Aditya Kulkarni
Multi-criteria Optimisation of Molecular Models of Water
Scientific Report Series Volume 48
2024

Scientific Report Series
Laboratory of Engineering Thermodynamics (LTD)
RPTU Kaiserslautern
P.O. Box 3049
67663 Kaiserslautern
Germany

ISSN 2195-7606
ISBN 978-3-944433-47-9

© LTD all rights reserved

# Multi-criteria Optimisation of Molecular Models of Water

Vom Fachbereich Maschinenbau und Verfahrenstechnik
der Rheinland-Pfälzischen Technischen Universität
Kaiserslautern-Landau
zur Verleihung des akademischen Grades

**Doktor-Ingenieur (Dr.-Ing.)**

genehmigte

**Dissertation**

von

M. Sc. Aditya Kulkarni
aus Dombivli, Indien

| | |
|---|---|
| Dekan: | Prof. Dr. rer. nat. Roland Ulber |
| Berichterstatter: | Prof. Dr.-Ing. Hans Hasse |
| | Jun. Prof. Dr.-Ing. Maximilian Kohns |
| | Prof. Dr. Michael Bortz |

Tag der mündlichen Prüfung: 2 August 2024

D 386

# Danksagung

This work was done during my activity as Research Assistant at Laboratory of Engineering Thermodynamics (LTD) at Rheinland-Pfälzische Technische Universität Kaiserslautern, under the grant for KSB foundation, grant No. 1323. I would like to thank the KSB foundation for the funding. I am also indebted to everyone who enabled my access to various supercomputers for my work. I would also take this opportunity to thank the people at the supercomputing centres in Kaiserslautern (Elwe) and Garching (SuperMUC), for making my work possible. Also, I would like to thank all those people, who have supported me professionally during my time at LTD, and to whom I am greatly indebted.

I want to express my gratitude towards Prof. Hans Hasse for making this work possible. He put his trust in me and gave me the opportunity to be a part of the wonderful group at LTD. He taught, challenged, and supported me throughout the my time at LTD. His guidance, advice and vast knowledge were very inspiring. I highly appreciate his leadership, efficiency and attention to detail.

I would also like to thank Jun. Prof. Maximilian Kohns, who has always been available to guide and support me. His easy going nature made me comfortable in handling even difficult tasks, which was crucial in completion of this work. I would also like to thank Prof. Michael Bortz, who has always been available to address my mathematical queries.

I am very grateful to my family, without whom none of this would ever have been possible. Thank you to my mother, father, brother, sister, and niece for the support in everything I have done.

I would also like to thank all my wonderful colleagues at LTD, for help, support and also for the various events we enjoyed at LTD. Your enthusiasm for several activities and events made the work much lighter than it otherwise would have been. A special thanks to those (Michael Schappals, Edder Garcia, Angelo Damone, Martin Lautenschläger and Simon Stephan) whose help was crucial for getting me introduced to the topic of molecular simulation, and helping me with understanding the technical details of different simulation packages. Also, I would like to thank Daniel Fröscher who was always available on a short notice to solve any of my computer related queries (which, to be frank, I should have solved myself).

Finally, I am thankful to all those who knowingly or unknowingly helped me to stay focussed in completing the thesis and who have encouraged me to continue when the going got tough.

Kaiserslautern, February 2024

Aditya Kulkarni

# Abstract

Water is one of the most important substances and many molecular models of water exist. Widely used molecular models of water are built from a single Lennard-Jones site on which three point charges are positioned, one negative and two positive ones. Models from that class, denoted LJ3PC here, are computationally efficient, but it is well-known that they cannot represent all relevant properties of water simultaneously with good accuracy. Therefore, multi-criteria optimisation (MCO) was used in the present work to parametrise molecular models of water from the LJ3PC class. The set of the best possible compromises between different competing objectives, the Pareto set (PS), as well as its picture in the parameter space (PSPS) were determined. Calculating PS for optimisation problems involving molecular simulations is computationally expensive. Therefore, a novel and highly efficient method was used, which is based on the fact that numerical results from molecular simulations can be interpreted as dimensionless numbers. Hence, they carry information on an entire class of models in physical units. Despite the importance of the LJ3PC water model class, its inherent limitations in simultaneously describing different properties of water have never been studied systematically. This was achieved here by calculating PS for this model class.

In the first MCO study carried out in this work, the objectives were the quality of the description of the vapour pressure, liquid density and enthalpy of vaporisation. The results were compared to models from the literature. Significant improvements were observed. The new optimisation method for the development of molecular models was shown to be efficient and robust.

Having demonstrated the applicability of the new MCO concept, in the second MCO study of the present work, the new optimisation method was applied in a hierarchical scheme to optimize LJ3PC water models taking into account five objectives: the representation of the vapour pressure, saturated liquid density, self-diffusion coefficient, shear viscosity, and relative permittivity. Of the six parameters of the LJ3PC models, five were varied; only the H–O–H bond angle, which is usually chosen based on physical arguments, was kept constant. Our hierarchical approach yields a Pareto set that contains attractive new water models. Furthermore, the results give an idea of what can be achieved by molecular modelling of water with models from the LJ3PC class.

In both MCO studies of the present work, we have found characteristic topological features of the PS and PSPS, which depend on the dimensionality of the parameter space $M$ and the objective space $N$. E.g., $M = 2$ and $N = 3$ yields triangles with needle-like extensions. The reasons for these topological features were unknown so far. It has been shown that they are to be expected if all objective functions of the MCO satisfy two conditions: (a) they can be approximated by quadratic functions and (b) one of the eigenvalues of the Hessian matrix evaluated at the function's minimum is small compared to the other eigenvalues. Objective functions which meet conditions (a) and (b) have a valley-like topology, for which the valley lies in the direction of the eigenvector corresponding to the lowest eigenvalue. The PSPS can be estimated by starting at the minimum of an objective function, following the valley, and combining these lines for all objective functions. The PS is obtained by evaluating the objective functions. We believe that the conditions (a) and (b) are met in many practical problems and discuss an example from molecular modelling. The improved understanding of the features of these MCO problems opens a route for designing methods for swiftly finding estimates of their PS and PSPS.

# Kurzfassung

Wasser ist einer der wichtigsten Stoffe, daher existieren viele molekulare Modelle für Wasser. Die am weitesten verbreiteten molekularen Modelle für Wasser bestehen aus einem Lennard-Jones Zentrum, auf dem drei Punktladungen angeordnet sind: eine negative und zwei positive. Modelle dieser Modellklasse, die hier als LJ3PC bezeichnet wird, sind in Bezug auf die Rechenzeit effizient, aber es ist bekannt, dass diese nicht alle relevanten Stoffdaten von Wasser gleichzeitig mit guter Genauigkeit abbilden können. Daher wird in der vorliegenden Arbeit die multikriterielle Optimierung (MCO) zur Parametrisierung von molekularen Modellen für Wasser eingesetzt. Dabei wird die Menge der bestmöglichen Kompromisse zwischen verschiedenen Zielgrößen, das Pareto-Set (PS), sowie sein Bild im Parameterraum (PSPS) ermittelt. Die Berechnung von PS bei Optimierungsproblemen mit molekularen Simulationen ist rechenzeitintensiv. Daher wird eine neuartige und hocheffiziente Methode verwendet, die darauf beruht, dass numerische Ergebnisse molekularer Simulationen als dimensionslose Zahlen interpretiert werden können. Daher enthalten sie Informationen über eine ganze Klasse von Modellen in physikalischen Einheiten. Trotz der Bedeutung der LJ3PC-Wassermodellklasse wurden ihre inhärenten Grenzen bei der gleichzeitigen Beschreibung verschiedener Stoffeigenschaften noch nie systematisch untersucht. Dies wurde in der vorliegenden Arbeit erreicht, in dem das PS für diese Modellklasse ermittelt wurde.

In der ersten MCO-Studie, die im Rahmen dieser Arbeit durchgeführt wurde, wurden der Dampfdruck, die Flüssigkeitsdichte und die Verdampfungsenthalpie als Zielgrößen betrachtet. Beim Vergleich der Ergebnisse mit Modellen aus der Literatur zeigten sich deutliche Verbesserungen. Es wird gezeigt, dass die neue Optimierungsmethode für die Entwicklung molekularer Modelle effizient und robust ist.

Im Anschluss an die Etablierung dieses neuen MCO-Konzepts wurde in der zweiten MCO-Studie der vorliegenden Arbeit die neue Optimierungsmethode in einem hierarchischen Schema zur Optimierung von LJ3PC-Wassermodellen eingesetzt. Dabei wurden fünf Zielgrößen betrachtet: die Beschreibung des Dampfdrucks, der Flüssigkeitsdichte, des Selbstdiffusionskoeffizienten, der Scherviskosität und der relativen Permittivität. Von den sechs Parametern der LJ3PC-Modelle wurden fünf variiert; nur der H–O–H-Bindungswinkel, der in der Regel auf Basis physikalischer Argumente gewählt wird,

wurde konstant gehalten. Der hierarchische Ansatz führt zu einem Pareto-Set, welches attraktive neue Wassermodelle enthält. Darüber hinaus zeigen die Ergebnisse die Modellgüte auf, die bei der molekularen Modellierung von Wasser mit Modellen aus der LJ3PC-Klasse für die fünf betrachteten Zielgrößen erreichbar ist.

In beiden MCO-Studien der vorliegenden Arbeit wurden charakteristische topologische Merkmale des PS und PSPS beobachtet, die von der Dimensionalität des Parameterraums $M$ und des Zielraums $N$ abhängen. So ergeben sich z.B. für $M = 2$ und $N = 3$ Dreiecke mit nadelartigen Erweiterungen. Die Gründe für diese topologischen Merkmale waren bislang unbekannt. In dieser Arbeit wird gezeigt, dass diese topologischen Besonderheiten zu erwarten sind, wenn alle Zielfunktionen des MCO zwei Bedingungen erfüllen: (a) sie können durch quadratische Funktionen approximiert werden; (b) werden die am Funktionsminimum ausgewerteten Eigenwerte der Hesseschen Matrix betrachtet, muss einer der Eigenwerte deutlich kleiner sein als die anderen. Zielfunktionen, die die Bedingungen (a) und (b) erfüllen, haben eine talartige Topologie. Dabei liegt das Tal in Richtung des Eigenvektors mit dem niedrigsten Eigenwert. Die Topologie des PSPS kann dann abgeschätzt werden, indem man am Minimum einer Zielfunktion beginnt, dem Tal folgt und die so erhaltenen Linien für alle Zielfunktionen kombiniert. Das PS wird durch Auswertung der Zielfunktionen auf dem PSPS ermittelt. Es ist zu erwarten, dass die Bedingungen (a) und (b) bei vielen Optimierungsproblemen in der Praxis erfüllt sind. Ein Beispiel aus der molekularen Modellierung wird diskutiert. Das verbesserte Verständnis der Eigenschaften von MCO-Problemen ebnet einen Weg für die Entwicklung von Methoden zur schnellen und numerisch effizienten Ermittlung von PS und PSPS.

# Contents

**Abstract** **IX**

**Kurzfassung** **XI**

**List of Symbols** **XVII**

**1 Introduction** **1**

**2 Fundamentals** **5**
2.1 Molecular models . . . . . 5
2.2 Reduced units method . . . . . 6

**3 Multi-criteria optimisation of Molecular Models of Water using a Reduced Units Approach** **9**
3.1 Introduction . . . . . 9
3.2 Molecular Models and Simulation Details . . . . . 11
3.2.1 Molecular Models . . . . . 11
3.2.2 Simulation Details . . . . . 12
3.3 Multicriteria optimisation . . . . . 12
3.3.1 Objective function . . . . . 12
3.3.2 Optimisation task . . . . . 13
3.4 Results and Discussion . . . . . 14
3.4.1 New VLE data for water models from the literature . . . . . 14
3.4.2 Overview of the optimisation studies . . . . . 21
3.4.3 Single-objective optimisation . . . . . 21
3.4.4 MCO with two objectives . . . . . 24
3.4.5 MCO with three objectives . . . . . 26
3.4.6 Selected MCO water model . . . . . 27
3.5 Conclusions . . . . . 28

**4 Hierarchical Multi-criteria Optimisation of Molecular Models of Water 29**
4.1 Introduction . . . . . 29
4.2 Studied model class and literature models . . . . . 31

4.3 Molecular simulations . . . 32
4.3.1 Simulation details for $p^s$ and $\rho^{liq}$ . . . 32
4.3.2 Details of Simulations for $D$, $\eta$ and $\epsilon_r$ . . . 32
4.3.3 Simulation outputs and empirical correlations . . . 33
4.4 Optimisation technique . . . 34
4.4.1 Objective function . . . 34
4.4.2 Grid of reduced parameters . . . 35
4.4.3 Reduced units method (RUM) . . . 36
4.4.4 Hierarchical optimisation scheme . . . 36
4.5 Results and discussion . . . 39
4.5.1 Simulation results . . . 39
4.5.2 First hierarchical MCO step: $p^s$ and $\rho^{liq}$ . . . 39
4.5.3 Second hierarchical MCO step: $p^s$, $\rho^{liq}$, $D$, $\eta$ and $\epsilon_r$ . . . 41
4.5.4 Selection of models from Pareto set . . . 43
4.6 Conclusions . . . 48

**5 Regularities of Pareto Sets in Low-dimensional Practical Multi-criteria Optimisation Problems: Analysis, Explanation, and Exploitation 51**
5.1 Introduction . . . 51
5.2 MCO with quadratic objective functions . . . 55
5.2.1 Quadratic objective functions . . . 55
5.2.2 Method of Augusto et al. . . . 55
5.2.3 Analysis . . . 57
5.2.4 Estimation of the Pareto set . . . 58
5.3 Application to non-quadratic MCO problems . . . 60
5.4 Examples . . . 61
5.4.1 Synthetic quadratic objective functions . . . 61
5.4.2 Parametrisation of a molecular model of water . . . 64
5.5 Conclusions . . . 69

**6 Conclusions 71**

**Literature 73**

**Appendix 83**

**A Supporting Information for Chapter 3 83**
A.1 Three-objective Pareto study . . . 83
A.2 Comparison between optimised and original models . . . 90
A.3 Comparison of the simulation results to literature data . . . 92

A.4 Comparison of simulation results to literature data for the Huang model 95

**B Supporting Information for Chapter 4** **97**

**C Appendix for Chapter 5** **98**

C.1 Mathematical analysis for arbitrary values of $M$ and $N$ . . . . . 98

**Declaration** **100**

**Student Theses** **102**

**Curriculum Vitae** **103**

# List of Symbols

## Latin symbols

| | |
|---|---|
| Å | Angstrom |
| $D$ | Self-diffusion coefficient |
| $\mathbf{H}$ | Hessian matrix |
| $k_{\mathrm{B}}$ | Boltzmann's constant |
| $l$ | Length |
| $m$ | Mass |
| $M$ | Number of parameters in multi-criteria optimisation |
| $N$ | Number of objectives in multi-criteria optimisation |
| $N_{\mathrm{A}}$ | Avogadro's number |
| O | Optimised |
| $p$ | pressure |
| $q$ | Point charge |
| $r_{ij}$ | Distance between centres $i$ and $j$ |
| $T$ | Temperature |
| $u_{ij}^{\mathrm{LJ}}$ | Lennard-Jones potential between LJ centres $i$ and $j$ |
| $u_{ij}^{\mathrm{E}}$ | Electrostatic potential between charge centres $i$ and $j$ |
| $V$ | Volume |
| $w$ | Weight |
| $\underline{\mathrm{x}}$ | Vector of model parameters |
| $Z$ | Thermophysical property |

## Subscripts and Superscripts

| | |
|---|---|
| exp | Experimetnal |
| liq | Liquid |
| r | Relative |
| s | Saturated |
| sim | Simulation |

| | |
|---|---|
| vap | Vaporisation |
| * | Reduced |

## Greek symbols

| | |
|---|---|
| $\alpha$ | Bond Angle |
| $\delta$ | Difference |
| $\epsilon_\mathrm{r}$ | Relative permittivity |
| $\epsilon_0$ | Vacuum permittivity |
| $\varepsilon$ | Lennard-Jones energy parameter |
| $\eta$ | Shear viscosity |
| $\lambda$ | Thermodynamic integration scaling parameter |
| $\mu$ | Chemical potential |
| $\xi$ | Eigenvalue |
| $\rho$ | density |
| $\sigma$ | Lennard-Jones size parameter |

## Abbreviations

| | |
|---|---|
| a.m.u. | Atomic mass unit |
| EoS | Equation of State |
| LJ | Lennard-Jones |
| LJ3PC | Lennard-Jones + 3 point charges |
| MC | Monte Carlo |
| MCO | Multi-criteria optimisation |
| PS | Pareto set |
| PSPS | Picture of Pareto set in parameter space |
| REFPROP | NIST Reference Fluid Thermodynamic and Transport Properties Database |
| RUM | Reduced units method |
| SGM | Selected grid model |
| VLE | Vapour-liquid equilibrium |

# 1 Introduction

Molecular simulation is an attractive approach for predicting thermodynamic properties of fluids. Molecular simulations help in digitally modelling the expensive and time-consuming laboratory experiments by relatively cheap and fast computer simulations for the chemical industry [1, 2]. The increase in computational capacity over the last decades has led to an increased use of simulations, including molecular simulations. To obtain accurate predictions of thermodynamic properties from molecular simulations, appropriate molecular models are needed. Water, being one of the most important substances, has been studied extensively from the perspective of molecular simulations. Yet, due to the intricate thermodynamic behaviour of water, no molecular model for water exists to date that describes all its properties satisfactorily.

Many molecular models of water have been described in the literature, a review has been given by Shvab and Sadus [3]. In the most widely used class of water models, water is described by a single Lennard-Jones (LJ) site (accounting for repulsion and dispersion) and three or four point charges (accounting for polarity and H-bonding) [4]. Introducing polarisability in water models can improve the description of thermodynamic properties of water [5, 6], however, at considerable computational cost and an increased complexity in the model development, including the risk of over-fitting. To assess the improvements that can be achieved by introducing polarisability, it is necessary to get an overview of what can be achieved with non-polarisable models.

In the development of molecular models of water, usually some of the model parameters are set based on results from quantum chemistry, while the others are adjusted to experimental data of various types, such as the data on the liquid density, vapour pressure and enthalpy of vaporisation. But there is no perfect water model that could describe all these data sets equally well. A parameter set that is optimal for one property is usually not optimal for other properties. Hence, compromises have to be made. This is usually accomplished by lumping together different objectives, which correspond to the different types of data, into a single objective function using weights, which are chosen in advance. The resulting objective function is then minimised to yield the desired optimal parameter set. Different strategies have been developed for optimising molecular models based on this approach [7–11].

The approach that was used in the present work is different. The problem of optimising water models, as it is described above, is obviously one with competing objectives. From a mathematical standpoint, it belongs to the field of multi-criteria optimisation (MCO). Therefore, techniques from that field have been applied to the optimisation of molecular water models. To the best of our knowledge, this has never been done before.

A straightforward application of MCO algorithms for optimising water models is presently practically infeasible due to the high computational cost associated with it. Therefore, a simplified MCO problem is solved here. Merker et al. [12] have pointed out that the results from molecular simulations are always dimensionless numbers and that this fact can be exploited for optimising molecular models. They have demonstrated this for the models that consist of LJ sites with superimposed electrostatic interaction sites, as they are also used in the present work. The idea of Merker et al. [12] is simple: physical properties can be made dimensionless by referring to the LJ size parameter $\sigma$, the LJ energy parameter $\varepsilon$ and eventually the mass of the LJ site $m$. Assume that a simulation run has been carried out and result has been obtained for a given physical property, e.g. the vapour pressure $p^s$. That result is obtained from the simulation in dimensionless form. Then, by varying $\sigma$ and $\varepsilon$, the corresponding result in physical units can be obtained – without having to carry out a new simulation – for an entire class of models, which differ in $\sigma$ and $\varepsilon$ but have the same value of the dimensionless $p^s$. This concept for optimising molecular models based on dimensionless numbers is called reduced units method (RUM) in the following.

Applying this idea to the optimisation of water models is possible and makes the solution of the MCO problem feasible. However, there is a drawback: the water models from the class that is considered here have six model parameters, but there are only two parameters that can be varied for obtaining results without the need of new simulations. Hence, only a two-dimensional subspace of the six-dimensional parameter space of the original problem can be explored and the Pareto set (PS) is only determined for this restricted subspace of the parameter space.

In the first MCO study, this drawback was accepted as this method enables tackling an otherwise practically unsolvable problem. Moreover, with the proposed method, the MCO problem is solved swiftly, i.e. with very low computational effort. It is emphasised that the solution of the restricted problem is always dominated by the solution of the original unrestricted problem: the results for the latter cannot be worse than those for the former. All that is needed for applying the method of Merker et al. [12] is one set of simulation results for a given water model from which the numbers of the objectives can be determined. Based on these data, an MCO can be carried out yielding the PS for the restricted problem along with its picture in the parameter space. Not only one model has been used as a starting point but each of the seven water models from the literature

mentioned above served as a starting point for an MCO, i.e. seven PS were obtained and compared. The properties that were considered for the fit are two vapour-liquid equilibrium (VLE) properties – $p^{\mathrm{s}}$ and the saturated liquid density $\rho^{\mathrm{liq}}$. This alleviates the drawback of the restricted parameter space to some extent.

In the second MCO study, these two issues are addressed: firstly, the study is extended to a five-dimensional objective space – including, besides $\rho^{\mathrm{liq}}$ and $p^{\mathrm{s}}$, also the following properties of the saturated liquid: shear viscosity $\eta$, self-diffusion coefficient $D$, and relative permittivity $\epsilon_{\mathrm{r}}$. All five properties are considered over a wide temperature range. Secondly, the study is extended to a five-dimensional parameter space in which, besides $\varepsilon$ and $\sigma$ of the LJ site, also the positions of the charges and the magnitude of the charges are varied. Only the angle between the two H–O bonds is kept constant and set to the physically reasonable value of 104.52°. To make the MCO study possible, a hierarchical approach was taken, and the studied region in the parameter space was restricted. The restrictions were chosen in such a way that particularly interesting regions are included.

In both the above MCO studies, characteristic topologies of the Picture of the PS in Parameter space (PSPS) as well as the PS were have been observed, which depend on the numbers of the parameters $M$ and that of the objectives $N$, but not on the specific problem that is investigated. For two objectives, the PS and PSPS show two distinct branches, which are both almost straight lines. The region near the intersection of the branches is the so called *Pareto knee* region, where the values of both the objectives are typically low, while the branches far from the Pareto knee region have high value of one objective and low value of the other. For $N = 3$, the PSPS shows triangles in the centre, with needle-like extensions from the vertices, which are, however, not always observed. In this case, the central triangular region is the Pareto knee. However, the inner workings of that relation have not been unravelled yet, and the conditions under which these characteristic topologies occur were unknown. In the third and final step of this work, these issues were clarified. The central hypothesis behind the approach was that the characteristic topological features are observed if the topologies of the individual objective functions meet certain requirements, and the aim was to elucidate the corresponding conditions.

This work is structured as follows: The fundamentals of molecular models studied in the present work, as well as those of the reduced units method are introduced in Chapter 2. The first MCO study in the present work, with two objectives, is presented in Chapter 3. The second MCO study in the present work, with five objectives and a more extensive set of model parameters, is discussed in Chapter 4. Chapter 5 presents a mathematical rationale behind the peculiar topologies of the PS and the PSPS observed in the optimisation problems from the present work. Finally, conclusions are drawn in Chapter 6.

# 2 Fundamentals

The fundamentals of molecular modelling and simulation are covered extensively in the literature, for example in the books of Rapaport [13], Frenkel and Smit [14], and Allen and Tildesley [15]. In this chapter, only concepts that are useful in the context of the present work are briefly discussed.

## 2.1 Molecular models

Non-polarisable water models, consisting of a single 12-6 LJ site and three point charges were studied in the present work. The LJ potential between two LJ centres $i$ and $j$ is given by

$$u_{ij}^{\mathrm{LJ}} = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{1}$$

Here, $r_{ij}$ is the distance between the two centres $i$ and $j$, while $\sigma$ and $\varepsilon$ are the LJ size and energy parameter, respectively. The electrostatic potential between two point charge sites $i$ and $j$ is given by

$$u_{ij}^{\mathrm{E}} = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}} \tag{2}$$

where $q_i$ and $q_j$ are the charges on sites $i$ and $j$ respectively, $r_{ij}$ is the distance between the sites $i$ and $j$, and $\epsilon_0$ is the vacuum permittivity. Fig. 1 illustrates the geometry of the LJ3PC class of water models studied here.

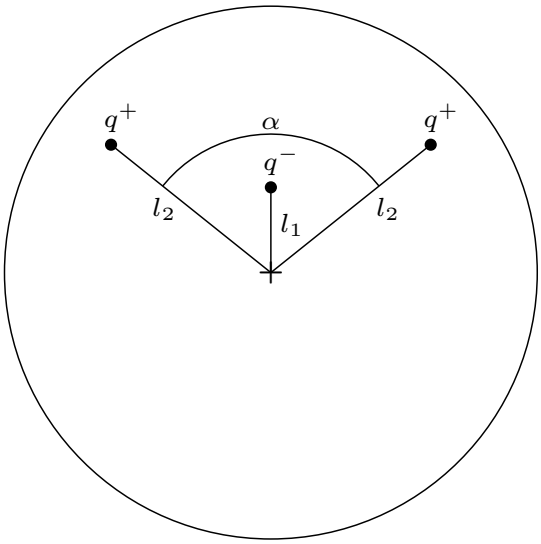



**Figure 1:** Sketch of the geometry of the class of water models studied in the present work. The diameter of the outer circle represents the LJ size parameter $\sigma$. Note that the figure is not to scale.

Due to their symmetry and the fact that a water molecule has no net charge, models from the LJ3PC class have six parameters: the two LJ parameters $\sigma$ and $\varepsilon$, the distance $l_1$ between the LJ centre and the negative point charge, the positive point charge $q^+$, its distance $l_2$ from the LJ centre, and the H–O–H bond angle $\alpha$ is usually based on the H–O–H bond angle of the real water molecule. The negative point charge is fixed to $q^- = -2q^+$ to preserve the electro-neutrality of the molecule. In some models, $l_1 = 0$, i.e. the negative point charge is placed directly on the LJ centre.

A comprehensive review of LJ3PC water models was presented by Vega and Abascal [16]. These models are computationally efficient and are, therefore, used in many scientific communities. The idea behind this modelling approach is that the simplifications, namely the lumping of polarity and H-bonding, as well as the omission of polarisability and internal degrees of freedom, is, at least partially, remedied by adjusting the model parameters to experimental thermophysical data. The parameterisation is, therefore, a crucial modelling step.

## 2.2 Reduced units method

Digital computers inherently work with dimensionless properties. The dimensionless results of physical simulations are re-converted into physical properties (carrying dimensions) in post-processing. The model parameters and thermophysical properties are represented by dimensionless numbers with respect to $\sigma$, $\varepsilon$ and the mass of the LJ site $m$. Reduced quantities are denoted by an asterisk ($*$) symbol. It is both insightful

and advantageous to take a look not only at six physical model parameters as discussed above, but also the reduced model parameters. They are obtained by making the physical parameters dimensionless using the LJ parameters $\sigma$ and $\varepsilon$ as reference according to the following equations:

$$l^* = \frac{l}{\sigma} \tag{3}$$

$$q^* = \frac{q}{\sqrt{4\pi\epsilon_0\sigma\varepsilon}} \tag{4}$$

The bond angle is the same in the physical and reduced representation. Thus, in addition to the bond angle, the LJ3PC models have three parameters in reduced units: $l_1^*$, $l_2^*$ and $q^*$.

The reduced temperature $T^*$, pressure $p^*$, density $\rho^*$, self-diffusion coefficient $D^*$ and reduced shear viscosity $\eta^*$ are defined as follows:

$$T^* = \frac{T}{\varepsilon/k_\mathrm{B}} \tag{5}$$

$$p^* = \frac{p\sigma^3}{\varepsilon} \tag{6}$$

$$\rho^* = \rho\sigma^3 N_\mathrm{A} \tag{7}$$

$$h^* = \frac{h}{\varepsilon N_\mathrm{A}} \tag{8}$$

$$D^* = \frac{D}{\sigma}\sqrt{\frac{m}{\varepsilon}} \tag{9}$$

$$\eta^* = \frac{\eta\sigma^2}{\sqrt{m\varepsilon}} \tag{10}$$

Here, $k_\mathrm{B}$ is Boltzmann's constant and $N_\mathrm{A}$ is Avogadro's number. The relative permittivity $\epsilon_\mathrm{r}$ is already a dimensionless number which does not change when applying the RUM. Here, $m$ is the mass of the water molecule, for which the physical value of 18.016 a.m.u. was used here.

The RUM [12] enables obtaining the information on an entire class of models, in which $\sigma$ and $\varepsilon$ are varied, from a single simulation run. The scaling of the results can be done in post-processing, i.e. without running additional simulations. However, together with $\sigma$ and $\varepsilon$, also the other model parameters $l_1$, $l_2$ and $q$ vary in the way defined by Eqs. (3) and (4). While $m$ can theoretically be varied in the RUM, it is kept constant in the present work, as varying $m$ will yield a molecular model that has different molecular mass than water. Moreover, the temperature also scales with $\varepsilon$. Hence, changing $\varepsilon$ has an influence on the points at which the model results can be compared to the reference data. This causes no problems, as the reference data are taken from a temperature-dependent correlation such that a change in $\varepsilon$ can be accounted for easily.

# 3 Multi-criteria optimisation of Molecular Models of Water using a Reduced Units Approach

## 3.1 Introduction

In the present work, it is shown that MCO can be applied for the optimisation of molecular water models. MCO is only briefly introduced here; the reader is referred to [17–21] for more comprehensive presentations. MCO yields not only a single solution but a set of optimal solutions, the so-called Pareto set. Thereby, the optimality is defined as follows: a solution of an optimisation problem with competing objectives is optimal, if any further improvement in one objective can only be achieved at the expense of a deterioration in at least one other objective [22]. The Pareto set can generally only be determined numerically; and different algorithms are available for this [23–26]. During the numerical solution of the MCO problem, usually many calls of a routine are needed, in which the values of the different objective functions are determined. This is extremely costly when the objectives are thermodynamic properties at various state points that have to be determined from molecular simulations. It is therefore not astonishing that MCO has never been used before for optimising molecular water models.

In MCO problems, the parameter space and the objective space are distinguished. For each point in the parameter space, there is a point in the objective space, representing the corresponding results for the different objectives. The Pareto set exists in the objective space and has an image in the parameter space. For a given application, a certain model has to be selected from the Pareto set. The choice will depend on the preferences regarding the different objectives. It is an instructed choice, as it is based on the knowledge of the entire Pareto set. The corresponding model parameters are found from the image of the chosen point of the Pareto set in the parameter space.

Only the most popular class of water models that consist of a 12-6 Lennard-Jones (LJ) site and three point charges is considered. A comprehensive overview of water models from this class has been given by Vega and Abascal [16]. For carrying out the

MCO, the objectives need to be defined. Here, three objectives that characterise the quality of the representation of the vapour-liquid equilibrium (VLE), namely $p^{\mathrm{s}}$, $\rho^{\mathrm{liq}}$ and $\Delta h^{\mathrm{vap}}$, are considered. These are basic properties that are important in many processes, and are therefore widely used in the development of molecular models of fluids [27, 28]. Furthermore, it is known that if these properties are described well, usually also other properties, like those in the homogeneous region, are predicted reasonably well by the molecular model [29–31]. The temperature range that was taken into account in the present work is 350-550 K. It is chosen such that $p^{\mathrm{s}}$, $\rho^{\mathrm{liq}}$ as well as $\Delta h^{\mathrm{vap}}$ can be determined with sufficient accuracy with molecular simulations. As reference data, results from an accurate multi-parametric equation of state (EoS) of water, the IAPWS-95 EoS [32], as implemented in the NIST Reference Fluid Thermodynamic and Transport Properties Database (REFPROP) [33], was used.

As a point of departure and for comparisons, following seven water models are considered: SPC [34], SPC/E [35], TIP3P [36], TIP4P [36], TIP4P/2005 [37], OPC [38] and the model of Huang et al. [39, 40], which will be abbreviated with 'Huang' for convenience. In the development of these models, the model parameters were adjusted to the experimental data of certain thermodynamic properties. In all cases, the classical optimisation approach was used, in which different objectives were lumped together in a single objective function. Molecular simulations with the seven water models from the literature were carried out in the present work and the numbers for the three objectives mentioned above were evaluated. These results are compared to a model obtained by MCO in the present work. In assessing the results, it has to be considered that the objectives that were used for training the literature models by their authors and those used for the present evaluation are not necessarily identical.

MCO has been applied before for the development of molecular models [29, 41–43]. However, in these works simple molecular models were studied, for which for all properties that were considered as objectives, dimensionless correlations were available. Such correlations are not available for the water model class considered here, and would be extremely tedious to supply, as for all properties, empirical functions of all model parameters would have to be established.

In this chapter, the first MCO study for parametrisation of water models is discussed. The RUM is applied in the present exercise to reduce the computational efforts required for MCO.

## 3.2 Molecular Models and Simulation Details

### 3.2.1 Molecular Models

Figure 2 gives an overview of the parameters of the seven water models from the literature studied in the present work. The parameters of the selected result of MCO from the present work (see Section 3.4.6) are also shown in Figure 2. The corresponding numerical values are given in Table 2. On the one hand, some parameters vary in a narrow range, i.e. the values of $\sigma$ vary from 3.08 Å to 3.17 Å. On the other hand, some of the model parameters vary in a wide range, i.e. the values of $\varepsilon/k_B$ vary from 76.54 K to 208.08 K, where $k_B$ is Boltzmann's constant. $\alpha$ has three distinct values, as different methods were used for the determination: quantum chemical calculations (for OPC), experimental value in the gas phase (for TIP4P/2005, TIP4P, TIP3P and Huang) and the tetrahedral angle (for SPC/E and SPC). Similarly, $l_2$ also has four distinct values, as different methods were used for the determination: quantum chemical calculations (for OPC), the experimental value (for TIP4P/2005, TIP4P and TIP3P) and preselected fixed values for SPC/E, SPC and Huang. The other parameters were adjusted to reproduce some experimental properties, hence their values do not show any pattern.

