This is the peer-reviewed version of the following article: *Forte E, Burger J, Langenbach K, Hasse H, and Bortz M Multi-criteria optimization for parameterization of SAFT-type equations of state for water. AIChE J. 2018; 64(1):226-237*, which has been published in final form at:

<http://onlinelibrary.wiley.com/doi/10.1002/aic.15857/abstract> with the following DOI: 10.1002/aic.15857. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Multi-criteria optimization for parametrization of SAFT-type equations of state for water

Esther Forte,^a Jakob Burger,^{a,*} Kai Langenbach^a Michael Bortz^b and Hans Hasse^a

Finding appropriate parameter sets for a given equation of state to describe different properties of a certain substance is an optimization problem with conflicting objectives. Such problem is commonly addressed by single-criteria optimization in which the different objectives are lumped into a single goal function. We show how multi-criteria optimization (MCO) can be beneficially used for parameterizing equations of state. The Pareto set which comprises a set of optimal solutions of the multi-criteria optimization problem is determined. As an example, the perturbed-chain statistical associating fluid theory (PC-SAFT) equation of state (EoS) is used and applied to the description of the thermodynamic properties of water, focusing on saturated liquid density and vapor pressure. Different options to describe the molecular nature of water by the PC-SAFT EoS are studied and for all variants, the Pareto sets are determined, enabling a comprehensive assessment. When compared to literature models, Pareto optimization yields improved models.

^a *Laboratory of Engineering Thermodynamics, University of Kaiserslautern, Erwin-Schrödinger-Str. 44, 67663 Kaiserslautern, Germany.* ^b *Fraunhofer Institute for Industrial Mathematics (ITWM), Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany.*

* Corresponding author's email: jakob.burger@mv.uni-kl.de

Introduction

Most applications in chemical engineering require an accurate description of more than one physical property. The development of physical property models for the simultaneous description of different physical properties is, however, not a simple task. The model adjustments to different physical properties are in conflict, as the optimal description of one property will usually not coincide with that for another property. Of course, for a truly ideal model, there would be no such conflict, but in reality, it is there for many reasons, among which is the unavoidable simplification of the real world by the model.

Already the definition of optimality is not trivial in the case of conflicting objectives. Leaving aside the problem of the ambiguity of defining goal functions and assuming that these are given for the different objectives, there will usually be no single optimal model, but rather a set of optimal models which represent best compromises between the different goals. This is well known to developers of thermodynamic models: once a certain stage of the modeling process is reached, further improvements in one objective are only possible at the expense of a deterioration in others.

The approach to find best compromises in a multi-objective or multi-criteria optimization (MCO) problem dealing with conflicting objectives is known as Pareto optimization, after the Italian economist Pareto¹. A Pareto-optimal solution of such a problem is one for which an improvement of any of the objectives is not possible without worsening at least one other objective. The set comprising all these solutions is the Pareto set, often referred to as the Pareto frontier of the MCO problem.

The use of MCO has received considerable attention in recent years in different areas of chemical engineering such as process design, operation and control²⁻⁵. There are many methods available for solving MCO problems and many of them involve the optimization of a “scalarization” function, or a function of the original objectives, by a suitable single-criteria optimization technique (SCO). A series of such SCO problems will have to be solved to find the Pareto-optimal solutions. The classical approach uses a weighting

scalarization approach, where the objectives are combined based on certain weighting factors, which have to be chosen in advance. The difficulty in the task is to define a systematic and efficient way to define those weights. Different weighting approaches can be found in the literature, such as the so-called weighted sum, the weighted global criterion, the weighted product and the exponential weighted⁶. Other common classical approaches are the ϵ -constraint, where the MCO problem is converted into an SCO only for one objective and the rest are treated as constraints, and goal programming methods, where the total deviation from goals defined for each objective is optimized. Other methods make use of stochastic optimization techniques including evolutionary approaches, such as genetic algorithms, swarm approaches and adaptations of simulated annealing⁶⁻⁸. In contrast to deterministic approaches, stochastic methods do not require gradient information. These approaches have the common characteristic of not requiring to solve a sequence of SCO problems, and are insensitive to the structure of constraints and objective functions. They may however sample large areas of the search space and therefore can be computationally expensive⁶⁻⁸.

The classical method for parameterizing thermodynamic models remains the use of single-objective optimization techniques⁹⁻¹¹. The different objectives are weighted and lumped into one goal function which is then minimized, yielding a single optimal model parameter set. MCO has however recently been used for parameterizing force fields to describe thermodynamic properties of fluids¹²⁻¹⁴ and there is a work on the application of MCO to parameterizing an excess Gibbs free energy model¹⁵.

The present work is, to the best of our knowledge, the first application of MCO to parameterizing equations of state. Equations of state are important tools for describing properties of fluids and widely used in chemical engineering and other fields. The accuracy of the description of the fluid properties by an equation of state crucially depends on its parametrization. The methodology presented here is also interesting for comparing different equations of state or molecular descriptions. As an example, we use

the perturbed-chain statistical associating fluid theory (PC-SAFT)^{16,17}, which is one of the different flavors of the statistical associating fluid theory (SAFT)^{18,19}, for reviews see^{20–24}. In PC-SAFT the molecules are modeled as chains of dispersive spherical segments with optional sites to mimic hydrogen-bonding interactions. An extension of the PC-SAFT theory incorporating a contribution accounting for dipolar interactions²⁵ is also considered here.

In the examples which we use to illustrate the MCO approach, the PC-SAFT equation of state is used to model two properties of water, namely the saturated liquid density and vapor pressure. Water is characterized by a peculiar behavior^{26,27}, which is a consequence of the complex interplay of intermolecular forces that define its structure^{28,29}. The choice of an adequate physical model is therefore not trivial (see^{30,31} for comprehensive reviews of water models appropriate for use in analytical theories), and understanding the influence of short-range attractive forces (hydrogen-bonding) and long-range electrostatic (primarily dipole-dipole) interactions on the structure and thermodynamic properties of water is essential³². Accounting for molecular features such as the presence of a permanent dipole moment and the formation of hydrogen bonds obviously requires complex models with different parameters which have to be suitably determined.

We first discuss an adaptive sampling scheme to calculate a finite set of Pareto points the interpolation of which optimally approximates the Pareto set with a given accuracy, and describe a numerical algorithm which is particularly suited for the present application. The PC-SAFT formalism (and a polar extension of the EoS) is then briefly introduced, including the adjustable parameters that can be employed to model water. The objective functions, which are used in the following to parametrize PC-SAFT for water, are then defined and the adaptive MCO algorithm is carried out to determine the Pareto sets for different PC-SAFT molecular descriptions of water. The results enable a comprehensive comparison and assessment of these variants. Finally we compare

the Pareto-optimal solutions found in the present work with PC-SAFT water models from the literature and show that for each studied molecular variant, better models were found than those which were previously available.

Methodology

Pareto terminology

The dimension of the objective space is given by the number of conflicting objective functions which are used. A point in the objective space is characterized by the values of the different objective functions. The Pareto frontier represents the set of optimal solutions in the objective space. In our example, it shows how well the studied equation of state can represent the studied set of properties. For each element in the objective space, there exists an element in the parameter space. The dimension of the parameter space is the number of parameters of the studied model and each element of the parameter space corresponds to a certain parametrization of the model. The Pareto frontier has an image in the parameter space which indicates where the optimal parameter sets lie. For parameters that have a physical meaning this is in itself interesting. As the relation between the points on the Pareto frontier and the points in the parameter space is known, the user can directly select a model of a desired quality (the compromise which is best suited for his application) in the objective space and get the corresponding parameters from the image of that point in the parameter space.

