J. R. Elliott, "Self-diffusivity estimation by molecular dynamics," *Ind. Eng. Chem. Res.*, vol. 49, no. 7, pp. 3411–3423, 2010.

[73] L. T. Novak, "Predictive corresponding-states viscosity model for the entire fluid region: n-alkanes," *Ind. Eng. Chem. Res.*, vol. 52, no. 20, pp. 6841–6847, 2013.

[74] E. H. Abramson, "Viscosity of carbon dioxide measured to a pressure of 8 GPa and temperature of 673 K," *Phys. Rev. E*, vol. 80, p. 021201, 2009.

[75] E. H. Abramson and H. West-Foyle, "Viscosity of nitrogen measured to pressures of 7 GPa and temperatures of 573 K," *Phys. Rev. E*, vol. 77, p. 041202, 2008.

[76] I. H. Bell and A. Laesecke, "Viscosity of refrigerants and other working fluids from residual entropy scaling," *16th International Refrigeration and Air Conditioning Conference at Purdue*, July 2016.

[77] E. H. Abramson, "Viscosity of water measured to pressures of 6 GPa and temperatures of 300 °C," *Phys. Rev. E*, vol. 76, p. 051203, 2007.

[78] M. Agarwal, M. Singh, R. Sharma, M. Parvez Alam, and C. Chakravarty, "Relationship between structure, entropy, and diffusivity in water and water-like liquids," *J. Phys. Chem. B*, vol. 114, no. 20, pp. 6995–7001, 2010.

[79] R. Chopra, T. M. Truskett, and J. R. Errington, "On the use of excess entropy scaling to describe the dynamic properties of water," *J. Phys. Chem. B*, vol. 114, no. 32, pp. 10 558–10 566, 2010.

[80] J. R. Errington, T. M. Truskett, and J. Mittal, "Excess-entropy-based anomalies for a waterlike fluid," *J. Chem. Phys.*, vol. 125, no. 24, p. 244502, 2006.

[81] M. E. Johnson and T. Head-Gordon, "Assessing thermodynamic-dynamic relationships for waterlike liquids," *J. Chem. Phys.*, vol. 130, no. 21, p. 214510, 2009.

[82] W. P. Krekelberg, M. J. Pond, G. Goel, V. K. Shen, J. R. Errington, and T. M. Truskett, "Generalized rosenfeld scalings for tracer diffusivities in not-so-simple fluids: Mixtures and soft particles," *Phys. Rev. E*, vol. 80, p. 061205, 2009.

[83] M. Hopp, J. Mele, and J. Gross, "Self-diffusion coefficients from entropy scaling using the PCP-SAFT equation of state," *Ind. Eng. Chem. Res.*, vol. 57, no. 38, pp. 12 942–12 950, 2018.

[84] G. Galliero and C. Boned, "Thermal conductivity of the Lennard-Jones chain fluid model," *Phys. Rev. E*, vol. 80, p. 061202, 2009.

[85] S. Delage-Santacreu, G. Galliero, H. Hoang, J.-P. Bazile, C. Boned, and J. Fernandez, "Thermodynamic scaling of the shear viscosity of Mie n-6 fluids and their binary mixtures," *J. Chem. Phys.*, vol. 142, no. 17, p. 174501, 2015.

[86] E. Voyiatzis, F. Müller-Plathe, and M. C. Böhm, "Do transport properties of entangled linear polymers scale with excess entropy?" *Macromolecules*, vol. 46, no. 21, pp. 8710–8723, 2013.

[87] S. Bastea, "Entropy scaling laws for diffusion," *Phys. Rev. Lett.*, vol. 93, p. 199603, 2004.

[88] A. Samanta, S. Ali, and S. Ghosh, "New universal scaling laws of diffusion and kolmogorov-sinai entropy in simple liquids," *Phys. Rev. Lett.*, vol. 92, no. 14, 2004.

[89] A. Samanta, S. Ali, and S. Ghosh, "Universal scaling laws of diffusion in a binary fluid mixture," *Phys. Rev. Lett.*, vol. 87, no. 24, pp. 245 901–1–245 901–4, 2001.

[90] S. Pieprzyk, D. M. Heyes, and A. C. Brańka, "Thermodynamic properties and entropy scaling law for diffusivity in soft spheres," *Phys. Rev. E*, vol. 90, p. 012106, 2014.

[91] G. Li, C. Liu, and Z. Zhu, "Excess entropy scaling for transport coefficients: diffusion and viscosity in liquid metals," *J. Non-Cryst. Solids*, vol. 351, no. 10-11, pp. 946–950, 2005.

[92] A. B. de Oliveira, E. Salcedo, C. Chakravarty, and M. C. Barbosa, "Entropy, diffusivity and the energy landscape of a waterlike fluid," *J. Chem. Phys.*, vol. 132, no. 23, p. 234509, 2010.