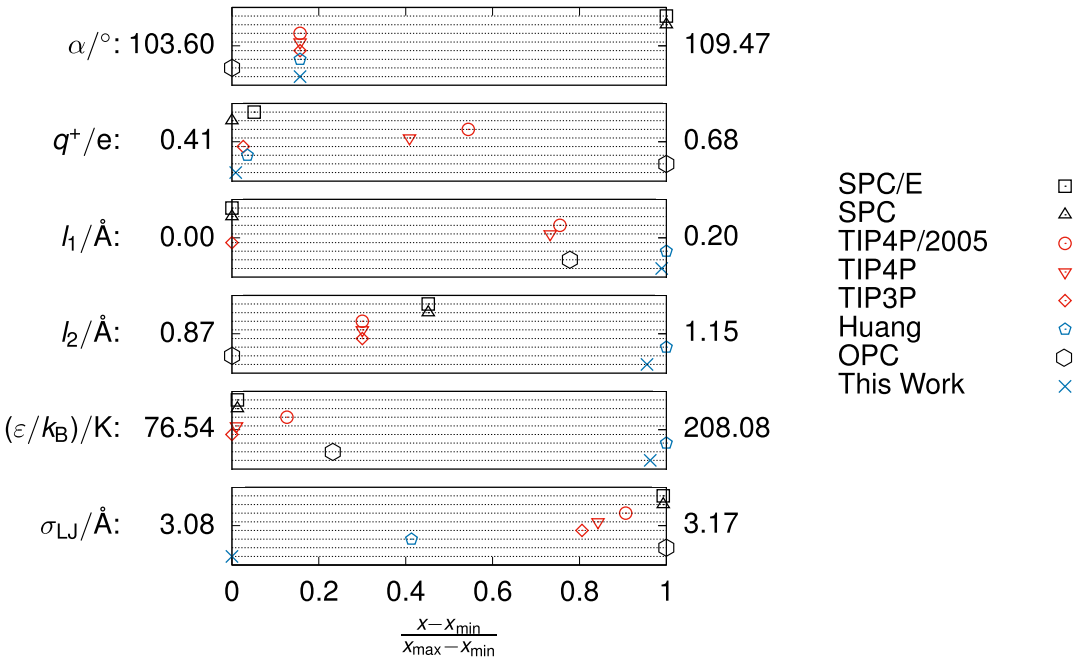



**Figure 2:** Overview of the six parameters of the seven water models from the literature that were studied in the present work. Also the parameters of the selected MCO water model from the present work (see Section 3.4.6) are shown. $x$ stands for the value of a parameter for a given model. The maximal value $x_{max}$ and the minimal value $x_{min}$ of the parameters of the models are given on the right and on the left.

### 3.2.2 Simulation Details

In this work, VLE data ($p^s$, $\rho^{liq}$ and $\Delta h^{vap}$ between 350 K and 550 K) of the seven water models from the literature were determined by molecular simulations. For TIP3P and TIP4P, the critical point is low and hence, VLE simulations were only carried out between 350 K and 525 K.

All simulations were carried out using the molecular simulation program *ms*2 (`https://www.ms-2.de/home.html`) [44]. The VLE of each water model was calculated with the Grand Equilibrium method [45]. In the Grand Equilibrium method, the VLE of a pure fluid is obtained in two subsequent steps: a Monte-Carlo (MC) $NpT$ run in which the chemical potential and its pressure derivative of the liquid are sampled, followed by a pseudo-$\mu VT$ vapour run in which the chemical potential is kept 'constant' as a function of pressure. First, MC simulations were performed in the $NpT$ ensemble for the liquid phase. Starting with 500 molecules in a face-centred cubic lattice, equilibration consisted of 20,000 MC cycles in the $NVT$ and 100,000 MC cycles in the $NpT$ ensemble. For production, 4.8 million MC cycles were performed at higher temperatures (500 K and 525 K/550 K), whereas 9.6 million cycles were performed at lower temperatures (450 K, 400 K and 350 K). Each MC cycle consisted of 500 translation moves, 500 rotation moves and 1 volume move. The chemical potential was calculated by thermodynamic integration [14]. 100 steps for $\lambda$ were used, the minimum and maximum values of $\lambda$ were 0.2 and 1.0 respectively. The maximum step for $\lambda$ was 0.1.

For the vapour phase, MC simulations were performed in the pseudo-$\mu VT$ ensemble. The simulation volume was adjusted to yield on average 864 molecules in the vapour phase. For equilibration, 10,000 cycles in the $NVT$ ensemble and 20,000 cycles in the pseudo-$\mu VT$ ensemble were performed, again starting from a face-centred cubic lattice.

## 3.3 Multicriteria optimisation

### 3.3.1 Objective function

In this work, $p^s$, $\rho^{liq}$ and $\Delta h^{vap}$ at five temperatures – 350 K, 400 K, 450 K, 500 K and 550 K (525 K for TIP3P and TIP4P) – are considered for optimisation. For each property $Z$, an objective function $\delta Z$ is defined as follows:

$$\delta Z(\underline{x}) = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\left[ \frac{Z_{sim}(\underline{x}, T_i) - Z_{exp}(T_i)}{Z_{exp}(T_i)} \right]^2} \tag{11}$$

Thus, for a given model, denoted by the vector of model parameters $\underline{x}$, $\delta Z$ contains the information on the deviation between the simulated value $Z_{\mathrm{sim}}$ and the experimental value $Z_{\mathrm{exp}}$ of the property $Z$, averaged over $N$ temperatures. For $Z_{\mathrm{sim}}$, the simulation results from the present work are used, while for $Z_{\mathrm{exp}}$, the numbers are taken from the IAPWS-95 EoS [32, 33].

### 3.3.2 Optimisation task

In the first step, for each of the seven water models from the literature, a set of molecular simulations was performed in which the VLE at five temperatures were determined and the corresponding values for $p^{\mathrm{s}}$, $\rho^{\mathrm{liq}}$ and $\Delta h^{\mathrm{vap}}$ were measured. In a post-processing step, the obtained simulation results were rescaled.

A grid was used for this purpose, on which both $\sigma$ and $\varepsilon$ were varied using 800 different values for each of the two parameters. This results in an 800×800 grid of model variants. The range covered by the grid was chosen such that it contains the picture of the entire Pareto set of the studied problem (see Section 3.4.4 and 3.4.5), which involved iterations in some cases. Calculating the results for $p^{\mathrm{s}}$, $\rho^{\mathrm{liq}}$ and $\Delta h^{\mathrm{vap}}$ for the 800×800 models from the grid can be done simply by using Eqs. (6) - (8), without additional simulations. From the results, the numbers for the objective functions $\delta p^{\mathrm{s}}$, $\delta \rho^{\mathrm{liq}}$ and $\delta(\Delta h^{\mathrm{vap}})$ can be determined easily.

As described above, the LJ parameters $\sigma$ and $\varepsilon$ were varied independently, while other model parameters vary depending on the choices for $\sigma$ and $\varepsilon$, as prescribed by Eqs. (3) and (4). Thus, this procedure is equivalent to investigating a two-dimensional subspace of the six-dimensional parameter space. The subspace that is investigated depends on the original model for which the simulations were carried out.

The results for the objective functions on the grid were then used in the MCO. In a preliminary step, the three objectives were considered individually. Then, the simultaneous representation of multiple objectives was considered. Pareto set was determined by a brute force enumeration: each model on the grid that is dominated by at least one other model was sorted out. The remaining set, i.e. the set of non-dominated models, is the Pareto set. This calculation, together with the reduced units scaling, was implemented in MATLAB®. A typical run took about 10 minutes. The lower and upper limits of the $\sigma - \varepsilon$ grid of each original model were chosen such that the picture of the entire Pareto set lies within the studied range of parameters, which required an iteration in some cases. A scheme illustrating the workflow of the optimisation procedure used in the present work, along with its comparison to that of a typical molecular simulation, is shown in Fig. 3.

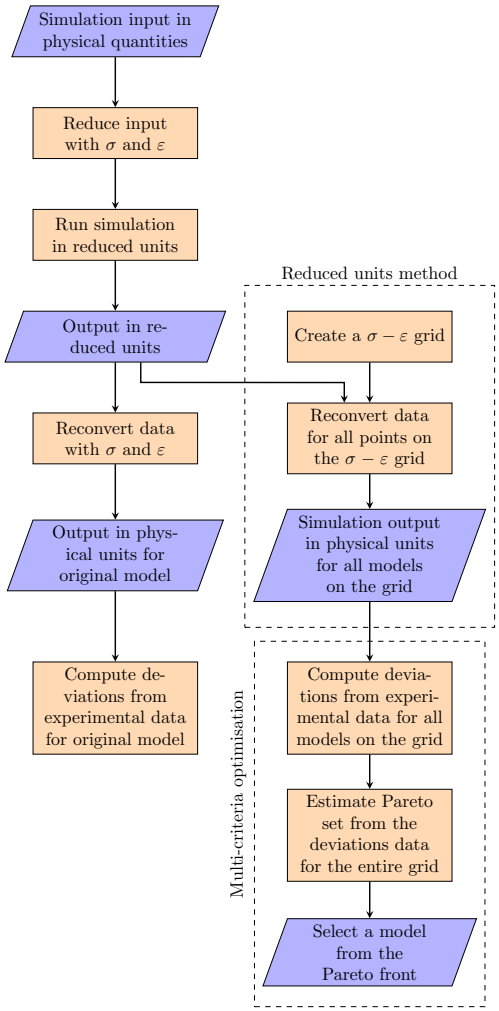



**Figure 3:** Scheme illustrating the workflows used in the present work. The left branch shows the typical workflow for carrying out a molecular simulation with a given molecular model and assessing its agreement with experimental data. The right branch shows the optimisation approach used in the present work, combining the reduced units method with multi-criteria optimisation.

## 3.4 Results and Discussion

### 3.4.1 New VLE data for water models from the literature

Figures 4, 5 and 6 present the simulation results for $p^{\mathrm{s}}$, $\rho^{\mathrm{liq}}$, $\rho^{\mathrm{vap}}$ and $\Delta h^{\mathrm{vap}}$ for the VLE of the seven studied water models from the literature. For comparison, the results of the model obtained from MCO in the present work are also presented. They are not discussed here, but only in Section 3.4.6. In Figures 4, 5 and 6, results for $p^{\mathrm{s}}$, $\rho^{\mathrm{liq}}$, $\rho^{\mathrm{vap}}$ and $\Delta h^{\mathrm{vap}}$ are shown along with the corresponding deviations from the reference data. The numerical simulation results are presented in Table 1. Numerical results for the

deviations between the simulation results for the literature models and the IAPWS-95 EoS ($\delta p^{\mathrm{s}}$, $\delta \rho^{\mathrm{liq}}$ and $\delta(\Delta h^{\mathrm{vap}})$) are presented in Table 2.

For SPC/E, TIP3P, TIP4P and TIP4P/2005, simulation data for $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$ are available on the NIST standard reference simulation website [46]. These data are also included in Figures 4, 5 and 6, but only in the deviation plots, for better clarity. Good agreement is obtained between the NIST data and our data for all models. As a further validation, a similar comparison with additional literature data [39, 47–49] is presented in Appendix A. This enables an assessment of the employed codes and procedures, which is not trivial [50–52]. Good agreement is found, except for the Huang model. For that model, the original data reported in ref. [39] , a previous work of our group, could not be reproduced in the present work. As this problem occurred only for the data from ref. [39] , it is concluded that there is an error in the data from ref. [39] and its further use is discouraged. Instead, the data from the present work for the Huang model reported in Table 1 should be used.

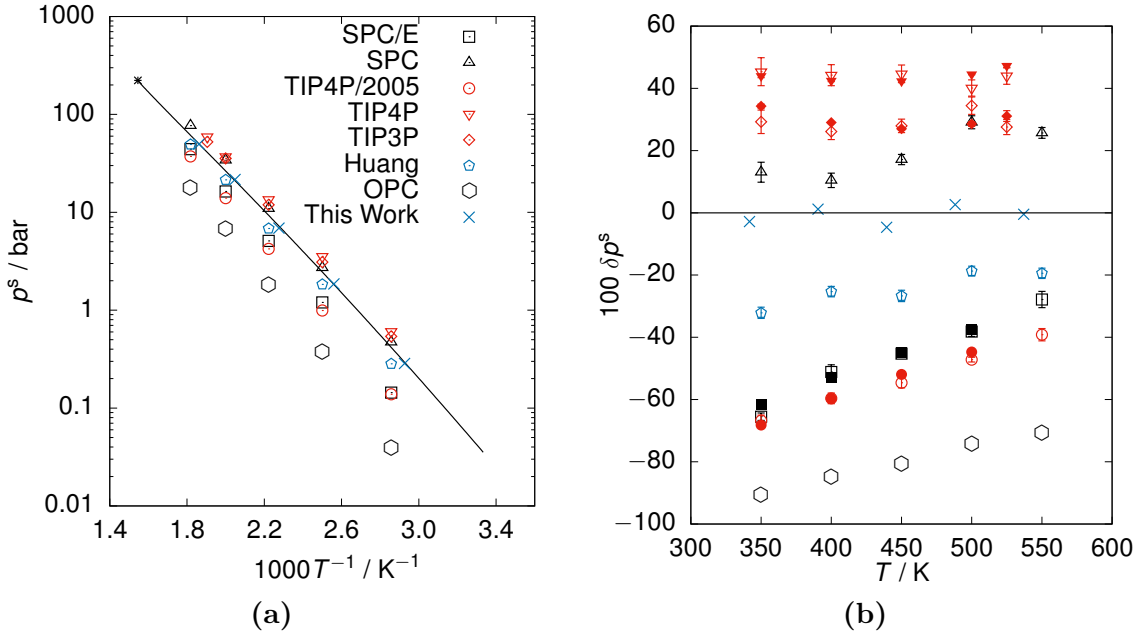
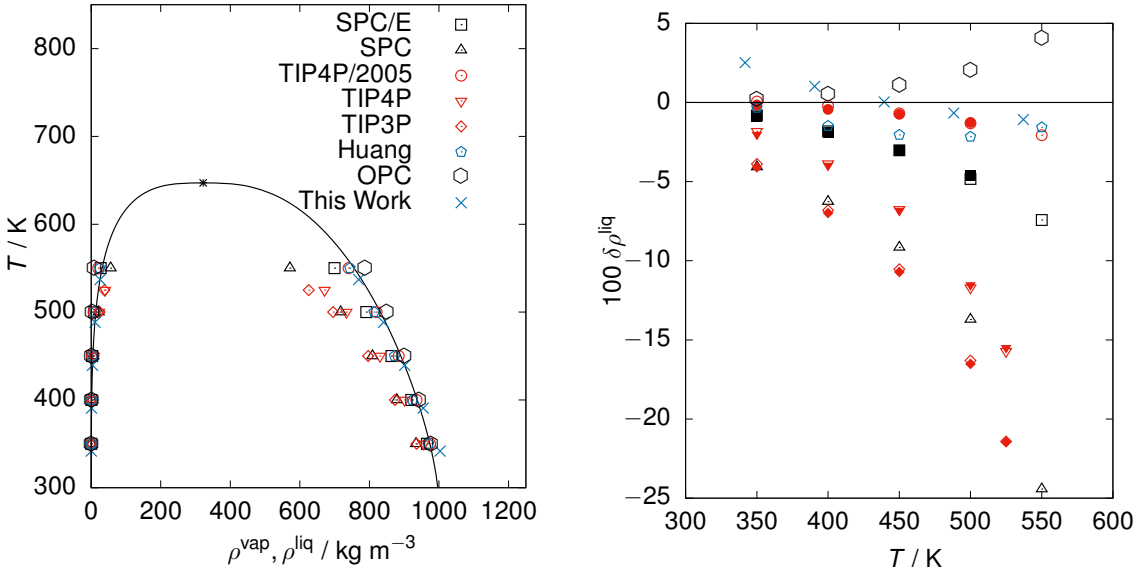
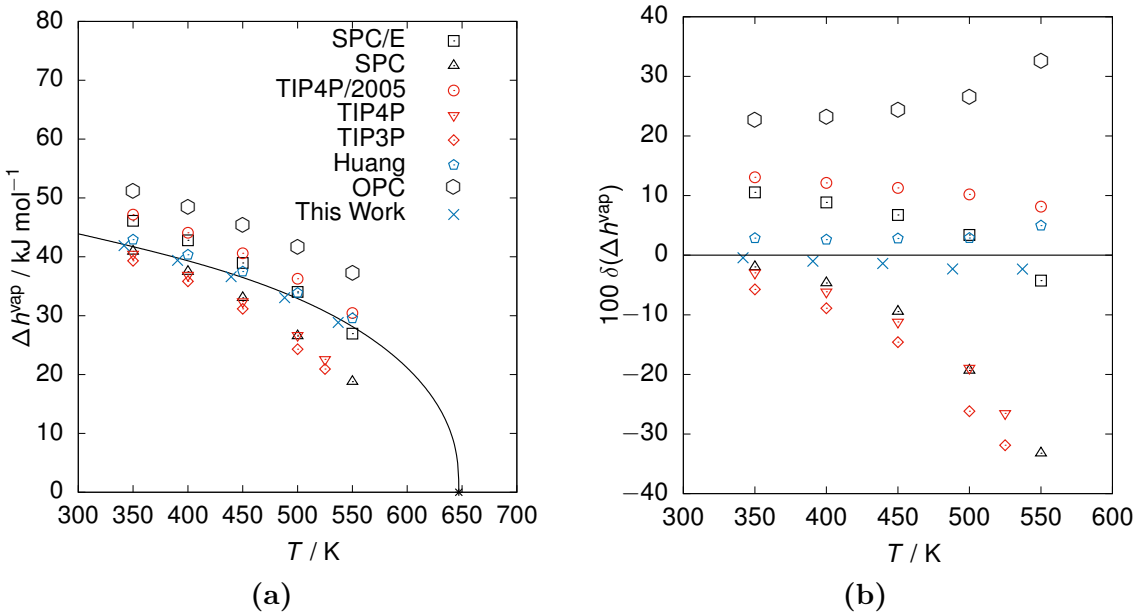



**Figure 4:** (a) Vapour pressure of water, obtained with different models from the literature as well as results from the selected MCO water model from the present work (see Section 3.4.6), as a function of temperature and (b) corresponding deviations from the IAPWS-95 EoS [32, 33]. Open symbols show the simulation results from the present work, filled symbols show simulation results from the literature for SPC/E, TIP4P/2005, TIP4P and TIP3P [46]. The line shows the results from the IAPWS-95 EoS. Statistical uncertainties are within symbol size where not indicated.




**(a)** **(b)**

**Figure 5:** (a) Saturated liquid and vapour densities of water, obtained with different models from the literature as well as results from the selected MCO water model from the present work (see Section 3.4.6), as a function of temperature and (b) corresponding deviations of the saturated liquid densities from the IAPWS-95 EoS [32, 33]. Open symbols show the simulation results from the present work, filled symbols show simulation results from the literature for SPC/E, TIP4P/2005, TIP4P and TIP3P [46]. The line shows the results from the IAPWS-95 EoS. Statistical uncertainties are within symbol size where not indicated.




**(a)** **(b)**

**Figure 6:** (a) Enthalpy of vaporisation of water, obtained with different models from the literature as well as results from the selected MCO water model from the present work (see Section 3.4.6), as a function of temperature and (b) corresponding deviations from the IAPWS-95 EoS [32, 33]. Symbols show the simulation results from the present work. The line shows the results from the IAPWS-95 EoS. Statistical uncertainties are within symbol size.

**Table 1:** VLE simulation results obtained in the present work for different water models.

| $T$/K | $p^s$/bar | $\rho^{liq}$/kg m$^{-3}$ | $\rho^{vap}$/kg m$^{-3}$ | $\Delta h^{vap}$/kJ mol$^{-1}$ |
|---|---|---|---|---|
| | | SPC | | |
| 350 | 0.472±0.02 | 934.07±0.1 | 0.302±0.009 | 40.879±0.02 |
| 400 | 2.778±0.06 | 875.29±0.1 | 1.634±0.036 | 37.466±0.02 |
| 450 | 11.103±0.20 | 808.17±0.1 | 6.413±0.113 | 32.975±0.02 |
| 500 | 31.571±0.59 | 716.63±0.3 | 19.421±0.364 | 27.221±0.05 |
| 550 | 80.627±1.45 | 576.62±2.5 | 64.371±1.154 | 17.787±0.13 |
| | | SPC/E | | |
| 350 | 0.156±0.01 | 965.66±0.1 | 0.097±0.004 | 46.902±0.01 |
| 400 | 1.111±0.03 | 920.20±0.1 | 0.631±0.018 | 42.938±0.02 |
| 450 | 5.010±0.11 | 863.18±0.1 | 2.701±0.057 | 39.005±0.03 |
| 500 | 16.140±0.38 | 792.31±0.3 | 8.765±0.207 | 34.196±0.04 |
| 550 | 40.485±0.81 | 698.12±0.4 | 23.637±0.468 | 27.933±0.07 |

Table 1 (continued)

| $T$/K | $p^{\mathrm{s}}$/bar | $\rho^{\mathrm{liq}}$/kg m$^{-3}$ | $\rho^{\mathrm{vap}}$/kg m$^{-3}$ | $\Delta h^{\mathrm{vap}}$/kJ mol$^{-1}$ |
|---|---|---|---|---|
| | | TIP3P | | |
| 350 | 0.553±0.01 | 935.62±0.1 | 0.355±0.010 | 39.296±0.02 |
| 400 | 3.183±0.06 | 873.13±0.1 | 1.913±0.040 | 35.758±0.03 |
| 450 | 11.336±0.19 | 795.95±0.2 | 6.683±0.112 | 31.448±0.03 |
| 500 | 33.250±0.66 | 694.75±0.4 | 21.817±0.432 | 25.034±0.05 |
| 525 | 52.327±0.99 | 625.65±0.9 | 37.942±0.719 | 20.930±0.06 |
| | | TIP4P | | |
| 350 | 0.587±0.02 | 955.53±0.1 | 0.372±0.012 | 40.519±0.01 |
| 400 | 3.535±0.08 | 900.66±0.1 | 2.072±0.049 | 36.948±0.02 |
| 450 | 13.290±0.25 | 829.47±0.1 | 7.692±0.142 | 32.454±0.02 |
| 500 | 38.856±0.73 | 734.60±0.4 | 24.682±0.468 | 26.192±0.05 |
| 525 | 59.074±1.12 | 670.99±0.7 | 39.855±0.755 | 22.572±0.07 |
| | | TIP4P/2005 | | |
| 350 | 0.118±0.01 | 973.46±0.1 | 0.074±0.003 | 47.183±0.04 |
| 400 | 0.988±0.03 | 934.42±0.1 | 0.558±0.017 | 44.101±0.02 |
| 450 | 4.454±0.09 | 883.88±0.1 | 2.342±0.047 | 40.545±0.02 |
| 500 | 15.132±0.36 | 820.68±0.2 | 7.927±0.180 | 35.917±0.02 |
| 550 | 38.606±0.69 | 739.25±0.3 | 21.313±0.378 | 30.157±0.04 |
| | | Huang | | |
| 350 | 0.268±0.01 | 970.32±0.1 | 0.167±0.007 | 42.905±0.01 |
| 400 | 1.762±0.06 | 923.54±0.1 | 0.973±0.031 | 40.373±0.01 |
| 450 | 7.419±0.13 | 871.78±0.1 | 3.855±0.067 | 37.375±0.01 |
| 500 | 21.065±0.36 | 813.08±0.1 | 10.521±0.182 | 33.904±0.01 |
| 550 | 49.450±0.87 | 743.57±0.2 | 25.042±0.450 | 29.479±0.02 |

| | | OPC | | |
|---|---|---|---|---|
| 350 | 0.039±0.01 | 976.25±0.1 | 0.023±0.001 | 51.176±0.01 |
| 400 | 0.377±0.01 | 942.51±0.1 | 0.209±0.006 | 48.448±0.01 |
| 450 | 1.818±0.04 | 899.92±0.1 | 0.924±0.022 | 45.506±0.02 |
| 500 | 6.816±0.19 | 848.57±0.2 | 3.315±0.094 | 41.656±0.03 |
| 550 | 18.005±0.38 | 786.72±0.2 | 8.684±0.186 | 37.315±0.04 |

Among the literature models, those with the lowest $\delta p^{\mathrm{s}}$ are SPC (19.1%), Huang (25.7%) and TIP3P (29.4%). The models with the lowest $\delta \rho^{\mathrm{liq}}$ are TIP4P/2005 (0.9%), Huang (1.5%) and OPC (1.5%). The models with the lowest $\delta(\Delta h^{\mathrm{vap}})$ are Huang (3.2%), SPC/E (6.7%) and TIP4P (9.8%). For the model with the lowest $\delta p^{\mathrm{s}}$, i.e. SPC, $\delta \rho^{\mathrm{liq}}$ is 11.5% and $\delta(\Delta h^{\mathrm{vap}})$ is 13.7%. For the model with the lowest $\delta \rho^{\mathrm{liq}}$, i.e. TIP4P/2005, $\delta p^{\mathrm{s}}$ is 53.5% and $\delta(\Delta h^{\mathrm{vap}})$ is 11%. For the model with the lowest $\delta(\Delta h^{\mathrm{vap}})$, i.e. Huang, $\delta p^{\mathrm{s}}$ is 25.7% and $\delta \rho^{\mathrm{liq}}$ is 1.5%. Altogether, for all models, the deviation in at least one of the three properties is high. This is consistent with the observations of Shvab and Sadus [53].

It is instructive to compare the results for all objectives obtained with all studied models. This is done in Figure 7. The plot confirms that none of the models studied here represents all the three properties satisfactorily. While Huang might be considered the best among the studied models (which is not astonishing as it was optimised for describing the properties that are also considered here), it still has a deviation of over 20% in $p^{\mathrm{s}}$. Thus, there appears to be ample room for an improvement in the simultaneous representation of the three VLE properties.

| Model | Parameters | | | | | | Deviations | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\sigma$/Å | $(\varepsilon/k_\mathrm{B})$/K | $l_1$/Å | $l_2$/Å | $q^+$/e | $\alpha$/deg | $100\delta p^\mathrm{s}$ | $100\delta\rho^\mathrm{liq}$ | $\delta(\Delta h^\mathrm{vap})$ |
| SPC Original | 3.1660 | 78.1775 | 0.0000 | 1.0000 | 0.4100 | 109.47 | 19.1 | 11.5 | 13.7 |
| Pareto Knee | 3.0821 | 80.3665 | 0.0000 | 0.9735 | 0.4102 | 109.47 | 9.0 | 5.2 | 9.9 |
| SPC/E Original | 3.1660 | 78.1775 | 0.0000 | 1.0000 | 0.4238 | 109.47 | 45.6 | 3.6 | 6.7 |
| Pareto Knee | 3.1066 | 74.1904 | 0.0000 | 0.9812 | 0.4090 | 109.47 | 8.0 | 3.3 | 5.5 |
| TIP3P Original | 3.1506 | 76.5418 | 0.0000 | 0.9572 | 0.4170 | 104.52 | 29.1 | 11.8 | 17.4 |
| Pareto Knee | 3.0876 | 78.7998 | 0.0000 | 0.9381 | 0.4189 | 104.52 | 2.2 | 3.0 | 8.6 |
| TIP4P Original | 3.1536 | 77.9369 | 0.1500 | 0.9572 | 0.5200 | 104.52 | 43.7 | 8.0 | 13.1 |
| Pareto Knee | 3.1268 | 81.2103 | 0.1487 | 0.9491 | 0.5285 | 104.52 | 5.5 | 2.1 | 5.3 |
| TIP4P/2005 Original | 3.1589 | 93.1996 | 0.1546 | 0.9572 | 0.5564 | 104.52 | 53.5 | 0.9 | 11.0 |
| Pareto Knee | 3.1170 | 86.9553 | 0.1543 | 0.9445 | 0.5339 | 104.52 | 9.7 | 2.3 | 4.4 |
| Huang Original | 3.1183 | 208.080 | 0.2048 | 1.1549 | 0.4195 | 104.52 | 25.7 | 1.5 | 3.2 |
| Pareto Knee* | 3.0843 | 203.190 | 0.2026 | 1.1423 | 0.4123 | 104.52 | 2.3 | 1.1 | 1.5 |
| OPC Original | 3.1666 | 107.079 | 0.1594 | 0.8724 | 0.6791 | 103.6 | 80.1 | 1.6 | 25.9 |
| Pareto Knee | 3.1128 | 92.6235 | 0.1567 | 0.8576 | 0.6262 | 103.6 | 4.3 | 2.1 | 2.9 |

**Table 2:** Overview of water models from the literature and results of the selected water models obtained from MCO (at Pareto knee), in which the corresponding literature models were taken as starting points. The model parameters and the results for the deviations in the vapour pressure ($\delta p^\mathrm{s}$), the saturated liquid density ($\delta\rho^\mathrm{liq}$) and the enthalpy of vaporisation ($\delta(\Delta h^\mathrm{vap})$) are given. In the MCO, only $\delta p^\mathrm{s}$ and $\delta\rho^\mathrm{liq}$ were used as objectives. The results for $\delta(\Delta h^\mathrm{vap})$ are only reported for completeness. The recommended model from the present work, for which the results are also shown in Figures 4-6, is marked with an asterisk.

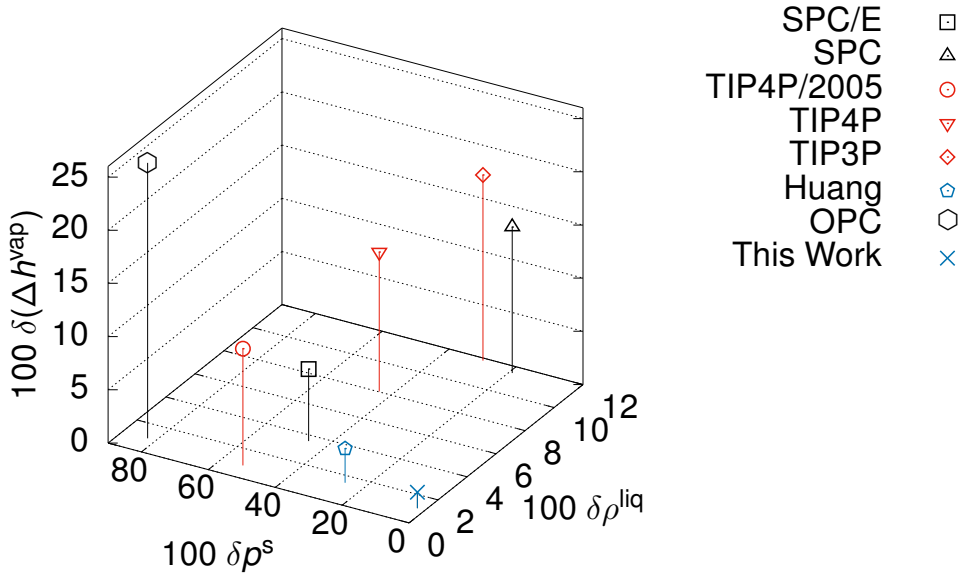



**Figure 7:** Overview of the deviations in vapour pressure, saturated liquid density and enthalpy of vaporisation of all literature models studied in the present work, as well as the result of the selected MCO water model from the present work (see Section 3.4.6).

### 3.4.2 Overview of the optimisation studies

As described above, all optimisation studies from the present work were carried out using the reduced units method, i.e. by scaling the results from a given model. This scaling is done by varying the parameters $\sigma$ and $\varepsilon$; all other parameters vary accordingly as given by Eqs. (3) and (4). Each of the seven literature models discussed above was used as starting point for this optimisation.

Among the seven models studied in the present work, Huang results in the best description of the VLE properties. Hence, the results that were obtained with the Huang model as a starting point are discussed here as an example. The results from the analogous studies with the other six water models are presented in Appendix A.

In a first step, each objective was considered separately. Hence a single objective optimisation was carried out, albeit in a way that prepares the ground for the MCO. The MCO was then carried out in two variants: one using two objectives (viz. $\delta p^{\mathrm{s}}$ and $\delta\rho^{\mathrm{liq}}$) and a second, in which all three objectives were considered (viz. $\delta p^{\mathrm{s}}$, $\delta\rho^{\mathrm{liq}}$ and $\delta(\Delta h^{\mathrm{vap}})$).

### 3.4.3 Single-objective optimisation

Figure 8 shows the dependence of $\delta p^{\mathrm{s}}$, $\delta\rho^{\mathrm{liq}}$ and $\delta(\Delta h^{\mathrm{vap}})$ on the LJ parameters $\sigma$ and $\varepsilon$ of the scaled variants of Huang, obtained by the reduced units method [12]. The

contour plots give information on deviations for the respective property averaged over five temperatures (see Eq. (11)).

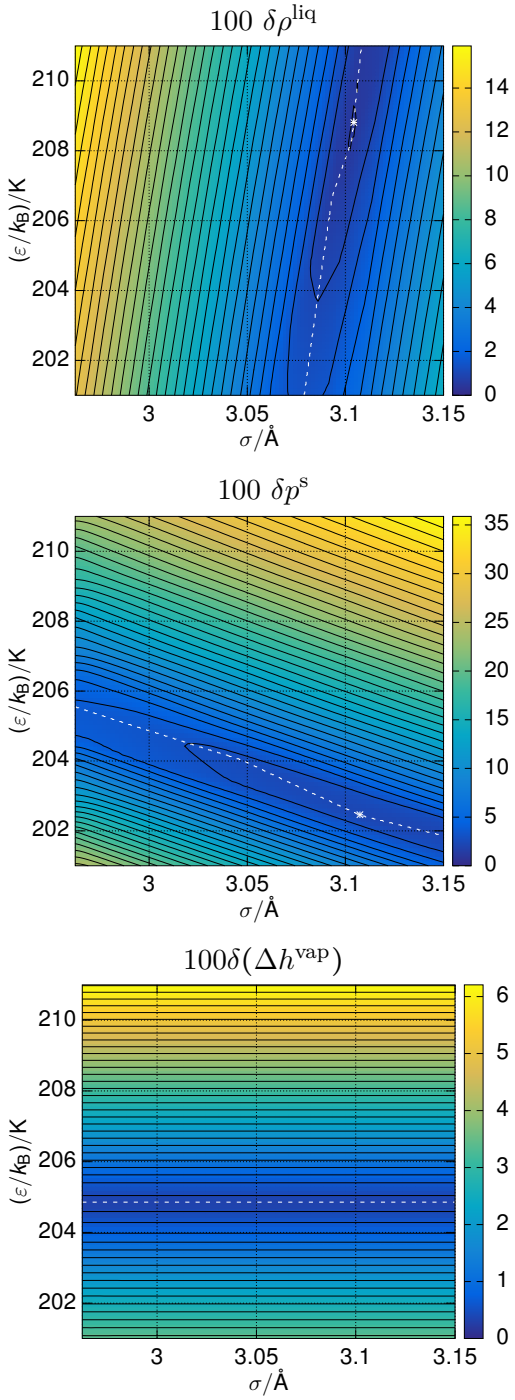



**Figure 8:** Dependence of the deviations of saturated liquid density (top), vapour pressure (middle) and enthalpy of vaporisation (bottom) on the LJ parameters $\sigma$ and $\varepsilon$ for scaled variants of the Huang model. The minima for the single property optimisation are marked by white stars and the paths of lowest ascent/descent through the minima, i.e. the 'valleys' in the 'landscape', of the respective objectives are indicated by white dashed lines. There is no distinct minimum for $\delta(\Delta h^{\mathrm{vap}})$, as the numbers for this objective do not depend on $\varepsilon$.

The contour plots also show the paths of the lowest ascent/descent, i.e. 'valleys' in the

'landscape' of the deviations (indicated by white dashed lines), as well as the global minima of the single-objective optimisations (indicated by white crosses). Excellent agreement with the experimental data can be obtained, with $\delta p^{\mathrm{s}} = 2.0\%$, $\delta \rho^{\mathrm{liq}} = 0.5\%$ and $\delta(\Delta h^{\mathrm{vap}}) = 0.3\%$ in the respective minima. As $\sigma$ has no influence on the scaling of the enthalpy (see Eq. (8)), $\delta(\Delta h^{\mathrm{vap}})$ varies only with $\varepsilon$. Hence, $\delta(\Delta h^{\mathrm{vap}})$ has no distinct minimum. For the optimum in $p^{\mathrm{s}}$, the deviations in the other properties are $\delta \rho^{\mathrm{liq}} = 2.2\%$ and $\delta(\Delta h^{\mathrm{vap}}) = 2.1\%$, and for the optimum in $\rho^{\mathrm{liq}}$, the corresponding numbers are $\delta p^{\mathrm{s}} = 25.2\%$ and $\delta(\Delta h^{\mathrm{vap}}) = 3.5\%$. Hence, the optima for a given single property are not necessarily close to the optima for the other properties. A compromise between the objectives is obviously needed.

