Computers and Chemical Engineering 167 (2022) 108015

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/cace

Multi-objective optimization of equation of state molecular parameters: SAFT-VR Mie models for water

Edward J. Graham [a], Esther Forte [b,1], Jakob Burger [c], Amparo Galindo [a], George Jackson [a], Claire S. Adjiman [a,*]

[a] Department of Chemical Engineering, Sargent Centre for Process Systems Engineering, Institute for Molecular Science and Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom
[b] Laboratory of Engineering Thermodynamics (LTD), University of Kaiserslautern, Erwin-Schrödinger-Str. 44, Kaiserslautern 67663, Germany
[c] Laboratory of Chemical Process Engineering, Campus Straubing for Biotechnology and Sustainability, Technical University of Munich, Schulgasse 16, 94315 Straubing, Germany

ARTICLE INFO

Keywords:
Multi-objective optimization
Equation of state
Parameter estimation
Sandwich algorithm
SAFT-VR Mie
Water

ABSTRACT

The determination of a suitable set of molecular interaction parameters for use with an equation of state (EoS) can be viewed as a multi-objective optimization (MOO) problem, where each objective quantifies the quality of the description for a particular type of thermodynamic property. We outline a methodology for the determination of a set of Pareto-optimal interaction parameters. The Pareto front is generated efficiently using a sandwich algorithm where one solves a sequence of weighted-sum scalarized single objective optimization problems. The algorithm presented can be used for any number of objective functions, allowing for the consideration of multiple thermodynamic property types as competing objectives in the MOO. The methodology is applied to the determination of suitable parameter sets for models of water within the SAFT-VR Mie framework. Three competing property targets are considered as objective functions: saturated-liquid density, vapour pressure and isobaric heat capacity. Two different types of molecular models are considered: spherical models of water, and non-spherical model of water. We analyse the two- and three-dimensional Pareto surfaces and parameter sets obtained for different property combinations in the MOO. The proposed methodology can be used to provide a rigorous comparison between different model types. Numerous Pareto-optimal parameter sets for SAFT-VR Mie water models are documented, and we recommend two new models (one spherical model and one non-spherical model) with an appropriate compromise between the competing objectives.

1. Introduction

The typical approach to determining the intermolecular parameters for use with an equation of state (EoS) is to solve a single objective optimization, where the objective function is constructed using a weighted-sum of various error functions, each representing the deviation between the prediction of a target thermodynamic property and the experimental measurement. The weighted-sum optimization, (WSP($\boldsymbol{w}$)), is defined as

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \boldsymbol{w}^T \boldsymbol{f}(\boldsymbol{x}) = \sum_{i=1}^{N_{\text{dim}}} w_i f_i(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in X, \end{aligned} \qquad \text{(WSP}(\boldsymbol{w})\text{)}$$

where f_i is the error function for property type i, N_{dim} is the number of property types, w_i is the weight given to property type i, and $\boldsymbol{x}$ is the vector of molecular parameters for the model which are to be estimated. X refers to the feasible set of parameters, which implicitly includes any inequality constraints, such as lower and upper bounds on the molecular parameters, or equality constraints, e.g., with the use of combining rules. There are various factors to consider in constructing a (WSP($\boldsymbol{w}$)), and the final molecular EoS model (to be used in e.g., a process simulation) will depend on the choice of the specific form of (WSP($\boldsymbol{w}$)). In particular, the definition of f_i, the choice of which property types to use, and the choice of $\boldsymbol{w}$.