Approximation of the Pareto frontier

An adaptive algorithm to calculate the Pareto points is used. It has previously been applied to product and process design^{33–37} and parameter estimation for molecular simulations¹⁴, and consists of a hybrid approach^{33,38} which aims at calculating a minimal number of Pareto points, the linear interpolation of which approximates the Pareto

frontier within a predefined tolerance or quality. The hybrid method is based on a sandwiching algorithm to sample convex regions and a hyperboxing scheme to sample non-convex regions of the Pareto set. In both cases a scalarization of the objectives is performed to calculate each Pareto-optimal point by a single-objective optimization. Within the sandwiching scheme a weighted-sum scalarization is employed³⁹ whereas within the hyperboxing approach a Pascoletti-Serafini scalarization⁴⁰ is used. In this work only problems with two objectives are considered. The algorithm for this case, and assuming that the objectives are to be minimized, is briefly summarized here. For a more comprehensive explanation the reader is referred to the original works^{33,38}.

In broad terms the algorithm first uses the sandwiching approximation method to create successively inner and outer approximations to the Pareto frontier. The inner approximation is found from the bounding facets (line segments in two dimensions) of the convex hull (set of convex -linear- combinations) of the Pareto points. See Figure 1 a) for a sketch: each facet coincides with the line segment connecting two neighboring points and the set determines the corners of the convex hull. The latter constitutes the approximation to the Pareto frontier sketched. The outer approximation is represented by supporting tangential hyperplanes (line segments in two dimensions) of the Pareto frontier at the calculated points. Each new Pareto point is calculated using the weighted sum scalarization which, for objective functions denoted generically as f_i , is formulated as

$$\min_{\Omega} \sum_{i=1}^Q w_i f_i(\Omega), \quad (1)$$

where Ω represents the vector of parameters to be estimated, w_i are certain weights and Q is the dimension of the problem ($Q = 2$ in this case). The weight vectors coincide with the normals to the supporting tangential hyperplanes of the Pareto set at each calculated point. Pareto points are sequentially added as long as the difference between

the outer and the inner approximation (the sandwich) is beyond the threshold fixed by the desired approximation quality. The progress of the algorithm, including the selection of weights and the calculation of the approximation quality is explained in more detail along the sketches shown in Figure 1 as follows:

1. Obtain the extreme compromises (Pareto points P_1 and P_2 in Figure 1 b)) optimizing the weighted sum with weight vectors $\mathbf{w}_1=(1,0)$ and $\mathbf{w}_2=(0,1)$, respectively. These two Pareto points are used to set the Nadir point, given as the worst value for each objective.
2. Find the inner and outer approximation to the Pareto frontier (see Figure 1 b)).
3. The Nadir point is employed as a reference to determine the quality of the current approximation σ_s . This is calculated as the ratio between the distances from the Nadir point (N) to the outer (O) and inner (I) intersection points (see Figure 1 c)) as $\sigma_s = \overline{ON}/\overline{IN}$. When more Pareto points have been added, the ratio between outer and inner distances is computed for each facet. The maximum value is taken as the quality of the current approximation and the corresponding facet is selected for step 3.
4. If the quality of the current approximation is worse than the predefined tolerance, continue with step 5, otherwise stop.
5. Calculate a new Pareto point within the selected facet. This Pareto point will be determined in a further weighted-sum optimization using a weight vector $\mathbf{w}_3$ (see Figure 1 c)) normal to the inner approximation. The facet will be replaced by two new facets around the Pareto point (cf. Figure 1 d)). The algorithm continues again with step 2.

Once the predefined approximation quality for the sandwiching is achieved, the convexity of the facets obtained is checked and non-convex regions are then treated within

the hyperboxing scheme. Again new Pareto points are sequentially added as long as the difference between the outer and inner approximation to the non-convex regions is worse than the predefined approximation quality for this scheme. For an explanation of the convexity check and the hyperboxing scheme the reader is referred to the original works^{33,38}.

PC-SAFT equation of state

In PC-SAFT the Helmholtz free energy is expressed as a sum of the following contributions^{16,17}, including an additional polar term A^{DP} ²⁵,:

$$A = A^{\text{Ideal}} + A^{\text{HC}} + A^{\text{Disp}} + A^{\text{Assoc}} + A^{\text{DP}}, \quad (2)$$

where A^{Ideal} is the free energy of an ideal gas, A^{HC} is the free energy of a hard-chain reference, A^{Disp} is the contribution due to dispersion interactions between the chain-like molecules and A^{Assoc} is the contribution due to association interactions. A^{DP} is incorporated following the work of Gross and Vrabec²⁵ to account for dipolar interactions among non-spherical molecules. During the execution of this work some inconsistencies in the formulation of the PC-SAFT association term in the literature were found. Some details about the particular formulation of the EOS used here are given in the Supplementary material.

An associating molecule is described within PC-SAFT by the number of segments m , the diameter of the segments σ , the energy of the dispersion interactions ε , the energy of the hydrogen-bonding interaction $\varepsilon_{a,b}^{\text{HB}}$ between association site types a and b, and the reduced volume available for bonding $\kappa_{a,b}^{\text{HB}}$ between those. The polar term does not necessarily require any additional parameter to be adjusted, since experimental values for the dipole moment μ can be used, following the original work²⁵. Here we will consider

models in which the dipole moment is neglected (i.e., the original PC-SAFT EoS is used), as well as scenarios where the dipole moment is set to experimental values (for the gas phase, as will be commented later) or used as an additional parameter.

Definition of the optimization problem

Two competing objective functions are used in the present work. They quantify deviations between the model and experimental data in two important properties, which are often used for estimating model parameter, namely the saturated liquid density ρ_{liq} and the vapor pressure p_v . The objective functions are defined as follows,

$$\delta_\rho(\mathbf{\Omega}) = \frac{1}{N_\rho} \sum_{i=1}^{N_\rho} \left| \frac{\rho_{\text{liq}}^{\text{calc}}(T; \mathbf{\Omega}) - \rho_{\text{liq}}^{\text{exp}}(T)}{\rho_{\text{liq}}^{\text{exp}}(T)} \right|, \quad (3)$$

and

$$\delta_p(\mathbf{\Omega}) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left| \frac{p_v^{\text{calc}}(T; \mathbf{\Omega}) - p_v^{\text{exp}}(T)}{p_v^{\text{exp}}(T)} \right|, \quad (4)$$

where N_ρ is the number of experimental values of saturated liquid density and N_p of vapor pressure.

Analytical equations of state are known to fail reproducing the critical region due to the large-scale density fluctuations present there, which results in an overestimation of the predicted critical temperature and pressure. Non-analytical corrections to equations of state have been developed^{41–43} based on renormalization-group theory but such make the equation rather cumbersome and they are out of the scope of this work. We focus our efforts on determining models which are accurate for regions far enough from the critical region and to avoid any bias we restrict the calculation of deviations in Equations (3) and (4) to 95% of the experimental critical temperature. We choose to determine only models where the critical temperature is above the experimental value. For this purpose

the constraint

$$T_c^{\text{exp}} - T_c^{\text{calc}}(\boldsymbol{\Omega}) \leq 0, \quad (5)$$

is added to the optimization problem, where T_c^{exp} is the experimental critical temperature and T_c^{calc} is the calculated one.

An additional constraint is employed to restrict the maximum value of each objective function to 10 %, i.e.,

$$\delta_\rho(\boldsymbol{\Omega}) \leq 0.1, \quad (6)$$

and

$$\delta_p(\boldsymbol{\Omega}) \leq 0.1, \quad (7)$$

as we are not interested in Pareto points with absurdly high deviations.