### 3.4.4 MCO with two objectives

In the present section, the results from the MCO with two objectives ($\delta p^{\mathrm{s}}$ and $\delta \rho^{\mathrm{liq}}$) are presented. Again, the reduced units method is used and the Huang model is taken as a starting point. Figure 9 shows the Pareto set, along with its picture in the $\sigma - \varepsilon$ parameter space.

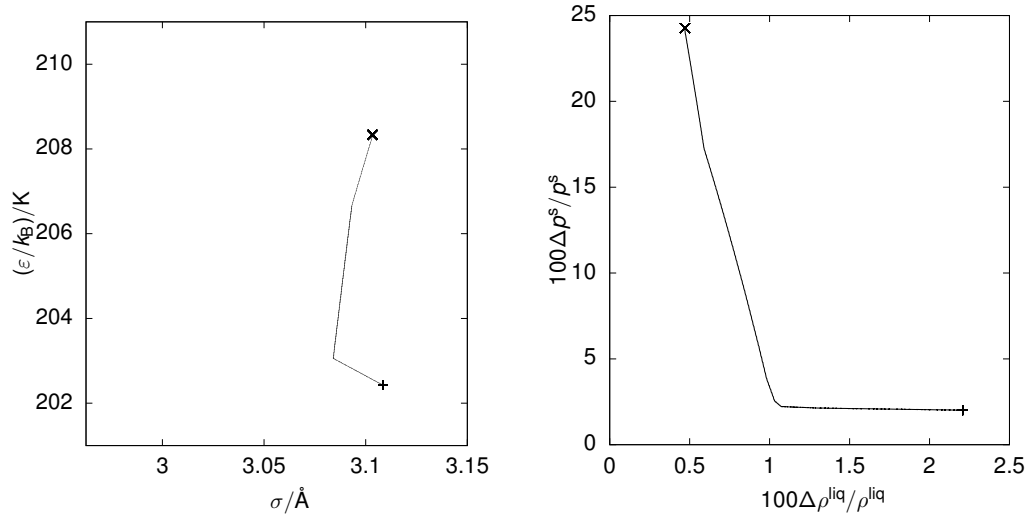



**Figure 9:** Results from the MCO with the two objectives $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$ using the reduced units method with Huang as starting point. (a) Picture of the Pareto set in parameter space. (b) Pareto set in objective space. The extreme compromises are marked by crosses.

The Pareto set and its picture in the parameter space are L-shaped lines. These lines end in the extreme compromises, i.e. in the optima of the single parameter optimisation. The two branches of the L-shaped line meet in the so called "Pareto knee". Obviously, solutions from the Pareto knee will often be particularly attractive as on each of the two

branches, the results for one objective get quickly worse, while the gains in the other objective are small.

It is remarkable that the valleys in the contour plots from the single-objective optimisation problems shown in Figure 8 coincide almost exactly with the branches of the Pareto sets of the MCO problem shown in Figure 9a. A similar phenomenon has been observed in the literature [41, 54].

The Pareto sets using all seven water models as starting points obtained in this manner are compared in Figure 10.

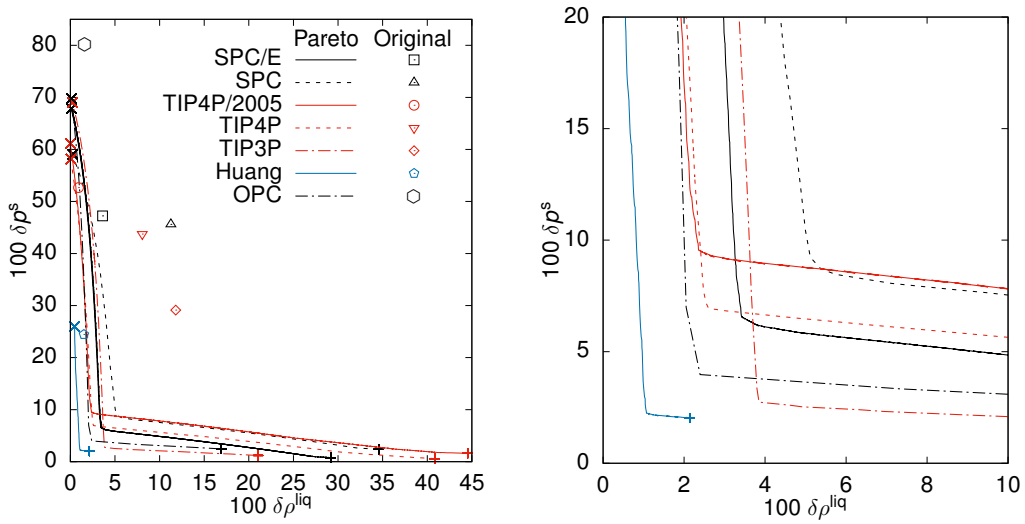



**Figure 10:** Results from the MCO with two objectives using the reduced units method with different water models from the literature as starting points. The Pareto sets are shown together with the results from the literature models, which are dominated by the Pareto set in all cases. Crosses mark extreme compromises. (a) Full picture of the Pareto sets. (b) Magnified view of the regions of the Pareto knees.

All seven Pareto sets have a similar L-type shape. Furthermore, Figure 10 also shows the results for the two objectives that were obtained for the original models. As expected, in all cases, the Pareto set that was obtained starting from a certain model dominates the result for that model. The results for the original TIP4P/2005 and Huang models lie close to the corresponding Pareto sets, while the results for some of the other literature models lie far away from their corresponding Pareto sets. All models lie in the upper region of the plot, which means that they can all be improved significantly regarding $\delta p^s$. The Pareto set that was obtained starting from the Huang model dominates all other Pareto sets, i.e. it lies closest to both axes in the plot shown in Figure 10. The Pareto sets obtained starting from SPC/E, TIP4P and TIP3P also have relatively low deviations in both objectives in their respective Pareto knees.

In Table 2, the original models from the literature are compared to the corresponding results obtained from the MCO using the reduced units method. For the comparison, the model from the Pareto knee is selected from the Pareto set. The model parameters as well as the results for the two objectives $\delta p^{\mathrm{s}}$ and $\delta\rho^{\mathrm{liq}}$ are compared. For completeness, also results for $\delta(\Delta h^{\mathrm{vap}})$ are shown. For most properties, significant improvements are obtained by the MCO, in some cases even by more than an order of magnitude. Furthermore, for most models, the results for all the three properties are improved by MCO. Only for TIP4P/2005 and OPC, $\delta\rho^{\mathrm{liq}}$ is slightly increased using the MCO result, while, however, the result for $\delta p^{\mathrm{s}}$ is considerably improved. The best results were obtained with the model obtained with MCO using the Huang model as starting point.

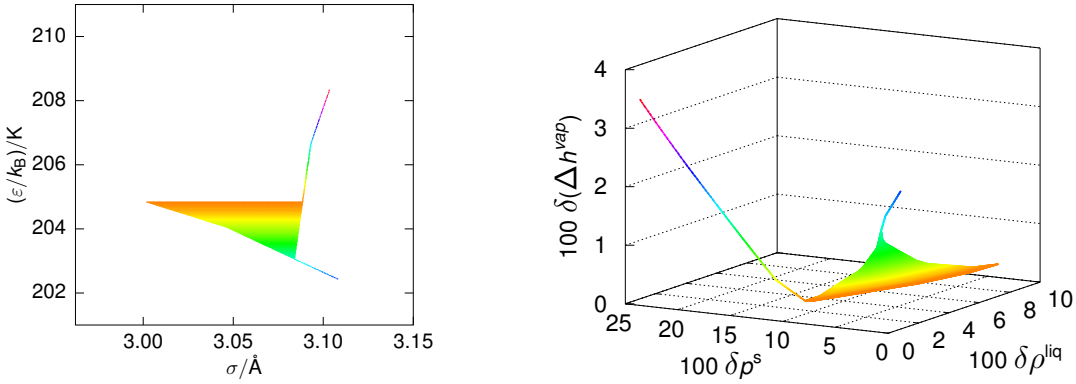
It is interesting to note that the parameters of the MCO models from the Pareto knee do not differ strongly from those of the literature models that were used as starting points of the reduced units method. This may imply that further substantial improvements could be achieved by a rigorous MCO, in which all model parameters would be varied independently. A more detailed comparison of the results from the original models and those from the MCO is given in Appendix A.

### 3.4.5 MCO with three objectives

In the following, the results from MCO of water models are discussed, in which three objectives were used: $\delta p^{\mathrm{s}}$, $\delta\rho^{\mathrm{liq}}$ and $\delta(\Delta h^{\mathrm{vap}})$. Hence, the parameter space is two-dimensional, as in the previous section, but the objective space is now three-dimensional. Only the results that were obtained with the reduced units method with the Huang model as starting point are discussed in detail, while the results obtained using the other water models as starting points are shown in Appendix A.

Figure 11 shows the Pareto set for this three-objective optimisation and its picture in the parameter space.

The Pareto set as well as its picture in the parameter space have a peculiar topology: they consist of a triangular surface with two tails starting from two of the three corners. The triangle has almost straight sides in the parameter space, while it is distorted in the objective space. As in the two-objective case, this topology can be understood using the results from the single-objective optimisation shown in Figure 8. Also here, the valleys in the contour plots and their traces in the parameter space play a decisive role. As there are three objectives now, there are also three valleys. Their traces intersect in three points. These are the three corner points of the triangular area in the picture of the Pareto set in the parameter space shown in Figure 11a. The tails are basically identical with the traces of the valleys in the contour plots. They are only observed when




**(a)** **(b)**

**Figure 11:** Results from the MCO with the three objectives $p^s$, $\rho^{liq}$ and $\Delta h^{vap}$, using the reduced units method with Huang as starting point. Colours indicate the mapping of a certain area from the parameter space to the corresponding area in the objective space. (a) Picture of the Pareto set in parameter space. (b) The Pareto set in objective space.

the parameters corresponding to the global optimum of the single objective optimisation (crosses in Figure 8) lie outside the triangular region. In that case, the tails end in the corresponding optimum of the single parameter optimisation, which correspond to the extreme compromises in the MCO.

Furthermore, the Pareto set from the MCO with two objectives is a subset of the Pareto set of the MCO with the additional third objective. A comparison of the pictures of the Pareto sets in the parameter space in Figures 9a and 11a shows that the triangular region in the result of the MCO with three objectives is basically an extension of the L-shaped result of the MCO with two objectives. By choosing models from the triangular region, improvements in $\delta(\Delta h^{vap})$ can be obtained, as compared to the result from the MCO with two objectives, in which $\delta(\Delta h^{vap})$ was not considered. But an inspection of the results for the objective space presented in Figure 11b reveals that these improvements are only small. This is not unexpected, as the predictions that were obtained for $\delta(\Delta h^{vap})$ with the model from the Pareto knee of the two parameter MCO were already excellent, see Table 2. Furthermore, the enthalpy of vaporisation, the vapour pressure and the saturated liquid density are coupled by the Clausius-Clapeyron equation, such that it is not unexpected that good results for $\delta p^s$ and $\delta\rho^{liq}$ entrain results for $\delta h^{vap}$ that are at least not bad.

### 3.4.6 Selected MCO water model

The MCO water model that was selected here for the comparisons shown in the Figures 4-6 is that from the Pareto knee obtained in the two-parameter optimisation using Huang as a starting point. It is indicated by an asterisk in Table 2. That model lies very close

to the bottom of the valleys of the contour plot from the single parameter optimisation of $\delta(\Delta h^{\mathrm{vap}})$ shown in Figure 8. Hence, no significant improvements can be achieved by including $\delta(\Delta h^{\mathrm{vap}})$ as third objective in the MCO, as described in the previous section. The selected MCO model performs clearly better than any of the previously available models (see Figures 4-6), in all the three objectives that are studied here.

## 3.5 Conclusions

This chapter reports on the optimisation of non-polarisable molecular models of water using RUM. Seven water models (SPC/E, SPC, TIP3P, TIP4P, TIP4P/2005, Huang and OPC) were used as starting values. A subset of the parameter space, starting from these seven models, was studied to obtain the Pareto sets of the MCO problem considering two VLE properties – $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$. Thus, seven Pareto sets corresponding to the seven models were obtained, and the seven optimal compromises from the individual Pareto knees were selected.

In all cases, not only the vapour pressure and saturated liquid density are described well by the models from the Pareto knee, but also the enthalpy of vaporisation. Nevertheless, also MCO with three objectives was carried out, which confirms that no significant further improvements in the enthalpy of vaporisation are possible. Also the interesting topologies of the Pareto sets and their pictures in the parameter space for the MCO problems with three objectives and two parameters can be understood from the results of the individual single-objective optimisations.

The selected MCO model that was obtained starting from Huang yields a good representation of the VLE data of water in the studied temperature range, with deviations in $p^{\mathrm{s}}$, $\rho^{\mathrm{liq}}$ and $\Delta h^{\mathrm{vap}}$ of 2.3%, 1.1% and 1.5% respectively.

Further improvements of the water models should be possible, if MCO could be applied to all six model parameters independently. The present work also points at ways how this could be achieved: from the solutions of single objective optimisation problems, valuable information can be gained on the Pareto set of the MCO problem that is obtained when the objectives that were previously considered individually are considered simultaneously.

The present work shows that the reduced units method can be combined in a fruitful way with MCO and that it makes MCO of complex molecular models feasible.

# 4 Hierarchical Multi-criteria Optimisation of Molecular Models of Water

## 4.1 Introduction

In the first MCO study (see Chapter 3), an overview of two vapour-liquid equilibrium (VLE) properties, viz. vapour pressure ($p^{\mathrm{s}}$) and saturated liquid density ($\rho^{\mathrm{liq}}$), of several water models from the LJ3PC class, was given. Most of these models, including the popular ones such as SPC/E [35] and TIP4P/2005 [37], were parametrised by a fit to experimental data for a few bulk liquid phase properties, such as density and enthalpy of vaporisation, at near ambient conditions. There are also reports on the development of new water models, in which the data base used for parametrisation is not or only partially disclosed. Generally, the data used for parametrisation are reproduced well by the model, while for other types of data or when the conditions differ strongly from those for which the model was optimised, large deviations may occur. This is in particular true for the vapour pressure curve of water [55]. The water model by Huang et al. [39, 56] (which will be abbreviated as "Huang" for convenience) is a notable exception, since it has been parametrised to $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$ of water over a broad temperature range. However, the Huang model gives poor results for properties such as the self-diffusion coefficient and the shear viscosity of liquid water. [53, 57]

These findings suggest that describing the thermodynamic properties of water with models from the LJ3PC class requires making compromises. A parameter set that is optimal for one property is usually not the optimal one for another property, i.e., there are conflicting objectives. This is a typical multi-criteria optimization (MCO) problem. Using MCO for the development of thermodynamic models is highly attractive, as it gives an overview of what can be achieved with a given model class regarding the representation of a given data set [29, 41, 42, 55, 58, 59]. Furthermore, MCO provides not only a single "best model" – where the question what is "best" depends on arbitrary choices (namely the weights assigned to the objectives), but provides a whole set of

models representing the best compromises.

There are different methods for determining Pareto sets [23–26], all of which require many evaluations of the objective functions. This is extremely costly when the objective function has to be evaluated by molecular simulations. Therefore, most of the previous works on MCO of molecular models [29, 41, 42, 58, 59] considered only model classes for which the dependence of the thermodynamic properties of interest from the model parameters had been explored before. Hence, in these cases, the MCO could be based on corresponding empirical correlations.

In the first MCO study of the present study (see chapter 3) [55], a different route is taken in which the RUM is used in the context of MCO to keep the computational efforts needed for the molecular simulations at a tractable level. While the results from the first MCO step [55] demonstrate the feasibility and usefulness of combining the RUM with MCO for molecular modelling, the study has some drawbacks: firstly, the objective space was only two-dimensional, i.e. only two thermodynamic properties were included in the study ($\rho^{\mathrm{liq}}$ and $p^{\mathrm{s}}$); the models optimized on that basis tend to give poor results for other properties, such as transport properties. It is, hence, interesting to extend the approach to include a larger variety of properties and to investigate the limits of the necessary compromises. Secondly, as the approach relied entirely on the RUM, the dimension of the parameter space was therefore also only two, while the studied model class has six parameters. Hence, not the Pareto set of the full problem was obtained, but only that of a heavily restrained problem. The second MCO study discussed in the present chapter addresses these issues by extending the study to a five-dimensional objective space as well as a five-dimensional parameter space.

Firstly, the study is extended to a five-dimensional objective space – including, besides $\rho^{\mathrm{liq}}$ and $p^{\mathrm{s}}$, also the following properties of the saturated liquid: shear viscosity $\eta$, self-diffusion coefficient $D$, and relative permittivity $\epsilon_{\mathrm{r}}$. All five properties are considered over a wide temperature range. Secondly, the study is extended to a five-dimensional parameter space in which, besides $\varepsilon$ and $\sigma$ of the LJ site, also the positions of the charges and the magnitude of the charges are varied. Only the angle between the two H–O bonds is kept constant and set to the physically reasonable value of 104.52°. To make the MCO study possible, a hierarchical approach was taken, and the studied region in the parameter space was restricted. The restrictions were chosen in such a way that particularly interesting regions are included.

## 4.2 Studied model class and literature models

In this chapter, the models in reduced units are referred to as *grid models*. Fig. 12 shows the parameters of seven grid models corresponding to popular water models from the literature that are members of the LJ3PC class. It is interesting that the grid models corresponding to some of the literature models are very similar: the grid models of TIP4P and TIP4P/2005 are very close to each other, and also those of SPC [34], SPC/E and TIP3P all lie in a narrow region of the parameter space. Hence, from the reduced units point of view, the difference between different models in the same group arises mainly from a different choice of the reference values $\sigma$ and $\varepsilon$. The Huang and OPC models are outliers, having quite unique reduced model parameters. For the Huang model, the reduced O–H bond length is distinctly larger than that of all the other models. For the OPC model, that length is shorter than for other models, which is accompanied by a larger magnitude of the reduced charge.

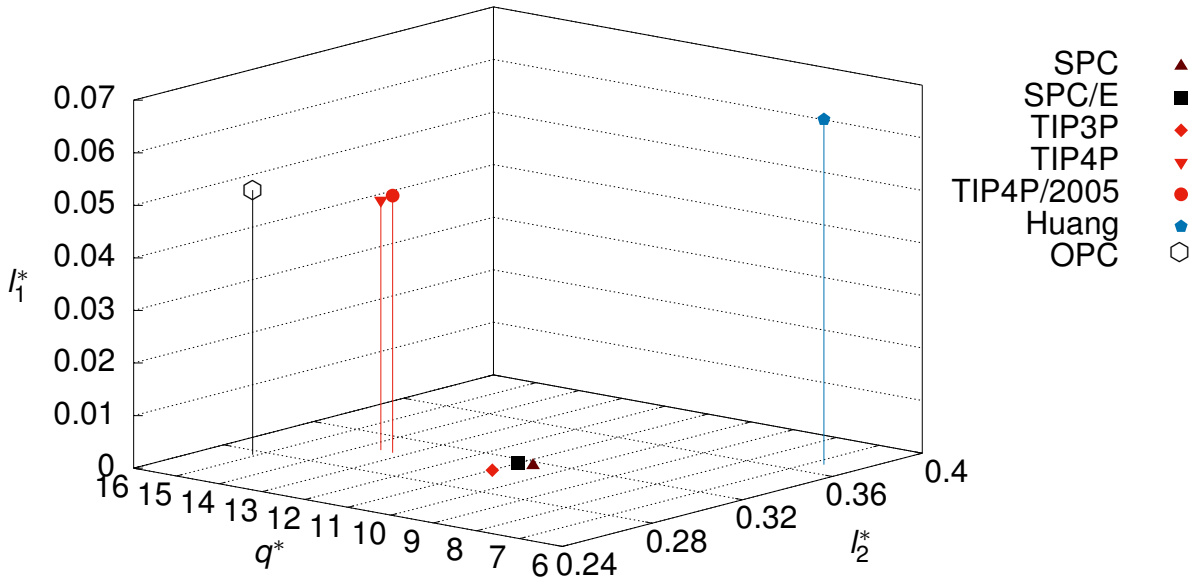



**Figure 12:** Location of selected literature models for water in the reduced parameter space. Different symbols denote the different models, and lines are added to show the projection of a point onto the $q^* - l_2^*-$ space for models with $l_1^* > 0$. Note the differences between angles as follows: $\alpha = 109.47°$ for SPC and SPC/E, $\alpha = 103.6°$ for OPC, while $\alpha = 104.52°$ for all other models.

Furthermore, Fig. 12 shows that the LJ3PC water models from the literature lie in a quite narrow part of the parameter space, but that even in this part, there are important unexplored gaps between different clusters of models. In the present work, the region in the vicinity of the two clusters (TIP4P-TIP4P/2005) and (SPC-SPC/E-TIP3P) is explored. The region around the outlier models (Huang and OPC) is disregarded, since preliminary studies showed that it is impossible to find a decent compromise between

the different objectives studied here near these models. Also, while the angle $\alpha$ = 109.47° for SPC and SPC/E is different from that for other models selected, i.e. $\alpha$ = 104.52° for TIP3P, TIP4P and TIP4P/2005, this is disregarded for the further analysis. All grid models that are studied in the following use the angle $\alpha$ = 104.52°, this value was chosen since it is the experimental value.

## 4.3 Molecular simulations

All simulations were carried out using the molecular simulation program *ms*2. [44, 60–62]

### 4.3.1 Simulation details for $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$

The vapour pressure $p^{\mathrm{s}}$ and the saturated liquid density $\rho^{\mathrm{liq}}$ were calculated with the Grand Equilibrium method [45]. In the Grand Equilibrium method, the VLE of a pure fluid is obtained in two subsequent steps: a Monte-Carlo (MC) $NpT$ run in which the chemical potential of the liquid $\mu$ and its pressure derivative at constant temperature, i.e. $(\partial\mu/\partial p)_T$, are sampled, followed by a pseudo-$\mu VT$ vapour run in which the chemical potential is kept "constant" as a function of pressure. First, MC simulations were performed in the $NpT$ ensemble for the liquid phase. Starting with 500 molecules in a face-centred cubic lattice, equilibration consisted of 20 000 MC cycles in the $NVT$ ensemble and 100 000 MC cycles in the $NpT$ ensemble. The cutoff radius was set to 12 Å throughout and a centre of mass cutoff scheme was employed. Analytical tail corrections were applied for LJ interactions, while the reaction field method was applied for the electrostatic interactions. For production, 9.6 million MC cycles were performed. Each MC cycle consisted of 500 translation, 500 rotation and 1 volume move. The chemical potential was calculated by thermodynamic integration. [14] A total of 100 steps for the scaling parameter $\lambda$ were used, and the minimum and maximum values of $\lambda$ were 0.2 and 1.0, respectively. The maximum step for $\lambda$ was 0.1. For the vapour phase, MC simulations were performed in the pseudo-$\mu VT$ ensemble. The simulation volume was adjusted to yield on average 864 molecules in the vapour phase. For equilibration, 15 000 cycles in the $NVT$ ensemble and 25 000 cycles in the pseudo-$\mu VT$ ensemble were performed, again starting from a face-centred cubic lattice. For production, 200 000 cycles in the pseudo-$\mu VT$ ensemble were performed.

### 4.3.2 Details of Simulations for $D$, $\eta$ and $\epsilon_{\mathrm{r}}$

The self-diffusion coefficient $D$, the shear viscosity $\eta$ and the relative permittivity $\epsilon_{\mathrm{r}}$ were calculated with molecular dynamics (MD) simulations in the $NVT$ ensemble. The

simulations were carried out for the liquid phase, using the same temperatures as the simulations for $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$, and prescribing the values of the liquid densities at VLE. Newton's equations of motion were solved using a fifth order Gear predictor-corrector integrator, and the time-step was 1 fs in all simulations. Again starting from a face-centred cubic lattice containing 4 000 molecules of water, the system was first equilibrated for 400 000 time steps and the following production run was 8 million time steps long. $D$ and $\eta$ were calculated with the Green-Kubo formalism [63, 64], and $\epsilon_{\mathrm{r}}$ was calculated by sampling Kirkwood's fluctuation formula. [65] The cutoff and long range corrections are the same as those in the simulations for $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$.

### 4.3.3 Simulation outputs and empirical correlations

The simulation output consisted of reduced properties, viz. $p^*$, $\rho^*$, $D^*$, $\eta^*$, and $\epsilon_{\mathrm{r}}$. For all studied grid models, all of the above properties were determined as a function of the reduced temperature $T^*$, i.e. functions $p^{\mathrm{s}*}(T^*)$, $\rho^{\mathrm{liq}*}(T^*)$, $D^*(T^*)$, $\eta^*(T^*)$, and $\epsilon_{\mathrm{r}}(T^*)$ were obtained by fitting individual temperature-dependent correlations to the simulation data. The reduced critical temperatures $T_{\mathrm{c}}^*$ were determined by fitting the Lotfi correlations [66] to the saturated liquid and vapour densities.

$$\rho^{\mathrm{liq}*} = \rho_{\mathrm{c}}^* + A_\rho (T_{\mathrm{c}}^* - T^*)^{\frac{1}{3}} + B_\rho (T_{\mathrm{c}}^* - T^*) + C_\rho (T_{\mathrm{c}}^* - T^*)^{\frac{3}{2}} \tag{12}$$

$$\rho^{\mathrm{vap}*} = \rho_{\mathrm{c}}^* - A_\rho (T_{\mathrm{c}}^* - T^*)^{\frac{1}{3}} + D_\rho (T_{\mathrm{c}}^* - T^*) + E_\rho (T_{\mathrm{c}}^* - T^*)^{\frac{3}{2}} \tag{13}$$

The extended Antoine correlation [67] (see Eq. (14)) was used for describing the reduced vapour pressure $p^{\mathrm{s}*}$.

$$\ln(p^{\mathrm{s}*}) = A_p + \frac{B_p}{T^* + C_p} \tag{14}$$

For the reduced self-diffusion coefficient $D^*$, reduced shear viscosity $\eta^*$ and the relative permittivity $\epsilon_{\mathrm{r}}$, the following correlations were used:

$$D^*(T^*) = A_D + B_D T^* + C_D (T^*)^2 + D_D (T^*)^3 \tag{15}$$

$$\eta^*(T^*) = A_\eta \exp \frac{B_\eta}{T^*} \tag{16}$$

$$\epsilon_{\mathrm{r}}(T^*) = A_\epsilon + B_\epsilon (T^*) + C_\epsilon (T^*)^2 + D_\epsilon (T^*)^3 \tag{17}$$

In Eq. (12)-(17), the parameters $A_i$ up to $E_i$ as well as $\rho_{\mathrm{c}}^*$ and $T_{\mathrm{c}}^*$ are adjustable parameters. Using the values of properties in reduced units obtained from these correlations at different temperatures, and converting them to physical units (see Sec. 4.4.3) allows for calculating the objectives used in the studied MCO problem.

## 4.4 Optimisation technique

### 4.4.1 Objective function

In the present work, five properties – $p^s$, $\rho^{liq}$, $D$, $\eta$, and $\epsilon_r$ – were considered. For a certain parametrisation of a water model of the LJ3PC class denoted by the vector of parameters $\underline{x}$, the relative deviation between the simulation result (sim) and the experimental data (exp) for a property $Z$ at temperature $T$ is defined as follows:

$$\delta Z(\underline{x}, T) = \frac{Z_{sim}(\underline{x}, T) - Z_{exp}(T)}{Z_{exp}(T)} \tag{18}$$

With that, an objective function for property $Z$ is defined that averages the relative deviations over a set of $N$ considered temperatures:

$$\overline{\delta Z}(\underline{x}) = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\delta Z(\underline{x}, T_i)^2} \tag{19}$$

To ensure a fair comparison between different model parametrisations $\underline{x}$, it is important to choose the temperatures for which the comparison with experimental data is carried out such that no bias is introduced. This was done here by using the results for a set of $N = 8$ fixed values for $T^*/T_c^* \in \{0.55, \ldots, 0.90\}$. The simulation values $Z_{sim}$ at these temperatures were calculated by using the individual correlations of the simulation data for each studied grid model and the information on $T_c^*$. Then, these dimensionless results were reconverted to the corresponding physical properties as described in Sec. 4.3.3. The comparison with the experimental data was done for the temperatures resulting from that conversion.

The experimental reference values $Z_{exp}$ for four of the properties, viz. $p^s$, $\rho^{liq}$, $\eta$ and $\epsilon_r$, were taken from REFPROP [33], in which the correlations recommended by the international association for the properties of water and steam (IAPWS) [32, 68] are implemented. For the self-diffusion coefficient $D$, no IAPWS recommendation is available. Hence, instead the data from Yoshida et al. [69] was and an empirical correlation to them has been adjusted, using the same mathematical function as for adjusting the simulation data (see Eq. (15)):

$$D/(10^{-10}\mathrm{m}^2\ \mathrm{s}^{-1}) = a_D + b_D(T/\mathrm{K}) + c_D(T/\mathrm{K})^2 + d_D(T/\mathrm{K})^3 \tag{20}$$

The resulting parameters are $a_D = 52.98$, $b_D = 0.2373$, $c_D = 9.265 \times 10^{-4}$, and $d_D = 3.385 \times 10^{-6}$. The correlations for properties taken from REFPROP were used in the temperature range 279 K – 647 K. The lower limit was selected at 279 K to avoid

the temperature range with anomalous behaviour of water. The correlations for $D$ were fitted to the available experimental data in the temperature range 303.16 K – 623.16 K. [69]

### 4.4.2 Grid of reduced parameters

As pointed out in Sec. 4.2 and depicted in Fig. 12, only small parts of parameter space of the LJ3PC water models have been explored to date. A grid is used to explore a region of this space which lies in the vicinity of successful models, so that it may be assumed that it contains further interesting models.

This grid is shown in Fig. 13. The varied parameters are the geometric parameters $l_1^*$ and $l_2^*$ (see Fig. 1) as well as the reduced charge $q^*$. The H–O–H bond angle $\alpha = 104.52°$ was chosen for all grid models.

The grid of grid models in parameter space was constructed using TIP3P as the starting point, and placing the adjacent grid points close enough such that the properties of TIP3P could be used as an initial guess for starting the simulations for obtaining $p^s$ and $\rho^{liq}$ of its neighbours. The size of the grid was $(4 \times 4 \times 6)$ in the $(q^* \times l_2^* \times l_1^*)$ directions, i.e. a total of 96 grid models were studied.

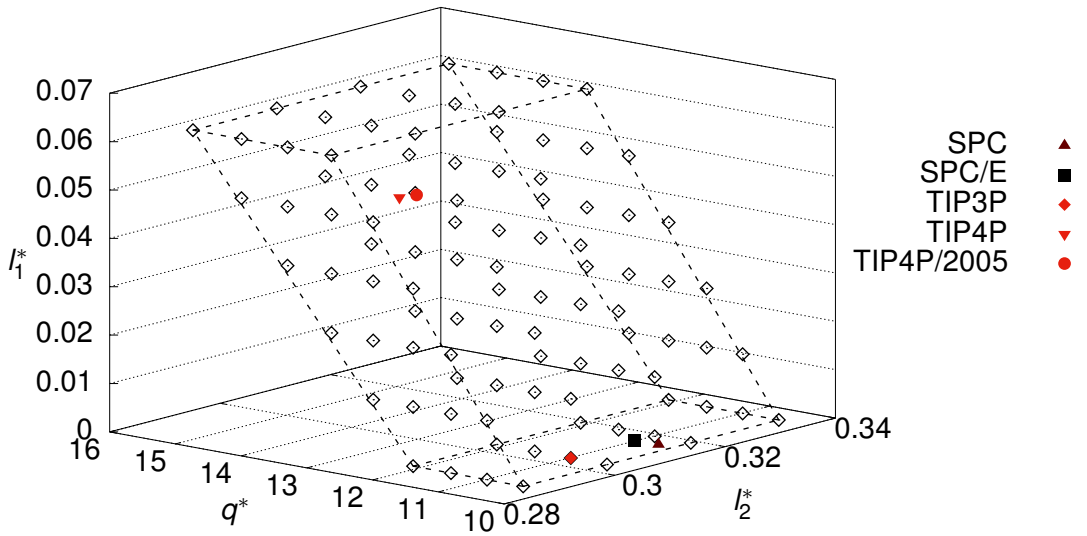



**Figure 13:** Grid of model parameter variations studied in this work. Diamonds are distinct parameter combinations for which simulations for $p^s$ and $\rho^{liq}$ were carried out, the dashed lines are guides to the eye. Literature models whose $q^*$, $l_1^*$ and $l_2^*$ parameter values lie inside the grid are also shown.

It would be highly desirable to develop universal correlations of the different thermophysical properties sampled in this work as functions of the three model parameters $l_1^*$, $l_2^*$, and $q^*$. However, $p^s$, $\rho^{liq}$ and critical properties of the different models on the

grid vary substantially and in a way that is difficult to correlate, so that after some preliminary attempts, such correlations have not been established.

### 4.4.3 Reduced units method (RUM)

For a general description of the RUM, the reader is referred to the original paper [12]. In the RUM, the reduced (dimensionless) simulation data are related to data in physical units by the choice of $\sigma$ and $\varepsilon$, see Eqs. (5)–(10). The values of $\sigma$ and $\varepsilon$ were varied continuously for each of the 96 studied grid models within the bounds that were chosen so that the highest and lowest studied temperatures were 640 K and 280 K respectively. Thus, in the present work, the dependence of the LJ3PC model properties on the model parameters was studied analytically for $\sigma$ and $\varepsilon$, whereas the influence of remaining three parameters was only studied in a discrete manner by varying $l_1^*$, $l_2^*$ and $q^*$ of the grid models.

### 4.4.4 Hierarchical optimisation scheme

Based on the simulation results determined as described above, a hierarchical MCO of LJ3PC water models was performed, in which two subsequent optimisation steps were carried out. In the first step, an MCO with two objective functions was carried out considering each of the 96 grid models independently. The two objectives were the representation of $p^\mathrm{s}$ and $\rho^\mathrm{liq}$. Only those grid models that yielded acceptable results for both $p^\mathrm{s}$ and $\rho^\mathrm{liq}$ (for certain choices of $\sigma$ and $\varepsilon$) were considered in the second step. The thresholds were chosen to be $\delta p^\mathrm{s} < 0.07$ and $\delta\rho^\mathrm{liq} < 0.03$. In addition, three of the literature models were considered in the second step, namely SPC, SPC/E and TIP4P/2005, which do not meet the criteria in representing the $p^\mathrm{s}$ and $\rho^\mathrm{liq}$, but were nevertheless kept as a reference as they are widely used. The reason for putting the emphasis on $p^\mathrm{s}$ and $\rho^\mathrm{liq}$ is that these properties are not only highly relevant but also that it is often found that molecular models for real fluids that represent these two properties well will also give reasonable predictions of other thermodynamic properties of fluid states.[51]

In the second step, simulations were carried out for obtaining the additional properties, i.e. $D$, $\eta$ and $\epsilon_\mathrm{r}$, for the models that passed the first step. The results from all selected models were then considered together, and a common Pareto set was determined, considering five objectives. Thus, the resulting Pareto set is five dimensional. The parameter space of the MCO problem is special: it contains two continuous variables ($\varepsilon$ and $\sigma$) and three distinct discrete variables ($l_1$, $l_2$ and $q$). Furthermore, following the arguments given above, only models with $\delta p^\mathrm{s} < 0.1$ and $\delta\rho^\mathrm{liq} < 0.05$ were considered. (Remark: The

fact that these threshold values slightly differ from those in step 1 is simply due to the history of the project and not motivated by some higher rationale; it is not expected that better results would have been obtained had the larger threshold been used also in step 1.) From the Pareto set of this MCO problem, eight exemplary models were selected to illustrate different compromises between the objectives that can be made. Fig. 14 summarizes the workflow of the hierarchical optimisation procedure used in the present work.