The precise definition of f_i is important as it not only affects the values of the parameters but also their statistical properties (Englezos and Kalogerakis, 2000). An example of f_i is a relative least-squares

* Corresponding author.
E-mail address: c.adjiman@imperial.ac.uk (C.S. Adjiman).
[1] Present Address: Evonik Technology & Infrastructure GmbH, Roden-bacher Chaussee 4, 63457 Hanau-Wolfgang, Germany.

https://doi.org/10.1016/j.compchemeng.2022.108015
Received 23 February 2022; Received in revised form 23 September 2022; Accepted 23 September 2022
Available online 28 September 2022
0098-1354/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

objective function,

$$f_i(\boldsymbol{x}) = \sum_{j=1}^{N_i} \left[\frac{Y_{i,j}^{\text{exp.}}(\boldsymbol{k}) - Y_{i,j}^{\text{model.}}(\boldsymbol{k}, \boldsymbol{x})}{Y_{i,j}^{\text{exp.}}(\boldsymbol{k})} \right]^2, \tag{1}$$

where N_i is the number of data points for each property type i, $\boldsymbol{k}$ is a vector of independent variables that are fixed in the model and the experiment (e.g., temperature and pressure), $Y_{i,j}^{\text{exp.}}$ is the experimental value for property type i and data point j, and $Y_{i,j}^{\text{model.}}$ is obtained with the EoS with the corresponding molecular parameters for property type i and data point j. Typically, f_i is averaged over N_i such that f_i represents the average deviation for each property type, e.g.,

$$f_i(\boldsymbol{x}) = \frac{1}{N_i} \sum_{j=1}^{N_i} \left[\frac{Y_{i,j}^{\text{exp.}}(\boldsymbol{k}) - Y_{i,j}^{\text{model.}}(\boldsymbol{k}, \boldsymbol{x})}{Y_{i,j}^{\text{exp.}}(\boldsymbol{k})} \right]^2. \tag{2}$$

One should note that the objective functions defined in Eqs. (1) and (2) lead to the same functional form of (WSP($\boldsymbol{w}$)), so the constant prefactor $(1/N_i)$ present in Eq. (2) may either be used in the definition of f_i or subsumed into the weight w_i.

Due to the highly nonlinear behaviour of the more sophisticated thermodynamic models, finding the solution to (WSP($\boldsymbol{w}$)) can be challenging because the objective function may have a non-convex behaviour and exhibit many local optima. The use of global optimization routines in the model development can therefore lead to better solutions than local optimization methods (Costa et al., 2000). Furthermore, the calculation of fluid-phase equilibrium properties can be challenging as it is important to ensure that the results correspond to stable equilibria during the parameter estimation procedure. This becomes particularly challenging in the case of multi-component mixtures. Sophisticated methods have been developed to deal with this problem (see, e.g., Glass et al., 2018).

The values of the parameters and the individual f_i at the optimal solution to (WSP($\boldsymbol{w}$)) depend on the choice of weights, but it is generally difficult to decide which weight to assign to each property *a priori*. One class of methods suited for the determination of a suitable weight vector are Bayesian-type approaches. Common methods include the error-in-variables-measured (EVM), the weighted least squares (WLS) approach, and the maximum-likelihood estimator (Bard, 1974; Wang et al., 2018), where one seeks to find the set of parameters that are most statistically significant with respect to the experimental data. These methods require prior knowledge of the experimental uncertainty, such as the variance and distribution of measurement errors, which are sometimes not known or reported. Furthermore, if the equation of state does not predict the experimental data to within the range of experimental uncertainty, then weighting properties according to the inverse of their variance can lead to an unfair weighting towards certain thermodynamic properties.

For these reasons, the weight vector is often chosen arbitrarily in the development of EoS model parameters. In the majority of cases, an equal weight is given to each property type. In other cases, a small number of weight vectors are chosen and the thermodynamic model is picked by looking at the f_i obtained from the different models and judging which one gives the best balance between the different properties (see, e.g., Lafitte et al., 2007). Systematically choosing the most appropriate weight vectors is however not straightforward. It is often not apparent how changes in the weight vector will affect the individual f_i, particularly when the number of property types increases, as the dependence of f_i on the weights can be highly non-linear (Marler and Arora, 2009). Furthermore, there may not be an 'optimal' set of parameters for a particular compound or mixture. This will depend upon the intended use of the model, and a single set of parameters may not suffice.