[93] M. Hopp and J. Gross, "Thermal conductivity of real substances from excess entropy scaling using PCP-SAFT," *Ind. Eng. Chem. Res.*, vol. 56, no. 15, pp. 4527–4538, 2017.

[94] R. Grover, W. G. Hoover, and B. Moran, "Corresponding states for thermal conductivities via nonequilibrium molecular dynamics," *J. Chem. Phys.*, vol. 83, no. 3, pp. 1255–1259, 1985.

[95] R. Srivastava, D. Dwivedee, and K. Khanna, "Scaling laws for transport coefficients of a hard sphere fluid," *J. Mol. Liq.*, vol. 139, no. 1-3, pp. 29–34, 2008.

[96] L. T. Novak, "Predicting natural gas viscosity with a mixture viscosity model for the entire fluid region," *Ind. Eng. Chem. Res.*, vol. 52, no. 45, pp. 16 014–16 018, 2013.

[97] M. S. Wertheim, "Fluids with highly directional attractive forces. I. statistical thermodynamics," *J. Stat. Phys.*, vol. 35, no. 1, pp. 19–34, 1984.

[98] M. S. Wertheim, "Fluids with highly directional attractive forces. II. thermodynamic perturbation theory and integral equations," *J. Stat. Phys.*, vol. 35, no. 1, pp. 35–47, 1984.

[99] M. S. Wertheim, "Fluids with highly directional attractive forces. III. multiple attraction sites," *J. Stat. Phys.*, vol. 42, no. 3, pp. 459–476, 1986.

[100] M. S. Wertheim, "Fluids with highly directional attractive forces. IV. equilibrium polymerization," *J. Stat. Phys.*, vol. 42, no. 3, pp. 477–492, 1986.

[101] M. S. Wertheim, "Thermodynamic perturbation theory of polymerization," *J. Chem. Phys.*, vol. 87, no. 12, pp. 7323–7331, 1987.

[102] W. Zmpitas and J. Gross, "Detailed pedagogical review and analysis of wertheim's thermodynamic perturbation theory," *Fluid Phase Equilib.*, vol. 428, pp. 121–152, 2016.

[103] W. Chapman, K. Gubbins, G. Jackson, and M. Radosz, "SAFT: Equation-of-state solution model for associating fluids," *Fluid Phase Equilib.*, vol. 52, pp. 31–38, 1989.

[104] J. Gross and G. Sadowski, "Application of the perturbed-chain SAFT equation of state to associating systems," *Ind. Eng. Chem. Res.*, vol. 41, no. 22, pp. 5510–5515, 2002.

[105] J. Gross and G. Sadowski, "Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules," *Ind. Eng. Chem. Res.*, vol. 40, no. 4, pp. 1244–1260, 2001.

[106] F. J. Blas and L. F. Vega, "Prediction of binary and ternary diagrams using the statistical associating fluid theory (SAFT) equation of state," *Ind. Eng. Chem. Res.*, vol. 37, no. 2, pp. 660–674, 1998.

[107] F. J. Blas and L. F. Vega, "Thermodynamic behaviour of homonuclear and heteronuclear Lennard-Jones chains with association sites from simulation and theory," *Mol. Phys.*, vol. 92, no. 1, pp. 135–150, 1997.

[108] A. Gil-Villegas, A. Galindo, P. J. Whitehead, S. J. Mills, G. Jackson, and A. N. Burgess, "Statistical associating fluid theory for chain molecules with attractive potentials of variable range," *J. Chem. Phys.*, vol. 106, no. 10, pp. 4168–4186, 1997.

[109] A. Galindo, L. A. Davies, A. Gil-Villegas, and G. Jackson, "The thermodynamics of mixtures and the corresponding mixing rules in the SAFT-VR approach for potentials of variable range," *Mol. Phys.*, vol. 93, no. 2, pp. 241–252, 1998.

[110] J. Gross and J. Vrabec, "An equation-of-state contribution for polar components: Dipolar molecules," *AIChE J.*, vol. 52, no. 3, pp. 1194–1204, 2006.

[111] J. Gross, "An equation-of-state contribution for polar components: Quadrupolar molecules," *AIChE J.*, vol. 51, no. 9, pp. 2556–2568, 2005.

[112] J. Vrabec and J. Gross, "Vapor-liquid equilibria simulation and an equation of state contribution for dipole-quadrupole interactions," *J. Phys. Chem. B*, vol. 112, no. 1, pp. 51–60, 2008.

[113] H. A. Lorentz, "über die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase," *Ann. Phys-leipzig.*, vol. 248, no. 1, pp. 127–136, 1881.