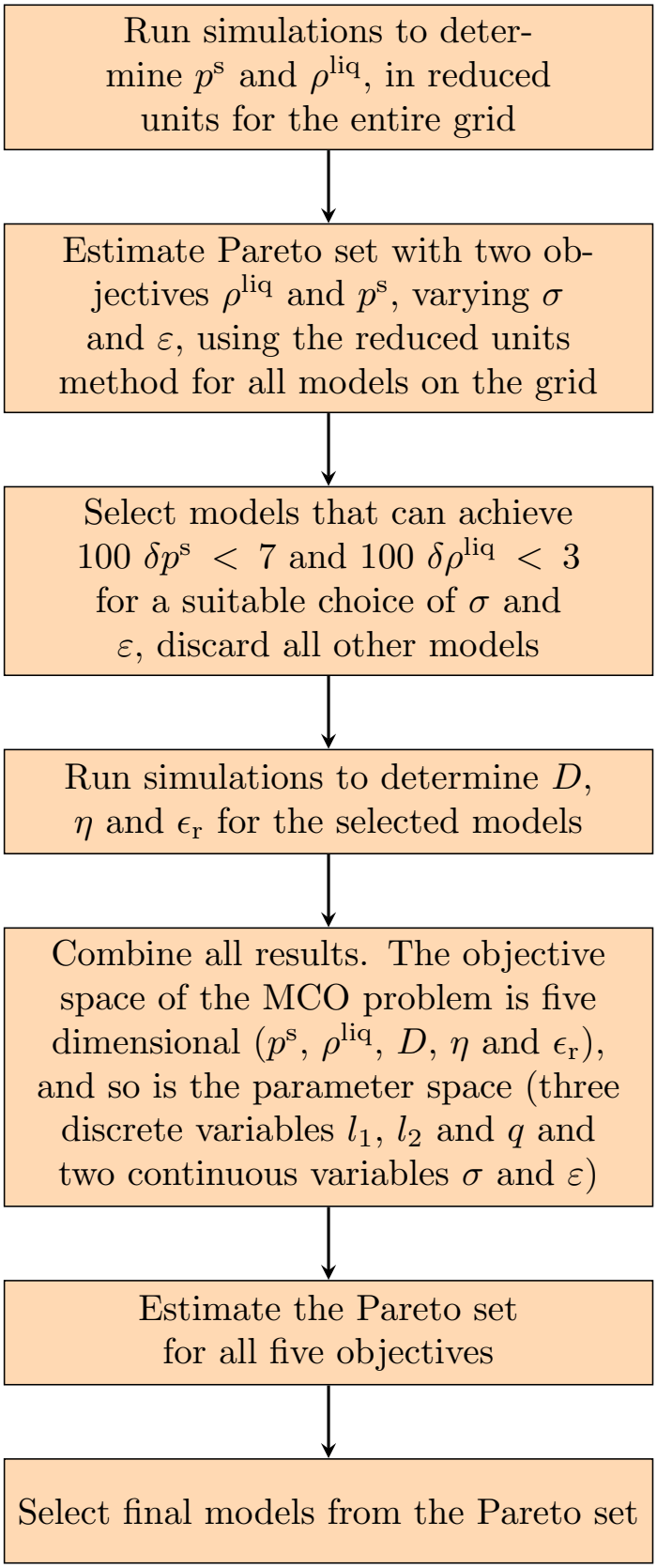



**Figure 14:** Scheme illustrating the hierarchical optimisation workflow used in the present work.

The Pareto set was determined by brute force enumeration. To this end, the continuous variables $\varepsilon$ and $\sigma$ were discretised using increments of 0.1 K for $\varepsilon/k_\mathrm{B}$ and 0.003 Å for $\sigma$. Each model on the grid that is dominated by at least one other model was sorted out. The remaining set, i.e. the set of non-dominated models, is the Pareto set. The calculation of the Pareto set, together with the application of the reduced units method, was implemented in MATLAB.

## 4.5 Results and discussion

### 4.5.1 Simulation results

Simulation results for $p^\mathrm{s}$ and $\rho^\mathrm{liq}$ for all 96 grid models in reduced units are provided as a spreadsheet document in Appendix B. Additionally, the simulation results for $D$, $\eta$ and $\epsilon_\mathrm{r}$ for the shortlisted models are also tabulated there. The spreadsheet document also contains the parameters of the empirical correlations to the simulation results, see Eq. (12)-(17)

### 4.5.2 First hierarchical MCO step: $p^\mathrm{s}$ and $\rho^\mathrm{liq}$

Fig. 15 presents the Pareto sets from the first MCO step, in which the objectives were $p^\mathrm{s}$ and $\rho^\mathrm{liq}$ and the parameters were $\varepsilon$ and $\sigma$, and in which an MCO was carried out for each of the 96 individual grid models on the $l_1^* \times l_2^* \times q^*$ grid (see Sec. 4.4.2). Additionally, the envelope of the curves is shown, which represents the Pareto set that is obtained when combining the results for all 96 grid models.

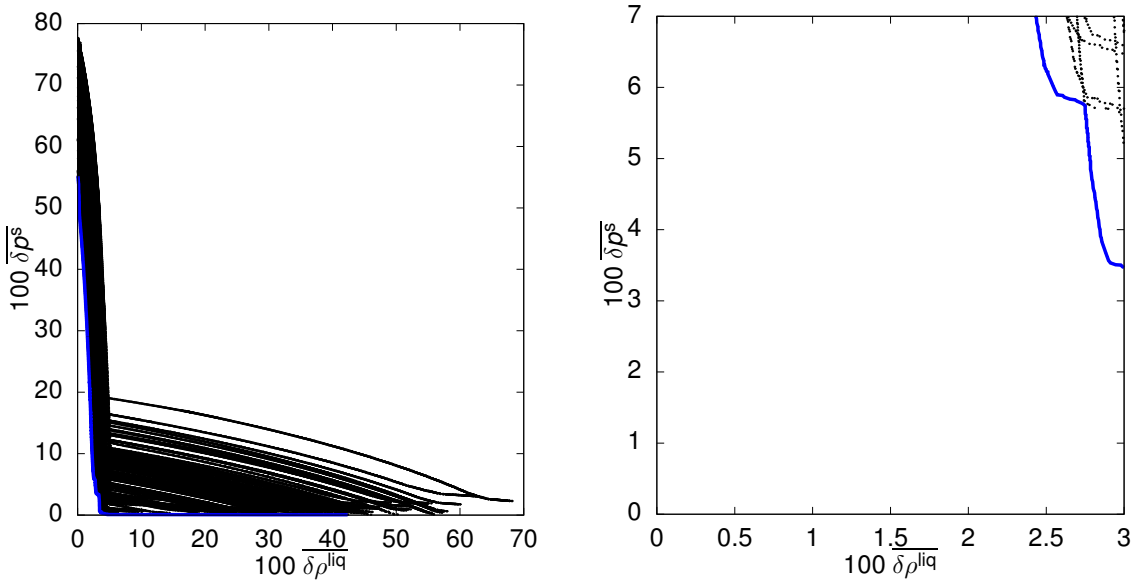



**Figure 15:** (a) Pareto sets with objectives $p^s$ and $\rho^{liq}$, and parameters $\sigma$ and $\varepsilon$, for all models on the grid shown in Fig. 13. (b) Zoom into the region of panel (a) for which $100\ \overline{\delta p^s} < 7$ and $100\ \overline{\delta \rho^{liq}} < 3$. The fact that points are visible in panel (b) rather than continuous lines arises from the discretisation of the $\sigma - \varepsilon$ grid. The blue line represents the envelope, i.e. the Pareto set of the combined results.

Most of the models on the grid cannot represent $p^s$ and $\rho^{liq}$ well, irrespective of the values of $\sigma$ and $\varepsilon$ that are chosen. Only for eight grid models, a part of their Pareto sets lies within the targets specified in Sec. 4.4.4, i.e. $100\ \overline{\delta p^s} < 7$ and $100\ \overline{\delta \rho^{liq}} < 3$, and only three of those models have a Pareto knee region that is within these thresholds. These eight selected grid models (SGM) were considered for further analysis in the second hierarchical MCO step.

In addition, three popular literature models – SPC, SPC/E and TIP4P/2005 – were included in the second hierarchical step. Thus, a total of eleven grid models were considered for the second hierarchical MCO step. Fig. 16 shows the reduced parameters of the grid models selected for the second hierarchical MCO step.

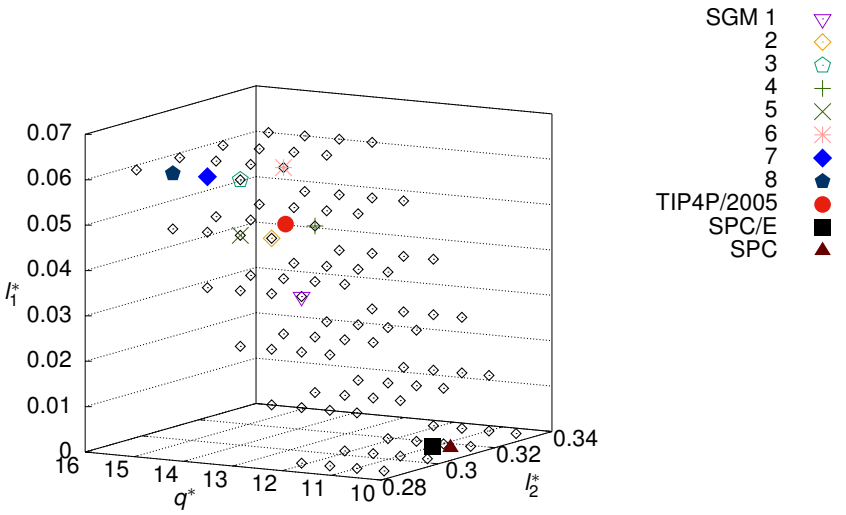



**Figure 16:** Grid of model parameter variations superimposed with eight models selected for the second MCO step. The selected grid models (SGM) are marked with coloured symbols as indicated in the legend. In addition, three popular literature models are considered further for the second MCO step even though they cannot achieve the prescribed limits with respect to the accuracy of $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$ (see text).

On the grid, all selected models are quite close to TIP4P/2005 (and, thereby, also to the original TIP4P model, cf. Fig. 12). This could either be a consequence of the slightly different angles between the grid models and the SPC and SPC/E models, but it could also indicate that for modelling $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$ the choice $l_1^* = 0$ (as is the case for SPC and SPC/E) is not a good one.

### 4.5.3 Second hierarchical MCO step: $p^{\mathrm{s}}$, $\rho^{\mathrm{liq}}$, $D$, $\eta$ and $\epsilon_{\mathrm{r}}$

Fig. 17 shows results from the second hierarchical MCO step, in which all five objectives were considered. There are six different panels in Fig. 17. In the upper left panel, the picture of the Pareto set (PSPS) in the $\sigma$, $\varepsilon$ - parameter space is shown for all eleven grid models (the eight SGMs and the three from the literature). Fig. 17 does not show the full PSPS obtained from the individual grid models. All results that were dominated by results from other grid models were eliminated. Hence, the results shown in Fig. 17 can be understood as the PSPS of a global Pareto set. While in the upper left panel of Fig. 17 an overview is given, and the colour indicates the grid model from which the PSPS was obtained, the remaining five panels give information on the quality of the description of the five studied properties $p^{\mathrm{s}}$, $\rho^{\mathrm{liq}}$, $D$, $\eta$, $\epsilon_{\mathrm{r}}$, again in a plot of the $\sigma$, $\varepsilon$ – space, in which the colour indicates the average deviation. In the plot, seven patches can be distinguished. They are labelled with roman numerals I-VII. Six of which are related to Pareto fronts from individual grid models (SGMs or grid models

corresponding to literature models), while the final patch (the one numbered IV) results from a superposition of the Pareto fronts of five different SGMs.

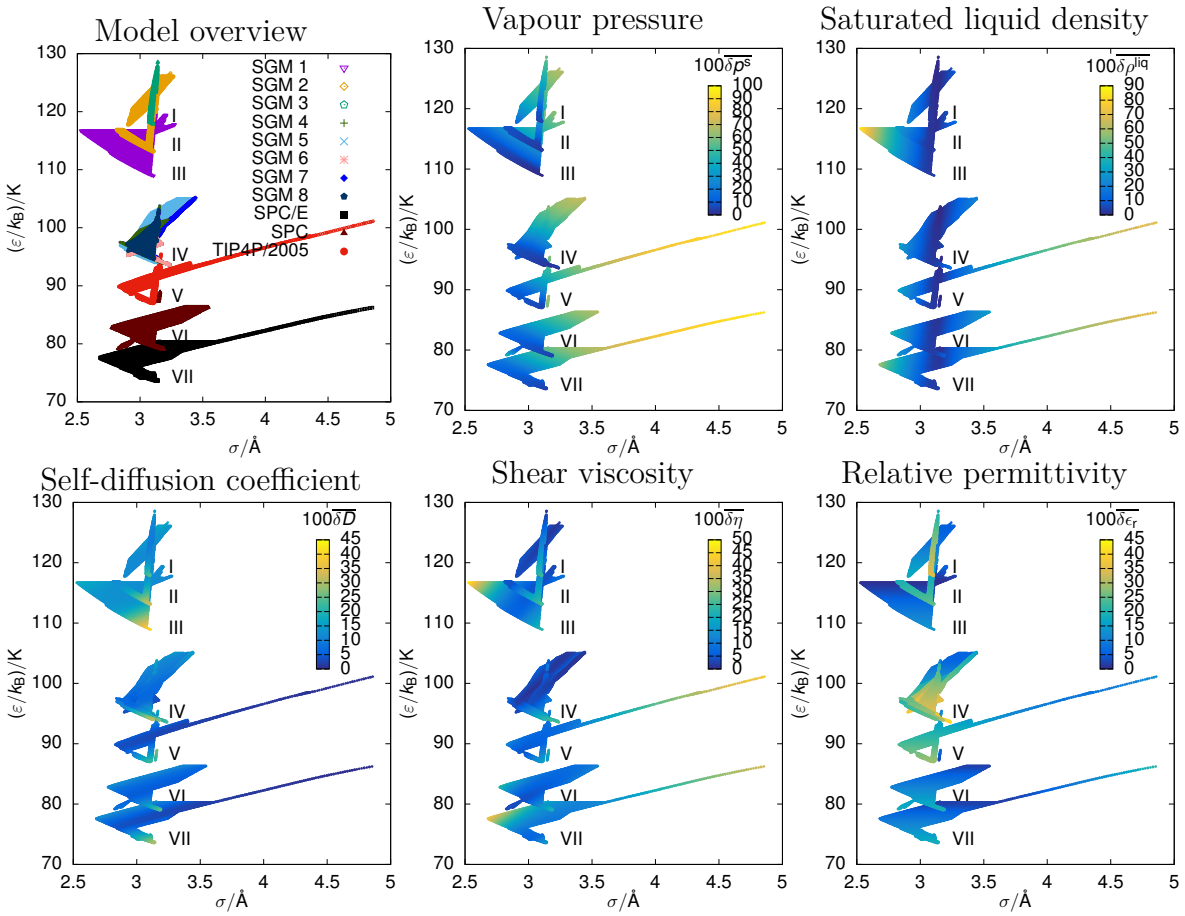



**Figure 17:** Results from the second MCO step. The objective space is five dimensional, and so is the parameter space. The objectives are $p^s$, $\rho^{liq}$, $D$, $\eta$ and $\epsilon$. The pictures show the projection of the picture of the five-dimensional Pareto set onto the $\sigma - \varepsilon$ plane of the parameter space. Roman numerals I-VII are for identifying the individual patches in the Pareto set, which stem from different models on the grid, compare also Fig. 16.

The individual patches in the PSPS can be described as surfaces with tails that are extensions of the edges of the surfaces. These geometrical features can be interpreted in a similar manner as in our previous work. [41, 55] The boundaries of the patches arises from singular regions where there is low deviation for one of the objectives. The fact that the results for $\sigma$ and $\varepsilon$ of Pareto optimal models obtained from 5 different grid models (i.e. models with different parameters for $l_1^*$, $l_2^*$ and $q^*$) fall basically together in the $\sigma$, $\varepsilon$ plane (see patch IV in Fig. 17) is astonishing. This is only stated as a finding, and is not interpreted physically here, which would involve detailed considerations on how the different contributions to the interactions are described by the different models, together with technical issues such as the fact that the numbers of $\sigma$ and $\varepsilon$ are used in the normalization of $l_1^*$, $l_2^*$, and $q^*$, see Eqs. (3) and (4).

### 4.5.4 Selection of models from Pareto set

The information on the Pareto set and PSPS as it is depicted in Fig. 17 allows for making an educated choice of a model for a certain application. Already in the first step of our hierarchical procedure, the models that yield gross deviations for the basic properties $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$ were discarded. It is continued here along this line and further impose the following constraints: $100\ \overline{\delta p^{\mathrm{s}}} \leq 10$ and $100\ \overline{\delta \rho^{\mathrm{liq}}} \leq 5$. The result of this filtering is shown in Fig. 18 which has the same layout as Fig. 17.

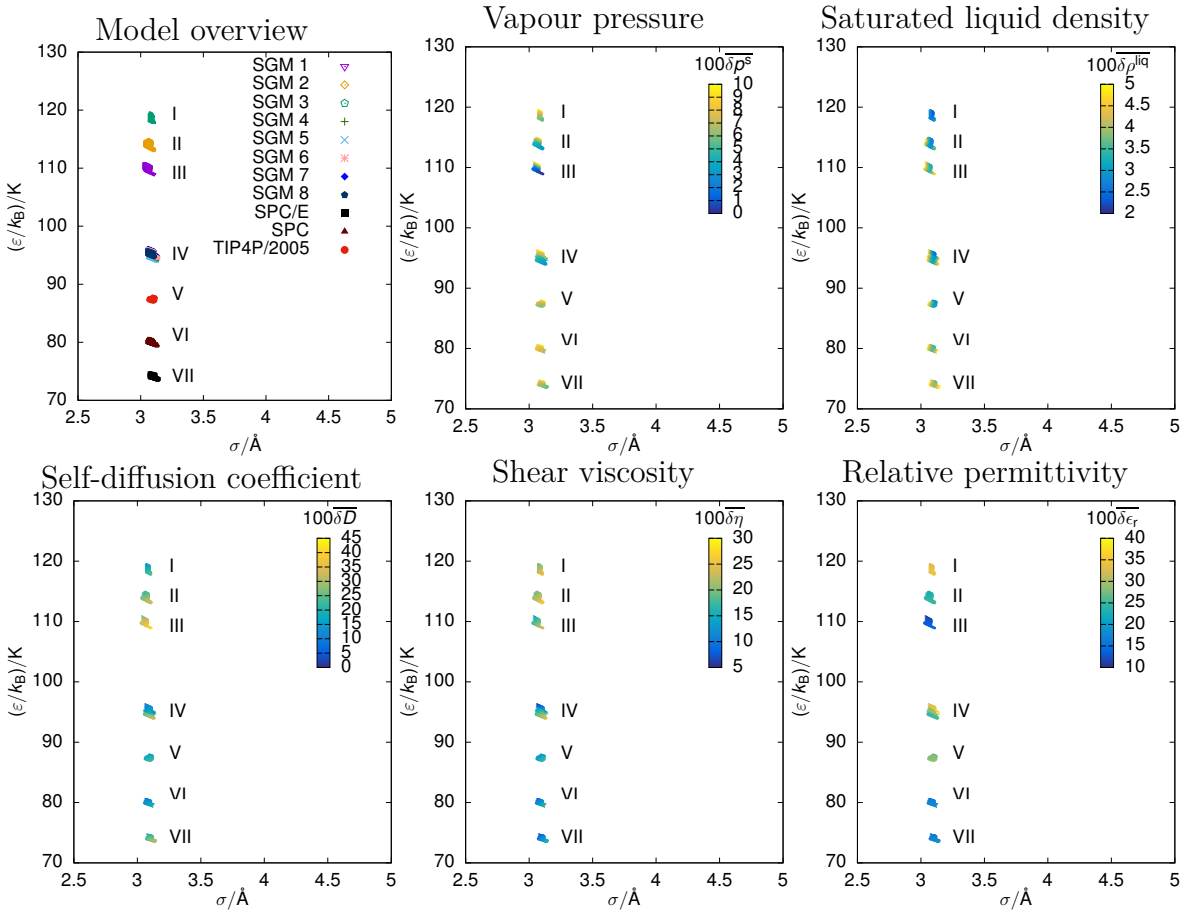



**Figure 18:** Projection of the picture of the five-dimensional Pareto set onto $\sigma - \epsilon$ plane of the parameter space. Colours indicate the deviations in the respective properties. Only the parts of the projections for which $100\ \overline{\delta p^{\mathrm{s}}} < 10$ and $100\ \overline{\delta \rho^{\mathrm{liq}}} < 5$ are shown. Roman numerals I – VII are for identifying the individual patches.

Only a small region from each patch in Fig. 17 remains after this filtering. Again, patch IV contains the Pareto optimal models corresponding to five different SGMs. The first striking feature is that the values of $\sigma$ are virtually the same for all Pareto-optimal models. The number that is found for $\sigma$ is always close to $\sigma = 3.1$ Å, which is a consequence of the fact that the density, which is expected to be described accurately depends predominantly and very strongly on $\sigma$. While the values for $\varepsilon/k_{\mathrm{B}}$ vary between

75 and 120 K, there are some models for which the values of $\varepsilon$ are similar and hence they collapse onto a single patch. This indicates different descriptions of the attractive interactions in the different models. There are two types of attractive interactions for models of the LJ3PC class: dispersion and electrostatics. With the electrostatic part, not only the polarity is described, but also H-bonding. As shown in [4] depending on the position of the charges on the LJ-spheres, results can be obtained that have many features of H-bonding. The models in the different patches differ in the weights that are assigned to the three contributions to the attractive interactions, in which high values for $\varepsilon$ indicate a high weight of the dispersion.

Figure 18 also shows that there is no single patch (corresponding to a single grid model) that would be clearly better in all properties than the other patches. One can distinguish two groups: patches I – III, with high numbers for $\varepsilon/k_\mathrm{B}$, that yield good results for $p^\mathrm{s}$ and $\rho^\mathrm{liq}$ but poorer results for $D$ and $\eta$. Within this group, Patch I works particularly well for the density, patch III for the vapor pressure, while patch II represents a compromise. The second group are the patches IV – VIII, which generally yield better results for $D$, $\eta$, but poorer results for $p^\mathrm{s}$ and $\rho^\mathrm{liq}$. Within this group, patch VI is a good compromise regarding the representation of the properties $D$, $\eta$ and $\epsilon_\mathrm{r}$. Also several other patches yield good compromises between some of the studied properties, but no patch shows a good performance for all of them. Self-diffusion and shear viscosity show similar patterns, i.e. a model that is good for one of them will also be good for the other, which is as expected.

The discussion above shows that the final selection of a model form the Pareto set depends on what is required. To illustrate different possibilities, a few models are selected and their performance is studied in more detail: two models were selected form the patches resulting from the second optimization step (Model A from patch III, and Model B from patch VI). Additionally, one model from each of the Pareto sets found based on the literature models SPC, SPC/E and TIP4P/2005, was also selected. These will be called SPC O, SPC/E O and TIP4P/2005 O in the following. All selected models will be compared to the three original literature models SPC, SPC/E and TIP4P/2005.

The parameters of all selected models as well as the literature models are given in Table 3 together with information on the average deviations in the description of the experimental data of the five considered properties. More detailed results of the comparison with the experimental data are presented in Fig. 19 and 20.

| Model | $100\ \overline{\delta p^{\mathrm{s}}}$ | $100\ \overline{\delta \rho^{\mathrm{liq}}}$ | $100\ \overline{\delta D}$ | $100\ \overline{\delta \eta}$ | $100\ \overline{\delta \epsilon_{\mathrm{r}}}$ | $\sigma$ / Å | $(\varepsilon/k_{\mathrm{B}})$ / K | $l_1$ / Å | $l_2$ / Å | $q^+$ / e |
|---|---|---|---|---|---|---|---|---|---|---|
| Model A | 0.44 | 3.67 | 38.50 | 21.86 | 14.07 | 3.0873 | 109.1541 | 0.1133 | 0.8911 | 0.5435 |
| Model B | 8.18 | 2.90 | 16.22 | 17.47 | 28.40 | 3.1096 | 95.2350 | 0.1522 | 0.9447 | 0.5354 |
| SPC O | 9.71 | 4.96 | 9.15 | 7.99 | 15.84 | 3.0552 | 80.5228 | 0.0000 | 0.9650 | 0.4088 |
| SPC/E O | 6.51 | 3.89 | 26.84 | 12.89 | 17.60 | 3.1185 | 73.8777 | 0.0000 | 0.9850 | 0.4089 |
| TIP4P/2005 O | 6.27 | 3.18 | 22.08 | 16.63 | 28.83 | 3.1020 | 86.9552 | 0.1518 | 0.9400 | 0.5326 |
| SPC | 17.1 | 8.5 | 29.41 | 23.47 | 20.82 | 3.1660 | 78.1775 | 0.0000 | 1.0000 | 0.4100 |
| SPC/E | 45.96 | 3.74 | 4.17 | 11.59 | 6.90 | 3.1660 | 78.1775 | 0.0000 | 1.0000 | 0.4238 |
| TIP4P/2005 | 51.15 | 1.14 | 8.35 | 8.22 | 16.26 | 3.1589 | 93.1996 | 0.1546 | 0.9572 | 0.5564 |

**Table 3:** Deviations of literature models and selected Pareto optimal models. The H–O–H bond angle for SPC, SPC/E, SPE O and SPC/E O is 109.47°, while that for all the other models is 104.52°.

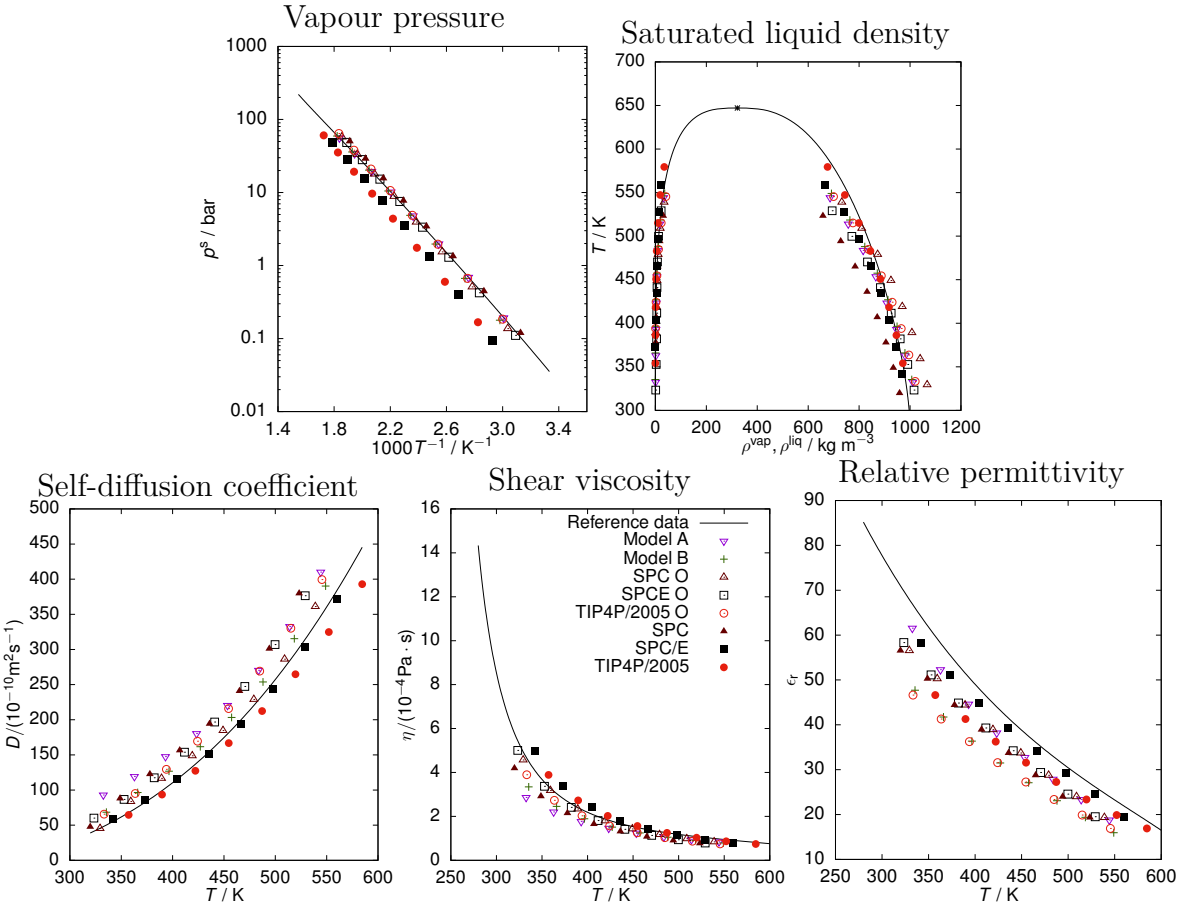



**Figure 19:** Physical properties of selected models compared to correlations of experimental data, also see Table 3. Lines show reference data obtained from a correlation fitted to the data from Yoshida et al. [69] for self-diffusion coefficients, and REFPROP [33] data for all other properties. Coloured symbols show different models as indicated in the legend.

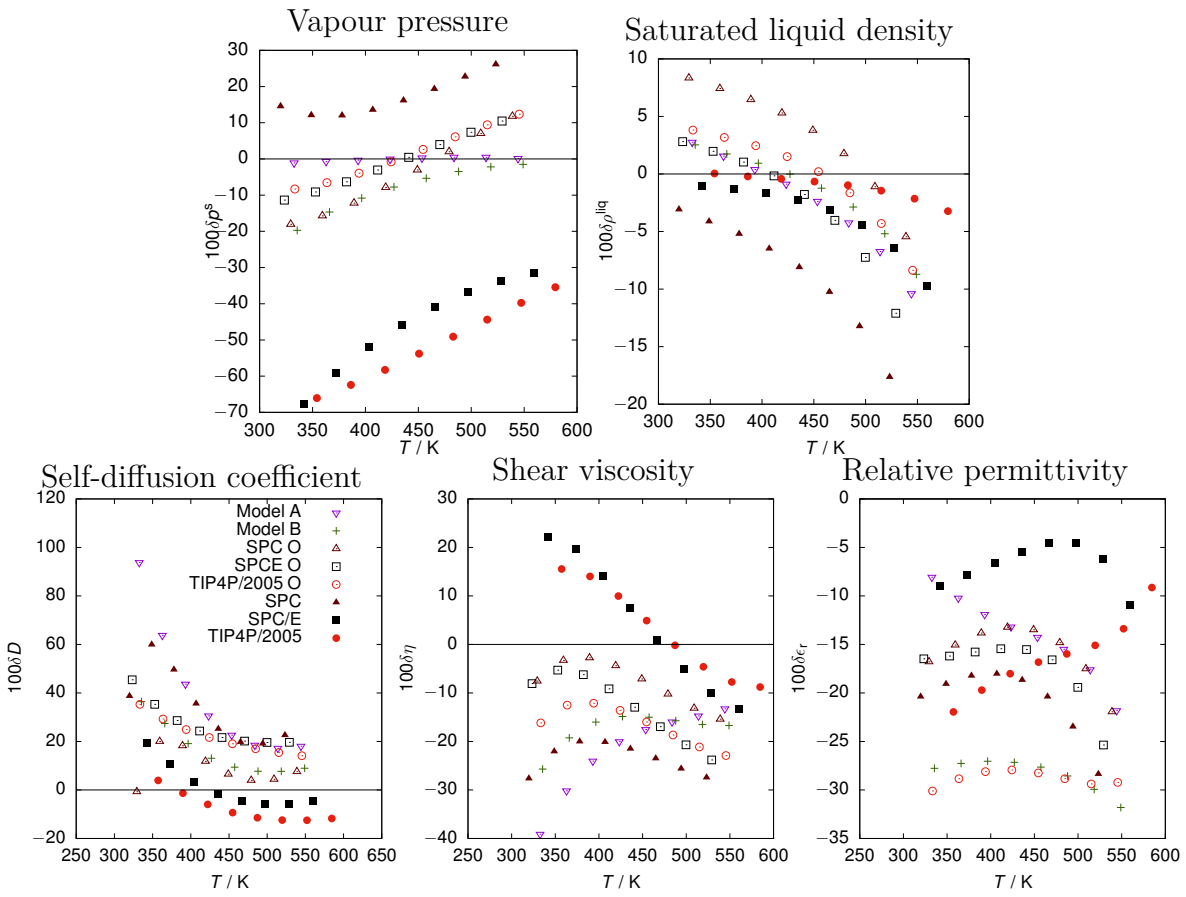



**Figure 20:** Relative deviations of physical properties of selected models with respect to the experimental data, also see Table 3.

Let us start the discussion by taking a look at the results from the three literature models that were considered in this work in Table 3 (compare also Figs. 19 and 20). Overall, SPC/E and TIP4P 2005 show a similar performance, that can be characterized as follows: good results are achieved for 4 of the 5 considered properties, however, at the expense of an extremely poor description of $p^{\mathrm{s}}$. Model A from the present work could be characterised as complementary to this as it gives an excellent description of $p^{\mathrm{s}}$. Also the description of $\rho^{\mathrm{liq}}$ and $\epsilon_{\mathrm{r}}$ is good, but the results for $D$ and $\eta$ are poor. Among the literature models, SPC shows a more balanced performance: the representation of $p^{\mathrm{s}}$ is distinctly better than for SPC/E and TIP4P 2005, but the results for the other four properties are poorer. However, SPC is the only model that is fully dominated by other models in Table 3: both SPC O (that was derived from SPC) as well as SPC/E O show a better performance in all indicators listed in Table 3. Basically, all models derived from the literature models as well as the Model B can be considered as variants resulting from a balanced approach that aims at limiting the maximal deviation in all considered properties. They are similar, with some variations in the strengths and weaknesses. It is interesting to note that over the entire temperature range, $D$ is overestimated by most of the models, while $\eta$ is typically underestimated. Finally, $\epsilon_{\mathrm{r}}$ is underestimated by all

models.

The results shown indicate that the main conflict in describing the considered properties of water with LJ3PC models is between getting a good description of $p^s$ on the one side and ($D$ and $\eta$) on the other. The best compromise that was found in this regard was SPC O, which yields average deviations in $p^s$, $D$, and $\eta$ that are all between 8 and 9%. In contrast, it is possible to find models that represent both $\rho^{liq}$ and $p^s$ well (e.g. Model A) and it is also possible to find models that represent both $\rho^{liq}$ and ($D$ and $\eta$) well (e.g. SPC/E). The results for $\epsilon_r$ are not clearly correlated with those for any of the other considered properties. SPC/E gives the best result for $\epsilon_r$ and good results for all other properties except for $p^s$, while Model A gives good results for $p^s$ and the second-best result for $\epsilon_r$.