An alternative approach is to formulate the parameter estimation as a multi-criteria (or multi-objective) optimization problem. Within this approach, the different thermodynamic properties are treated as competing objectives, and the output is a set of non-dominated (or Pareto-optimal) models, from which a model may be chosen after the parameter estimation procedure. Importantly, the multi-objective optimization approach removes any arbitrariness in the selection of the weight vector. Examples are parameter estimation for force-field models used in molecular simulation (Stöbener et al., 2014, 2016; Kulkarni et al., 2020), Gibbs energy models (Forte et al., 2020), or molecular parameters used in equations of state (Forte et al., 2017). The latter authors have developed Pareto-optimal models for water using the PC-SAFT EoS (Gross and Sadowski, 2000, 2001), using experimental data for the saturated-liquid density and vapour pressure as two competing objectives. They considered models for water which differ in their association scheme (2-site, 3-site and 4-site models), and whether or not to include a dipolar term (PCP-SAFT (Gross and Vrabec, 2006)). It was shown that knowledge of the Pareto front provides a useful means of comparison between different model types, as one is able to simultaneously compare the full set of Pareto-optimal solutions. They concluded that the 2- and 4-site association schemes provide very similar predictions, and that the addition of a dipolar term offers only a small improvement in the Pareto front if the literature value of the dipole moment is used, while a vast improvement in the Pareto front is seen if one treats the dipole moment of water as an adjustable parameter. Recently, Rehner and Gross (2020) used multi-objective optimization to develop PCP-SAFT parameters for water and alcohols. Two competing objective functions were considered, one incorporating data for the saturated-liquid density and vapour pressure with equal weights, and another incorporating interfacial-tension data modelled using predictive density gradient theory (pDGT (Rehner and Gross, 2018)). The description of properties was found to be improved significantly when a dipole moment is included in the molecular model. The Pareto fronts obtained using different association schemes for water (2-, 3- and 4-site) were compared, and it was found that every one of the three schemes considered dominates the other two in some part of the diagram, thus it is inconclusive which scheme is best for this particular set of objectives.

Typically, it is difficult to determine the complete set of Pareto points since a closed-form solution to the multi-objective optimization problem is not always available. The single objective optimization problem (WSP($\boldsymbol{w}$)) can be highly nonlinear and non-deterministic-polynomial (NP) hard. For such problems and for approaches that incorporate a scalarization method, the MOO is at best an infinite set of single-objective-optimization (SOO) problems. Thus, various methods have been developed to efficiently approximate the non-dominated set, the outputs of which are a finite number of Pareto points. The two main classes of approaches are stochastic and deterministic. Stochastic methods such as evolutionary algorithms work by sampling a large area of the parameter space and are not typically gradient based (Rangaiah, 2009). They can therefore be applied to any problem, regardless of the nature of the objective functions and constraints (Marler and Arora, 2004). A recent comparative study of various MOO methods applied to mixed-integer nonlinear problems has shown that stochastic methods do not always lead to the identification of the true Pareto front when non-convexities are present (Lee et al., 2020). Most deterministic methods make use of a particular type of scalarization, whereby the MOP is converted into an SOO, allowing for gradient-based techniques as solution methods. Common examples of scalarization techniques include: the ε-constraint method (Haimes, 1971), whereby inequality constraints are added to the individual objective functions; the Pascoletti–Serafini scalarization (Pascoletti and Serafini, 1984) which relates the objectives via an equality constraint; and the weighted-sum method (Zadeh, 1963). The latter scalarization method, as described in (WSP($\boldsymbol{w}$)) is commonly used for the parametrization of thermodynamic models.

In addition to specifying suitable weight vectors as an input to (WSP($\boldsymbol{w}$)), the choice of the particular data types used is of crucial importance. In the context of modelling of industrial processes, it is preferable that the equation of state can be used to predict