[114] J. Vijande, M. M. Pineiro, D. Bessieres, H. Saint-Guirons, and J. L. Legido, "Description of PVT behaviour of hydrofluoroethers using the PC-SAFT EOS," *Phys. Chem. Chem. Phys.*, vol. 6, pp. 766–770, 2004.

[115] A. Tihic, G. Kontogeorgis, N. von Solms, M. Michelsen, and L. Constantinou, "A predictive group-contribution simplified PC-SAFT equation of state: Application to polymer systems," *Ind. Eng. Chem. Res.*, vol. 47, no. 15, pp. 5092–5101, 2008.

[116] L. Constantinou and R. Gani, "New group contribution method for estimating properties of pure compounds," *AIChE J.*, vol. 40, no. 10, pp. 1697–1710, 1994.

[117] S. Tamouza, J. Passarello, P. Tobaly, and J. de Hemptinne, "Group contribution method with SAFT EOS applied to vapor liquid equilibria of various hydrocarbon series," *Fluid Phase Equilib.*, vol. 222, pp. 67–76, 2004.

[118] E. Sauer, M. Stavrou, and J. Gross, "Comparison between a homo- and a heterosegmented group contribution approach based on the perturbed-

chain polar statistical associating fluid theory equation of state," *Ind. Eng. Chem. Res.*, vol. 53, no. 38, pp. 14 854–14 864, 2014.

[119] A. Lymperiadis, C. S. Adjiman, A. Galindo, and G. Jackson, "A group contribution method for associating chain molecules based on the statistical associating fluid theory (SAFT-γ)," *J. Chem. Phys.*, vol. 127, p. 234903, 2007.

[120] Y. Peng, K. D. Goff, M. dos Ramos, and C. McCabe, "Developing a predictive group-contribution-based SAFT-VR equation of state," *Fluid Phase Equilib.*, vol. 277, no. 2, pp. 131–144, 2009.

[121] K. Paduszyński and U. Domańska, "Heterosegmented perturbed-chain statistical associating fluid theory as a robust and accurate tool for modeling of various alkanes. 1. pure fluids," *Ind. Eng. Chem. Res.*, vol. 51, no. 39, pp. 12 967–12 983, 2012.

[122] F. T. Peters, F. S. Laube, and G. Sadowski, "Development of a group contribution method for polymers within the PC-SAFT model," *Fluid Phase Equilibr.*, vol. 324, no. 0, pp. 70–79, 2012.

[123] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed., ser. Springer Series in Statistics. New York: Springer, 2009.

[124] W. M. Thorburn, "Occam's razor," *Mind*, vol. XXIV, no. 2, pp. 287–288, 1915.

[125] D. Draper, "Assessment and propagation of model uncertainty," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 57, no. 1, pp. 45–97, 1995.

[126] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: A tutorial," *Stat. Sci.*, vol. 14, no. 4, pp. 382–401, 1999.

[127] H. Jeffreys, *Scientific Inference.* Cambridge University Press; 3rd ed. edition, 1931.

[128] A. E. Raftery, "Bayesian model selection in social research," *Sociol. Methodol.*, vol. 25, pp. 111–163, 1995.

[129] J. P. Huelsenbeck, B. Larget, and M. E. Alfaro, "Bayesian phylogenetic model selection using Reversible Jump Markov Chain Monte Carlo," *Mol. Biol. Evol.*, vol. 21, no. 6, pp. 1123–1133, 2004.

[130] N. K. Ajami and C. Gu, "Complexity in microbial metabolic processes in soil nitrogen modeling: a case for model averaging," *Stoch. Env. Res. Risk. A.*, vol. 24, no. 6, pp. 831–844, 2010.

[131] D. Seifert, T. O. Sonnenborg, J. C. Refsgaard, A. L. Hojberg, and L. Troldborg, "Assessment of hydrological model predictive ability given multiple conceptual geological models," *Water Resour. Res.*, vol. 48, no. 6, 2012.

[132] T. Wöhling, A. Schöniger, S. Gayler, and W. Nowak, "Bayesian model averaging to explore the worth of data for soil-plant model selection and prediction," *Water Resour. Res.*, vol. 51, no. 4, pp. 2825–2846, 2015.

[133] R. Rojas, L. Feyen, O. Batelaan, and A. Dassargues, "On the value of conditioning data to reduce conceptual model uncertainty in groundwater modeling," *Water Resour. Res.*, vol. 46, 2010.

[134] S. P. Neuman, L. Xue, M. Ye, and D. Lu, "Bayesian analysis of data-worth considering model and parameter uncertainties," *Adv. Water Resour.*, vol. 36, pp. 75–85, 2012.