Table 3 also gives information on the model parameters. There are two obvious messages: firstly, there is not much choice regarding $\sigma$, it has to be about 3.1 Å, which can be explained by the very strong influence of this parameter on the density. Secondly, models that give a good representation of the vapour pressure usually seem to have a high value for $\varepsilon$. Both findings can also be seen in Fig. 18. The explanation for the second finding is difficult, as in our approach that is based on the RUM, the choice of $\varepsilon$ influences not only the dispersion but also the electrostatic contribution with which the polarity and the H-bonding are modelled. Even though the data set provided is very large, it is difficult to discern other correlations of the choice of individual parameters and features of the representation of the different properties from this data, as the parameters are highly correlated, namely in regard to the description of the attractive interactions which is influenced not only by $\varepsilon$ and $q$ but also by all geometrical parameters.

## 4.6 Conclusions

A comprehensive study on the performance of the LJ3PC model class (LJ3PC: Lennard-Jones plus 3 partial charges) for describing thermophysical properties of water is carried out. Models form that class include SPC, SPC/E and TIP4P/2005 and are widely used for molecular simulations of water and aqueous systems. We have used an MCO approach to obtain optimal parametrisation of water models, which accurately describes five thermophysical properties of water over a broad temperature range, i.e. vapour pressure ($p^s$), saturated liquid density ($\rho^{liq}$), self-diffusion coeffcient ($D$), shear viscosity ($\eta$) and relative permittivity ($\epsilon_r$) over temperature range between 55% and 90% of the critical temperature.

Some models from the Pareto set were selected and studied further and compared to prominent models from the literature, namely SPC, SPC/E and TIP4P/2005. It turns

out that models can be found that perform clearly better than SPC in all regards the objective space studied here. SPC/E and TIP4P/2005 are good in describing most of the properties studied here but yield extremely large deviations in the vapor pressure. More balanced compromises can be selected easily from the provided Pareto set.

The most important outcome of the present research is probably that it clearly shows the limitations of describing a substance as complex as water by the simple LJ3PC modelling approach. It is conceptually too simple to represent the data considered here simultaneously well. The problem is not so much the wide temperature range that was studied, but rather the large variety of the studied properties. Among the different simplifications of describing the interactions of water with LJ3PC models one stands out: the point charge sites must represent both polarity and H-bonding. This can in principle be achieved to a certain extent, as discussed in more detail in [4], but the approach is obviously too crude to capture the thermodynamically relevant characteristics of the complex three-dimensional H-bonding network in water. The adjustment of the parameters to heal the deficiencies of the model works only to some extent; the limits are shown here. Another issue could be the neglect of polarisability. Experience from modelling many other pure substances indicates, however, that this may be easier to heal. The observed systematic deviations of the results of all models for $\epsilon_r$ are in line with this interpretation. Hence, the results from the present study also indicate routes that should be taken to develop better, but still computationally efficient water models. The MCO methodology is clearly the best choice to carry out such developments.

# 5 Regularities of Pareto Sets in Low-dimensional Practical Multi-criteria Optimisation Problems: Analysis, Explanation, and Exploitation

## 5.1 Introduction

Optimisation problems with conflicting objectives are encountered in many practical applications, viz. in engineering; and multi-criteria optimisation (MCO) is an appropriate way to deal with them. In MCO, the Pareto set (PS) is determined, which is the set of the best possible compromises between the objectives [22]. For any point on the PS, an improvement in one objective is only possible at the cost of a decline in at least one other objective; hence, the PS represents the optimal solutions of the MCO problem. For a detailed mathematical discussion, the reader is referred to Ref. [70]. In practical MCO problems, typically only one solution can be realised, which should be chosen from the PS. In MCO, not only the PS is determined, but also its picture in the parameter space (PSPS). Having selected a point on the PS, the PSPS shows which parameters are needed to get to that point. Determining PS is far from trivial as a brute force enumeration of the parameter space is usually infeasible from a practical standpoint. Hence, dedicated methods for determining PS have been developed, see e.g. [25, 26], which, however usually still require a large number of evaluations of the objective function. The present work deals with the topology of low-dimensional PS, as they are encountered in many practical problems. The results enable swift estimations of the PS in these cases.

MCO has been used in many fields of engineering, e.g. the design of energy systems [71, 72], semiconductors [73] and chemical processes [17, 18, 21, 74–76]. In the field of thermodynamics, to give just one example, MCO has been used before for modelling

pure compounds and mixtures [29, 41, 42, 54, 55] with equations of state [77–79] and excess Gibbs energy models [58]. In such applications, both the number of objectives $N$ as well as the number of parameters $M$ is usually low, with typical values below five for both numbers. However, if the evaluation of the objective functions is costly, such as in the development of molecular models, the determination of PS may pose a severe problem even for such low-dimensional MCO problems.

In many previous works on MCO in thermodynamics [41, 54, 55], characteristic topologies of the PSPS and the PS were observed, which depend on the numbers for $N$ and $M$, but not on the specific problem that is investigated. A sketch of these topologies for some cases is shown in Fig. 21.

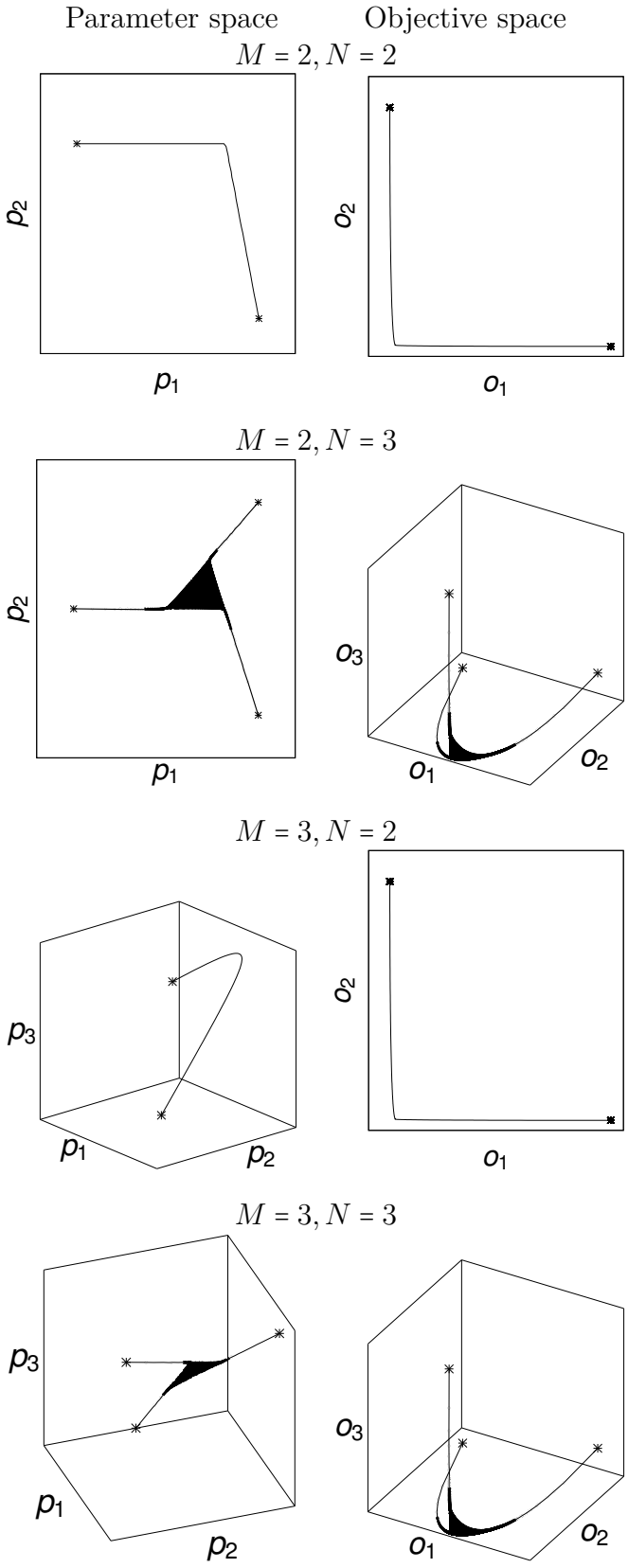



**Figure 21:** Sketches of typical topologies of the Pareto set (PS, right panels) in the objective space and their pictures in the parameter space (PSPS, left panels) for low dimensional MCO problems with $N$ objectives and $M$ parameters as observed in previous studies of our group [41, 54, 55, 58]. The stars indicate the extreme compromises, i.e. the global minima of the $N$ individual objective functions.

It has been shown previously [41, 54, 55] that these characteristic topologies are closely related to properties of the individual objective functions of the MCO problem. In the present chapter, the inner workings of this relation, and the condition under which these topologies occur, are investigated.

The results published by Augusto et al. [80] turned out to be a key for tackling this challenge. Augusto et al. [80] have derived analytical expressions for the PSPS for arbitrary numbers for $M$ and $N$ for the case that all objective functions are quadratic. The topologies sketched in Fig. 21 can be obtained from their results if certain conditions are met. Basically, the lines in Fig. 21 correspond to valleys in the contours of the individual objective functions. Mathematically, the valleys can be related to Eigenvalues and Eigenvectors of the individual objective functions at their minima (crosses in Fig. 21). From this, mathematical conditions can be derived which, if met, lead to the occurrence of the characteristic topologies.

Such valley-like topologies are found in the objective functions of many practical MCO problems. For example, such topologies were observed in many previous studies on MCO in the field of thermodynamics, independent of the type of thermodynamic model that was used, i.e. for molecular models [41, 54, 55], equations of state [59], as well as for models of the Gibbs excess energy [58]. Also, in MCO problems from various other fields, e.g., from process design [76, 81, 82], energy systems engineering [83], and from quantitative spectroscopy [84], PS were found that indicate valley-like structures of the topology of the objective functions. It is referred here to the existence of so-called "Pareto knees" in the PS [85], which may be interpreted as the result of valleys in the topology that intersect each other. If these valleys are steep, the resulting Pareto knee will be sharp.

Hence, there is good reason to believe that what is being studied in the present work is not just some special case, but one which is of high practical interest. The insights gained in our study explain not only empirical observations from the literature (e.g. regarding the Pareto knees), they can also be used for designing new methods for determining PS and the PSPS in practical problems. The conditions for which the interesting and practically relevant type of behaviour studied here has to be expected are also specified.

To illustrate these findings, a practical MCO problem is discussed: the development of molecular models of water, where the aim is to describe different thermodynamic properties well, viz. the vapour pressure and liquid density. The objectives are conflicting, as the considered class of water models (which is the most common one) is fairly simple, so that not all properties of water can be simultaneously modelled with high accuracy. Therefore, compromises have to be made and it is highly desirable to know the PS, which is the set of the best compromises. It is shown that for this MCO problem, the valley-like topologies occur and the mathematical conditions that characterise them are

met. The example is also used to discuss how the insights from the present work can be used for estimating the Pareto set of the MCO problem, based on the analysis of the individual objective functions. It is also clarified that the MCO problems studied here contain no constraints, neither for the objectives nor for the parameters. Considering the influence of adding constraints was out of the scope of the present study, especially as it would have required individual considerations for different types of constraints (e.g. constraints on the PS and the PSPS, respectively).

This chapter is organised as follows: Sec. 5.2 presents a mathematical analysis of properties of Pareto sets in the case of quadratic objective functions, including a discussion of constructing estimates of these sets based on properties of the individual objective functions. In Sec. 5.3, how these findings can be applied to systems, for which the objective functions are not quadratic, but can be approximated by quadratic functions, is discussed. The results are then illustrated by two types of examples in Sec. 5.4: the first one is a study on systems with synthetic quadratic objective functions and the second one is a practical MCO problem from thermodynamics, the optimisation of molecular model of water. Finally, conclusions are drawn in Sec. 5.5.

## 5.2 MCO with quadratic objective functions

### 5.2.1 Quadratic objective functions

In general, a quadratic objective function $f_i$ in an $M$-dimensional parameter space can be written as

$$f_i(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_i)^\intercal \mathbf{H}_i (\mathbf{x} - \mathbf{x}_i)$$
$$\mathbf{H}_i = \begin{pmatrix} h_{11,i} & \dots & h_{1M,i} \\ \vdots & \ddots & \vdots \\ h_{M1,i} & \dots & h_{MM,i} \end{pmatrix} \tag{21}$$

where $i = 1, \dots, N$, and $N$ is the dimensionality of the objective space. The vector $\mathbf{x} \in \mathbb{R}^M$ is a point in the parameter space, $\mathbf{x}_i \in \mathbb{R}^M$ indicates the coordinates of the minimum of $f_i$, $\mathbf{H}_i$ is the Hessian matrix of $f_i$. The Hessian matrix $\mathbf{H}_i$ is a symmetric $M \times M$ matrix, thus $h_{jk,i} = h_{kj,i} \ \forall \ j \neq k$.

### 5.2.2 Method of Augusto et al.

Augusto et al. [80] have described an analytical method for determining the PS for an MCO containing an arbitrary number of quadratic objective functions described

by Eq. (21). The full method, which works for any $M$ and $N$, is described in the Appendix C. Here, it is introduced using the simple but vivid case $M = N = 2$ for illustration.

Let the two objective functions be $f_\mathrm{A}(\mathbf{x})$ and $f_\mathrm{B}(\mathbf{x})$, their Hessian matrices $\mathbf{A}$ and $\mathbf{B}$, and their minima at $\mathbf{x}_\mathrm{A}$ and $\mathbf{x}_\mathrm{B}$. $f_\mathrm{A}$ can be written as

$$\begin{aligned} f_\mathrm{A}(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}_\mathrm{A})^\top \mathbf{A} (\mathbf{x} - \mathbf{x}_\mathrm{A}) \\ \mathbf{A} &= \begin{pmatrix} a_{11} & a_\mathrm{nd} \\ a_\mathrm{nd} & a_{22} \end{pmatrix} \end{aligned} \tag{22}$$

where, $\mathbf{x}, \mathbf{x}_\mathrm{A} \in \mathbb{R}^2$. The Hessian matrix $\mathbf{A}$ is symmetric, and hence both the non-diagonal elements are equal, thus $a_{12} = a_{21} = a_\mathrm{nd}$.

In principle, the Hessian $\mathbf{A}$ as well as the corresponding minimum $\mathbf{x}_\mathrm{A}$ are arbitrary. However, without loss of generality, a coordinate transformation can be applied such that (*i*) $\mathbf{x}_\mathrm{A}$ lies at the origin, and (*ii*) the eigenvectors of $\mathbf{A}$ are parallel to the coordinate axes, so that the non-diagonal elements of $\mathbf{A}$ vanish. Additionally, for the purpose of the present discussion, one of the eigenvalues of $\mathbf{A}$ is scaled to 1, which does not impact the topology of the objective function. Denoting the other eigenvalue by $\xi$, $\mathbf{A}$ can then be written as

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & \xi \end{pmatrix} \tag{23}$$

Varying the parameter $\xi$ yields different ratios of eigenvalues. A valley-like topology of $f_\mathrm{A}$ is obtained if either $\xi \gg 1$ or $\xi \ll 1$. The former case will be discussed, for which the valley is oriented in the $x_1$-direction.

For two quadratic objective functions, following Augusto et al. [80] (Eq. (27) in their paper), any point $\mathbf{x}^*$ on the PSPS can be calculated as follows:

$$\mathbf{x}^* = (w_\mathrm{A}\mathbf{A} + w_\mathrm{B}\mathbf{B})^{-1}(w_\mathrm{A}\mathbf{A}\mathbf{x}_\mathrm{A} + w_\mathrm{B}\mathbf{B}\mathbf{x}_\mathrm{B}) \tag{24}$$

$$w_\mathrm{A} + w_\mathrm{B} = 1 \tag{25}$$

Here, $w_\mathrm{A}$ and $w_\mathrm{B}$ are the weights corresponding to the objective functions $f_\mathrm{A}(\mathbf{x})$ and $f_\mathrm{B}(\mathbf{x})$ respectively. Each choice of the weights $(w_\mathrm{A}, w_\mathrm{B})$ gives a unique point $\mathbf{x}^*$ of the PSPS, so that the evaluation of Eq. (24) as a function of $w_\mathrm{A}$ yields the entire PSPS. The PS is then found by evaluating the objective functions $f_\mathrm{A}$ and $f_\mathrm{B}$ for all values of $\mathbf{x}^*$. When $w_\mathrm{A} \approx 1$, the point $\mathbf{x}^*$ is close to $\mathbf{x}_\mathrm{A}$ (the coordinates of the minimum of $f_\mathrm{A}(\mathbf{x})$), and when $w_\mathrm{B} \approx 1$, the point $\mathbf{x}^*$ is close to $\mathbf{x}_\mathrm{B}$ (the coordinates of the minimum of $f_\mathrm{B}(\mathbf{x})$).

### 5.2.3 Analysis

In the present section, the results of Augusto et al. [80] are used for an analysis of the topology of the PSPS, and, as a consequence, also of the PS. For simplicity and clarity, the case $M = N = 2$ is discussed here, but it is also emphasised that the argument can be generalised, as shown in the Appendix C.

Eq. (24) can be rearranged to give:

$$\mathbf{x}^* = \left[\left(\mathbf{I} + \frac{w_\mathrm{B}}{w_\mathrm{A}}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}(w_\mathrm{A}\mathbf{A})^{-1}\right] w_\mathrm{B}\mathbf{B}\mathbf{x}_\mathrm{B} \tag{26}$$

where $\mathbf{I}$ is the unit matrix of dimension $M \times M$.

Let us consider a point $\mathbf{x}^*$ in the vicinity of $\mathbf{x}_\mathrm{A}$, so that $w_\mathrm{A} \gg w_\mathrm{B}$. Applying a Taylor series expansion in $\frac{w_\mathrm{B}}{w_\mathrm{A}}$ to the first term in brackets in Eq. (26) yields the following approximation:

$$\mathbf{x}^* = \left(\frac{w_\mathrm{B}}{w_\mathrm{A}}\right)\mathbf{A}^{-1}\mathbf{B}\mathbf{x}_\mathrm{B} - \left(\frac{w_\mathrm{B}}{w_\mathrm{A}}\right)^2(\mathbf{A}^{-1}\mathbf{B})^2\mathbf{x}_\mathrm{B} + \mathcal{O}\left[\left(\frac{w_\mathrm{B}}{w_\mathrm{A}}\right)^3\right] \tag{27}$$

In the following, the leading term is kept, i.e.,

$$\mathbf{x}^* = \frac{w_\mathrm{B}}{w_\mathrm{A}}\mathbf{A}^{-1}\mathbf{B}\mathbf{x}_\mathrm{B} + \mathcal{O}\left[\left(\frac{w_\mathrm{B}}{w_\mathrm{A}}\right)^2\right] \tag{28}$$

Substituting the values for $\mathbf{x}_\mathrm{B} = [x_{1,\mathrm{B}}, x_{2,\mathrm{B}}]^\intercal$ and the matrices $\mathbf{B}$ and $\mathbf{A}^{-1}$ yields:

$$\mathbf{x}^* = \begin{pmatrix} \frac{w_\mathrm{B}}{w_\mathrm{A}}(b_{11}x_{1,\mathrm{B}} + b_\mathrm{nd}x_{2,\mathrm{B}}) \\ \frac{w_\mathrm{B}}{w_\mathrm{A}}\frac{1}{\xi}(b_\mathrm{nd}x_{1,\mathrm{B}} + b_{22}x_{2,\mathrm{B}}) \end{pmatrix} \tag{29}$$

Note that the coordinate transformation was carried out such that the minimum of $f_\mathrm{A}$ is shifted to the origin and $\mathbf{A}$ becomes a diagonal matrix, while the minimum of $f_\mathrm{B}$ is generally at some other position $\mathbf{x}_\mathrm{B} = (x_{1,\mathrm{B}}, x_{2,\mathrm{B}})^\intercal$ and $\mathbf{B}$ is not a diagonal matrix, i.e. $b_\mathrm{nd} \neq 0$. Eq. (29) shows directly that for $\frac{w_\mathrm{B}}{w_\mathrm{A}} \ll 1$, the PSPS approaches the origin, i.e. the coordinates of minimum of $f_\mathrm{A}$.

If $\xi \gg 1$, Eq. (29) can be approximated by:

$$\mathbf{x}^* = \begin{pmatrix} \frac{w_\mathrm{B}}{w_\mathrm{A}}(b_{11}x_{1,\mathrm{B}} + b_\mathrm{nd}x_{2,\mathrm{B}}) \\ \mathcal{O}(\frac{w_\mathrm{B}}{w_\mathrm{A}}\frac{1}{\xi}) \end{pmatrix} \tag{30}$$

Hence, for $\xi \gg 1$, the $x_2$-component of $\mathbf{x}^*$ is small, i.e. the PSPS near the minimum of $f_\mathrm{A}$ lies close to the $x_1$-axis.

The eigenvalues of $f_A$ are 1 and $\xi$, and the eigenvector belonging to the eigenvalue 1 is the vector in the $x_1$-direction. Hence, Eq. (30) can be interpreted as follows: for $\xi \gg 1$, the minimum of $f_A$ together with the eigenvector corresponding to the smaller eigenvalue of $\mathbf{A}$ define a linear approximation of the PSPS in the vicinity of the minimum of $f_A$. The geometric interpretation of this finding in the $f_A(x_1, x_2)$ space is that for $\xi \gg 1$, the minimum of $f_A$ lies at the bottom of a valley which goes in the $x_1$-direction (the direction of the eigenvector corresponding to the eigenvalue 1). The deviations of the PSPS from that linear approximation increase with increasing distance from the minimum of $f_A$ (with increasing $\frac{w_B}{w_A}$) and with lower ratios of the eigenvalues $\xi$. The same arguments can be applied to $f_B$.

The findings that were discussed above for the case $M = N = 2$ can be generalised to arbitrary values of $M$ and $N$. For each objective $f_i$ $(i = 1, \ldots, N)$, an analysis as the one described above can be carried out. If one of the eigenvalues of the $M \times M$ diagonal Hessian matrix $\mathbf{H}_i$ defining $f_i$ (see Eq. (21)) is much smaller than all the other eigenvalues, then the minimum of $f_i$ together with the eigenvector corresponding to that eigenvalue define a linear approximation of the PSPS, in the vicinity of the minimum. The corresponding equations for the general case are given in the Appendix C.

### 5.2.4 Estimation of the Pareto set

Now the case of quadratic objective functions is considered and it is assumed that for each objective function, one eigenvalue of the Hessian matrix is much smaller than the others, i.e. the objective functions have valley-like topologies. It then follows from Sec. 5.2.3 that estimates for the Pareto set of the MCO problem in the vicinity of the minima can be obtained by considering only the valleys of the individual objective functions $f_i$ $(i = 1, \ldots, N)$. Two alternative ways for achieving this are described.

a) *Linear approximation*
The linear approximation consists of the following three steps to be carried out for all $i = 1, \ldots, N$ objectives:

1) Find the minimum $\mathbf{x}_i$ of $f_i$.

2) Find the eigenvector $\mathbf{u}_i$ belonging to the smallest eigenvalue of the matrix $\mathbf{H}_i$.

3) Combine the results from 1) and 2) to obtain a linear approximation of the PSPS near $\mathbf{x}_i$.

This process yields $N$ straight lines that can be combined to get an estimate of the PSPS. From this, the PS can be found by evaluating the objective functions of the

MCO problem at each point on the PSPS. The way these individual lines have to be combined to form the estimate of the PSPS depends on $M$ and $N$.

This will be discussed by referring to Fig. 21. Let us consider the first case, $M = N = 2$: there are two lines which will intersect in general. The intersection point in the objective space is known as the *Pareto knee*. The arguments given here explain why often very sharp Pareto knees are found. Not the entire lines belong to the PSPS, but only their respective sections between the minima and the Pareto knee.

For $M = 2$ and $N = 3$, there are three lines in the two-dimensional parameter space. Not only these three lines belong to the PSPS, but also the triangular region between these lines (see Fig. 21). The picture of this triangular region in the objective space corresponds to the Pareto knee. This area connecting the lines (both in the PS as well as the PSPS) will be referred to as the *Pareto knee region*.

For the case considered here ($M = 2$, $N = 3$), the area between the lines in the PSPS is in general triangular and easy to construct. For other cases, defining this area is not so straightforward; e.g. for $M = 3$ and $N = 3$, there are three lines which, however, will in general not intersect. It was not in the scope of the present work to establish a general theory on how to construct the Pareto knee region from the knowledge of the lines. Intuitively, the Pareto knee region is expected to be found where lines are close to each other and the solution on how to construct it will depend on the quality of the linear approximation of the PSPS in the regions where the lines are close.

The linear approximation of the PSPS will only be good near the minima $\mathbf{x}_i$, while deviations will occur far from the minima. The size of these deviations depends on the details of the problem, including the distance of the different minima $\mathbf{x}_i$ ($i = 1, \ldots, N$) from each other. It would be possible to get better predictions by taking into account higher order terms in the Taylor series expansion mentioned above, but this would require the knowledge of derivatives of the objective function up to an order higher than two, which may be difficult to obtain in practical problems. The idea of the valley-like topology of the objective functions enables devising a different strategy instead, which is called *valley approximation* in the following.

b) *Valley approximation*

The valley approximation is a modification of the linear approximation, in which the step 2 in the scheme shown above is replaced by a different scheme for finding the lines which start at the minima $\mathbf{x}_i$ of the objective functions $f_i$. The geometrical picture of the objective functions for which one of the eigenvalues is much smaller than the others is a valley-like topology. The lowest point of the valley is the minimum of $f_i$, located at $\mathbf{x}_i$. The valley is first approximated well by linear approximation, in the vicinity of $\mathbf{x}_i$, using the approach described above under a). However, with increasing

distance from $\mathbf{x}_i$, deviations between the valley of $f_i$ and the linear approximation will occur. It is suggested to use the bottom line of the valley as an approximation of the PSPS rather than using the linear approximation also for large distances away from $\mathbf{x}_i$. While a formal proof that this is an improvement over the linear approximation is not given here, but some examples below that confirm the according expectation have been discussed.

## 5.3 Application to non-quadratic MCO problems

The topologies shown in Fig. 21 were observed in practical MCO problems from chemical engineering, and they are fully in line with the theory developed in the previous section for the quadratic objective functions. In particular, the theory explains the occurrence of the different cases shown in Fig. 21. For $N > 2$, needle-like extensions of the PSPS are only found if the minima of the objective functions $f_i$ lie outside the Pareto knee region.

This suggests that the following holds for the objective functions $f_i$ in the underlying MCO problem:

(i) The $f_i$ can be approximated reasonably by quadratic functions.

(ii) One of the eigenvalues of the Hessian matrix of $f_i$, evaluated at the minimum of $f_i$ is much smaller than the other eigenvalue.

If these conditions are met, it is expected that a) the linear approximation as well as, alternatively, b) the valley approximation can be used for obtaining estimates of the PSPS, and, accordingly of the PS, as described in Sec. 5.2.4. Furthermore, for real MCO problems, there is a further option:

c) *Quadratic approximation*

In the first step, for each goal function $f_i$, quadratic approximations around their minima are constructed. Then, the method of Augusto et al. [80] is applied to each of these quadratic functions, so that an estimate of the PSPS is obtained.

Near the minimum, the quadratic approximation yields the same results as the linear approximation, but in the region of the Pareto knee, it may deviate strongly from the linear approximation. Furthermore, the linear approximation as well as the valley approximation yield only curves, while the quadratic approximation can yield objects of higher dimensions, which is important in the Pareto knee region.

## 5.4 Examples

### 5.4.1 Synthetic quadratic objective functions

In this section, synthetic quadratic objective functions are considered, as defined in Eq. (22) in two-dimensional parameter space, i.e. $M = 2$. Three different quadratic objective functions, $f_\mathrm{A}$, $f_\mathrm{B}$ and $f_\mathrm{C}$, are defined in the following. An example with a two-dimensional objective space, i.e. $N = 2$, using only the two objectives $f_\mathrm{A}$ and $f_\mathrm{B}$, as well as a second example with a three-dimensional objective space, i.e. $N = 3$, using all three objectives $f_\mathrm{A}$, $f_\mathrm{B}$ and $f_\mathrm{C}$, will be considered.

As in Sec. 5.2.2, the objective function $f_\mathrm{A}(\mathbf{x})$ was chosen to have its minimum at the origin, its eigenvectors parallel to the coordinate axes, and the two eigenvalues 1 and $\xi$. As a consequence, $\mathbf{A}$ is a diagonal matrix with its diagonal entries as 1 and $\xi$. The other two objective functions, $f_\mathrm{B}(\mathbf{x})$ and $f_\mathrm{C}(\mathbf{x})$, are obtained by rotating $f_\mathrm{A}(\mathbf{x})$ about the origin and translating it by different amounts, such that the minima are shifted. $f_\mathrm{B}(\mathbf{x})$ is obtained by rotating $f_\mathrm{A}(\mathbf{x})$ by -70° and positioning its minimum at $\mathbf{x}_\mathrm{B} = [10, -5]^\intercal$, while $f_\mathrm{C}(\mathbf{x})$ is obtained by rotating $f_\mathrm{A}(\mathbf{x})$ by 45° and positioning its minimum at $\mathbf{x}_\mathrm{C} = [10, 5]^\intercal$. Thus, the ratio of eigenvalues $\xi$ is the same for all three objectives. Choosing the same ratio of eigenvalues for all objective functions is not a requirement for the current analysis to be valid; it was only done for convenience.

Two different ratios of eigenvalues, viz. 10 and 100, are used to assess the accuracy of the linear approximation and the valley approximation of the PSPS and the PS (see Sec. 5.2.4). The exact solution was calculated using the method of Augusto et al. [80] For $\xi = 100$, the valley-like landscape is "steeper" than that for $\xi = 10$. Fig. 22 illustrates the three objective functions $f_\mathrm{A}$, $f_\mathrm{B}$, and $f_\mathrm{C}$, for $\xi = 10$.

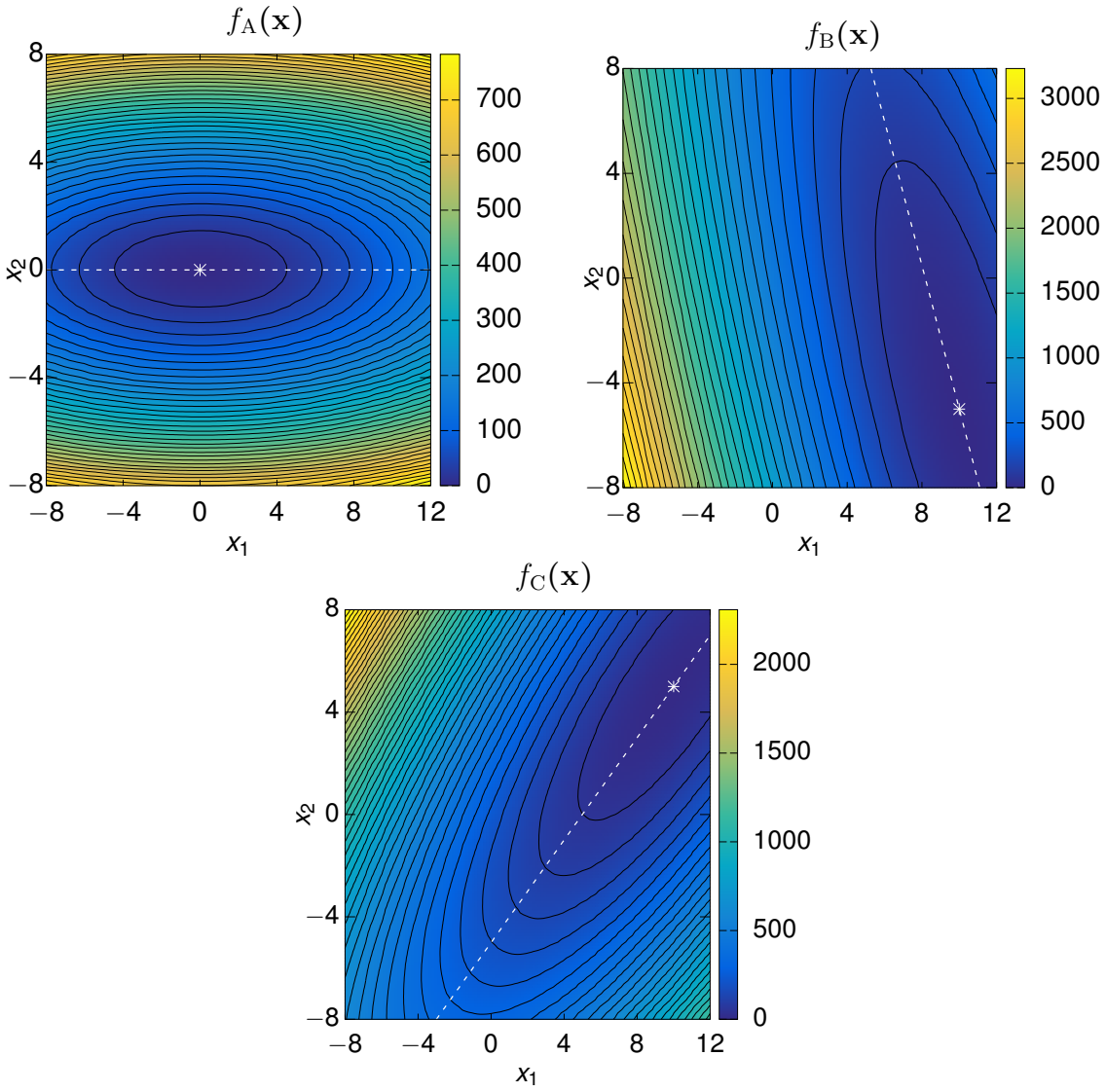



**Figure 22:** Contour plots of the three synthetic quadratic objective functions considered in the present work. The white stars indicate the global minima of the functions, while the white dashed lines indicate the valleys (paths of lowest ascent) that pass through the minima.

Fig. 23 presents the Pareto sets for the MCO when the two objectives $f_\mathrm{A}(\mathbf{x})$ and $f_\mathrm{B}(\mathbf{x})$ are considered. The figures show the PSPS on the left and the PS on the right. Additionally, in the parameter space, the valleys of the individual functions as well as their minima (analogous to Fig. 22) are shown for reference.