Calculation details

The objective functions δ_ρ and δ_p (cf. Equations (3) and (4)) for saturated liquid densities and vapor pressures, respectively, are in this work calculated through comparison with correlated data according to the IAPWS-IF97 formulation for water⁴⁴ in temperature intervals of 10 K in a temperature range from $T = 373$ K to 613 K (i.e., ensuring temperatures below 95 % the experimental critical temperature). The low temperature bound is chosen given the difficulty of physically-sound equations of state to reproduce with the standard state of the art developments the density anomaly of water^{30,45–48}. This approach is also employed to calculate the deviations in these properties for the literature models, for purposes of comparison (i.e., the same correlated data is used). As a second comparison, we also calculate deviations in the whole temperature range of

vapor-liquid equilibrium, from $T = 273\text{ K}$ to 647 K , also in intervals of temperature of 10 K , and denote such deviations as δ'_ρ and δ'_p . The uncertainties of the IAPWS-IF97 formulation for water have been reported to be, in the region of interest, not higher than $\pm 0.05\%$ for both saturated liquid densities and pressures⁴⁴.

The bounds on the PC-SAFT parameters (and, where relevant, the dipole moment) employed during the optimizations carried out in this work are collected in Table 1.

The sandwiching and hyperboxing algorithm are carried out with an approximation quality of 1.01 and 1.09, respectively, for all the Pareto sets calculated. Other implementation details are found in the Supplementary material.

Application of MCO to Parameter Estimation

Review of literature models

An overview of the PC-SAFT models of water described in the literature is given in Table 2. It includes also the results of a comparison of the models with experimental data: deviations are reported for saturated liquid densities and vapor pressures, calculated based on Equations (3) and (4) (cf. Section “Calculation details”), as well as for the predicted critical parameters. If the authors present more than one model, only that with the lowest deviation is discussed here. As a first observation, it can be seen water is in all studies modeled as a non-spherical core, i.e., m differs from 1.

Water associates through hydrogen-bonding interactions. This is accounted for in models of the SAFT family considering short-range association sites of two different types: one type mimics lone pairs of electrons (acceptor) and the other type mimics hydrogens (donor). Water has been typically modeled using two different association schemes. The first one is the so-called 4C association model, after the nomenclature of Huang and Radosz⁴⁹, which assumes two association sites of each type; this model therefore enables that the molecule forms up to four hydrogen-bonds, as found in ice²⁸.

The second is the 2B association model (also after the same classification scheme⁴⁹), which assumes one association site of each type. The PC-SAFT EoS has been applied to modeling water using either a 2B^{50–55} or a 4C^{51,53,56–58} association scheme. Although less frequent, a three-site model denoted as 3B⁴⁹ has also been used in the literature⁵¹ (see Table 2).

Water molecules possess a considerable dipole moment. The value of the dipole moment of water in the gas phase has been determined experimentally to be $\mu = 1.86 \text{ D}$ ^{59,60}. It is known that the dipole moment is larger in condensed phases, for which values ranging from $\mu = 2.0 \text{ D}$ to $\mu = 3.0 \text{ D}$ are reported for water by different theoretical studies and extracted from x-ray data^{61,62}. There are two models in the literature^{54,55} that have also considered the dipolar nature of water through the extension of PC-SAFT of Gross and Vrabec²⁵. In both studies, a 2B association site model and a dipole moment of $\mu = 1.85 \text{ D}$ are employed (cf. Table 2).

To obtain the parameters of their water models, all different authors followed a similar strategy. In general they employed saturated liquid densities and vapor pressures and one included also mole fractions of non-associated molecules⁵⁶. In the work of Niño-Amézquita *et al.*⁵⁷ the model obtained was a posteriori successfully tested against IR spectroscopy data on the mole fraction of non-associated molecules as well as surface tension data and second derivative properties⁵⁸. In a number of works a temperature range from about 273 K to 640 K^{50,53,54} was used in the parameter estimation. Smaller temperature ranges were used in the work of Aparicio and Hall⁵¹, from 273 K to 634 K, and Cameretti *et al.*⁵², from 278 K to 393 K. In the rest of studies^{55–57} the temperature range was not specified.

Models for water using other variants of the original PC-SAFT EoS, such as the simplified PC-SAFT^{63–65} the group-contribution polar GC-PPC-SAFT EoS⁶⁶, or other polar versions⁶⁷ are not discussed here.

Results

Non-polar models using different association schemes

Pareto-optimal solutions have been calculated (see the Supplementary material for implementation details) for the three association schemes used for water in the literature (i.e., 4C, 2B and 3B) using the objective functions defined in Section “Definition of the optimization problem”. In Figure 2 the objective space of these Pareto frontiers is plotted along with the values for the literature models (note that one of the literature models, leading to high deviations in pressure, has been omitted from the graph for the purposes of plotting within a small scale). As seen the Pareto frontiers calculated present a sharp knee (the strong curvature near the origin) and relatively large almost linear parts in which one objective can be improved at practically no cost. It is seen that some of the 4C literature models lead to deviations that are close to the Pareto frontiers calculated, but in general none of the literature models appears to be in the Pareto-knee of the frontiers calculated. This may be due to a different definition of objective functions, which are often only vaguely specified. Comparing the three Pareto frontiers it is possible to see that the three types of models lead to similar deviations. The 3B association scheme yields, however, slightly poorer results than the other two models in the region of interest (i.e., the region where both objectives reach low values). 4C models may lead to lower δ_ρ values if the description of vapor pressures is compromised. The optimal 4C and 2B models in the region of interest (the Pareto-knee) are however equivalent. From that perspective, it can be concluded that modeling water as either a four associating-site model or as a two associating-site model leads to practically no differences if the aim is to minimize the deviations of both saturated liquid density and vapor pressure.

The Pareto-optimal solutions for the 4C and 2B schemes are analyzed in more detail in Figures 3 and 4. In these figures, for each Pareto point both the parameter space and

the objective space are plotted, along with the deviations in the critical temperature and pressure. Not only the objective functions δ_ρ and δ_p are given but also δ'_ρ and δ'_p (see Section “Calculation details”). A comparison of the Pareto points to the existing literature models is also included in the figures. In Figure 3, which corresponds to the 4C scheme, it is seen that both the chain length m and the bonding volume $\kappa_{a,b}^{\text{HB}}$ peak in the region in which small values for both saturated liquid densities and vapor pressures are attained. The segment diameter σ stays rather constant for the various Pareto models, whereas the dispersion-energy parameter ε and the energy of association $\varepsilon_{a,b}^{\text{HB}}$ present monotonic trends along the Pareto frontier which may indicate a more direct relation of the effect of those on the trade-offs observed for the deviations in the two properties. In this regard it is observed that lower deviations in density are accompanied by a decrease in the values of the dispersion-energy parameter. On the right-hand side of Figure 3 a comparison with all 4C literature models can be seen, confirming that there is always a Pareto-optimal solution leading to a more accurate estimation of both saturated liquid densities and vapor pressures in the range of temperatures used for the estimation. Some literature models lead however to slightly lower deviations in both properties in the whole saturation range, but this is not surprising considering our models were not fitted for this purpose. It is worth wondering whether achieving lower deviations in saturated liquid densities and vapor pressures in the temperature range of interest leads to higher deviations in the estimated critical temperature. The models rendered by the Pareto frontier are in principle seen to present critical temperatures in a similar range as those of the literature models, but this topic will be analyzed in more detail in the next section.

In Figure 4, which corresponds to the 2B scheme, it is seen that the trends in the parameters are similar to those shown in Figure 3 but the peaks in m and $\kappa_{a,b}^{\text{HB}}$ are smaller. The chain length and the bonding volume fluctuate less. The fact that in this case only two association sites are employed for water leads to models with higher energies of association compared to the 4C models. Pareto-optimal models are found

which not only lead to lower deviations than the literature models in (δ_ρ, δ_p) , but also to lower deviations of the critical temperature and pressure and it is comforting to note that some of them further present lower deviations in the whole temperature range (i.e. δ'_ρ and δ'_p) than some of the literature models.