[135] P. C. Leube, A. Geiges, and W. Nowak, "Bayesian assessment of the expected data impact on prediction confidence in optimal sampling design," *Water Resour. Res.*, vol. 48, no. 2, 2012.

[136] M. Bjørner, G. Sin, and G. Kontogeorgis, "Uncertainty analysis of the CPA and a quadrupolar CPA equation of state - with emphasis on CO2," *Fluid Phase Equilib.*, vol. 414, pp. 29–47, 2016.

[137] A. C. Atkinson, A. N. Donev, and R. D. Tobias, *Optimum experimental designs, with SAS*, ser. Oxford Statistical Science Series 34. Oxford: Oxford University Press, 2007.

[138] I. O. Igwe, "The effects of temperature on the viscosity of vegetable oils in solution," *Ind. Crops Prod.*, vol. 19, no. 2, pp. 185–190, 2004.

[139] A. Abolle, L. Kouakou, and H. Planche, "The viscosity of diesel oil and mixtures with straight vegetable oils: Palm, cabbage palm, cotton, groundnut, copra and sunflower," *Biomass Bioenergy*, vol. 33, no. 9, pp. 1116–1121, 2009.

[140] A. Karunanithi, L. Achenie, and R. Gani, "A computer-aided molecular design framework for crystallization solvent design," *Chem. Eng. Sci.*, vol. 61, no. 4, pp. 1247–1260, 2006.

[141] A. I. Papadopoulos, M. Stijepovic, and P. Linke, "On the systematic design and selection of optimal working fluids for organic rankine cycles," *Appl. Therm. Eng.*, vol. 30, no. 6-7, pp. 760–769, 2010.

[142] E. Marcoulaki and A. Kokossis, "On the development of novel chemicals using a systematic optimisation approach. part II. solvent design," *Chem. Eng. Sci.*, vol. 55, no. 13, pp. 2547–2561, 2000.

[143] A. Bardow, K. Steur, and J. Gross, "Continuous-molecular targeting for integrated solvent and process design," *Ind. Eng. Chem. Res.*, vol. 49, no. 6, pp. 2834–2840, 2010.

[144] E. Hendriks, G. M. Kontogeorgis, R. Dohrn, J.-C. de Hemptinne, I. G. Economou, L. F. Žilnik, and V. Vesovic, "Industrial requirements for thermodynamics and transport properties," *Ind. Eng. Chem. Res.*, vol. 49, no. 22, pp. 11 131–11 141, 2010.

[145] E. Lemmon, M. McLinden, and M. Huber, "Refprop: Reference fluid thermodynamic and transport properties, version 8.0," *NIST standard reference database*, vol. 23, no. 8.0, 2007.

[146] E. Lemmon and R. Jacobsen, "Viscosity and thermal conductivity equations for nitrogen, oxygen, argon, and air," *Int. J. Thermophys.*, vol. 25, no. 1, pp. 21–69, 2004.

[147] F. Llovell, O. Vilaseca, N. Jung, and L. Vega, "Water+1-alkanol systems: Modeling the phase, interface and viscosity properties," *Fluid Phase Equilib.*, vol. 360, no. 0, pp. 367–378, 2013.

[148] T. Sun and A. S. Teja, "Correlation and prediction of the viscosity and thermal conductivity of dense fluids," *J. Chem. Eng. Data*, vol. 54, no. 9, pp. 2527–2531, 2009.

[149] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular theory of gases and liquids.* Wiley-VCH, 1954.

[150] D. A. McQuarrie, *Statistical mechanics.* University science books, 2000.

[151] P. D. Neufeld, A. R. Janzen, and R. A. Aziz, "Empirical equations to calculate 16 of the transport collision integrals $\omega(l, s)*$ for the Lennard-Jones (12-6) potential," *J. Chem. Phys.*, vol. 57, no. 3, pp. 1100–1102, 1972.

[152] J. Mittal, J. R. Errington, and T. M. Truskett, "Relationship between thermodynamics and dynamics of supercooled liquids," *J. Chem. Phys.*, vol. 125, no. 7, p. 076102, 2006.

[153] DDBST Dortmund Data Bank Software & Separation Technology GmbH, Oldenburg, Germany, 2014.

[154] D. W. Marquardt, "An algorithm for least-squares estimation of non-linear parameters," *SIAM J. Appl. Math.*, vol. 11, no. 2, pp. 431–441, 1963.

[155] W. Wagner and A. Pruss, "The iapws formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use," *J. Phys. Chem. Ref. Data*, vol. 31, no. 2, pp. 387–535, 2002.

[156] W. Wagner and U. Overhoff, *Software FLUIDCAL for the Calculation of Thermodynamic and Transport Properties of Water*, Ruhr-Universität Bochum, Lehrstuhl für Thermodynamik.