accurately both the fluid-phase equilibria and second-derivative thermodynamic properties, in particular, the caloric properties such as the heats of vaporization and isobaric heat capacities (Hendriks et al., 2010). The more recent versions of molecular-based SAFT EoSs that incorporate an adjustable repulsive interaction have been shown to provide an excellent simultaneous description of the fluid-phase equilibria and second-derivative properties, allowing for the possibility of including more data types in the model parametrization (Lafitte et al., 2007, 2013). A good representation of the experimental properties included in the objective function is, however, not the only concern. In molecular-based EoSs, where the parameters characterize the intermolecular potential, one needs to ensure that the model correctly captures the physics of the interactions in order for the equation to have any predictive capability with respect to, e.g., properties at other thermodynamic conditions, mixture properties, or the transferability of some potential parameters to chemically similar compounds. When SAFT parameters are estimated for real molecules, the identification of molecular models that provide a good approximation of a physically sound intermolecular potential becomes challenging if there is degeneracy in the parameters; in this case the experimental data do not provide enough information to distinguish between different models. Typically the extent of degeneracy increases with the number of parameters of the molecular model.

When chemical association is present and is explicitly accounted for in the molecular model, e.g., in the SAFT family of equations of state, it is clear from the literature that data types in addition to the vapour pressure and saturated-liquid density alone are required in order to decouple the degeneracy between the energy due to dispersive interactions from that due to association (hydrogen bonding or other short range directional interactions). This is apparent in the work of Clark et al. (2006) for the SAFT-VR SW equation of state, and Dufal et al. (2015) for the SAFT-VR Mie equation of state, when developing SAFT models for water. Using a weighted-sum of saturated-liquid density and vapour-pressure data, the authors observed large shallow regions in the objective function space when plotted against discretized pairs of parameters, particularly the dispersive energy and association energy. This indicates a high degree of degeneracy between these two parameters. Their discretization method allows for a visualization of the parameter space and allows one to generate a variety of models that do not necessarily minimize the objective function but are within a certain acceptable tolerance. By analysing these water models, Clark et al. (2006) found that the heat of vaporization was not a suitable data type to discriminate between the different SAFT models due to the similar prediction of this property for models along the "optimal" valley. The authors found that the use of the vapour–liquid surface tension calculated using the SAFT-VR DFT of Gloor et al. (2004) allowed one to discriminate between the different models, but the theoretical description overestimated the experimental surface tension in all cases due to the overestimate of the critical temperature with the theory. The predictions of the fraction of association sites not bonded between the different models were much more varied, and hence the authors were able to discriminate between models by comparison with the spectroscopic data of Luck (1980). Similarly, Dufal et al. (2015) were able to validate their water models (regressed to vapour–liquid equilibrium (VLE) data only) by comparison to the degree of association. In this case the authors used the degree of association predicted with Monte Carlo simulations of the SPC/E and TIP4P/2005 force fields as pseudo-experimental data, given issues with the interpretation of the spectroscopic data of Luck (1980). Cripwell et al. (2018) used a similar discretization approach to determine models for the SAFT-VR Mie-GV (polar variant of SAFT-VR) equation of state, whereby near-optimal solutions were discriminated with respect to their ability to predict vapour–liquid equilibria. Gloor (2003) found that the vapour–liquid interfacial tension is a useful property to determine the balance between dispersive and associative forces when coupling SAFT-VR SW with a suitable density functional theory (SAFT-VR DFT). Oliveira et al. (2016) demonstrated that in the case of strongly associating molecules, using only saturated-liquid density and vapour pressure as target properties (using the soft-SAFT equation of state (Blas and Vega, 1997)), one is unable to obtain a good prediction of various derivative properties (the speed of sound, isochoric heat capacity, thermal expansion coefficient, isothermal compressibility, and isobaric heat capacity). In their work they showed that the isobaric heat capacity, C_P, is much better described by a water model that accounts for non-sphericity.

The isobaric liquid heat capacity is a promising data type to use in the objective function as it can be measured accurately, and its accurate prediction is vitally important for the modelling of chemical processes. It also provides specific information on the nature of the liquid due to its relationship with entropy ($C_P = T(dS/dT)_P$). Cerdeiriña et al. (2007) showed that a simple two-state association model, TSAM (Cerdeiriña et al., 2004) is able to capture the various trends of the temperature dependence of $C_P(T)$, for specific isobars, and observed that this dependence for associating molecules is only sensitive to the ideal and association contributions to the isobaric heat capacity. This provides a good indication that one may be able to use this property to decouple the effect of the dispersive and associative forces. Note that in when one uses C_P to regress EoS models one also requires experimental information of the ideal gas heat capacity (or some means to estimate it accurately).