Similar trends are seen for the parameters of the 3B Models (cf. Supplementary material).

Optimization including the critical temperature

The definition of the objectives in Pareto optimization is flexible. Additional objectives, such as the deviation in the critical temperature, can always be added in a straightforward manner. An alternative to this approach is the consideration of requirements by introducing constraints. This is illustrated here using the critical temperature calculated for 4C models as an example. The additional constraint included in the optimization problem is formulated as follows

$$\frac{T_c^{\text{calc}}(\mathbf{\Omega}) - T_c^{\text{exp}}}{T_c^{\text{exp}}} \leq \epsilon. \quad (8)$$

Different values of ϵ are used, namely $\epsilon = 0.005, 0.02, 0.05, 0.075$ and 0.10 . The resulting Pareto frontiers are shown in Figure 5. The optimization of the critical temperature conflicts with the optimization of the original objectives (i.e., relaxing the constraint leads to Pareto frontiers with lower values for δ_ρ and δ_p) up to a certain limit, further than which the frontiers are seen to overlap. The constraint is active only up to that point. The Pareto frontiers calculated present a knee which is more or less beaked depending on the constraint. The literature models can be compared to the Pareto frontiers of the same color (corresponding to the same overestimation of the critical temperature). Among them that of von Solms *et al.*⁵⁶ is the one that approaches closer the corresponding Pareto frontier, but in general all the literature models are dominated

by the corresponding Pareto frontiers (i.e., there are Pareto-optimal solutions with lower values for both objective functions than those presented by the literature models for a given overestimation of the critical temperature).

Models considering the dipolar nature of water

A more detailed model of water accounting for the dipole moment of the molecule is now considered. In our first study the dipole moment is fixed to the experimental value in the gas phase ($\mu = 1.855 \text{ D}^{60}$), so that no additional parameter is introduced. The resulting Pareto frontier is compared to that of the corresponding non-polar model in Figure 6. The frontier for the polar model dominates that for the non-polar one, even though in the region of the Pareto knee, which is sharper for the non-polar case, both frontiers are very close. The polar models found in the literature^{54,55} show higher deviations in both properties than the best Pareto-optimal models. More details of the polar Pareto-optimal solutions are given in the Supplementary material. A third Pareto frontier in which the dipole moment is also treated as an adjustable parameter is also shown in Figure 6. This results in further improvements in the region of the Pareto knee. Additional detail regarding those Pareto-optimal solutions is presented in Figure 7. The values obtained for the dipole moment along the frontier are found in the range of the experimental data for that property.

If a 4C association scheme is used a Pareto frontier rendering also very low deviations is obtained. The results are given in the Supplementary material.

Decision support in model selection

The objective space of the Pareto frontiers and the respective parameter values can be used to arrive at a more transparent choice of a model. Clearly one would focus only on Pareto-optimal models and ignore dominated models in the decision. The preferred models are located in the knee of the Pareto frontier. Which model to choose depends

on the purpose of the model. One of the great benefits of the Pareto optimization is the knowledge of the trade-off one has to accept when using a certain model type. Knowing the Pareto frontier is also advantageous from the perspective that preferences and weights may change during the decision process when the decision maker sees all the available options. Eventually a model can be chosen based on a given scenario, but if the requirements change the Pareto frontier is still valid.

For the best Pareto frontiers obtained in this work, a Pareto-optimal model is selected and shown in Table 3 for comparison. Five models are presented: M_{4C} is a 4C non-polar model, M_{2B} is a 2B non-polar model, M_{P2B} is a 2B polar model, and $M_{P2B\mu}$ and $M_{P4C\mu}$ are respectively a 2B and a 4C polar model using also μ as a parameter in the optimization. The models have been selected based on visual inspection of the Pareto knees and with no preference given to minimizing one property over the second one. Overall it can be observed that the polar models obtained fitting also the dipole moment of water (i.e., $M_{P4C\mu}$ and $M_{P2B\mu}$) are the most spherical; the values for the dispersive energy are in a similar range for all models and higher differences are seen for the association energy, where the non-polar models are characterized by higher values (comparing within the same association scheme). It is also interesting to note that the polar models obtained using the dipole moment as an extra adjustable parameter present also rather lower values of the bonding volume, as is deemed consistent with the bonding constraints on which SAFT approaches are based^{68,69}. The agreement of these models with experimental data is depicted in Figure 8. On a first inspection of Figure 8 a) it is seen that most differences between the models are in the critical region, whereas at low temperature all models fail to reproduce the density anomaly of water. The latter is known to be a challenging feature not easily reproduced by models adjusted to capture the global behavior of this compound. The performance of the models to describe vapor pressure is not distinguished easily in Figure 8 b). A closer look can be taken analyzing the deviations in Figures 8 c) and d); the region of temperature

used to adjust the parameters of the models is indicated by the vertical lines in these plots. The models $M_{P4C\mu}$ and $M_{P2B\mu}$ lead to low deviations in the whole temperature range considered in the estimation, and also to a lower over-prediction of the critical temperature (see Figure 8). The PC-SAFT equation of state (and its polar extension) therefore demonstrates to be capable of providing a very good estimation of the VLE of water. Model $M_{P4C\mu}$ is also seen to outperform in the region of low temperatures, which was not included in the optimization problem, demonstrating good extrapolation capabilities of the equation. In particular, with the latter model it is remarkable that it is possible to achieve deviations lower than 0.1% for both liquid densities and vapor pressures (close to the uncertainties of the IAPWS-IF97 correlation) in the interval of temperatures from $T = 373$ K to 613 K, and of about 2% for liquid densities and lower than 1% for vapor pressures from the triple point to the critical point (i.e, from $T = 273$ K to 647 K).

As a final check the models are assessed in the prediction of the enthalpy of vaporization. This is shown in Figure 9. Overall all models are seen to provide a good description of this property, even though they were adjusted to reproduce saturated liquid densities and vapor pressures. Expressed in MCO terminology, the objective of minimizing the deviations in the enthalpy of vaporization is not conflicting with the two objectives used in the fit. On the other side, the good representation of this property could be expected given the fact that the models were also seen to provide a good description of saturated gas densities. The best description is again provided by model $M_{P4C\mu}$ over all the temperature range, confirming the enhanced performance of this model.

Conclusions

We present the first study on the application of multi-criteria optimization based on the determination of Pareto sets in the field of parameter estimation for equations of state. An adaptive hybrid algorithm combining a sandwiching and a hyperboxing technique is

employed and the capabilities of the methodology are illustrated using PC-SAFT models of water as an example. Pareto-optimal solutions for PC-SAFT models with different level of detail in the description of the molecular characteristics of water are determined and compared to models that were published in the literature. New water models which lead to an improved agreement in the description of both saturated liquid densities and vapor pressures of water in comparison with existing ones are found. The overview which is obtained from inspection of the Pareto sets shows that PC-SAFT models with a four association-site scheme and those with a two association-site scheme lead to only minor differences in their capabilities to describe the present objectives. Accounting for the permanent dipole of the molecule through the incorporation of a polar contribution leads to a small improvement if the dipole moment is simply taken from literature and not adjusted. If however the dipole moment is included as an additional parameter in the optimization substantial improvements are achieved. In the latter case it is possible to obtain models which not only provide very low deviations for both equilibrium properties considered, but also lead to a reduced over-prediction of the critical temperature and an enhanced prediction of enthalpies of vaporization. In all studied cases Pareto-optimal models are found which are better than all previously published models of the same type in the range of temperatures of interest.