[157] S. H. Huang and M. Radosz, "Equation of state for small, large, polydisperse, and associating molecules," *Ind. Eng. Chem. Res.*, vol. 29, no. 11, pp. 2284–2294, 1990.

[158] A. Singh, S. Mishra, and G. Ruskauff, "Model averaging techniques for quantifying conceptual model uncertainty," *Ground Water*, vol. 48, no. 5, pp. 701–715, 2010.

[159] M. Ye, K. F. Pohlmann, J. B. Chapman, G. M. Pohll, and D. M. Reeves, "A model-averaging method for assessing groundwater conceptual model uncertainty," *Ground Water*, vol. 48, no. 5, pp. 716–28, 2010.

[160] W. A. Link and R. J. Barker, "Model weights and the foundations of multimodel inference," *Ecology*, vol. 87, no. 10, pp. 2626–2635, 2006.

[161] X. Li and F. T.-C. Tsai, "Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multimethod," *Water Resour. Res.*, vol. 45, no. 9, 2009.

[162] M. Troldborg, W. Nowak, N. Tuxen, P. L. Bjerg, R. Helmig, and P. J. Binning, "Uncertainty evaluation of mass discharge estimates from a contaminated site using a fully Bayesian framework," *Water Resour. Res.*, vol. 46, no. 12, 2010.

[163] A. Schöniger, T. Wöhling, and W. Nowak, "A statistical concept to assess the uncertainty in Bayesian model weights and its impact on model ranking," *Water Resour. Res.*, vol. 51, no. 9, pp. 7524–7546, 2015.

[164] H. Jeffreys, *Theory of Probability*, 1st ed. Oxford: Clarendon Press, 1939.

[165] S. F. Gull, *Bayesian inductive inference and maximum entropy.* Dordrecht / Boston / London: Kluwer Academic Publishers, 1988, vol. 1.

[166] R. E. Kass and A. E. Raftery, "Bayes factors," *J. Am. Stat. Assoc.*, vol. 90, no. 430, pp. 773–795, 1995.

[167] S. P. Neuman, "Maximum likelihood Bayesian averaging of uncertain model predictions," *Stoch. Env. Res. Risk. A.*, vol. 17, no. 5, pp. 291–305, 2003.

[168] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.

[169] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki, Eds. Budapest: Akademiai Kiado, 1973, pp. 367–281.

[170] M. Ye, P. D. Meyer, and S. P. Neuman, "On model selection criteria in multimodel analysis," *Adv. Water Resour.*, vol. 44, no. 3, 2008.

[171] F. T.-C. Tsai and X. B. Li, "Inverse groundwater modeling for hydraulic conductivity estimation using Bayesian model averaging and variance window," *Water Resour. Res.*, vol. 44, no. 9, 2008.

[172] E. Morales-Casique, S. P. Neuman, and V. V. Vesselinov, "Maximum likelihood Bayesian averaging of airflow models in unsaturated fractured tuff using Occam and variance windows," *Stoch. Env. Res. Risk. A.*, vol. 24, no. 6, pp. 863–880, 2010.

[173] L. Foglia, S. W. Mehl, M. C. Hill, and P. Burlando, "Evaluating model structure adequacy: The case of the Maggia Valley groundwater system, southern Switzerland," *Water Resour. Res.*, vol. 49, no. 1, pp. 260–282, 2013.

[174] A. Schöniger, T. Wöhling, L. Samaniego, and W. Nowak, "Model selection on solid ground: Rigorous comparison of nine ways to evaluate Bayesian model evidence," *Water Resour. Res.*, vol. 50, no. 12, pp. 9484–9513, 2014.

[175] H. Jeffreys, *Theory of Probability*, 3rd ed. Oxford University Press, 1961.

[176] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice.* New York, NY: Springer New York, 2001, ch. An Introduction to Sequential Monte Carlo Methods, pp. 3–14.

[177] J. V. Beck and K. A. Woodbury, "Inverse problems and parameter estimation: integration of measurements and analysis," *Meas. Sci. Technol.*, vol. 9, no. 6, pp. 839–847, 1998.

[178] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 2008.

[179] A. Davison and D. Hinkley, *Bootstrap Methods and Their Application*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

[180] R. L. Rowley, W. V. Wilding, J. L. Oscarson, Y. Yang, N. A. Zundel, T. E. Daubert, and R. P. Danner, *DIPPR Data Compilation of Pure Chemical Properties.*

[181] A. Bardow, "Optimal experimental design of ill-posed problems: The meter approach." *Comput. Chem. Eng.*, vol. 32, no. 1-2, pp. 115–124, 2008.

[182] A. Schöniger, W. A. Illman, T. Wöhling, and W. Nowak, "Finding the right balance between groundwater model complexity and experimental effort via Bayesian model selection," *J. Hydrol.*, vol. 531, Part 1, pp. 96–110, 2015.