In summary, different property types (beyond the saturated-liquid density and vapour pressure alone) may be required for the estimation of equation-of-state model parameters from target experimental data. By constructing an optimization problem of the form described by (WSP($\boldsymbol{w}$)) for the regression, it is clear that increasing the number of property types makes it even more difficult for the user to define which weights should be specified in the optimization problem. In the current work, we extend the application of the multi-objective approach to EoS model parameter estimation as proposed by Forte et al. (2017) by considering more than two data types (or two objectives, f_i). The methods required to approximate the Pareto front are significantly more complex when dealing with more than two objective functions. We therefore provide a detailed description of the MOO problem and the algorithm used to approximate the Pareto front, along with graphical explanations to aid the explanation. The method is applied to the generation of a number of Pareto-optimal models for water using the SAFT-VR Mie (Lafitte et al., 2013; Dufal et al., 2015) and SAFT-γ Mie (Papaioannou et al., 2014) equations of state. We first discuss in detail the theory behind multi-objective optimization and methodologies for providing an efficient approximation of the Pareto front. We propose an algorithm in which a series of single objective optimizations (SOOs) is solved. It is ideally suited for the parametrization of EoS models, and can be retrofitted with existing methods and code bases, provided that the objective function in the SOOs is of the form described by (WSP($\boldsymbol{w}$)). The sandwich algorithm we implement for the determination of an efficient sequence of weight vectors used in (WSP($\boldsymbol{w}$)) is closely related to that of Bokrantz and Forsgren (2013), and is applicable to any number of dimensions (N_{dim}). We include three data types in the parameter estimation procedure: saturated-liquid density, vapour pressure, and isobaric heat capacity. The thermodynamic properties of water are notoriously difficult to predict with any EoS due to the complex interplay between associative, dispersive, and polar forces. This means the objectives are indeed conflicting, so that Pareto fronts are generated. Finally, we choose a particular model for water that represents a good compromise between the objectives, and we assess the ability of the model to predict properties that are not included in the objective function.

2. Methodology

2.1. Definition of the multi-objective optimization problem and non-dominance

The following multi-objective optimization problem (MOP) is considered:

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \dots, f_{N_{\text{dim}}}(\boldsymbol{x})] \\ \text{subject to} \quad & \boldsymbol{x}^{\text{LB}} \le \boldsymbol{x} \le \boldsymbol{x}^{\text{UB}}, \end{aligned} \tag{MOP}$$

where $\boldsymbol{f}$ is a vector of individual objective functions, in our case, some scalar functions $f_i, i = 1, \dots, N_{\text{dim}}$, that characterize the deviations between model and experiment for property type i; $\boldsymbol{x}$ is a vector of equation-of-state parameters that describe the molecular interactions, which we refer to as a "model" throughout; $N_{\text{dim}} \ge 2$ is the number of objectives; and superscripts LB and UB represent the lower and upper bound constraints on the model parameters, respectively. Throughout our work, the vector inequalities imply that the inequalities hold at the component-wise level. If it is possible to minimize all f_i simultaneously, i.e., the objective functions do not conflict, then the solution is a single optimal model. Otherwise, the solution is a (potentially infinite) set of models that are non-dominated. A model with parameters $\tilde{\boldsymbol{x}}$ is non-dominated if there exists no other feasible set of parameters $\boldsymbol{x}$ such that:

$$\boldsymbol{f}(\tilde{\boldsymbol{x}}) \in \boldsymbol{f}(\boldsymbol{x}) + C \setminus \{0\}, \tag{3}$$

where C is an ordering cone that is closed and pointed.