In future work it could be useful to extend the methodology to a three-dimensional (or larger) objective space, so that other properties could be included in the study, such as the degree of hydrogen-bonding considered by other authors^{30,48,56}.

In summary this work shows the usefulness of applying Pareto optimization in the field of parametrization of EoS and more generally in the development of thermodynamic models.

Acknowledgements

We gratefully acknowledge support by the German Research Foundation (Deutsche Forschungsgemeinschaft) in the Reinhart Koselleck program.

References

- [1] Pareto V. *Manual of political economy*. Oxford: Oxford University Press. 2014. English translation. First Edition dated of 1906.
- [2] Clark PA, Westerberg AW. Optimization for design problems having more than one objective. *Comput Chem Eng*. 1983;7(4):259 – 278.
- [3] Bhaskar V, Gupta SK, Ray AK. Applications of multiobjective optimization in chemical engineering. *Rev Chem Eng*. 2000;16(1):1–54.
- [4] Masuduzzaman, Rangaiah GP. Multi-Objective Optimization Applications in Chemical Engineering. In: *Multi-Objective Optimization. Techniques and Applications in Chemical Engineering*, edited by Rangaiah GP, pp. 27–59. World Scientific. 2009;.
- [5] Sharma S, Rangaiah GP. Multi-Objective Optimization Applications in Chemical Engineering. In: *Multi-Objective Optimization in Chemical Engineering: Developments and Applications*, edited by Rangaiah GP, Bonilla-Petriciolet A, pp. 35–102. John Wiley and Sons, Ltd. 2013;.
- [6] Marler RT, Arora JS. Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*. 2004;26(6):369–395.
- [7] Rangaiah GP, ed. *Multi-Objective Optimization. Techniques and Applications in Chemical Engineering*. World Scientific. 2009.

- [8] Rangaiah GP, Bonilla-Petriciolet A, eds. *Multi-Objective Optimization in Chemical Engineering: Developments and Applications*. John Wiley and Sons, Ltd. 2013.
- [9] Anderson TF, Abrams DS, Grens EA. Evaluation of parameters for nonlinear thermodynamic models. *AIChE J.* 1978;24(1):20–29.
- [10] Salazar-Sotelo D, Boiveaut Boudehen A, Renon H. Computer calculation of the optimal parameters of a model for the simultaneous representation of experimental binary and ternary vapor-liquid and liquid-liquid equilibrium data. *Fluid Phase Equilib.* 1986;27:383 – 403.
- [11] Englezos P, Kalogerakis N. *Applied parameter estimation for chemical engineers*. CRC Press. 2000.
- [12] Stöbener K, Klein P, Reiser S, Horsch M, Küfer KH, Hasse H. Multicriteria optimization of molecular force fields by Pareto approach. *Fluid Phase Equilib.* 2014; 373:100 – 108.
- [13] Werth S, Stöbener K, Klein P, Küfer KH, Horsch M, Hasse H. Molecular modelling and simulation of the surface tension of real quadrupolar fluids. *Chem Eng Sci.* 2015;121:110 – 117.
- [14] Stöbener K, Klein P, Horsch M, Küfer KH, Hasse H. Parametrization of two-center Lennard-Jones plus point-quadrupole force field models by multicriteria optimization. *Fluid Phase Equilib.* 2016;411:33 – 42.
- [15] Punnapala S, Vargas FM, Elkamel A. Parameter estimation in phase equilibria calculations using multi-objective evolutionary algorithms. In: *Multi-Objective Optimization in Chemical Engineering: Developments and Applications*, edited by Rangaiah GP, Bonilla-Petriciolet A, pp. 249–266. John Wiley and Sons, Ltd. 2013;.

- [16] Gross J, Sadowski G. Application of perturbation theory to a hard-chain reference fluid: an equation of state for square-well chains. *Fluid Phase Equilib.* 2000; 168(2):183–199.
- [17] Gross J, Sadowski G. Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. *Ind Eng Chem Res.* 2001;40(4):1244–1260.
- [18] Chapman WG, Gubbins KE, Jackson G, Radosz M. SAFT: Equation-of-state solution model for associating fluids. *Fluid Phase Equilib.* 1989;52:31–38.
- [19] Chapman WG, Gubbins KE, Jackson G, Radosz M. New reference equation of state for associating liquids. *Ind Eng Chem Res.* 1990;29(8):1709–1721.
- [20] Müller EA, Gubbins KE. Chapter 12: Associating Fluids and Fluid Mixtures. In: *Equations of State for Fluids and Fluid Mixtures*, edited by Sengers JV, Kayser RF, Peters CJ, White HJ. Amsterdam: Elsevier. 2000;.
- [21] Müller EA, Gubbins KE. Molecular-Based Equations of State for Associating Fluids: A Review of SAFT and Related Approaches. *Ind Eng Chem Res.* 2001; 40(10):2193–2211.
- [22] Economou IG. Statistical Associating Fluid Theory: A Successful Model for the Calculation of Thermodynamic and Phase Equilibrium Properties of Complex Fluid Mixtures. *Ind Eng Chem Res.* 2002;41(5):953–962.
- [23] Tan SP, Adidharma H, Radosz M. Recent Advances and Applications of Statistical Associating Fluid Theory. *Ind Eng Chem Res.* 2008;47(21):8063–8082.
- [24] McCabe C, Galindo A. Chapter 8: SAFT Associating Fluids and Fluids Mixtures. In: *Applied Thermodynamics of Fluids*, edited by Goodwin ARH, Sengers JV. London: Royal Society of Chemistry. 2010;.

- [25] Gross J, Vrabec J. An equation-of-state contribution for polar components: Dipolar molecules. *AIChE J.* 2006;52(3):1194–1204.
- [26] Guillot B. A reappraisal of what we have learnt during three decades of computer simulations on water. *J Molec Liq.* 2002;101(13):219 – 260. Molecular Liquids. Water at the New Millenium.
- [27] Stillinger FH. Theory and Molecular Models for Water. In: *Advances in Chemical Physics*, pp. 1–101. John Wiley & Sons, Inc. 2007;.
- [28] Stillinger FH. Water revisited. *Science.* 1980;209(4455):451–457.
- [29] Finney JL. The water molecule and its interactions: the interaction between theory, modelling, and experiment. *J Molec Liq.* 2001;90(13):303 – 312.
- [30] Clark GNI, Haslam AJ, Galindo A, Jackson G. Developing optimal Wertheim-like models of water for use in Statistical Associating Fluid Theory (SAFT) and related approaches. *Mol Phys.* 2006;104(22-24):3561–3581.
- [31] Vega LF, Llovel F. Review and new insights into the application of molecular-based equations of state to water and aqueous solutions. *Fluid Phase Equilib.* 2016; 416:150–173.
- [32] Nezbeda I, Kolafa J. Effect of short- and long-range forces on the structure of water: temperature and density dependence. *Mol Phys.* 1999;97(10):1105–1116.
- [33] Bortz M, Burger J, Asprion N, Blagov S, Böttcher R, Nowak U, Scheithauer A, Welke R, Küfer KH, Hasse H. Multi-criteria optimization in chemical process design and decision support by navigation on Pareto sets. *Comput Chem Eng.* 2014;60:354 – 363.
- [34] Burger J, Asprion N, Blagov S, Böttcher R, Nowak U, Bortz M, Welke R, Küfer