[183] J. F. Ely and H. J. M. Hanley, "Prediction of transport properties. 1. viscosity of fluids and mixtures," *Ind. Eng. Chem. Fundam.*, vol. 20, no. 4, pp. 323–332, 1981.

[184] K. S. Pedersen, A. Fredenslund, P. L. Christensen, and P. Thomassen, "Viscosity of crude oils," *Chem. Eng. Sci.*, vol. 39, no. 6, pp. 1011–1016, 1984.

[185] W. T. Ashurst and W. G. Hoover, "Dense-fluid shear viscosity via nonequilibrium molecular dynamics," *Phys. Rev. A*, vol. 11, pp. 658–678, 1975.

[186] O. Lötgering-Lin, A. Schöniger, W. Nowak, and J. Gross, "Bayesian model selection helps to choose objectively between thermodynamic models: A demonstration of selecting a viscosity model based on entropy scaling," *Ind. Eng. Chem. Res.*, vol. 55, no. 38, pp. 10 191–10 207, 2016.

[187] C. R. Wilke, "A viscosity equation for gas mixtures," *J. Chem. Phys.*, vol. 18, no. 4, pp. 517–519, 1950.

[188] W. G. Chapman, G. Jackson, and K. E. Gubbins, "Phase equilibria of associating fluids," *Mol. Phys.*, vol. 65, no. 5, pp. 1057–1079, 1988.

[189] M. Stavrou, A. Bardow, and J. Gross, "Estimation of the binary interaction parameter kij of the PC-SAFT equation of state based on pure component parameters using a QSPR method," *Fluid Phase Equilibr.*, vol. 416, pp. 138–149, 2016.

[190] M. S. Green, "Markoff random processes and the statistical mechanics of time-dependent phenomena," *J. Chem. Phys.*, vol. 20, no. 8, pp. 1281–1295, 1952.

[191] M. S. Green, "Markoff random processes and the statistical mechanics of time-dependent phenomena. II. irreversible processes in fluids," *J. Chem. Phys.*, vol. 22, no. 3, pp. 398–413, 1954.

[192] R. Kubo, "Statistical-mechanical theory of irreversible processes. I. general theory and simple applications to magnetic and conduction problems," *J. Phys. Soc. Jpn.*, vol. 12, no. 6, pp. 570–586, 1957.

[193] R. Kubo, M. Yokota, and S. Nakajima, "Statistical-mechanical theory of irreversible processes. II. response to thermal disturbance," *J. Phys. Soc. Jpn.*, vol. 12, no. 11, pp. 1203–1211, 1957.

[194] M. Schoen and C. Hoheisel, "The shear viscosity of a Lennard-Jones fluid calculated by equilibrium molecular dynamics," *Mol. Phys.*, vol. 56, no. 3, pp. 653–672, 1985.

[195] K. Meier, A. Laesecke, and S. Kabelac, "Transport coefficients of the Lennard-Jones model fluid. I. viscosity," *J. Chem. Phys.*, vol. 121, no. 8, pp. 3671–3687, 2004.

[196] E. Sauer and J. Gross, "Classical density functional theory for liquid-fluid interfaces and confined systems: A functional for the perturbed-chain polar statistical associating fluid theory equation of state," *Ind. Eng. Chem. Res.*, vol. 56, no. 14, pp. 4119–4135, 2017.

[197] M. Thol, G. Rutkai, A. Köster, R. Lustig, R. Span, and J. Vrabec, "Equation of state for the Lennard-Jones fluid," *J. Phys. Chem. Ref. Data*, vol. 45, no. 2, p. 023101, 2016.

[198] J. L. E. Chevalier, P. J. Petrino, and Y. H. Gaston-Bonhomme, "Viscosity and density of some aliphatic, cyclic, and aromatic hydrocarbons binary liquid mixtures," *J. Chem. Eng. Data*, vol. 35, no. 2, pp. 206–212, 1990.

[199] A. Aucejo, M. Cruz Burguet, R. Munoz, and J. L. Marques, "Densities, viscosities, and refractive indices of the binary liquid systems n-alkanes + isomers of hexane at 298.15 K," *J. Chem. Eng. Data*, vol. 40, no. 4, pp. 871–874, 1995.

[200] M. J. Assael, E. Charitidou, J. H. Dymond, and M. Papadaki, "Viscosity and thermal conductivity of binary n-heptane + n-alkane mixtures," *Int. J. Thermophys.*, vol. 13, no. 2, pp. 237–249, 1992.

[201] E. F. Cooper and A. F. A. Asfour, "Densities and kinematic viscosities of some C6-C16 n-alkane binary liquid systems at 293.15 K," *J. Chem. Eng. Data*, vol. 36, no. 3, pp. 285–288, 1991.