We follow the approach of Bokrantz and Forsgren (2013) by considering polyhedral ordering cones of the form $C = \{\mathbf{Q}\boldsymbol{\mu} : \boldsymbol{\mu} \ge 0\}$, whereby we specify the maximum admissible trade-offs between pairs of objectives to obtain $\mathbf{Q}$. Let t_{ij} ($t_{ij} \ge 0$) be the reciprocal of the maximum admissible increase in f_i per unit decrease in f_j. Then C can be represented by a set of inequality constraints, $C = \{\boldsymbol{z} : \mathbf{T}\boldsymbol{z} \ge 0\}$, where $\mathbf{T}$ is a matrix with dimensions N_{dim}, with ones on the diagonal and off-diagonal elements t_{ij}. The dual cone to C, C^*, is the cone generated by $\mathbf{T}$: $C^* = \{\mathbf{T}\boldsymbol{\mu} : \boldsymbol{\mu} \ge 0\}$.

To find the generating matrix $\mathbf{Q}$, we use Minkowski's theorem for closed convex pointed cones (Blekherman et al., 2012; Fawzi, 2017), which requires that any point in C can be described by the conical hull of its extreme rays. The extreme rays of C can be found by setting $(N_{\text{dim}} - 1)$ inequality constraints that define C to be active. In our work, we set all maximal admissible trade-offs between objective pairs to be equal and remove the subscripts: $t_{ij} = t$. This allows us to simplify the extreme ray representation of C. In three dimensions:

$$C = \left\{ \boldsymbol{z} : \begin{bmatrix} 1 & t & t \\ t & 1 & t \\ t & t & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \ge \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}. \tag{4}$$

Taking the first two inequality constraints to be active, we obtain $z_1 = z_2 = \frac{-t}{1+t} z_3$. Setting $z_3 = 1$ (the length of the extreme ray is arbitrary), we obtain an extreme ray and row of $\mathbf{Q}$: $[\frac{-t}{1+t}, \frac{-t}{1+t}, 1]$. The other extreme rays can be found in an equivalent manner by appealing to symmetry:

$$\mathbf{Q} = \begin{bmatrix} 1 & \frac{-t}{1+t} & \frac{-t}{1+t} \\ \frac{-t}{1+t} & 1 & \frac{-t}{1+t} \\ \frac{-t}{1+t} & \frac{-t}{1+t} & 1 \end{bmatrix}. \tag{5}$$

Using similar arguments for N_{dim} dimensions, $\mathbf{Q}$ is given by a matrix with ones along the diagonal and $\frac{-t}{1+(N_{\text{dim}}-2)t}$ as the off-diagonal elements. If $t = 0$ then $\mathbf{Q}$ and $\mathbf{T}$ are both identity matrices, which leads to the conventional Pareto ordering, i.e., a model $\boldsymbol{x}$ is Pareto-optimal if there is no other feasible model $\tilde{\boldsymbol{x}}$ such that $\boldsymbol{f}(\tilde{\boldsymbol{x}}) \le \boldsymbol{f}(\boldsymbol{x})$ with at least one strict inequality. In Fig. 1a we demonstrate this concept: a model is dominated if it lies anywhere above and to the right of another model.

Throughout, we refer to the complete set of Pareto-optimal models as the Pareto front, and refer to a discrete set of non-dominated models, e.g., as generated by an algorithm, by a matrix $\boldsymbol{\Theta}$, where the kth row ($k = 1, \dots, N_{\text{P}}$) corresponds to a non-dominated point which we refer to as $\boldsymbol{\theta}_k$. N_{P} is the number of Pareto points in $\boldsymbol{\Theta}$. The algorithm proposed in our paper will always produce points that are Pareto-optimal in the conventional sense, but some Pareto-optimal points may be discarded when $t > 0$ due to the dominance criterion, which removes parts of the Pareto front where trade-offs are unfavourable. While in Fig. 1a, we provide a visual interpretation of non-dominance and Pareto dominance in two dimensions for an ordering cone C generated by $t = 0$, in Fig. 1b, we illustrate how Pareto-optimal models are dominated when a broader ordering cone is chosen.