- KH, Hasse H. Multi-Objective Optimization and Decision Support in Process Engineering Implementation and Application. *Chem Ing Tech.* 2014;86(7):1065–1072.
- [35] Bortz M, Maag V, Schwientek J, Benfer R, Böttcher R, Burger J, von Harbou E, Asprion N, Küfer KH, Hasse H. Computer Aided Chemical Engineering, 37. In: *Decision Support by Multicriteria Optimization in Process Development: An Integrated Approach for Robust Planning and Design of Plant Experiments*, edited by Huusom JK, Gani R, Gernaey KV, pp. 2063–2068. Elsevier. 2015;.
- [36] Burger J, Papaioannou V, Gopinath S, Jackson G, Galindo A, Adjiman CS. A hierarchical method to integrated solvent and process design of physical CO₂ absorption using the SAFT- Mie approach. *AIChE J.* 2015;61(10):3249–3269.
- [37] Asprion N, Benfer R, Blagov S, Böttcher R, Bortz M, Berezhnyi M, Burger J, Harbou Ev, Küfer KH, Hasse H. INES An Interface Between Experiments and Simulation to Support the Development of Robust Process Designs. *Chem Ing Tech.* 2015;87(12):1810–1825.
- [38] Hernández JIS. Multi-objective optimization in mixed integer problems: With application to the beam selection optimization problem in IMRT. Ph.D. thesis, TU Kaiserslautern. 2012.
- [39] Ehrgott M. *Multicriteria optimization*, vol. 2nd ed. Berlin: Springer. 2005.
- [40] Pascoletti A, Serafini P. Scalarizing vector optimization problems. *J Optimiz Theory App.* 1984;42(4):499–524.
- [41] Bymaster A, Emborsky C, Dominik A, Chapman WG. Renormalization-Group Corrections to a Perturbed-Chain Statistical Associating Fluid Theory for Pure Fluids Near to and Far from the Critical Region. *Ind Eng Chem Res.* 2008;47(16):6264–6274.

- [42] Tang X, Gross J. Renormalization-Group Corrections to the Perturbed-Chain Statistical Associating Fluid Theory for Binary Mixtures. *Ind Eng Chem Res.* 2010; 49(19):9436–9444.
- [43] Forte E, Llovel F, Vega LF, Trusler JPM, Galindo A. Application of a renormalization-group treatment to the statistical associating fluid theory for potentials of variable range (SAFT-VR). *J Chem Phys.* 2011;134(15):154102.
- [44] Wagner W, Cooper JR, Dittmann A, Kijima J, Kretzschmar HJ, Kruse A, Mareš R, Oguchi K, Sato H, Stöcker I, Šifner O, Takaishi Y, Tanishita I, Trübenbach J, Willkommen T. The IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam. *J Eng Gas Turbines Power.* 2000;122(1):150–184.
- [45] Müller EA, Gubbins KE. An Equation of State for Water from a Simplified Intermolecular Potential. *Ind Eng Chem Res.* 1995;34(10):3662–3673.
- [46] Cameretti LF, Sadowski G. Modeling of aqueous amino acid and polypeptide solutions with PC-SAFT. *Chem Eng Process Process Intensif.* 2008;47(6):1018 – 1025.
- [47] Vega LF, Llovel F, Blas FJ. Capturing the Solubility Minima of n-Alkanes in Water by Soft-SAFT. *J Phys Chem B.* 2009;113(21):7621–7630.
- [48] Dufal S, Lafitte T, Galindo A, Jackson G, Haslam AJ. Developing intermolecular-potential models for use with the SAFT-VRMie equation of state. *AIChE J.* 2015; 61(9):2891–2912.
- [49] Huang SH, Radosz M. Equation of state for small, large, polydisperse, and associating molecules. *Ind Eng Chem Res.* 1990;29(11):2284–2294.
- [50] Gross J, , Sadowski G. Application of the Perturbed-Chain SAFT Equation of State to Associating Systems. *Ind Eng Chem Res.* 2002;41(22):5510–5515.

- [51] Aparicio-Martínez S, Hall KR. Phase equilibria in water containing binary systems from molecular based equations of state. *Fluid Phase Equilib.* 2007;254(1-2):112 – 125.
- [52] Cameretti LF, Sadowski G, Mollerup JM. Modeling of Aqueous Electrolyte Solutions with Perturbed-Chain Statistical Associated Fluid Theory. *Ind Eng Chem Res.* 2005;44(9):3355–3362.
- [53] Diamantonis NI, Economou IG. Evaluation of Statistical Associating Fluid Theory (SAFT) and Perturbed Chain-SAFT Equations of State for the Calculation of Thermodynamic Derivative Properties of Fluids Related to Carbon Capture and Sequestration. *Energy Fuels.* 2011;25(7):3334–3343.
- [54] Al-Saifi NM, Hamad EZ, Englezos P. Prediction of vapor-liquid equilibrium in water-alcohol-hydrocarbon systems with the dipolar perturbed-chain SAFT equation of state . *Fluid Phase Equilib.* 2008;271(12):82 – 93.
- [55] Kleiner M, , Sadowski G. Modeling of polar systems using PCP-SAFT: An approach to account for induced-association interactions. *J Phys Chem C.* 2007; 111(43):15544–15553.
- [56] von Solms N, Michelsen ML, Passos CP, Derawi SO, Kontogeorgis GM. Investigating Models for Associating Fluids Using Spectroscopy. *Ind Eng Chem Res.* 2006; 45(15):5368–5374.
- [57] Niño-Amézquita G, van Putten D, Enders S. Phase equilibrium and interfacial properties of water + CO₂ mixtures. *Fluid Phase Equilib.* 2012;332:40 – 47.
- [58] Langenbach K, Enders S. Cross-association of multi-component systems. *Mol Phys.* 2012;110(11-12):1249–1260.
- [59] Clough SA, Beers Y, Klein GP, Rothman LS. Dipole moment of water from Stark measurements of H₂O, HDO, and D₂O. *J Chem Phys.* 1973;59(5):2254–2259.

- [60] Lide DR. *CRC Handbook of Chemistry and Physics*. Boca Raton: CRC Press/Taylor and Francis, 86th ed. 2005.
- [61] Silvestrelli PL, Parrinello M. Water Molecule Dipole in the Gas and in the Liquid Phase. *Phys Rev Lett*. 1999;82:3308–3311.
- [62] Gubskaya AV, Kusalik PG. The total molecular dipole moment for liquid water. *J Chem Phys*. 2002;117(11):5290–5302.
- [63] Grenner A, Schmelzer J, Nicolas von Solms N, Kontogeorgis GM. Comparison of Two Association Models (ElliottSureshDonohue and Simplified PC-SAFT) for Complex Phase Equilibria of HydrocarbonWater and Amine-Containing Mixtures. *Ind Eng Chem Res*. 2006;45(24):8170–8179.
- [64] Fouad WA, Ballal D, Cox KR, Chapman WG. Examining the Consistency of Water Content Data in Alkanes Using the Perturbed-Chain Form of the Statistical Associating Fluid Theory Equation of State. *J Chem Eng Data*. 2014;59(4):1016–1023.
- [65] Liang X, Tsivintzelis I, Kontogeorgis GM. Modeling Water Containing Systems with the Simplified PC-SAFT and CPA Equations of State. *Ind Eng Chem Res*. 2014;53(37):14493–14507.
- [66] Nguyen-Huynh D, de Hemptinne JC, Lugo R, Passarello JP, Tobaly P. Modeling Liquid-Liquid and Liquid-Vapor Equilibria of Binary Systems Containing Water with an Alkane, an Aromatic Hydrocarbon, an Alcohol or a Gas (Methane, Ethane, CO₂ or H₂S), Using Group Contribution Polar Perturbed-Chain Statistical Associating Fluid Theory. *Ind Eng Chem Res*. 2011;50(12):7467–7483.
- [67] Karakatsani EK, Spyriouni T, Economou IG. Extended statistical associating fluid theory (SAFT) equations of state for dipolar fluids. *AIChE J*. 2005;51(8):2328–2342.
- [68] Docherty H, Galindo A. A study of Wertheim’s thermodynamic perturbation theory

(TPT1) for associating fluids with dispersive interactions: the importance of the association range. *Mol Phys.* 2006;104(22):3551–3560.