[202] D. L. Wakefield, "Viscosities of nonelectrolyte liquid mixtures. III. selected binary and quaternary mixtures," *Int. J. Thermophys.*, vol. 9, no. 3, pp. 365–381, 1988.

[203] Y. Tanaka, H. Hosokawa, H. Kubota, and T. Makita, "Viscosity and density of binary mixtures of cyclohexane with n-octane, n-dodecane, and n-hexadecane under high pressures," *Int. J. Thermophys.*, vol. 12, no. 2, pp. 245–264, 1991.

[204] A. M. Awwad and M. A. Salman, "Volume of mixtures of nonane isomers with normal nonane and normal hexadecane at 298.15 K; an interpretation in terms of the flory-patterson theory," *Fluid Phase Equilibr.*, vol. 31, no. 1, pp. 105–115, 1986.

[205] P. Dauge, X. Canet, A. Baylaucq, and C. Boned, "Measurements of the density and viscosity of the tridecane + 2,2,4,4,6,8,8-heptamethylnonane mixtures in the temperature range 293.15 - 353.15 K at pressures up to 100 MPa," *High Temp. - High Pressures*, vol. 33, pp. 213–230, 2001.

[206] A. Laesecke, "Viscosity measurements and model comparisons for the refrigerant blends R-410A and R-507A," *ASHRAE Trans.*, vol. 110, no. 2, pp. 503–521, 2004.

[207] D. C. Landaverde-Cortes, G. A. Iglesias-Silva, M. Ramos-Estrada, and K. R. Hall, "Densities and viscosities of MTBE + nonane or decane at p = 0.1 MPa from (273.15 to 363.15) k," *J. Chem. Eng. Data*, vol. 53, no. 1, pp. 288–292, 2008.

[208] C. D. Holcomb and S. L. Outcalt, *Design of a Vapor-Liquid Equilibrium, Surface Tension, and Density Apparatus*, 1997.

[209] A. Laesecke, R. Hafer, and D. Morris, "Saturated-liquid viscosity of ten binary and ternary alternative refrigerant mixtures. part I: Measurements," *J. Chem. Eng. Data*, vol. 46, no. 2, pp. 433–445, 2001.

[210] J. Kestin and S. T. Ro, "The viscosity of nine binary and two ternary mixtures of gases at low density," *Ber. Bunsenges. Phys. Chem.*, vol. 78, no. 1, pp. 20–24, 1974.

[211] J. Kestin and J. Yata, "Viscosity and diffusion coefficient of six binary mixtures," *J. Chem. Phys.*, vol. 49, no. 11, pp. 4780–4791, 1968.

[212] S. W. Campbell, R. A. Wilsak, and G. Thodos, "(vapor + liquid) equilibrium behavior of (n-pentane + ethanol) at 372.7, 397.7, and 422.6 K," *J. Chem. Thermodyn.*, vol. 19, no. 5, pp. 449–460, 1987.

[213] B. Orge, M. Iglesias, A. Rodríguez, J. Canosa, and J. Tojo, "Mixing properties of (methanol, ethanol, or 1-propanol) with (n-pentane, *n*-hexane, *n*-heptane and *n*-octane) at 298.15 K," *Fluid Phase Equilibr.*, vol. 133, no. 1, pp. 213–227, 1997.

[214] C. Berro, "Excess volume, liquid-vapor equilibrium, and excess gibbs energy of 1-alkanols (C2-C4)+ some normal alkanes," *Int. DATA Ser., Sel. Data Mixtures, Ser. A*, vol. 15, pp. 62–81, 1987.

[215] E. Jiménez, L. Segade, C. Franjo, H. Casas, J. Legido, and M. P. Andrade, "Viscosity deviations of ternary mixtures di-*n*-butyl ether + 1-propanol + *n*-octane at several temperatures," *Fluid Phase Equilibr.*, vol. 149, no. 1, pp. 339–358, 1998.

[216] A. Estrada-Baltazar, G. A. Iglesias-Silva, and C. Caballero-Cerón, "Volumetric and transport properties of binary mixtures of *n*-octane + ethanol, + 1-propanol, + 1-Butanol, and + 1-pentanol from (293.15 to 323.15) K at atmospheric pressure," *J. Chem. Eng. Data*, vol. 58, no. 12, pp. 3351–3363, 2013.

[217] J. Schilling, M. Lampe, J. Gross, and A. Bardow, "1-stage CoMT-CAMD: An approach for integrated design of ORC process and working fluid using PC-SAFT," *Chem. Eng. Sci.*, vol. 159, no. Supplement C, pp. 217–230, 2017.