2.2. The weighted-sum scalarization and sandwich algorithms

In this section we first show how weighted-sum scalarizations of the MOP can be used to generate Pareto-optimal solutions and provide outer approximations to the Pareto front. We then show how inner approximations to the Pareto front can be constructed based on a discrete set of Pareto points $\boldsymbol{\Theta}$, and the ordering cone C, in the case that the MOP is a convex problem. Finally, we describe sandwich algorithms and how they can be used to generate Pareto points and inner (in the case of convex Pareto fronts) and outer approximations to the Pareto front efficiently.

2.2.1. The weighted-sum scalarization

We define the region Z as the set of all feasible vectors $\boldsymbol{z}$ in objective space, $Z = \{\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{x}) \mid \boldsymbol{x} \in X\}$. The minimization of a weighted-sum scalarization of the MOO can be interpreted geometrically as shifting a hyperplane (a line in 2 dimensions, a plane in 3 dimensions, etc.), with a normal vector which is equal to the weight vector, as close as possible to the origin in the feasible objective space. Solving the weighted-sum scalarized problem with the kth weight vector $\boldsymbol{w}_k$ that lies in C^* will give an optimal solution $\boldsymbol{x}^*$, corresponding to the kth non-dominated Pareto point, $\boldsymbol{\theta}_k = \boldsymbol{f}(\boldsymbol{x}^*)$. We can then define a hyperplane that has normal vector $\boldsymbol{w}_k$, and passes through $\boldsymbol{\theta}_k$, by $\boldsymbol{w}_k^T \boldsymbol{f} = \boldsymbol{w}_k^T \boldsymbol{\theta}_k = b_k$, and its associated positive half space, $\{\boldsymbol{z} \mid \boldsymbol{w}_k^T \boldsymbol{z} \ge b_k, \boldsymbol{z} \in Z\}$.

2.2.2. Definition of the outer approximation

By solving the weighted-sum scalarization with a single weight vector to global optimality, not only do we always obtain a Pareto-optimal solution, but we also obtain a positive half-space that provides a lower bound on the feasible region, and thus a lower bound on the Pareto front. Equivalently, there is no feasible solution that lies in the associated negative half-space. With multiple Pareto points and their associated weight vectors, we define an outer approximation to the feasible objective space, $Z \subseteq Z^{\text{out}}$, by the set of positive half-spaces:

$$Z^{\text{out}} = \{\boldsymbol{z} \mid \mathbf{W}^T \boldsymbol{z} \ge \boldsymbol{b}\}, \tag{6}$$

where $\mathbf{W}$ is a matrix with each row corresponding to a particular weight vector, and $\boldsymbol{b}$ is a column vector where the kth element is obtained by the scalar product between the kth row of $\mathbf{W}$ and the kth row of $\boldsymbol{\Theta}$. One should note that Z^{out} is an outer approximation even if the Pareto front is non-convex.

2.2.3. Definition of the inner approximation

An inner approximation to the Pareto front can be found via polyhedral approximations if the Pareto front is convex. Sufficient conditions for the convexity of (MOP) are that all $f_i, i = 1, \dots, N_{\text{dim}}$ are convex, the inequality constraint functions are convex, and the equality constraints are affine (Boyd and Vandenberghe, 2004). However, the convexity of the Pareto front only requires that the feasible region, Z, be *p-directionally convex* (Holtzman and Halkin, 1966), for any definitely negative vector $\boldsymbol{p}$ ($\boldsymbol{p} < \mathbf{0}$), as demonstrated by Lin (1976). This