[69] Jiménez G, Santillán S, Avendaño C, Castro M, Gil-Villegas A. Molecular Thermodynamics of Adsorption using Discrete-Potential Systems. *OGST - Rev IFP.* 2008; 63(3):329–341.

[70] MATLAB. <http://www.mathworks.com/>. 2015.

[71] Lemmon EW, McLinden MO, Friend DG. Thermophysical Properties of Fluid Systems. In: *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, edited by Linstrom PJ, Mallard WG. National Institute of Standards and Technology, Gaithersburg MD, 20899. 2015;.

List of Tables

1	Bounds chosen for the PC-SAFT-intermolecular model parameters (and the polar extension, where relevant) during the optimizations.	34
2	PC-SAFT literature models for water. Their corresponding intermolecular model parameters are shown along with their deviations in saturated liquid density and vapor pressure, as well as their deviations in critical temperature and pressure. T_c^{exp} and p_c^{exp} are respectively the experimental critical temperature and pressure ($T_c^{\text{exp}} = 647.1$ K and $p_c^{\text{exp}} = 22.1$ MPa ⁷¹). k_B is Boltzmann's constant. Models are non-polar, unless specified otherwise.	35
3	Selected Pareto-optimal PC-SAFT models: M_{4C} is a 4C non-polar model, M_{2B} is a 2B non-polar model, M_{P2B} is a 2B polar model, and $M_{P2B\mu}$ and $M_{P4C\mu}$ are respectively a 2B and a 4C polar model using also μ as a parameter in the optimization. Their corresponding intermolecular model parameters are shown along with their deviations in saturated liquid density and vapor pressure, as well as their deviations in critical temperature and pressure.	36

List of Figures

- 1 Sketch of the sandwiching method for two objectives f_1 and f_2 . Filled circles denote Pareto points, dashed blue lines and solid red lines represent respectively the inner and the outer approximation, dotted gray curves are the (unknown) Pareto frontier, open blue circles with an inner cross represent the Nadir point, open black circles are new Pareto points calculated in the step represented, and open blue and open red diamonds are respectively the inner and outer intersection points used to compute the approximation quality. 37
- 2 Deviations in saturated liquid density and vapor pressure for the Pareto frontiers calculated using 4C non-polar (blue circles), 2B non-polar (red diamonds) and 3B non-polar (green squares) water models. The solid lines are a guide to the eye joining the Pareto points. The dispersed symbols are the corresponding values for the literature models following the same color code for the different schemes (triangle⁵⁰, cross⁵³, upside-down triangle⁵², asp⁵¹, star⁵⁶). 38
- 3 Intermolecular parameters, deviations in saturated liquid density and vapor pressure, as well as in the critical parameter values for 4C non-polar water models: P_i are the Pareto points calculated (star symbols connected by solid lines) and L_i are the literature models (circles⁵⁷, diamonds⁵³, squares⁵¹ and triangles⁵⁶). Deviations in vapor pressure are shown in red and in saturated liquid density in green. The solid lines are a guide to the eye joining the Pareto points. The parameter m is out of the bounds of the plot for the literature model⁵⁶ and therefore not represented. . . . 39

- 4 Intermolecular parameters, deviations in saturated liquid density and vapor pressure, as well as in the critical parameter values for 2B non-polar water models: P_i are the Pareto points calculated (star symbols connected by solid lines) and L_i are the literature models (circles⁵⁰, diamonds⁵³, squares⁵² and triangles⁵¹). Deviations in vapor pressure are shown in red and in saturated liquid density in green. The solid lines are a guide to the eye joining the Pareto points. 40
- 5 Deviations in saturated liquid density and vapor pressure for the Pareto frontiers calculated for 4C non-polar water models restricting the over-prediction of the critical temperature to values below 0.5% (pink dashed line), 2% (green dashed line), 5% (blue line), 7.5% (red line) and 10% (gray line) the experimental one. The solid (or dashed) lines are joining the Pareto points, which are here not represented. Literature 4C models are shown for comparison with open symbols: blue circle⁵³, gray diamond⁵¹ and red right triangle⁵⁶ following the same color code according to their proximity to the experimental critical temperature. 41
- 6 Deviations in saturated liquid density and vapor pressure for the Pareto frontiers calculated using 2B non-polar (red diamonds), 2B polar (blue triangles) and 2B polar fitting the dipole moment (green squares) water models. The solid lines are a guide to the eye joining the Pareto points. The dispersed symbols are the corresponding values for the existing 2B polar literature models (circle⁵⁴, star⁵⁵). 42

7	Intermolecular parameters, deviations in saturated liquid density and vapor pressure, as well as in the critical parameter values for 2B polar water models obtained fitting also the dipole moment: P_i are the Pareto points calculated (star symbols connected by solid lines). Deviations in vapor pressure are shown in red and in saturated liquid density in green. The solid lines are a guide to the eye joining the Pareto points.	43
8	a) Saturated density, b) vapor pressure, c) deviation in saturated liquid density and d) deviation in vapor pressure versus temperature for water. The symbols correspond to correlated experimental data ⁴⁴ while the curves correspond to the PC-SAFT (or its extended polar version) for the models: M_{4C} (dashed black), M_{2B} (dotted blue), M_{P2B} (dash-dotted green), $M_{P2B\mu}$ (dashed light blue) and $M_{P4C\mu}$ (solid red) following the nomenclature specified in Table 3. The region between the gray lines in plots b) and d) indicates the temperature range used in the optimizations.	44
9	Enthalpy of vaporization of water versus temperature. The symbols correspond to correlated experimental data ⁴⁴ while the curves correspond to the PC-SAFT (or its extended polar version) for the models: M_{4C} (dashed black), M_{2B} (dotted blue), M_{P2B} (dash-dotted green), $M_{P2B\mu}$ (dashed light blue) and $M_{P4C\mu}$ (solid red) following the nomenclature specified in Table 3.	45

Table 1: Bounds chosen for the PC-SAFT-intermolecular model parameters (and the polar extension, where relevant) during the optimizations.

	m	$\sigma/\text{\AA}$	$\varepsilon/(k_{\text{B}}\text{K})$	$\varepsilon_{a,b}^{\text{HB}}/(k_{\text{B}}\text{K})$	$\kappa_{a,b}^{\text{HB}}$	μ/D
Lower bound	1.0	1.0	100	800	0.0001	1.8
Upper bound	2.5	4.0	600	2500 †	0.25 §	3.2

†This bound is increased to 2600 for the Pareto frontiers calculated assuming 2B and 3B models of water.

Table 2: PC-SAFT literature models for water. Their corresponding intermolecular model parameters are shown along with their deviations in saturated liquid density and vapor pressure, as well as their deviations in critical temperature and pressure. T_c^{exp} and p_c^{exp} are respectively the experimental critical temperature and pressure ($T_c^{\text{exp}} = 647.1 \text{ K}$ and $p_c^{\text{exp}} = 22.1 \text{ MPa}$ ⁷¹). k_B is Boltzmann’s constant. Models are non-polar, unless specified otherwise.