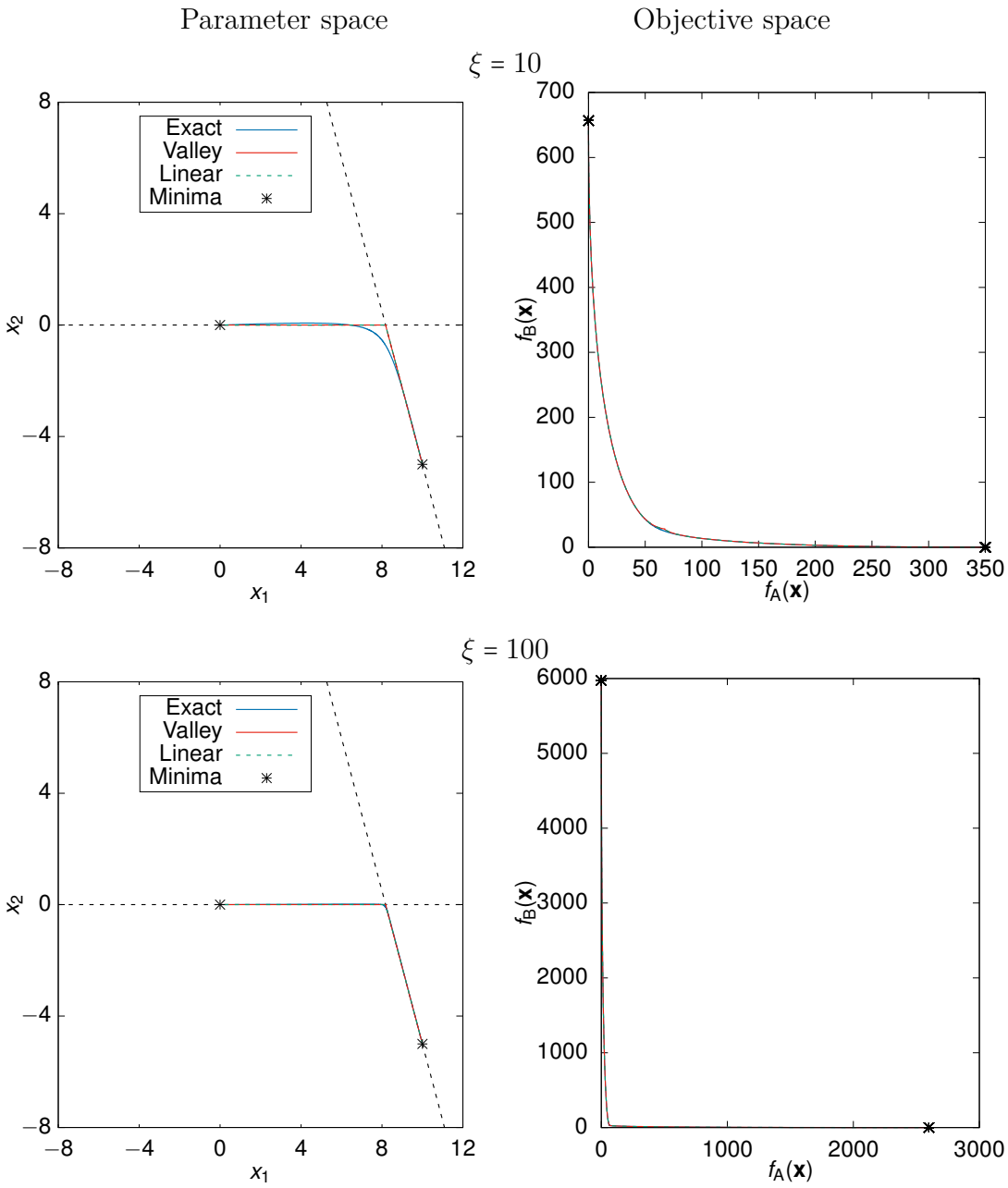



**Figure 23:** Pareto sets for the MCO with the two synthetic quadratic objective functions $f_A$ and $f_B$. Top row: $\xi = 10$, bottom row: $\xi = 100$. Left: parameter space (PSPS), right: objective space (PS). Blue lines: exact solution, red lines: valley approximation, green dotted lines: linear approximation(see Sec. 5.2.4), black dotted lines: valleys of the individual functions. The stars in the plots indicate the global minima of the individual objective functions. The valley approximation and the linear approximation are strictly identical here.

For the synthetic quadratic functions, the valleys are straight lines along the eigenvectors corresponding to the smaller eigenvalue of the Hessians of the objective functions. As a consequence, the valley approximations (red lines) and the linear approximations (green dotted lines) coincide with each other.

For $\xi = 10$, PSPS is accurately approximated, except for the region of the Pareto knee. However, even the differences in that region hardly show up in the PS, which is very well approximated. As expected, the approximation is even better for $\xi = 100$, where the differences between the approximation and the exact solution are basically negligible both for the PSPS and the PS.

Fig. 24 presents the Pareto sets for the synthetic quadratic objective functions for all three objectives $f_A(\mathbf{x})$, $f_B(\mathbf{x})$, and $f_C(\mathbf{x})$. Again, the two ratios of eigenvalues $\xi = 10$ and $\xi = 100$ are considered. As for the case with two objectives discussed above, the PS is approximated very well for both $\xi = 10$ as well as $\xi = 100$. Therefore, for brevity, only the results for the PSPS are shown in Fig. 24.

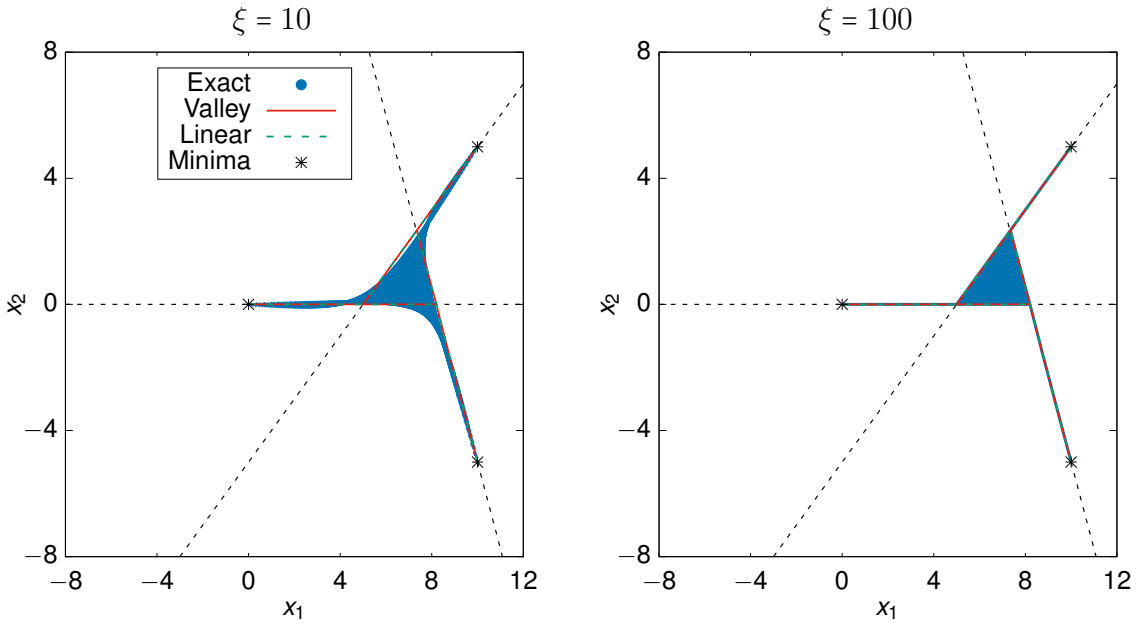



**Figure 24:** PSPS for the MCO with the three synthetic quadratic objective functions $f_A$, $f_B$ and $f_C$. Left: $\xi = 10$, right: $\xi = 100$. Blue: exact solution, red: valley approximation, green: linear approximation, black dotted lines: valleys of the individual functions. The stars in the plots indicate global minima of the individual objective functions. The valley approximation and the linear approximation are strictly identical here.

For $\xi = 10$, the approximations work very well near the minima, while in the region of the Pareto knee, deviations are observed. For $\xi = 100$, these deviations basically vanish, leading to an almost perfect approximation of the PSPS.

### 5.4.2 Parametrisation of a molecular model of water

As a practical example for the applying of the concepts for estimating Pareto sets developed in the present work, an MCO problem from thermodynamics, i.e. the parametrisa-

tion of a molecular model of water, is now discussed. Therefore, a popular water model, the so-called SPC/E model [35] is chosen as a starting point. The problem with applying MCO for the development of molecular models of fluids is that the evaluations of the goal functions are computationally extremely costly, as they usually require carrying out molecular simulations at many state points in each iteration step. This problem is circumvented here by using the so-called reduced units methods, which is explained below, together with simulation data for the SPC/E model from a previous study of our group [55]. This enabled us to carry out the analysis without carrying out additional molecular simulations. In the MCO problem, which is studied here, there are three objective functions which measure the difference between the results from the molecular model [55] and experimental data [32, 33] for three important thermodynamic properties of water: saturated liquid density ($\rho^{\mathrm{liq}}$), vapour pressure ($p^{\mathrm{s}}$) and enthalpy of vaporisation ($\Delta h^{\mathrm{vap}}$). The definition of the corresponding objective functions $\delta p^{\mathrm{s}}$, $\delta \rho^{\mathrm{liq}}$ and $\delta(\Delta h^{\mathrm{vap}})$ was adopted from our previous work [55] and is included in the Appendix C.

The *reduced units method* [12] is used here to obtain the relation between the parameters and the objectives. While SPC/E model has five parameters, only two of them were used as variables ($M = 2$), in order to be able to apply the reduced units method. These are the Lennard-Jones size parameter $\sigma$ and the Lennard-Jones energy parameter $\varepsilon$. Upon varying $\sigma$ and $\varepsilon$, the other model parameters vary in a prescribed way. Details are given in [12] and are not important for the present discussion. This is mentioned here only to emphasise that further imptovements would be possible if all parameters were used in the MCO, which is simply not done here for computational reasons, and as $M = 2$ is also convenient for the discussion. If simulation results for the original model, which is SPC/E here, are available, the reduced units method yields analytical expressions for obtaining the relation between the simulation results and the parameters $\sigma$ and $\varepsilon$. From these, the objectives of the MCO problem, $\delta p^{\mathrm{s}}$, $\delta \rho^{\mathrm{liq}}$ and $\delta(\Delta h^{\mathrm{vap}})$, can be calculated. The reader is referred to the original Ref. [12] for more details on the reduced units method.

Contour plots for the three objectives $\delta \rho^{\mathrm{liq}}$, $\delta p^{\mathrm{s}}$ and $\delta(\Delta h^{\mathrm{vap}})$ as functions of the parameters $\sigma$ and $\varepsilon$ are shown in Fig. 25. Both $\delta \rho^{\mathrm{liq}}$ and $\delta p^{\mathrm{s}}$ show global minima as well as valleys passing through the minima. Here, the valleys were estimated by selecting the lowest values of the individual objectives along their cross sections parallel to the coordinate axes. Note that $\delta(\Delta h^{\mathrm{vap}})$ does not depend on $\sigma$ here. As a consequence, the contour lines of this objective are parallel to the $\sigma$-axis. Thus, there is no distinct minimum for this objective, rather the minimum is "spread" across an entire line parallel to the $\sigma$-axis, i.e. a line of constant $\varepsilon$.

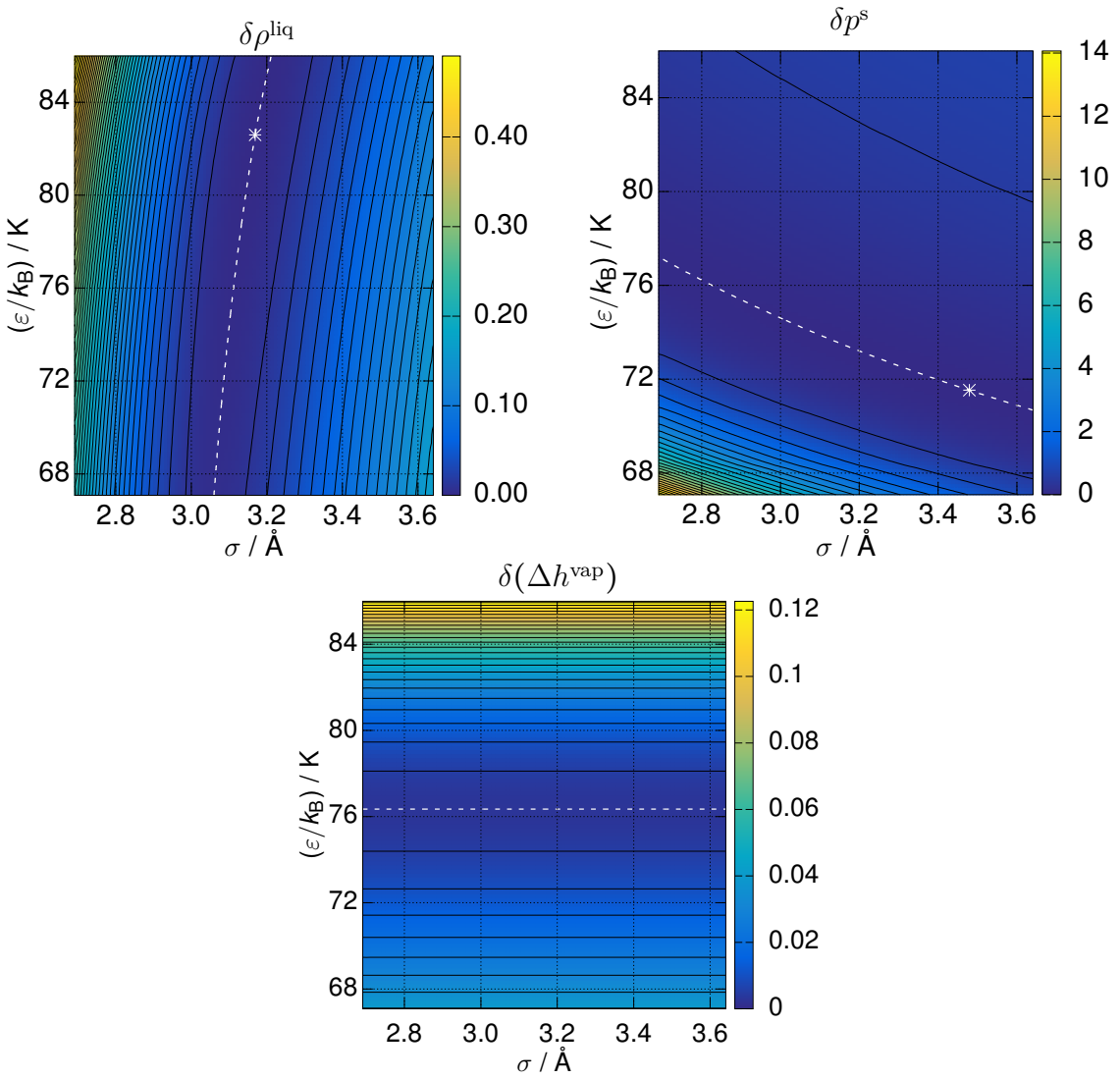



**Figure 25:** Contour plots of the three objective functions for the optimisation of the SPC/E water model. The objectives $\delta\rho^{\mathrm{liq}}$, $\delta p^{\mathrm{s}}$ and $\delta(\Delta h^{\mathrm{vap}})$ indicate the deviations between simulation results for SPC/E and the experimental data for saturated liquid density, vapour pressure and enthalpy of vaporisation respectively. The white stars indicate the minima of the functions, while the white dashed lines indicate the valleys (paths of lowest ascent) that pass through the minima. There is no distinct minimum for $\delta(\Delta h^{\mathrm{vap}})$ as this objective does not depend on $\sigma$.

Since analytical relations between the objectives and the parameters are available, it is possible to calculate their second derivatives at the respective minima. From these, the Hessian matrices can be calculated, and the quadratic approximations of the objective functions can be obtained. Also, the eigenvalues of the Hessian at the minima of the individual objectives were computed, which are shown in Table 4. The ratios of the eigenvalues are significantly larger than those considered for the synthetic quadratic functions, c.f. Sec. 5.4.1.

| Objective | $\delta\rho^{\mathrm{liq}}$ | $\delta p^{\mathrm{s}}$ | $\delta(\Delta h^{\mathrm{vap}})$ |
|---|---|---|---|
| Larger eigenvalue | 1.7922 | 1.5337 | $1.1147 \times 10^{-3}$ |
| Smaller eigenvalue | $9.3032 \times 10^{-5}$ | $2.7859 \times 10^{-3}$ | 0 |
| Ratio of eigenvalues | $1.9265 \times 10^{4}$ | 550.5430 | $\infty$ |

**Table 4:** Eigenvalues of the Hessian matrices for the simulation results of the three objective functions considered in the MCO of the molecular model of water.

For MCO in this case, similar to the synthetic quadratic objective functions, first, only two objectives, viz. $\delta\rho^{\mathrm{liq}}$ and $\delta p^{\mathrm{s}}$, are considered, so that $N = 2$. In a second step, three objectives, $\delta\rho^{\mathrm{liq}}$, $\delta p^{\mathrm{s}}$ and $\delta(\Delta h^{\mathrm{vap}})$, were considered, so that $N = 3$. For both cases, the PSPS and the PS obtained by three methods presented in Sec. 5.3, the linear approximation, the valley approximation and the quadratic approximation, were compared to the exact Pareto set. The latter was calculated by a brute-force enumeration of the numbers for the objective functions on a narrowly spaced parameter grid, followed by a selection of the points that were not dominated.

Fig. 26 presents the results obtained for the PSPS (left) and the PS (right) for the case $N = 2$. Also, a zoom into the Pareto knee region of the PS is provided. The three approximations are shown together with the exact Pareto set. Additionally, the bottom lines of the valleys in the two objective functions and the corresponding minima are shown.

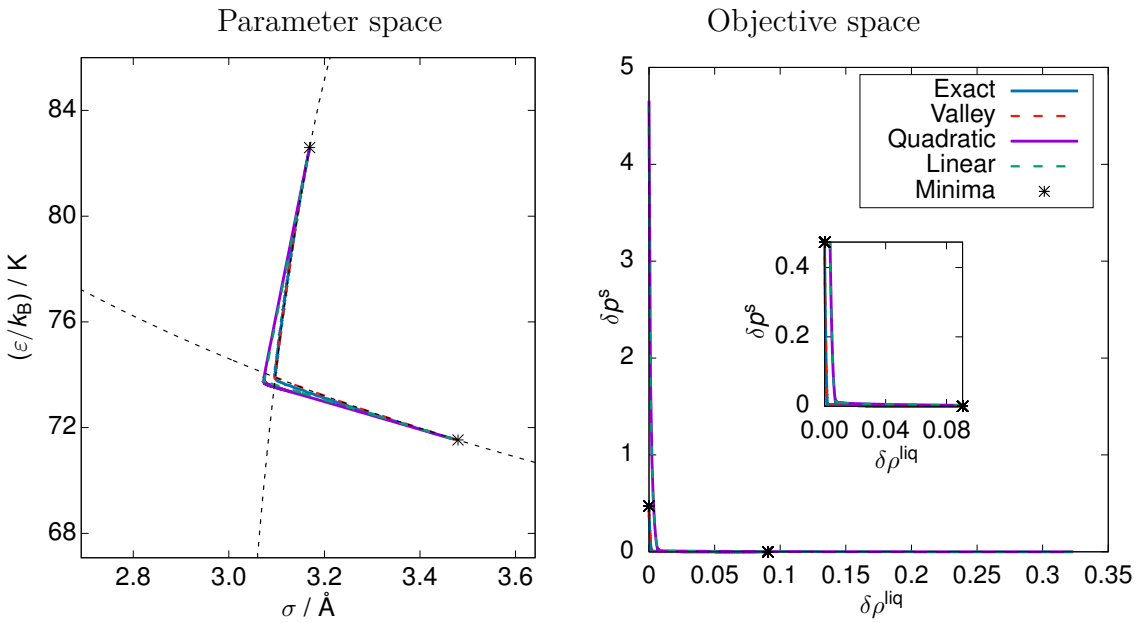



**Figure 26:** Pareto set for the MCO of the SPC/E water model with the two objective functions $\delta\rho^{\text{liq}}$ and $\delta p^{\text{s}}$. Left: parameter space (PSPS) and right: objective space (PS) along with the magnified view of the PS around the Pareto knee region. Blue: exact solution, red: valley approximation, purple: quadratic approximation and green: linear approximation of the Pareto set. Additionally, the valleys (see Fig. 25) are shown as black dashed lines for reference. Stars indicate the global minima of the individual objective functions. Exact solution and valley approximation of the Pareto set coincide with each other, while quadratic approximation and linear approximation of the Pareto set also coincide with each other.

The valley approximation matches the exact Pareto set almost perfectly. As the objective functions are no longer strictly quadratic, the linear approximation deviates from the bottom lines of the valleys, which are slightly non-linear. The linear approximation and the quadratic approximation show almost no differences, even in the region of the Pareto knee, which can be understood as a consequence of the topology of the individual objective functions with their sharp and deep valleys, and, hence, as a result from the wide spread of the Eigenvalues. However, these two coinciding approximations, while giving still fair results, are clearly less accurate than the (almost perfect) valley approximation.

Fig. 27 presents the results obtained for the PSPS (left) and the PS (right) for the case $N = 3$. As in Fig. 26, the results of the three approximations are compared to the exact solution. However, now there are three objective functions, and, hence, three valleys, only two of which have minima for the reasons discussed above.

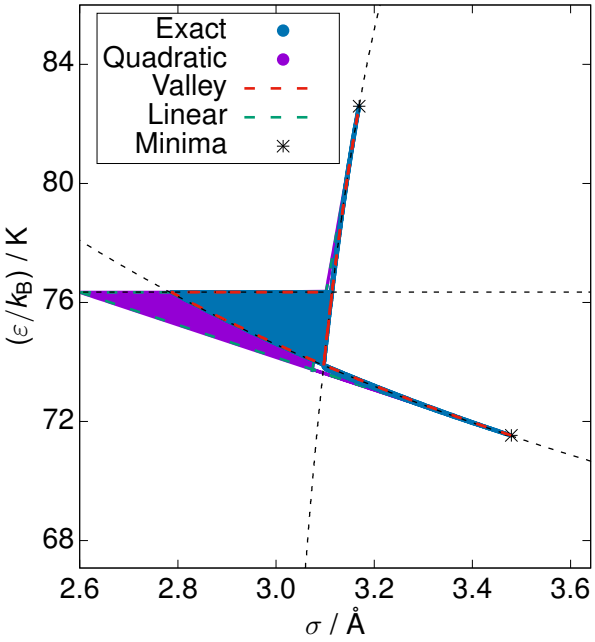



**Figure 27:** PSPS for the MCO of the SPC/E water model with the three objective functions $\delta\rho^{\mathrm{liq}}$, $\delta p^{\mathrm{s}}$ and $\delta(\Delta h^{\mathrm{vap}})$. Blue: exact solution, red: valley approximation, purple: quadratic approximation, and green: linear approximation. Stars indicate the global minima of the individual objective functions. Additionally, the valleys (see Fig. 25) are shown as black dashed lines for reference.

In this case, a triangular Pareto knee region with two extensions is expected, which is the case here. These two extensions end at the two minima, which lie outside the Pareto knee region. This topology is not only found in the brute force evaluation of the Pareto set but also by all considered approximation methods. The quadratic approximation yields the triangular surface and the extensions, while the other two methods yield only three intersecting lines, from which, it is, however, easy to construct the triangular Pareto knee region including the two extensions. As for $N = 2$, the predictions from the valley method are almost perfect, and as for $N = 2$, the differences between the results from the quadratic approximation and the linear approximation are negligible. Again, as for $N = 2$, these predictions are less accurate than those of the valley approximation.

Altogether, the results from the presented examples suggest that the valley approximation is an accurate and useful one in practical MCO problems.

## 5.5 Conclusions

The observed peculiar topologies of Pareto sets and their pictures in the parameter space are found when the individual objective functions have minima that lie in deep

and sharp valleys. In the present work, this has been analysed in depth considering quadratic objective functions as well as objective functions from practical engineering problems. The findings explain not only the peculiar topologies of the Pareto sets, they can also be used to construct the Pareto sets.

Questions left open by the present work include: (a) an analysis of more examples of low-dimensional practical MCO problems. The hypothesis is that many will be found, which show the peculiar topologies mentioned above. It would also be interesting to make generic statements on why this type of problems is so common. (b) The present method for estimating the Pareto set is based on finding branches of the Pareto set that start in the extreme compromises. Basically the question how to find out in which way these branches are linked in the Pareto set when this does not follow from simple geometric considerations such as intersection points is left open. Here the quadratic approximation method, in which all objective functions are approximated with quadratic functions around their respective minima, which was presented in the present work, could be useful, as it yields not only branches but also their connections. Finally, the new methods should be implemented in a robust and efficient program package.

# 6 Conclusions

In the present work, the challenging problem of applying multi-criteria optimisation (MCO) to the paramterisation of molecular models of water has been tackled.

As a starting point, molecular simulations were carried out, in which seven common water models (SPC, SPC/E, TIP3P, TIP4P, TIP4P/2005, Huang and OPC) of the type "Lennard-Jones (LJ) site + three partial charges" were considered. For these water models, the vapour pressure $p^{\mathrm{s}}$, saturated liquid density $\rho^{\mathrm{liq}}$ and the enthalpy of vaporisation $\Delta h^{\mathrm{vap}}$ were determined in the temperature range between 350 K and 550 K. The results were compared to reference data and the objective functions for the description of the three properties were defined: $\delta p^{\mathrm{s}}$, $\delta \rho^{\mathrm{liq}}$ and $\delta(\Delta h^{\mathrm{vap}})$. None of the existing models represent the three studied VLE properties satisfactorily. This leaves room for a further improvement, for which multi-criteria optimisation (MCO) was used.

Applying MCO to the optimisation of molecular models is challenging, as even when sophisticated MCO algorithms are applied, brute force strategies are infeasible due to exceedingly high computational effort. This problem is circumvented here by applying the reduced units method, which is based on the fact that the simulation yields results in dimensionless properties and contain information on an infinite number of solutions in physical variables. Starting only with a single set of simulations for a given molecular model, this enables MCO without having to run additional simulations. The drawback of this method is that only a subspace of the model parameter space can be used for the optimisation. In the present case, the independent model parameters are the two LJ parameters $\sigma$ and $\varepsilon$, while the other parameters vary in a prescribed way together with $\sigma$ and $\varepsilon$. This drawback of the restricted parameter space can be mitigated partially, by using different starting models in the reduced units methods.

In the first MCO study in the present work, the seven water models mentioned above were used as starting values. Using this strategy, Pareto sets were determined for an MCO problem in which two objectives ($\delta p^{\mathrm{s}}$ and $\delta \rho^{\mathrm{liq}}$) were considered. The Pareto set is the set of the best compromises between these conflicting objectives. As both the parameter space and the objective space are two-dimensional here, both have an L-shape, which can be understood from the consideration of the two single-criteria optimisations that yield the extreme compromises. A particularly attractive model is

obtained in the so-called Pareto knee, which corresponds here to the tip of the "L". Using the reduced units method, seven Pareto sets were determined starting from the seven models from the literature. In all cases, the MCO leads to significant improvements. The best performance is achieved for the model from the Pareto knee that was obtained with the reduced units method starting from the Huang model.

In the second MCO study in the present work, we have carried out a comprehensive study on the performance of the LJ3PC model class for describing thermophysical properties of water. Five properties (vapour pressure ($p^{\mathrm{s}}$), saturated liquid density ($\rho^{\mathrm{liq}}$), self-diffusion coeffcient ($D$), shear viscosity ($\eta$) and relative permittivity ($\epsilon_{\mathrm{r}}$)) were considered over a wide temperature range between 55% and 90% of the critical temperature. None of the literature models gives a good representation of this large and diverse data set of thermophysical properties.

Five parameters of the LJ3PC model class were used in a multicriteria optimization (MCO): $\sigma$, $\varepsilon$, the partial charge $q$, and two of the three parameters that determine the position of the partial charges $l_1$ and $l_2$, while the angle $\alpha$ was set using physical arguments. Similar to the first MCO study, due to the high cost of the molecular simulations, it is impossible to determine the full Pareto set for this MCO problem. We have therefore used a similar work-around, in which the reduced units method (RUM) was combined with a hierarchical approach. This allowed us to sample important parts of the five-dimensional parameter space of the LJ3PC model class and to determine the set of Pareto-optimal models in that subspace. The choice of the subspace was guided by an analysis of the parameters of the previously published literature models of water from the LJ3PC class, but in contrast to the previous study, a large region of the parameter space was covered. To the best of the authors' knowledge, it is by far the most comprehensive assessment of the performance of LJ3PC models for describing thermophysical properties of water that has ever been carried out.

The Pareto set determined in the present work gives an overview of what can be achieved with LJ3PC models regarding the description of the considered properties of water. We cannot guarantee that no substantially better solutions exist outside the studied range, but have designed our approach so that this is unlikely.

The results from the present study confirm that there are conflicts between the five goals to represent the five studied properties of water well with a single model. The exception is the representation of $D$ and $\eta$, which behave similarly. This confirms once more that these properties are closely related. The overall picture resulting from the MCO would not have changed substantially, had we taken into account only one of them. Conflicts exist between all other properties. However, in many cases good compromises can be found between several of the studied properties, but no model was found that would yield a good description of all of them. The most important conflict is that between $p^{\mathrm{s}}$

on the one side and ($D$ and $\eta$) on the other. There are models that represent either $p^s$ (and some other properties) well or ($D$ and $\eta$) (and some other properties); also balanced compromises exist, but only at the cost of increased deviations in all three properties.

In both MCO studies in the present work, the Pareto sets, as well as their pictures in the parameter space, were found to have peculiar topologies that are closely related to the topologies of the individual objective functions, which often have a minimum lying at the lowest point of a deep and sharp valley. Such structures are characterized mathematically by the fact that one of the Eigenvalues of the Hessian of the objective function, at its minimum, is much smaller than the others, which seems to be the case in many engineering problems. This finding can be used for estimating the Pareto set of the MCO problem: the minima of the individual objective functions are extreme compromises of the MCO problem, and, as such belong to the Pareto set. In the case described above, starting from the minimum, more points of the Pareto set can be found by following the valleys. We call this this valley approximation. The so-called Pareto knee region, which is usually the region where the most interesting compromises are located, is found where the valleys meet or at least approach each other. This is discussed in the present work first for MCO problems with quadratic objective functions, for which the Pareto set can be determined analytically using the method of Augusto et al. [80] Then, the analysis is extended to MCO problems with non-quadratic objective functions, that, however, are required to have the topology described above. Three schemes to determine the Pareto set of these MCO problems are discussed, of which the valley approximation gave the best results. Hence, the results not only explain the peculiar topologies of the Pareto sets observed earlier, but can also be used to construct approximations of Pareto sets.

The results from both MCO studies suggest that while a good representation of up to three properties can be obtained by the studied LJ3PC water model class, models of this class are not able to accurately represent all five studied objectives over a broad temperature range simultaneously. Additionally, the rationale behind the peculiar topologies of the Pareto sets and their pictures in the parameter space have been clarified, and the conditions necessary for their occurrence have been found.

# Literature

[1] E. Maginn, J. Elliott: Historical perspective and current outlook for molecular dynamics as a chemical engineering tool, Industrial & engineering chemistry research 49 (2010) 3059–3078.

[2] D. N. Theodorou: Progress and outlook in monte carlo simulations, Industrial & engineering chemistry research 49 (2010) 3047–3058.

[3] I. Shvab, R. J. Sadus: Intermolecular potentials and the accurate prediction of the thermodynamic properties of water, J. Chem. Phys. 139 (2013) 194505.

[4] K. Langenbach, C. Engin, S. Reiser, M. Horsch, H. Hasse: On the simultaneous description of h-bonding and dipolar interactions with point charges in force field models, AIChE J. 61 (2015) 2926–2932.

[5] P. T. Kiss, A. Baranyai: Testing the recent charge-on-spring type polarizable water models. II. vapor-liquid equilibrium, J. Chem. Phys. 137 (2012) 194103.

[6] E. A. Ploetz, A. S. Rustenburg, D. P. Geerke, P. E. Smith: To polarize or not to polarize? Charge-on-spring versus KBFF models for water and methanol bulk and vapor–liquid interfacial mixtures, J. Chem. Theory Comput. 12 (2016) 2373–2387.

[7] R. Faller, H. Schmitz, O. Biermann, F. Müller-Plathe: Automatic parameterization of force fields for liquids by simplex optimization, J. Comput. Chem 20 (1998) 100–109.

[8] D. Reith, H. Meyer, F. Müller-Plathe: CG-OPT: A software package for automatic force field design, Comput. Phys. Commun. 148 (2002) 299–313.

[9] J. Wang, P. A. Kollman: Automatic parameterization of force field by systematic search and genetic algorithms, J. Comput. Chem. 22 (2001) 1219–1228.

[10] E. Bourasseau, M. Haboudou, A. Boutin, A. H. Fuchs, P. Ungerer: New optimization method for intermolecular potentials: Optimization of a new anisotropic united atoms potential for olefins: Prediction of equilibrium properties, J. Chem. Phys. 118 (2003) 3020–3034.

[11] S. Deublein, P. Metzler, J. Vrabec, H. Hasse: Automated development of force fields for the calculation of thermodynamic properties: acetonitrile as a case study, Mol. Simul. 39 (2013) 109–118.

[12] T. Merker, J. Vrabec, H. Hasse: Engineering molecular models: efficient parameterization procedure and cyclohexanol as case study, Soft Mater. 10 (2012) 3–25.

[13] D. C. Rapaport: The art of molecular dynamics simulation, Cambridge university press, 2004.

[14] D. Frenkel, B. Smit: Understanding molecular simulation: from algorithms to applications, volume 1, Elsevier, 2001.

[15] M. P. Allen, D. J. Tildesley: Computer simulation of liquids, Oxford university press, 2017.

[16] C. Vega, J. L. Abascal: Simulating water with rigid non-polarizable models: a general perspective, Phys. Chem. Chem. Phys. 13 (2011) 19663–19688.

[17] P. A. Clark, A. W. Westerberg: Optimization for design problems having more than one objective, Comput. Chem. Eng. 7 (1983) 259–278.

[18] V. Bhaskar, S. K. Gupta, A. K. Ray: Applications of multiobjective optimization in chemical engineering, Rev. Chem. Eng. 16 (2000) 1–54.

[19] Masuduzzaman, G. P. Rangaiah: Multi-objective optimization applications in chemical engineering, in: G. P. Rangaiah (Ed.), Multi-Objective Optimization. Techniques and Applications in Chemical Engineering, World Scientific, 2009, pp. 27–59.

[20] S. Sharma, G. P. Rangaiah: Multi-objective optimization applications in chemical engineering, in: G. P. Rangaiah, A. Bonilla-Petriciolet (Eds.), Multi-Objective optimization in chemical engineering: developments and applications, John Wiley and Sons, Ltd., 2013, pp. 35–102.

[21] J. Höller, P. Bickert, P. Schwartz, M. von Kurnatowski, J. Kerber, N. Künzle, H.-M. Lorenz, N. Asprion, S. Blagov, M. Bortz: Parameter estimation strategies in thermodynamics, ChemEngineering 3 (2019) 56.

[22] V. Pareto: Manual of political economy: a critical and variorum edition, OUP Oxford, 2014.

[23] R. Vlennet, C. Fonteix, I. Marc: Multicriteria optimization using a genetic algorithm for determining a pareto set, Int. J. Syst. Sci. 27 (1996) 255–260.

[24] N. Srinivas, K. Deb: Muiltiobjective optimization using nondominated sorting in genetic algorithms, Evol. Comput. 2 (1994) 221–248.

[25] F. Logist, B. Houska, M. Diehl, J. Van Impe: Fast pareto set generation for nonlinear optimal control problems with multiple objectives, Struct. Multidiscip. Optim. 42 (2010) 591–603.

[26] J. I. S. Hernández: Multi-objective optimization in mixed integer problems: With application to the beam selection optimization problem in IMRT, mbv, Mensch-und-Buch-Verlag, 2012.