[218] G. Galliero, C. Boned, A. Baylaucq, and F. Montel, "Influence of the mass ratio on viscosity in Lennard-Jones mixtures: The one-fluid model revisited using nonequilibrium molecular dynamics," *Fluid Phase Equilib.*, vol. 234, no. 1, pp. 56–63, 2005.

[219] A. J. Queimada, S. E. Quiñones-Cisneros, I. M. Marrucho, J. A. P. Coutinho, and E. H. Stenby, "Viscosity and liquid density of asymmetric hydrocarbon mixtures," *Int. J. Thermophys.*, vol. 24, no. 5, pp. 1221–1239, 2003.

[220] M. A. Hernández-Galván, F. García-Sánchez, and R. Macías-Salinas, "Liquid viscosities of the ternary system benzene + cyclohexane + n-tetradecane from (313 to 393) K and pressures up to 60 MPa," *J. Chem. Eng. Data*, vol. 54, no. 4, pp. 1329–1333, 2009.

[221] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, pp. 107–113(6), 1993.

[222] R. A. Freeze, B. James, J. Massmann, T. Sperling, and L. Smith, "Hydrogeological decision analysis: 4. the concept of data worth and its use in the development of site investigation strategies," *Groundwater*, vol. 30, no. 4, pp. 574–588, 1992.

[223] L. Feyen and S. M. Gorelick, "Framework to evaluate the worth of hydraulic conductivity data for optimal groundwater resources management in ecologically sensitive areas," *Water Resour. Res.*, vol. 41, no. 3, 2005.

[224] R. M. Maxwell, W. E. Kastenberg, and Y. Rubin, "A methodology to integrate site characterization information into groundwater-driven health risk assessment," *Water Resour. Res.*, vol. 35, no. 9, pp. 2841–2855, 1999.

[225] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989.

[226] A. Bardow, S. Kossack, E. Kriesten, and W. Marquardt, "MEXA goes CAMD - computer-aided molecular design for physical property model building," in *18th European Symposium on Computer Aided Process*

Engineering, ser. Comput. Aided Chem. Eng. Elsevier, 2008, vol. 25, pp. 733 – 738.

[227] A. Kryukov, Z. Markhovskaya, I. Gabrielova, and V. Kiva, "Untersuchung und mathematische Beschreibung des Dampf-Flüssig-Gleichgewichtes im System Benzaldehyd-Phenole-Naphthalin bei 13,3kPa (100 mmHg)," *Osnovnoi Organ. Sintez. Neftekhim.*, vol. 22, pp. 47–52, 1986.

[228] A. Dejoz, V. Gonzalez-Alfaro, P. Miguel, and M. Vazquez, "Isobaric vapor-lliquid equilibria of trichloroethylene with 1-Butanol and 2-Butanol at 20 and 100 kPa," *J. Chem. Eng. Data*, vol. 41, pp. 89–92, 1996.

[229] J. Pardo, M. Lopez, J. Santafe, F. Royo, and J. Urirta, "Solubility of gases in butanols. 1. solubilities of nonpolar gases in 1-butanol from 263.15 to 303.15 K at 101.33 kPa partial pressure of gas," *Fluid Phase Equilib.*, vol. 109, pp. 29–37, 1995.

[230] A. Sousa, P. Fialho, C. N. D. Castro, R. T. R., and B. L. Neindre, "The thermal conductivity of 1-Chloro-1,1-difluoroethane (HCfC-142b)," *Int. J. Thermophys.*, vol. 13, pp. 383–399, 1992.

[231] A. Grishunin, M. Balashov, L. Serafimov, and E. Belikova, "Untersuchung des Dampf-Flüssig-Gleichgewichtes im System Methylacetat-Methylethylketon-Cyclohexan," *Tr. Mosk. Inst. Tonkoj Khim. Tekhnol.*, vol. 4, pp. 126–129, 1974.

[232] P. Oracz and S. Warycha, "Vapour-liquid equilibria. VII: The ternary system hexane-methanol-acetone at 313.15 K," *Fluid Phase Equilib.*, vol. 108, pp. 199–211, 1995.

[233] G. Egorochkin, Y. Semchikov, and Z. Tikhonova, "Dampf-Flüssig-Gleichgewicht in Systemen Acrylonitril-Styrol und Acrylonitril-Methylmethacrylat," *Termodin. Organic. Soedin.*, pp. 65–67, 1986.

[234] S. D. Santacreu, G. Galliero, M. Odunlami, and C. Boned, "Low density shear viscosity of Lennard-Jones chains of variable rigidities," *J. Chem. Phys.*, vol. 137, no. 20, p. 204306, 2012.

[235] M. Stavrou, A. Bardow, and J. Gross, "Parameters not published but developed during the preparation of the work by Stavrou at al.[189]."