Scheme	m	$\sigma/\text{\AA}$	$\varepsilon/(k_B\text{K})$	$\varepsilon_{a,b}^{\text{HB}}/(k_B\text{K})$	$\kappa_{a,b}^{\text{HB}}$	$\delta_\rho/\%$	$\delta_p/\%$	$\delta'_\rho/\%$	$\delta'_p/\%$	$\frac{T_c - T_c^{\text{exp}}}{T_c^{\text{exp}}}/\%$	$\frac{p_c - p_c^{\text{exp}}}{p_c^{\text{exp}}}/\%$	Ref.
2B	1.0656	3.0007	366.51	2500.7	0.034868	5.31	1.26	8.93	1.69	7.77	65.97	50
2B	1.9599	2.362	279.42	2059.28	0.1750	3.10	1.52	5.64	1.20	6.79	53.76	53
2B	1.0953	2.8898	365.956	2515.67	0.034868	4.90	4.68	8.10	3.90	8.10	79.89	52
2B	1.3112	2.7613	372.37	2123.1	0.048987	3.91	0.82	7.59	0.98	8.35	66.82	51
4C	1.0656	3.0007	366.51	1800.0	0.01	2.92	45.79	5.35	39.12	2.32	84.28	57
4C	2.1945	2.229	141.66	1804.17	0.2039	0.75	1.91	3.04	2.04	5.16	42.70	53
4C	2.7500	2.0794	183.61	1354.1	0.3374	1.59	0.35	4.56	0.59	7.63	59.95	56
4C	1.5725	2.6270	291.13	1334.2	0.074347	4.40	0.80	8.59	0.90	10.20	87.06	51
3B	1.7960	2.4697	327.62	1558.4	0.068277	3.71	0.67	7.25	0.84	8.21	67.57	51
2B†	1.2255	2.7920	203.00	2406.7	0.07172	0.69	3.46	1.98	3.97	2.65	26.02	54
2B†	1.6963	2.5144	311.19	1469.1	0.0635	1.83	0.17	4.15	0.37	4.14	33.23	55

†Polar model obtained using the extension of Gross and Vrabec²⁵ with dipole moment $\mu = 1.85 \text{ D}$.

Table 3: Selected Pareto-optimal PC-SAFT models: M_{4C} is a 4C non-polar model, M_{2B} is a 2B non-polar model, M_{P2B} is a 2B polar model, and $M_{P2B\mu}$ and $M_{P4C\mu}$ are respectively a 2B and a 4C polar model using also μ as a parameter in the optimization. Their corresponding intermolecular model parameters are shown along with their deviations in saturated liquid density and vapor pressure, as well as their deviations in critical temperature and pressure.

Model	m	$\sigma/\text{\AA}$	$\varepsilon/(k_B K)$	$\varepsilon_{a,b}^{HB}/(k_B K)$	$\kappa_{a,b}^{HB}$	(μ/D)	$\delta_\rho/\%$	$\delta_p/\%$	$\delta'_\rho/\%$	$\delta'_p/\%$	$\frac{T_c - T_c^{exp}}{T_c^{exp}}/\%$	$\frac{p_c - p_c^{exp}}{p_c^{exp}}/\%$
M_{4C}	2.5000	2.1201	147.917	1581.945	0.3939	-	0.54	0.37	3.35	3.19	5.67	43.30
M_{2B}	1.7854	2.3609	233.787	2506.794	0.2637	-	0.41	0.43	3.10	4.36	4.60	32.12
M_{P2B}	2.2159	2.2396	214.485	1543.672	0.5000	1.8550	0.31	0.47	2.69	2.89	3.34	26.55
$M_{P2B\mu}$	1.2256	2.8249	292.990	1760.237	0.0171	2.1870	0.07	0.02	2.20	1.57	1.76	12.26
$M_{P4C\mu}$	1.1990	2.8792	285.368	1117.327	0.0144	2.2088	0.09	0.05	2.07	0.62	1.72	12.37

Figure 1: Sketch of the sandwiching method for two objectives f_1 and f_2 . Filled circles denote Pareto points, dashed blue lines and solid red lines represent respectively the inner and the outer approximation, dotted gray curves are the (unknown) Pareto frontier, open blue circles with an inner cross represent the Nadir point, open black circles are new Pareto points calculated in the step represented, and open blue and open red diamonds are respectively the inner and outer intersection points used to compute the approximation quality.

Figure 2: Deviations in saturated liquid density and vapor pressure for the Pareto frontiers calculated using 4C non-polar (blue circles), 2B non-polar (red diamonds) and 3B non-polar (green squares) water models. The solid lines are a guide to the eye joining the Pareto points. The dispersed symbols are the corresponding values for the literature models following the same color code for the different schemes (triangle⁵⁰, cross⁵³, upside-down triangle⁵², asp⁵¹, star⁵⁶).

Figure 3: Intermolecular parameters, deviations in saturated liquid density and vapor pressure, as well as in the critical parameter values for 4C non-polar water models: P_i are the Pareto points calculated (star symbols connected by solid lines) and L_i are the literature models (circles⁵⁷, diamonds⁵³, squares⁵¹ and triangles⁵⁶). Deviations in vapor pressure are shown in red and in saturated liquid density in green. The solid lines are a guide to the eye joining the Pareto points. The parameter m is out of the bounds of the plot for the literature model⁵⁶ and therefore not represented.

Figure 4: Intermolecular parameters, deviations in saturated liquid density and vapor pressure, as well as in the critical parameter values for 2B non-polar water models: P_i are the Pareto points calculated (star symbols connected by solid lines) and L_i are the literature models (circles⁵⁰, diamonds⁵³, squares⁵² and triangles⁵¹). Deviations in vapor pressure are shown in red and in saturated liquid density in green. The solid lines are a guide to the eye joining the Pareto points.

Figure 5: Deviations in saturated liquid density and vapor pressure for the Pareto frontiers calculated for 4C non-polar water models restricting the over-prediction of the critical temperature to values below 0.5% (pink dashed line), 2% (green dashed line), 5% (blue line), 7.5% (red line) and 10% (gray line) the experimental one. The solid (or dashed) lines are joining the Pareto points, which are here not represented. Literature 4C models are shown for comparison with open symbols: blue circle⁵³, gray diamond⁵¹ and red right triangle⁵⁶ following the same color code according to their proximity to the experimental critical temperature.

Figure 6: Deviations in saturated liquid density and vapor pressure for the Pareto frontiers calculated using 2B non-polar (red diamonds), 2B polar (blue triangles) and 2B polar fitting the dipole moment (green squares) water models. The solid lines are a guide to the eye joining the Pareto points. The dispersed symbols are the corresponding values for the existing 2B polar literature models (circle⁵⁴, star⁵⁵).

Figure 7: Intermolecular parameters, deviations in saturated liquid density and vapor pressure, as well as in the critical parameter values for 2B polar water models obtained fitting also the dipole moment: P_i are the Pareto points calculated (star symbols connected by solid lines). Deviations in vapor pressure are shown in red and in saturated liquid density in green. The solid lines are a guide to the eye joining the Pareto points.

Figure 8: a) Saturated density, b) vapor pressure, c) deviation in saturated liquid density and d) deviation in vapor pressure versus temperature for water. The symbols correspond to correlated experimental data⁴⁴ while the curves correspond to the PC-SAFT (or its extended polar version) for the models: M_{4C} (dashed black), M_{2B} (dotted blue), M_{P2B} (dash-dotted green), $M_{P2B\mu}$ (dashed light blue) and $M_{P4C\mu}$ (solid red) following the nomenclature specified in Table 3. The region between the gray lines in plots b) and d) indicates the temperature range used in the optimizations.

Figure 9: Enthalpy of vaporization of water versus temperature. The symbols correspond to correlated experimental data⁴⁴ while the curves correspond to the PC-SAFT (or its extended polar version) for the models: M_{4C} (dashed black), M_{2B} (dotted blue), M_{P2B} (dash-dotted green), $M_{P2B\mu}$ (dashed light blue) and $M_{P4C\mu}$ (solid red) following the nomenclature specified in Table 3.