[27] S. Stephan, M. T. Horsch, J. Vrabec, H. Hasse: Molmod–an open access database of force fields for molecular simulations of fluids, Mol. Simul. 45 (2019) 806–814.

[28] M. G. Martin, J. I. Siepmann: Transferable potentials for phase equilibria. 1. united-atom description of n-alkanes, J. Phys. Chem. B 102 (1998) 2569–2577.

[29] K. Stöbener, P. Klein, M. Horsch, K.-H. Küfer, H. Hasse: Parametrization of two-center lennard-jones plus point-quadrupole force field models by multicriteria optimization, Fluid Phase Equilib. 411 (2016) 33 – 42.

[30] J. Fischer, B. Saager, M. Bohn, H. Oelschläger, J. M. Haile: Specific heat of simple liquids, Mol. Phys. 62 (1987) 1175–1185.

[31] B. Saager, J. Fischer: Prediction of thermodynamic properties for fluid oxygen with md simulations, Fluid Phase Equilib. 66 (1991) 103–111.

[32] W. Wagner, A. Pruß: The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use, J. Phys. Chem. Ref. Data 31 (2002) 387–535.

[33] E. W. Lemmon, I. Bell, M. L. Huber, M. O. McLinden: NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version 10.0, National Institute of Standards and Technology, 2018. doi:https://dx.doi.org/10.18434/T4JS3C.

[34] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans: Springer Netherlands, Dordrecht, 1981, pp. 331–342.

[35] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma: The missing term in effective pair potentials, J. Phys. Chem. 91 (1987) 6269–6271.

[36] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein: Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.

[37] J. L. F. Abascal, C. Vega: A general purpose model for the condensed phases of water: TIP4P/2005, J. Chem. Phys. 123 (2005) 234505.

[38] S. Izadi, R. Anandakrishnan, A. V. Onufriev: Building water models: A different approach, J. Phys. Chem. Lett. 5 (2014) 3863–3871.

[39] Y.-L. Huang, T. Merker, M. Heilig, H. Hasse, J. Vrabec: Molecular modeling and simulation of vapor-liquid equilibria of ethylene oxide, ethylene glycol, and water as well as their binary mixtures, Ind. Eng. Chem. Res. 51 (2012) 7428–7440.

[40] Y. M. Muñoz-Muñoz, M. Llano-Restrepo: Vapor–liquid equilibria for the binary systems ethylene+water, ethylene+ethanol, and ethanol+water, and the ternary system ethylene+water+ethanol from gibbs-ensemble molecular simulation, Fluid Phase Equilib. 394 (2015) 1 – 11.

[41] K. Stöbener, P. Klein, S. Reiser, M. Horsch, K.-H. Küfer, H. Hasse: Multicriteria optimization of molecular force fields by Pareto approach, Fluid Phase Equilib. 373 (2014) 100–108.

[42] S. Werth, K. Stöbener, P. Klein, K.-H. Küfer, M. Horsch, H. Hasse: Molecular modelling and simulation of the surface tension of real quadrupolar fluids, Chem. Eng. Sci. 121 (2015) 110–117.

[43] S. Werth, K. Stöbener, M. Horsch, H. Hasse: Simultaneous description of bulk and interfacial properties of fluids by the Mie potential, Mol. Phys. 115 (2017) 1017–1030.

[44] R. Fingerhut, G. Guevara-Carrion, I. Nitzke, D. Saric, J. Marx, K. Langenbach, S. Prokopev, D. Celnỳ, M. Bernreuther, S. Stephan, et al.: ms2: A molecular simulation tool for thermodynamic properties, release 4.0, Computer Physics Communications 262 (2021) 107860.

[45] J. Vrabec, H. Hasse: Grand equilibrium: vapour-liquid equilibria by a new molecular simulation method, Mol. Phys. 100 (2002) 3375–3383.

[46] V. Shen, D. Siderius, W. Krekelberg, H. Hatch: NIST Standard Reference Simulation Website, NIST Standard Reference Database Number 173, National Institute of Standards and Technology, Gaithersburg MD, 20899, 2020. doi:https://doi.org/10.18434/T4M88Q, retrieved 2020-02-07.

[47] G. Raabe, R. J. Sadus: Influence of bond flexibility on the vapor-liquid phase equilibria of water, The Journal of chemical physics 126 (2007) 044701.

[48] R. Sakamaki, A. K. Sum, T. Narumi, K. Yasuoka: Molecular dynamics simulations of vapor/liquid coexistence using the nonpolarizable water models, The Journal of chemical physics 134 (2011) 124708.

[49] C. Vega, J. Abascal, I. Nezbeda: Vapor-liquid equilibria from the triple point up to the critical point for the new generation of TIP4P-like models: TIP4P/Ew, TIP4P/2005, and TIP4P/ice, The Journal of chemical physics 125 (2006) 034503.

[50] M. Schappals, A. Mecklenfeld, L. Kröger, V. Botan, A. Köster, S. Stephan, E. J. García, G. Rutkai, G. Raabe, P. Klein, K. Leonhard, C. W. Glass, J. Lenhard, J. Vrabec, H. Hasse: Round robin study: Molecular simulation of thermodynamic properties from models with internal degrees of freedom, Journal of chemical theory and computation 13 (2017) 4270–4280.

[51] S. Stephan, M. Thol, J. Vrabec, H. Hasse: Thermophysical properties of the lennard-jones fluid: Database and data assessment, J. Chem. Inf. Model. (2019).

[52] I. Nezbeda: Simulations of vapor–liquid equilibria: Routine versus thoroughness, J. Chem. Eng. Data 61 (2016) 3964–3969.

[53] I. Shvab, R. J. Sadus: Atomistic water models: Aqueous thermodynamic properties from ambient to supercritical conditions, Fluid Phase Equilib. 407 (2016) 7 – 30.

[54] M. Kohns, M. Schappals, M. Horsch, H. Hasse: Activities in aqueous solutions of the alkali halide salts from molecular simulation, J. Chem. Eng. Data 61 (2016) 4068–4076.

[55] A. Kulkarni, M. Bortz, K.-H. Küfer, M. Kohns, H. Hasse: Multicriteria optimization of molecular models of water using a reduced units approach, Journal of chemical theory and computation 16 (2020) 5127–5138.

[56] A. Kulkarni, R. Fingerhut, M. Kohns, H. Hasse, J. Vrabec: Correction to "molecular modeling and simulation of vapor–liquid equilibria of ethylene oxide, ethylene glycol, and water as well as their binary mixtures", Industrial & Engineering Chemistry Research 59 (2020) 20232–20234.

[57] I. N. Tsimpanogiannis, O. A. Moultos, L. F. Franco, M. B. d. M. Spera, M. Erdős, I. G. Economou: Self-diffusion coefficient of bulk and confined water: a critical review of classical molecular simulation studies, Molecular Simulation 45 (2019) 425–453.

[58] E. Forte, A. Kulkarni, J. Burger, M. Bortz, K.-H. Küfer, H. Hasse: Multi-criteria optimization for parametrizing excess gibbs energy models, Fluid Phase Equilibria 522 (2020) 112676.

[59] E. Forte, J. Burger, K. Langenbach, H. Hasse, M. Bortz: Multi-criteria optimization for parameterization of saft-type equations of state for water, AIChE Journal 64 (2018) 226–237.

[60] G. Rutkai, A. Köster, G. Guevara-Carrion, T. Janzen, M. Schappals, C. W. Glass, M. Bernreuther, A. Wafai, S. Stephan, M. Kohns, S. Reiser, S. Deublein, M. Horsch, H. Hasse, J. Vrabec: ms2: A molecular simulation tool for thermodynamic properties, release 3.0, Comput. Phys. Commun. 221 (2017) 343 – 351.

[61] S. Deublein, B. Eckl, J. Stoll, S. V. Lishchuk, G. Guevara-Carrion, C. W. Glass, T. Merker, M. Bernreuther, H. Hasse, J. Vrabec: ms2: A molecular simulation tool for thermodynamic properties, Computer Physics Communications 182 (2011) 2350–2367.

[62] C. W. Glass, S. Reiser, G. Rutkai, S. Deublein, A. Köster, G. Guevara-Carrion, A. Wafai, M. Horsch, M. Bernreuther, T. Windmann, et al.: ms2: A molecular simulation tool for thermodynamic properties, new version release, Computer Physics Communications 185 (2014) 3302–3306.

[63] M. S. Green: Markoff random processes and the statistical mechanics of time-dependent phenomena. ii. irreversible processes in fluids, The Journal of Chemical Physics 22 (1954) 398–413.

[64] R. Kubo: Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems, Journal of the Physical Society of Japan 12 (1957) 570–586.

[65] J. G. Kirkwood: The dielectric polarization of polar liquids, The Journal of Chemical Physics 7 (1939) 911–919.

[66] A. Lotfi, J. Vrabec, J. Fischer: Vapour liquid equilibria of the lennard-jones fluid from the npt plus test particle method, Molecular Physics 76 (1992) 1319–1333.

[67] C. Antoine: Vapor pressure: a new relationship between pressure and temperature, CR Acad. Sci 107 (1888) 681–685.

[68] M. L. Huber, R. A. Perkins, A. Laesecke, D. G. Friend, J. V. Sengers, M. J. Assael, I. N. Metaxa, E. Vogel, R. Mareš, K. Miyagawa: New international formulation for the viscosity of h 2 o, Journal of Physical and Chemical Reference Data 38 (2009) 101–125.

[69] K. Yoshida, N. Matubayasi, M. Nakahara: Self-diffusion coefficients for water and organic solvents at high temperatures along the coexistence curve, The Journal of chemical physics 129 (2008) 214501.

[70] M. Ehrgott: Multicriteria optimization, volume 491, Springer Science & Business Media, Berlin, Heidelberg, 2005.

[71] A. Shirazi, B. Najafi, M. Aminyavari, F. Rinaldi, R. A. Taylor: Thermal–economic–environmental analysis and multi-objective optimization of an ice thermal energy storage system for gas turbine cycle inlet air cooling, Energy 69 (2014) 212–226.

[72] B. Najafi, A. Shirazi, M. Aminyavari, F. Rinaldi, R. A. Taylor: Exergetic, economic and environmental analyses and multi-objective optimization of an SOFC-gas turbine hybrid cycle coupled with an MSF desalination system, Desalination 334 (2014) 46–59.

[73] T. Ganesan, I. Elamvazuthi, P. Vasant: Multiobjective design optimization of a nano-CMOS voltage-controlled oscillator using game theoretic-differential evolution, Applied Soft Computing 32 (2015) 293–299.

[74] G. P. Rangaiah, A. Petriciolet: Multi-objective optimization in chemical engineering, Developments and applications/edited by Gade Pandu Rangaiah, Adrián Bonilla-Petriciolet (2013).

[75] M. von Kurnatowski, M. Bortz, A. Scherrer, A. Hoffmann, H.-M. Lorenz, M. Caraucan, T. Grützner, N. Künzle, K.-H. Küfer: Multi-criteria optimization of an industrial world-scale process, Chemie Ingenieur Technik 89 (2017) 1471–1478.

[76] M. Bortz, J. Burger, N. Asprion, S. Blagov, R. Böttcher, U. Nowak, A. Scheithauer, R. Welke, K.-H. Küfer, H. Hasse: Multi-criteria optimization in chemical process design and decision support by navigation on pareto sets, Computers & chemical engineering 60 (2014) 354–363.

[77] P. Rehner, J. Gross: Multiobjective optimization of PCP-SAFT parameters for water and alcohols using surface tension data, Journal of Chemical & Engineering Data 65 (2020) 5698–5707.

[78] E. J. Graham: Development of advanced thermodynamic models for $CO_2$ absorption: From numerical methods to process modelling (2020).

[79] E. Forte, E. von Harbou, J. Burger, N. Asprion, M. Bortz: Optimal design of laboratory and pilot-plant experiments using multiobjective optimization, Chemie Ingenieur Technik 89 (2017) 645–654.

[80] O. B. Augusto, F. Bennis, S. Caro: Multiobjective optimization involving quadratic functions, Journal of Optimization 2014 (2014) 406092.

[81] M. Bortz, J. Burger, E. von Harbou, M. Klein, J. Schwientek, N. Asprion, R. Böttcher, K.-H. Küfer, H. Hasse: Efficient approach for calculating pareto boundaries under uncertainties in chemical process design, Industrial & Engineering Chemistry Research 56 (2017) 12672–12681.

[82] J. Burger, N. Asprion, S. Blagov, R. Böttcher, U. Nowak, M. Bortz, R. Welke, K.-H. Küfer, H. Hasse: Multi-objective optimization and decision support in process engineering–implementation and application, Chemie Ingenieur Technik 86 (2014) 1065–1072.

[83] W.-Y. Chiu, J.-T. Hsieh, C.-M. Chen: Pareto optimal demand response based on energy costs and load factor in smart grid, IEEE Transactions on Industrial Informatics 16 (2019) 1811–1822.

[84] Y. Matviychuk, E. Steimers, E. von Harbou, D. J. Holland: Bayesian approach for automated quantitative analysis of benchtop NMR data, Journal of Magnetic Resonance 319 (2020) 106814.

[85] J. Branke, K. Deb, H. Dierolf, M. Osswald: Finding knees in multi-objective optimization, in: International conference on parallel problem solving from nature, Springer, 2004, pp. 722–731.

# Appendix

# A Supporting Information for Chapter 3

## A.1 Three-objective Pareto study

The Pareto sets obtained using Huang as starting point, in two- and three-dimensional objective space as well as their pictures in the parameter space, are presented in the main text. Results for all other studied models are shown here. Figures A.1-A.6 indicate the dependence of saturated liquid density $\rho^{\mathrm{liq}}$, saturated vapour pressure $p^{\mathrm{s}}$ and enthalpy of vaporisation $\Delta h^{\mathrm{vap}}$, of the scaled variants of the respective models obtained using reduced units method, on the Lennard-Jones parameters $\sigma$ and $\varepsilon$. The three-dimensional Pareto set in objective space for each model is also shown. Qualitatively, the results for all the models are analogous to those for the Huang model presented and discussed in the main text.

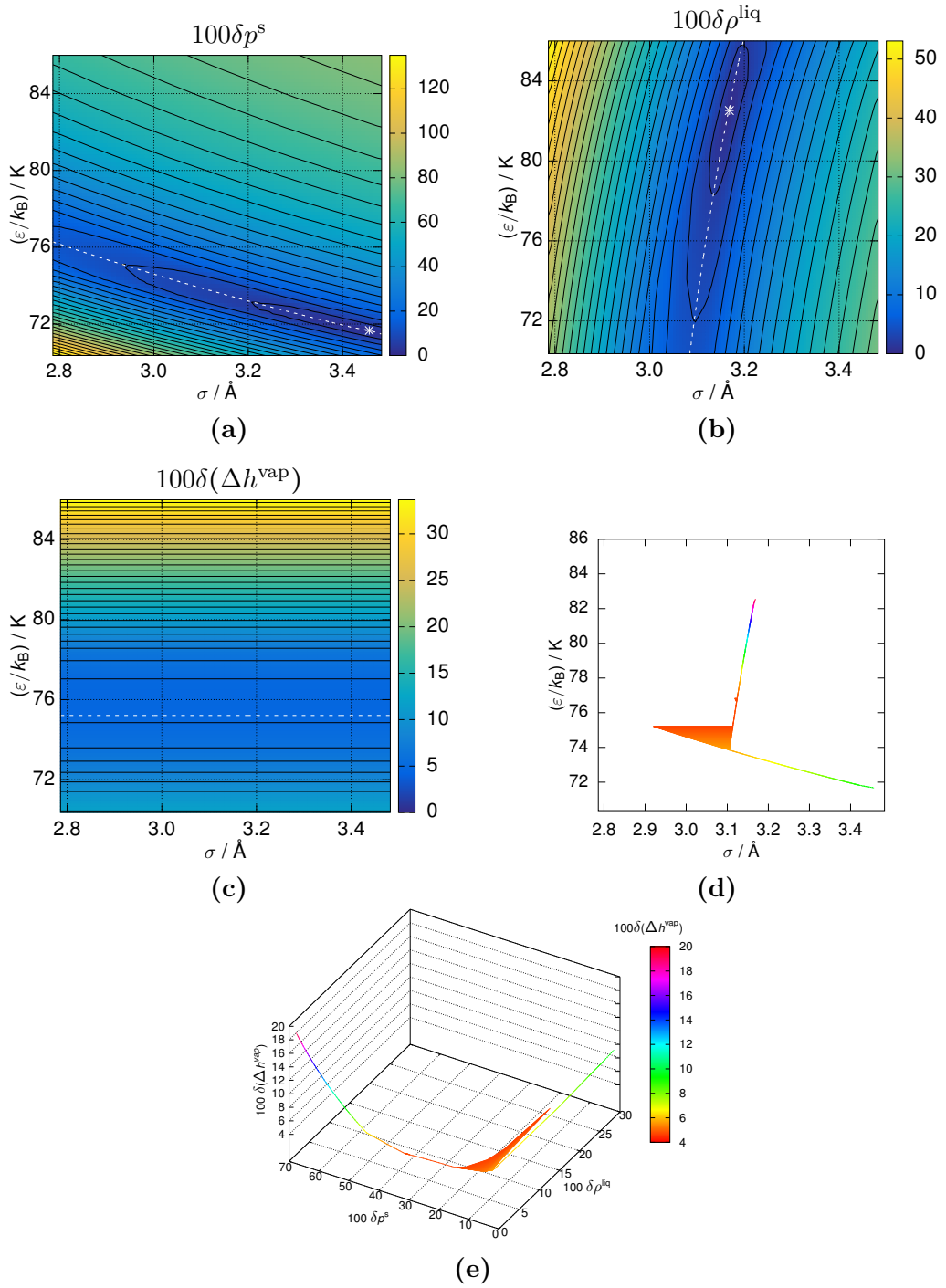



**Figure A.1:** (a), (b) and (c): Dependence of deviations of vapour pressure, saturated liquid density and enthalpy of vaporisation respectively on the LJ parameters $\sigma$ and $\varepsilon$ for for scaled variants of SPC/E model. The minima for single property optimisation are marked by white stars and the paths of lowest ascent/descent through the minima, i.e. the 'valleys' in the 'landscape', of the respective properties are indicated by white dashed lines. There is no distinct minimum for $\delta(\Delta h^{\mathrm{vap}})$, as the numbers for this objective do not depend on $\varepsilon$. (d) The picture of the Pareto set for MCO with the three objectives, using the reduced units method with SPC/E as starting point. (e) The Pareto set in objective space.

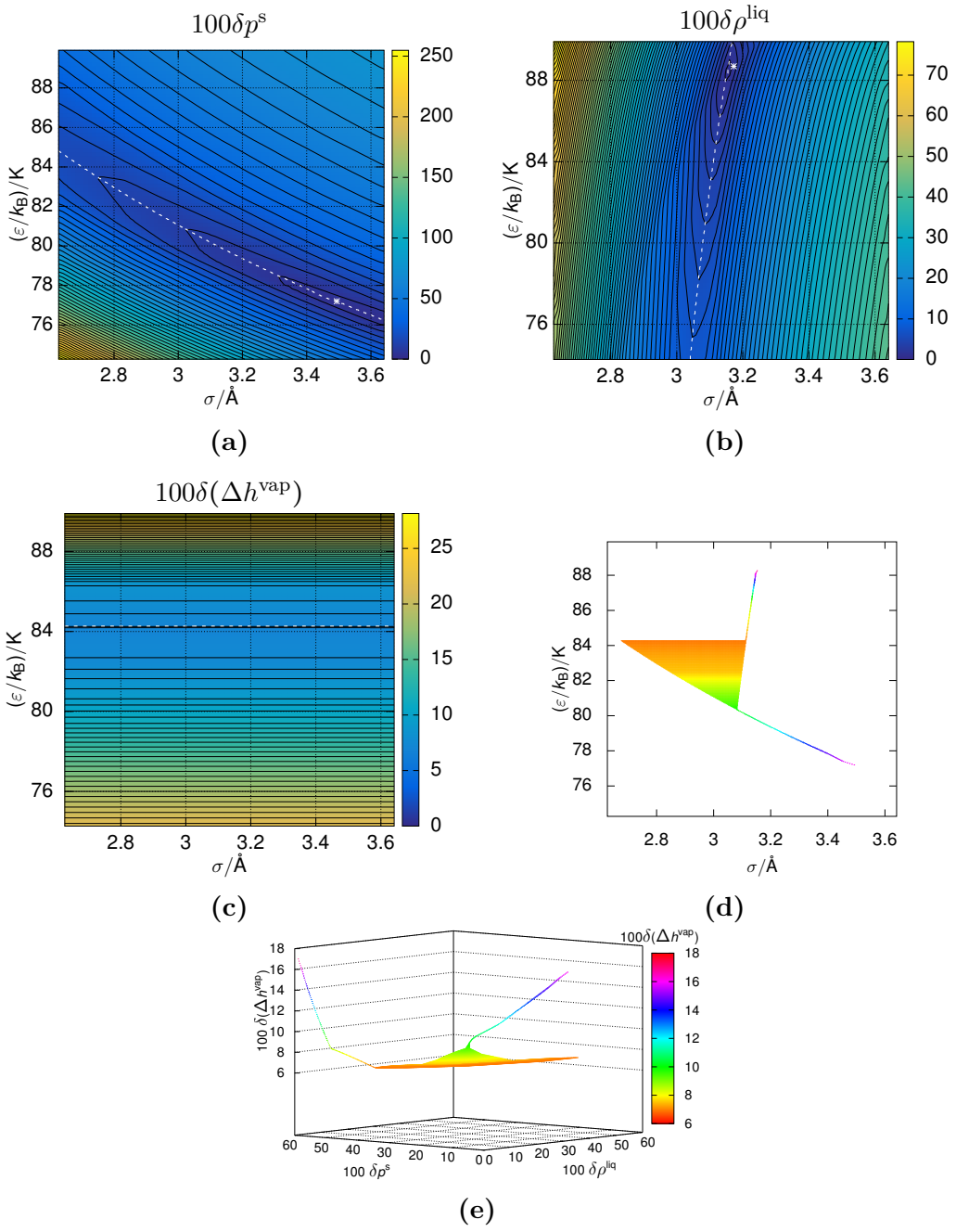



**Figure A.2:** (a), (b) and (c): Dependence of deviations of vapour pressure, saturated liquid density and enthalpy of vaporisation respectively on the LJ parameters $\sigma$ and $\varepsilon$ for for scaled variants of SPC model. The minima for single property optimisation are marked by white stars and the paths of lowest ascent/descent through the minima, i.e. the 'valleys' in the 'landscape', of the respective properties are indicated by white dashed lines. There is no distinct minimum for $\delta(\Delta h^{\mathrm{vap}})$, as the numbers for this objective do not depend on $\varepsilon$. (d) The picture of the Pareto set for MCO with the three objectives, using the reduced units method with SPC as starting point. (e) The Pareto set in objective space.

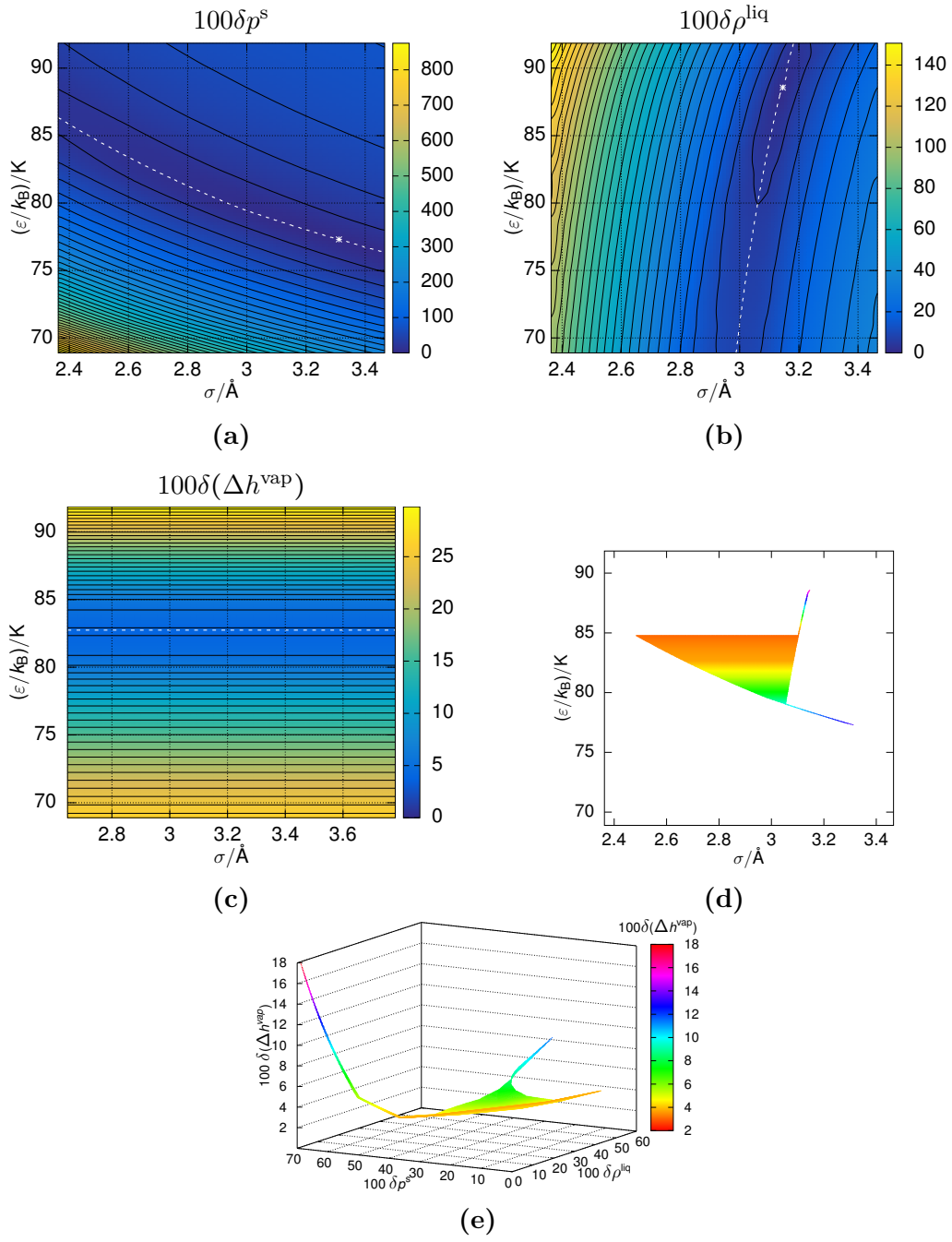



**Figure A.3:** (a), (b) and (c): Dependence of deviations of vapour pressure, saturated liquid density and enthalpy of vaporisation respectively on the LJ parameters $\sigma$ and $\varepsilon$ for for scaled variants of TIP3P model. The minima for single property optimisation are marked by white stars and the paths of lowest ascent/descent through the minima, i.e. the 'valleys' in the 'landscape', of the respective properties are indicated by white dashed lines. There is no distinct minimum for $\delta(\Delta h^{\mathrm{vap}})$, as the numbers for this objective do not depend on $\varepsilon$. (d) The picture of the Pareto set for MCO with the three objectives, using the reduced units method with TIP3P as starting point. (e) The Pareto set in objective space.

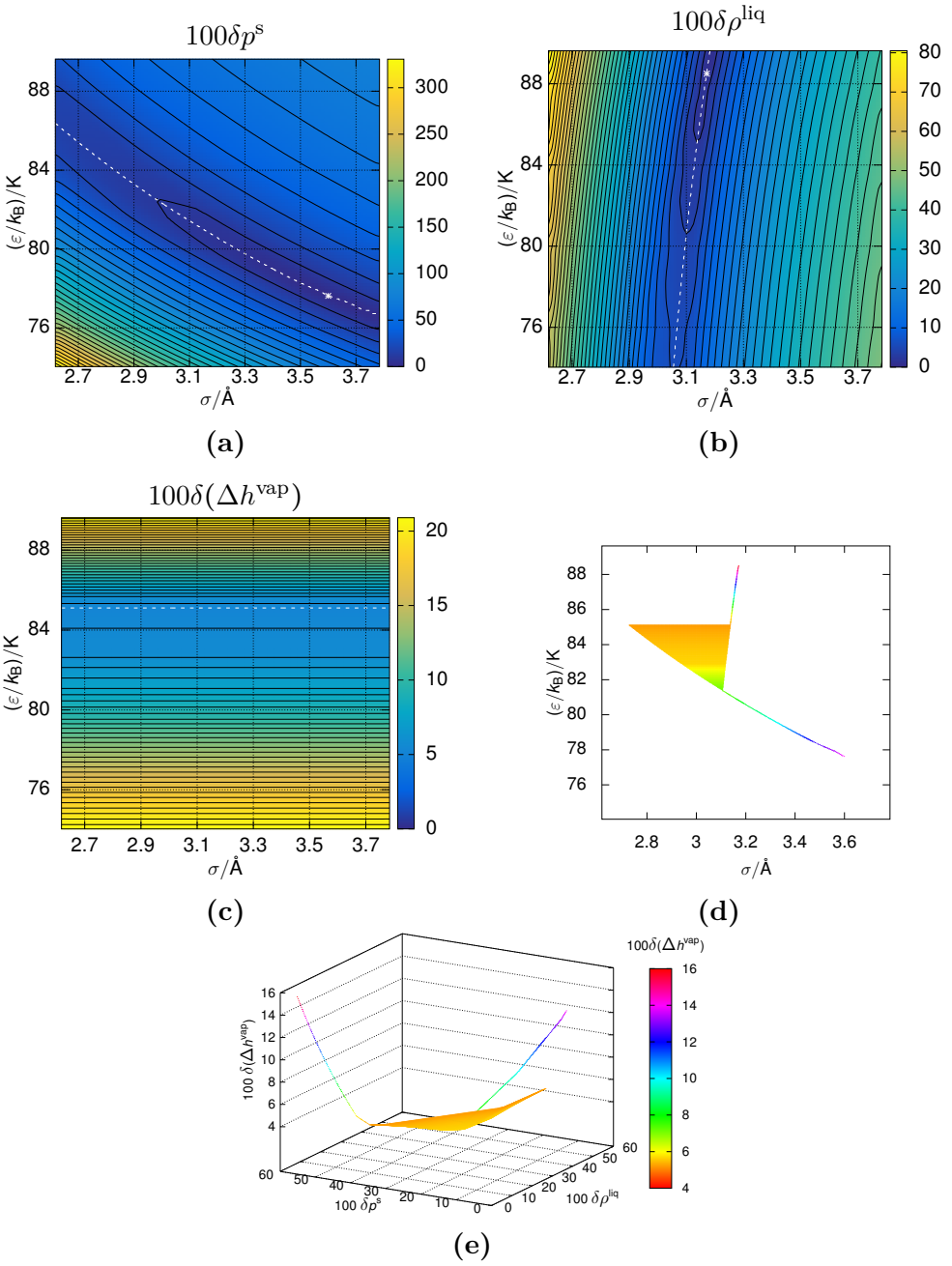



**Figure A.4:** (a), (b) and (c): Dependence of deviations of vapour pressure, saturated liquid density and enthalpy of vaporisation respectively on the LJ parameters $\sigma$ and $\varepsilon$ for for scaled variants of TIP4P model. The minima for single property optimisation are marked by white stars and the paths of lowest ascent/descent through the minima, i.e. the 'valleys' in the 'landscape', of the respective properties are indicated by white dashed lines. There is no distinct minimum for $\delta(\Delta h^{\mathrm{vap}})$, as the numbers for this objective do not depend on $\varepsilon$. (d) The picture of the Pareto set for MCO with the three objectives, using the reduced units method with TIP4P as starting point. (e) The Pareto set in objective space.

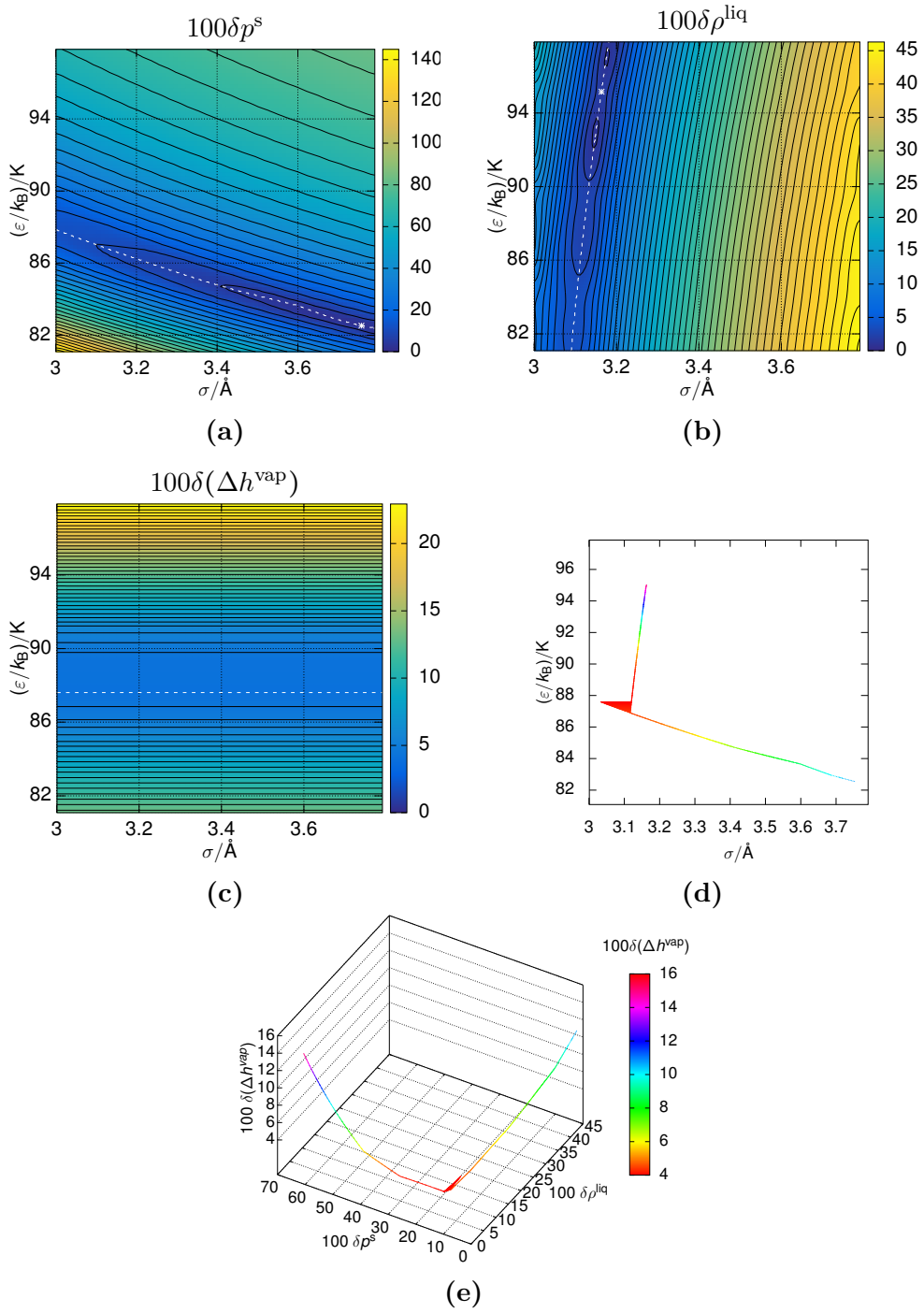



**Figure A.5:** (a), (b) and (c): Dependence of deviations of vapour pressure, saturated liquid density and enthalpy of vaporisation respectively on the LJ parameters $\sigma$ and $\varepsilon$ for for scaled variants of TIP4P/2005 model. The minima for single property optimisation are marked by white stars and the paths of lowest ascent/descent through the minima, i.e. the 'valleys' in the 'landscape', of the respective properties are indicated by white dashed lines. There is no distinct minimum for $\delta(\Delta h^{\mathrm{vap}})$, as the numbers for this objective do not depend on $\varepsilon$. (d) The picture of the Pareto set for MCO with the three objectives, using the reduced units method with TIP4P/2005 as starting point. (e) The Pareto set in objective space.

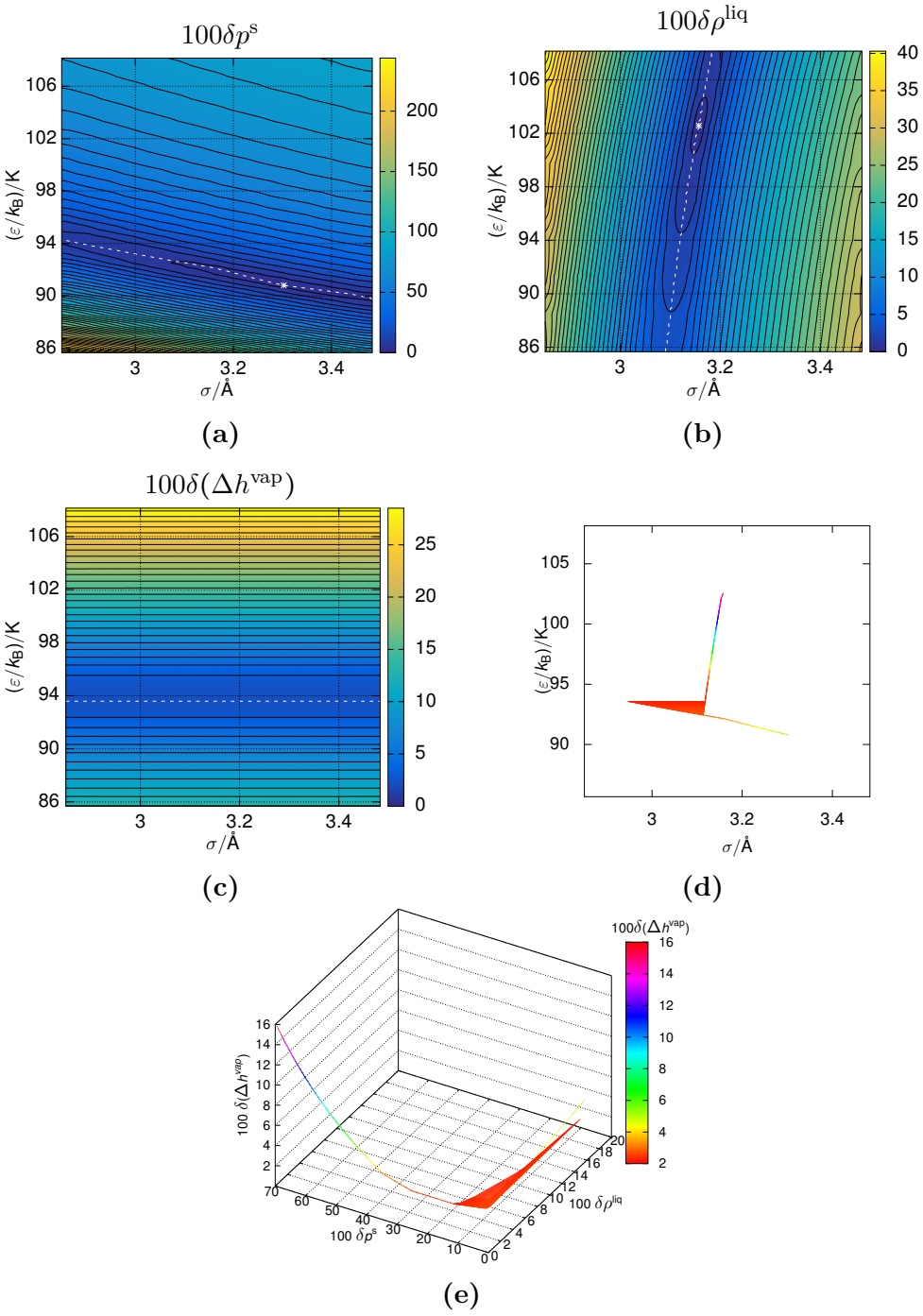
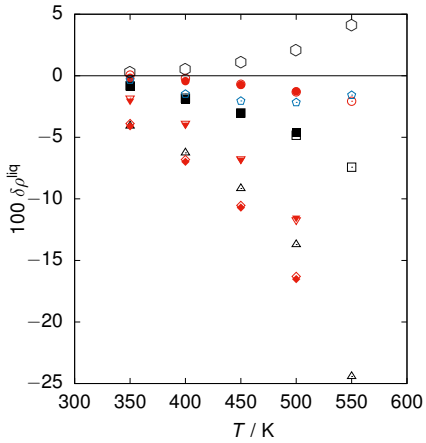


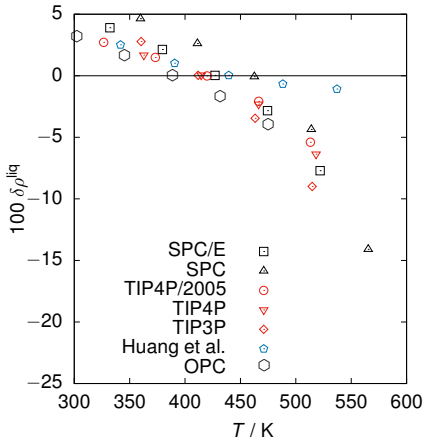

**Figure A.6:** (a), (b) and (c): Dependence of deviations of vapour pressure, saturated liquid density and enthalpy of vaporisation respectively on the LJ parameters $\sigma$ and $\varepsilon$ for for scaled variants of OPC model. The minima for single property optimisation are marked by white stars and the paths of lowest ascent/descent through the minima, i.e. the 'valleys' in the 'landscape', of the respective properties are indicated by white dashed lines. There is no distinct minimum for $\delta(\Delta h^{\mathrm{vap}})$, as the numbers for this objective do not depend on $\varepsilon$. (d) The picture of the Pareto set for MCO with the three objectives, using the reduced units method with OPC as starting point. (e) The Pareto set in objective space.

## A.2 Comparison between optimised and original models

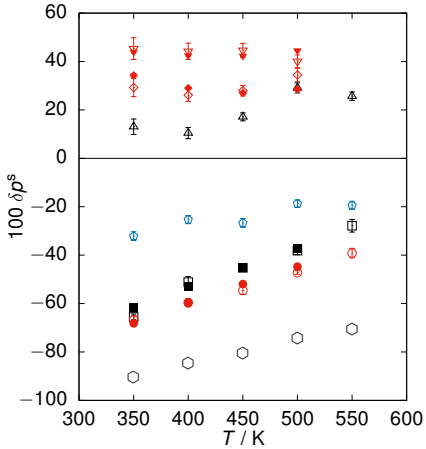
The deviations in VLE properties, with respect to IAPWS-95 EoS, for Pareto knee models obtained from two objective optimisation (considering of $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$) are presented here, along with the comparison to the deviations obtained from the original models. The optimised models have low deviations for all three VLE properties, in spite of the fact that they were selected from a two-objective Pareto front. The comparison of deviations of the VLE properties of original models and the optimised models is provided in Figure A.7. The optimised models show significant improvement in the representation of $p^{\mathrm{s}}$, while the improvements in the representation of $\rho^{\mathrm{liq}}$ are comparatively minor. The representation of $\Delta h^{\mathrm{vap}}$ improves significantly for some models, while the improvement for some models is minor.



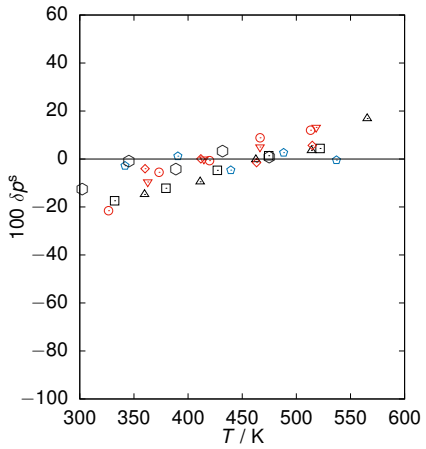

**(a)** $\delta\rho^{\mathrm{liq}}$ for original models



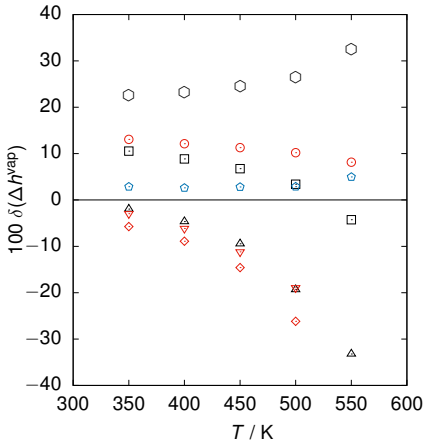

**(b)** $\delta\rho^{\mathrm{liq}}$ for optimised models



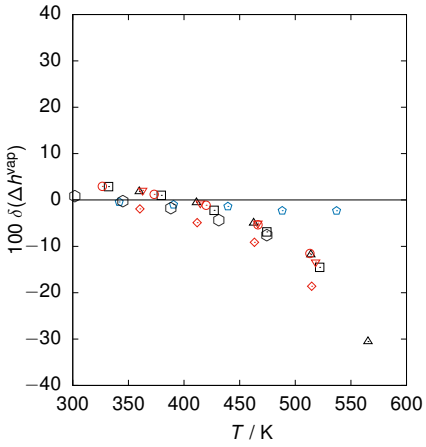

**(c)** $\delta p^{\mathrm{s}}$ for original models




**(d)** $\delta p^{\mathrm{s}}$ for optimised models




**(e)** $\delta(\Delta h^{\mathrm{vap}})$ for original models



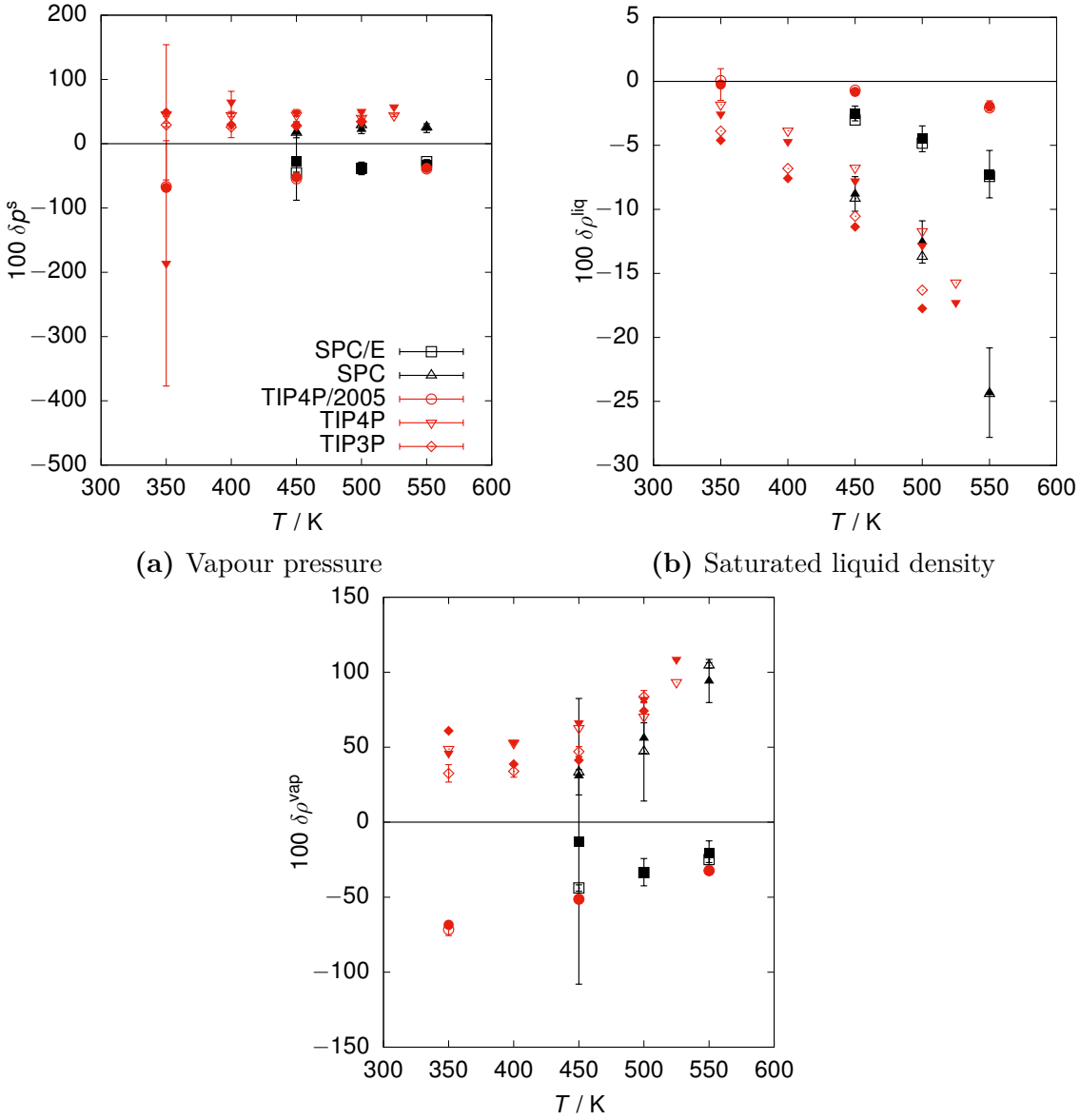

**(f)** $\delta(\Delta h^{\mathrm{vap}})$ for optimised models

**Figure A.7:** The comparison of deviations in the three VLE properties of seven original models and their variants at the Pareto knee. Panels (a), (c) and (e) are reproductions from Figures 3-5 from the main text.

## A.3 Comparison of the simulation results to literature data

In the main manuscript, the simulation results of the present work for SPC/E, TIP3P, TIP4P and TIP4P/2005 are compared to the literature data from the NIST Reference Simulations Database [46]. In Fig. A.8, we show a similar comparison of the simulation results from the present work with further literature sources for these four models and for SPC [47–49]. There, only the temperatures at which literature data are available are shown. The corresponding numerical values are given in Table A.1.

Fig. A.8 and Table A.1 show a good agreement between our data and the literature data [47–49], apart from the liquid density data by Sakamaki et al [48]. For the liquid densities, a slight and apparently constant offset is observed for all temperatures. For the vapour densities, some points are in good agreement while others disagree. Sakamaki et al. do not provide statistical uncertainties for their density data. However, from the large uncertainties in vapour pressures that they report, we assume that also the densities are rather uncertain. Hence, we consider their and our data sets to be in mutual agreement.




**(a)** Vapour pressure

**(b)** Saturated liquid density

**(c)** Saturated vapour density

**Figure A.8:** Comparison of the simulation results from the present work to literature data [47–49]. The comparison is shown as the deviation from the IAPWS-95 EoS [32, 33]. Open symbols show the simulation results from the present work and filled symbols show the literature data. In the plots, the temperature for the data by Sakamaki et al. [48] are raised by 2 K to match our temperatures.

| $T$/K | Present work | | | Literature data | | | |
|---|---|---|---|---|---|---|---|
| | $p^{s}$/bar | $\rho^{liq}$/kg m$^{-3}$ | $\rho^{vap}$/kg m$^{-3}$ | $p^{s}$/bar | $\rho^{liq}$/kg m$^{-3}$ | $\rho^{vap}$/kg m$^{-3}$ | Reference |
| | | | | SPC | | | |
| 450 | 11.103±0.20 | 808.17±0.1 | 6.413±0.113 | 11.02±0.99 | 812±11 | 6.3±0.8 | |
| 500 | 31.571±0.59 | 716.63±0.3 | 19.421±0.364 | 32.95±2.18 | 727±12 | 20.6±2.1 | Raabe and Sadus [47] |
| 550 | 80.627±1.45 | 576.62±2.5 | 64.371±1.154 | 76.67±5.22 | 572±20 | 61.1±8.8 | |
| | | | | SPC/E | | | |
| 450 | 5.010±0.01 | 863.18±0.1 | 2.701±0.057 | 6.73±4.05 | 868±5 | 4.2±4.0 | |
| 500 | 16.140±0.38 | 792.31±0.3 | 8.765±0.207 | 16.25±1.59 | 794±7 | 8.8±0.8 | Raabe and Sadus [47] |
| 550 | 40.485±0.81 | 698.12±0.4 | 23.637±0.467 | 41.02±3.33 | 701±13 | 24.9±2.1 | |
| | | | | TIP3P | | | |
| 350 | 0.553±0.01 | 935.62±0.1 | 0.355±0.010 | 0.57±0.60[a] | 930[a, b] | 0.42[a, b] | |
| 400 | 3.183±0.06 | 873.13±0.1 | 1.913±0.040 | 3.00±0.61[a] | 868[a, b] | 1.9[a, b] | Sakamaki et al. [48] |
| 450 | 11.336±0.19 | 795.95±0.2 | 6.683±0.112 | 11.49±0.61[a] | 791[a, b] | 6.8[a, b] | |
| 500 | 33.250±0.66 | 694.75±0.4 | 21.817±0.432 | 34.09±0.60[a] | 686[a, b] | 23[a, b] | |
| | | | | TIP4P | | | |
| 350 | 0.587±0.02 | 955.53±0.1 | 0.372±0.012 | −0.33±0.63[a] | 950[a, b] | 0.38[a, b] | |
| 400 | 3.535±0.08 | 900.66±0.1 | 2.072±0.049 | 3.81±0.64[a] | 895[a, b] | 2.1[a, b] | |
| 450 | 13.290±0.25 | 829.47±0.1 | 7.692±0.142 | 13.20±0.64[a] | 823[a, b] | 8.0[a, b] | Sakamaki et al. [48] |
| 500 | 38.856±0.73 | 734.60±0.4 | 24.682±0.468 | 38.18±0.62[a] | 727[a, b] | 24[a, b] | |
| 525 | 59.074±1.12 | 670.99±0.7 | 39.855±0.755 | 62.29±0.60[a] | 661[a, b] | 43[a, b] | |
| | | | | TIP4P/2005 | | | |
| 350 | 0.118±0.01 | 973.46±0.1 | 0.074±0.003 | 0.131[b] | 971.3±1.2 | 0.0823±0.0002 | |
| 450 | 4.454±0.09 | 883.88±0.1 | 2.342±0.047 | 4.46[b] | 883±2 | 2.342±0.010 | Vega et al. [49] |
| 550 | 38.606±0.69 | 739.25±0.3 | 21.313±0.378 | 38.01[b] | 741±3 | 21.3±0.3 | |
| | | | | Huang | | | |
| 350 | 0.268±0.01 | 970.32±0.1 | 0.167±0.007 | 0.46±0.02 | 971.24±0.5 | 0.304±0.013 | |
| 450 | 7.419±0.13 | 871.78±0.1 | 3.855±0.067 | 8.9±0.2 | 880.08±0.9 | 4.702±0.090 | Huang et al. [39] |
| 550 | 49.450±0.87 | 743.57±0.2 | 25.042±0.450 | 58.0±0.6 | 752.17±1.2 | 28.285±0.360 | |

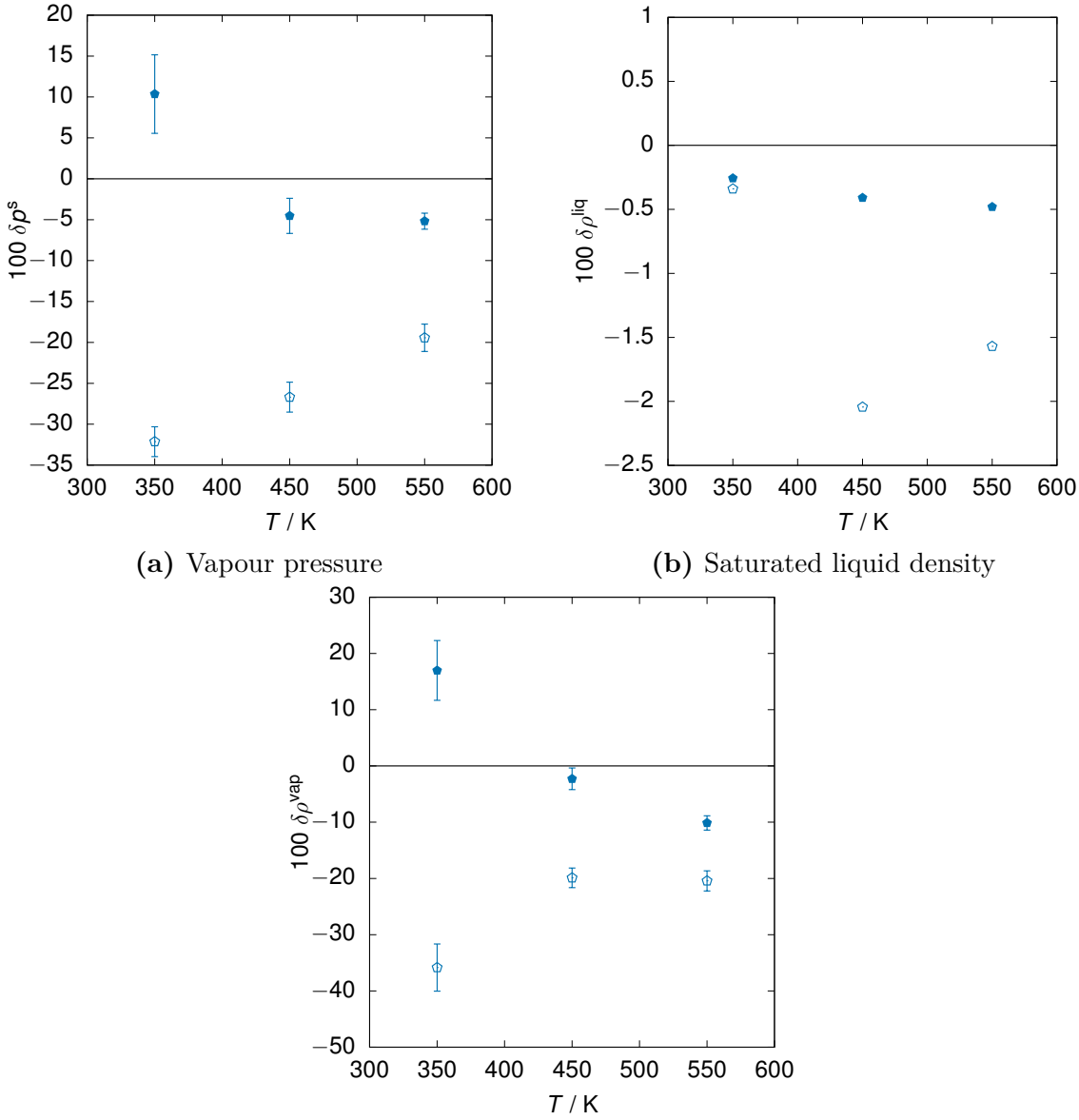
[a] The temperatures for the data are raised by 2 K to match our temperatures.

[b] Statistical uncertainty is not reported.

**Table A.1:** VLE simulation results obtained in the present work for different water models along with their comparison to select literature data.

## A.4 Comparison of simulation results to literature data for the Huang model

For all the models considered in the previous section, the simulation results from the present work and the literature agree well. For the Huang model, however, the original data reported in the original paper from our group [39] differ considerably from the results of the present work. The numerical data are shown in Table A.1. In Fig. A.9, the deviation plots are shown in the same manner as for the other models. Only the liquid density at the lowest temperature is in good agreement to the data from the original paper [39]. For the remaining two state points, the deviations are around 2%. For the vapour pressures and vapour densities, gross deviations are observed. Obviously, this must be the case for both these properties as these are strongly related. The good agreement obtained with various literature sources for all the other models (see main text and Section A.3 above) confirms our computational approach. Hence, we trust our present data and have to rate the original data [39] inaccurate. Unfortunately, we can only speculate about possible sources of error in the original paper [39]. Hence, we suggest that readers use the data presented here in favour of those from the original paper [39].




**(a)** Vapour pressure

**(b)** Saturated liquid density

**(c)** Saturated vapour density

**Figure A.9:** Comparison of the simulation results from the present work to literature data [39] for the Huang model. The comparison is shown as the deviation from the IAPWS-95 EoS [32, 33]. Open symbols show the simulation results from the present work and filled symbols show the simulation results from the literature.

# B Supporting Information for Chapter 4

The simulation results for $p^{\mathrm{s}}$ and $\rho^{\mathrm{liq}}$ and the correlation coefficients for all 96 grid models, the parameters of the empirical correlations to the simulation results, as well as the simulation results for $D$, $\eta$ and $\epsilon_{\mathrm{r}}$ are available online and can be found under the following link: `https://rptu.de/s/kulkarni`.

# C Appendix for Chapter 5

## C.1 Mathematical analysis for arbitrary values of $M$ and $N$

Eq. (30) was only derived in the main text for the case of two objective functions and two parameters ($M = N = 2$). The result is generalised here to arbitrary $N$ and $M$. The treatment is very similar to that shown in Sec. 5.2.3 for $M = N = 2$.

Consider $N$ quadratic objective functions $f_i \ \forall \ i = 1, \ldots, N$, as defined in Eq. (21). In the following, a line of arguments will be made for the objective $f_1$, but as in Sec. 5.2.3, similar arguments can be made for all other objectives. A coordinate transformation is applied similar to that explained in Sec. 5.2, such that the non-diagonal elements of $\mathbf{H}_1$ are equal to zero. Thus, $\mathbf{H}_1$ can be written as:

$$\mathbf{H}_1 = \begin{pmatrix} \xi_1 & 0 & \ldots & 0 \\ 0 & \xi_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \xi_M \end{pmatrix} \tag{C.1}$$

Without loss of generality, one of the eigenvalues of $\mathbf{H}_1$ can be scaled to 1, so that $\xi_1 = 1$. Also, let us assume that one of the eigenvalues is dominated by all the others, i.e. $\xi_2, \xi_3, \ldots, \xi_M \gg 1$. Following Augusto et al. [80] (Eq. (27) in their paper), any point $\mathbf{x}^*$ on the Pareto set can then be calculated as follows:

$$\mathbf{x}^* = \Big( \sum_{i=1}^{N} w_i \mathbf{H}_i \Big)^{-1} \Big( \sum_{j=1}^{N} w_j \mathbf{H}_j \mathbf{x}_j \Big) \tag{C.2}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{C.3}$$

Here, $w_i$ is the weight that corresponds to the objective $f_i$. Each choice of the weights $(w_1, w_2, \ldots, w_N)$ gives a unique point $\mathbf{x}^*$ on the Pareto set, so that the evaluation of Eq. (C.2) as a function of $(w_1, w_2, \ldots, w_N)$ yields the entire Pareto set.

Eq. (C.2) can be rearranged to give:

$$\mathbf{x}^* = \left[\left(\mathbf{I} + \sum_{i=2}^{N} \frac{w_i}{w_1}\mathbf{H}_1^{-1}\mathbf{H}_i\right)^{-1}(w_1\mathbf{H}_1)^{-1}\right]\left(\sum_{j=2}^{N} w_j\mathbf{H}_j\mathbf{x}_j\right) \tag{C.4}$$

Let us consider a point in the vicinity of $\mathbf{x}_1$, so that $w_1 \gg w_2, \ldots, w_N$. Applying a Taylor series expansion in $\frac{w_i}{w_1}$ to the first term in brackets in Eq. (C.4) yields the following approximation:

$$\mathbf{x}^* = \sum_{j=2}^{N} \frac{w_j}{w_1}\mathbf{H}_1^{-1}\mathbf{H}_j\mathbf{x}_j - \sum_{i=2}^{N}\sum_{j=2}^{N} \frac{w_i w_j}{w_1^2}\mathbf{H}^{-1}\mathbf{H}_i\mathbf{H}_1^{-1}\mathbf{H}_j\mathbf{x}_j + \mathcal{O}\left(\frac{\max_i(\{w_i^3\})}{w_1^3}\right) \tag{C.5}$$

Rearranging the above equation results in:

$$\mathbf{x}^* = \sum_{i=2}^{N} \frac{w_i}{w_1}\mathbf{H}_1^{-1}\mathbf{H}_i\mathbf{x}_i + \mathcal{O}\left(\frac{\max_i(\{w_i^2\})}{w_1^2}\right) \tag{C.6}$$

Substituting the values for $\mathbf{x}_i = [x_{1,i}, x_{2,i}, \ldots, x_{M,i}]^\intercal$ and the matrices $\mathbf{H}_i$ and $\mathbf{H}_1^{-1}$ yields:

$$\mathbf{x}^* = \sum_{i=2}^{N} \frac{w_i}{w_1} \begin{pmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \frac{1}{\xi_M} \end{pmatrix} \begin{pmatrix} h_{11,i} & \ldots & h_{1M,i} \\ \vdots & \ddots & \vdots \\ h_{M1,i} & \ldots & h_{MM,i} \end{pmatrix} \begin{pmatrix} x_{1,i} \\ \vdots \\ x_{M,i} \end{pmatrix} \tag{C.7}$$

Further rearrangement results in

$$\mathbf{x}^* = \sum_{i=2}^{N} \begin{pmatrix} \frac{w_i}{w_1} & 0 & \ldots & 0 \\ 0 & \mathcal{O}(\frac{w_i}{w_1}\frac{1}{\xi_2}) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & \ldots & \mathcal{O}(\frac{w_i}{w_1}\frac{1}{\xi_M}) \end{pmatrix} \begin{pmatrix} h_{11,i}x_{1,i} + h_{12,i}x_{2,i} + \cdots + h_{1M,i}x_{M,i} \\ h_{21,i}x_{1,i} + h_{22,i}x_{2,i} + \cdots + h_{2M,i}x_{M,i} \\ \vdots \\ h_{M1,i}x_{1,i} + h_{M2,i}x_{2,i} + \cdots + h_{MM,i}x_{M,i} \end{pmatrix} \tag{C.8}$$

Thus,

$$\mathbf{x}^* = \sum_{i=2}^{N} \begin{pmatrix} \frac{w_i}{w_1}(h_{11,i}x_{1,i} + h_{12,i}x_{2,i} + \cdots + h_{1M,i}x_{M,i}) \\ \mathcal{O}(\frac{w_i}{w_1}\frac{1}{\xi_2}) \\ \vdots \\ \mathcal{O}(\frac{w_i}{w_1}\frac{1}{\xi_M}) \end{pmatrix} \tag{C.9}$$

Eq. (C.9) shows that for $\xi_2, \ldots, \xi_M \gg 1$, all the components of $\mathbf{x}^*$ except the $x_1$-component are small, i.e. the PSPS near the minimum of $f_1$ lies close to the $x_1$-axis.

# Declaration

This dissertation contains material that has been published previously or that is included in submitted publications. In the following, these publications are listed together with a statement on the contributions of the author of the present dissertation.

- A. Kulkarni, M. Bortz, K.-H. Küfer, M. Kohns, H. Hasse: Hierarchical Multicriteria Optimization of Molecular Models of Water, Journal of Chemical Information and Modeling 64 (2024) 5077-5089,
  DOI: 10.1021/acs.jcim.4c00404
  *The author carried out the molecular simulations, performed the post-processing and multi-criteria optimisation, evaluated the results, and wrote the manuscript.*

- A. Kulkarni, M. Kohns, M. Bortz, K.-H. Küfer, H. Hasse: Regularities of Pareto Sets in Low-dimensional Practical Multi-criteria Optimisation Problems: Analysis, Explanation, and Exploitation, Optimization and Engineering 24 (2023) 1611-1632,
  DOI: 10.1007/s11081-022-09746-z.
  *The author contributed to developing the mathematical formulation, prepared the test cases and wrote the manuscript.*

- A. Kulkarni, M. Bortz, K.-H. Küfer, M. Kohns, H. Hasse: Multicriteria optimization of molecular models of water using a reduced units approach, Journal of Chemical Theory and Computation 16 (2020) 5127-5138,
  DOI: 10.1021/acs.jctc.0c00301.
  *The author carried out the molecular simulations, performed post-processing and multi-criteria optimisation, and wrote the corresponding parts of the manuscript.*

Furthermore, the author contributed to the following publications that are relevant for this thesis and cited therein:

- A. Kulkarni, R. Fingerhut, M. Kohns, H. Hasse, J. Vrabec: Correction to "Molecular Modeling and Simulation of Vapor–Liquid Equilibria of Ethylene Oxide, Ethylene Glycol, and Water as Well as their Binary Mixtures", Industrial & Engineering Chemistry Research 59 (2022) 20232-20234,

DOI: 10.1021/acs.iecr.0c04937.
*The author evaluated the results and performed the post-processing.*

- E. Forte, A. Kulkarni, J. Burger, M. Bortz, K.-H. Küfer, H. Hasse: Multi-criteria optimization for parametrizing excess Gibbs energy models, Fluid Phase Equilibria 522 (2020) 112676,
DOI: 10.1016/j.fluid.2020.112676.
*The author wrote parts of the manuscript.*

- A. Kulkarni, E. Garcia, A. Damone, M. Schappals, S. Stephan, M. Kohns, H. Hasse: A Force Field for Poly(oxymethylene) Dimethyl Ethers (OME$n$), Journal of Chemical Theory and Computation 16 (2020) 2517-2528,
DOI: 10.1021/acs.jctc.9b01106.
*The author contributed to the simulations and wrote corresponding parts of the manuscript.*

# Student theses

The following student theses were prepared under the supervision of the author of the present doctoral thesis in the frame of his research:

- M. L. Guay: Pareto optimization of molecular models using the reduced units method. Student thesis, Laboratory of Engineering Thermodynamics (LTD), TU Kaiserslautern (2022).

# Curriculum vitae

| | |
|---|---|
| Name: | Aditya Kulkarni |
| Nationality: | Indian |

**Education**

| | |
|---|---|
| 2015 – 2017 | Master studies at University of Massachusetts Amherst, USA<br>Program: Mechanical and Industrial Engineering<br>Degree: Master of Science in Mechanical Engineering |
| 2009 – 2013 | Bachelor studies at University of Pune, India<br>Program: Mechanical Engineering<br>Degree: Bachelor of Engineering |
| 2007 – 2009 | Maharashtra State Board, India<br>Degree: Higher Secondary School Certificate (Science) |
| 1997 – 2009 | Maharashtra State Board, India<br>Degree: Secondary School Certificate |

**Professional**

| | |
|---|---|
| since 09/2022 | Associate Director, Treasury Model Validation Quantitative Analyst at<br>UBS Business Solutions India Pvt. Ltd., Mumbai, India |
| 10/2017 – 06/2022 | Research Associate at<br>Laboratory of Engineering Thermodynamics (LTD)<br>RPTU Kaiserslautern<br>Prof. Dr.-Ing. Hans Hasse |
| 09/2016 – 05/2017 | Graduate Teaching Assistant at<br>University of Massachusetts Amherst, USA |
| 10/2013 – 07/2015 | Project Assistant at<br>Tata Institute of Fundamental Research, Mumbai, India |