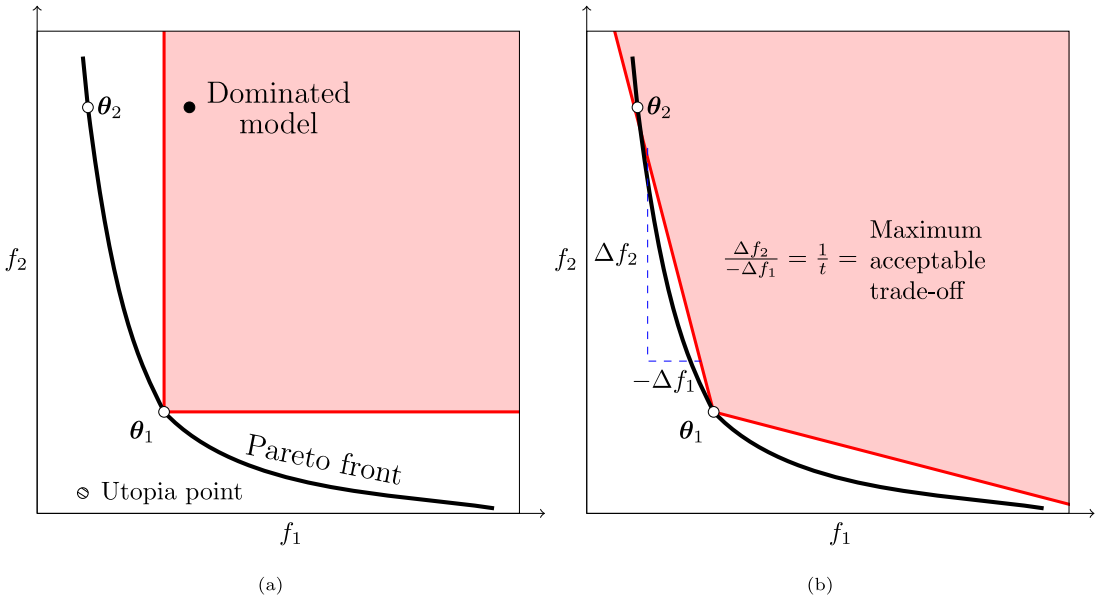


Fig. 1. Schematic illustrating Pareto-optimality and dominance in two dimensions. We show two Pareto points, θ_1 and θ_2 (open circles), that lie on the Pareto front (continuous black curve). The region shaded in red is $\theta_1 + C$; any model that lies in this region (e.g., the black circle) is dominated by θ_1. In (a), the ordering cone C ($t = 0$) implies the conventional Pareto ordering between objectives f_1 and f_2, where models are dominated if they are above and to the right of a Pareto-optimal model. In (b) we illustrate the dominance criterion for a broader ordering cone C when $t > 0$. We specify $1/t$, the maximum acceptable increase in f_2 (Δf_2) for a unit decrease in f_1 ($-\Delta f_1$), which corresponds to the gradient of the left-most extreme ray of C. This broader ordering cone leads to θ_2 becoming dominated.

is because the Pareto front is always on the boundary of Z (Marler and Arora, 2009). The definition of p–directionally convex is as follows (Marler and Arora, 2009):

Given a nonzero vector $\boldsymbol{p}$, Z is said to be p-directionally convex if given any two distinct points in Z, $\boldsymbol{z_1}$ and $\boldsymbol{z_2}$, and any two positive scalars, α_1 and α_2, with $\alpha_1 + \alpha_2 = 1$, there is a positive number β such that $\alpha_1 \boldsymbol{z_1} + \alpha_2 \boldsymbol{z_2} + \beta \boldsymbol{p} \in Z$.

The Pareto front may be convex even if $f_i, i = 1, \ldots, N_{\text{dim}}$, are non-convex functions. Assuming convexity of the Pareto front, an inner approximation can be defined by convex combinations of the Pareto points *plus* the ordering cone, $C = \{\mathbf{Q}\boldsymbol{\mu} : \boldsymbol{\mu} \geq 0\}$. The inner approximation is given by the set Z^{in} (Bokrantz and Forsgren, 2013):

$$Z^{\text{in}} = \{\boldsymbol{z} \mid \boldsymbol{\Theta}^T \boldsymbol{\lambda} + \mathbf{Q}^T \boldsymbol{\mu} : \ \boldsymbol{\lambda}, \boldsymbol{\mu} \geq 0 \quad \text{and} \quad \boldsymbol{e}^T \boldsymbol{\lambda} = 1\}, \tag{7}$$

where $\boldsymbol{e}$ is a vector of ones with the same dimensions as $\boldsymbol{\lambda}$. Any feasible choice of parameters $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ defines a point in objective space that must lie either above or on the Pareto front. In Fig. 2, we provide a geometric interpretation of Z^{in} and Z^{out}. For non-convex surfaces, the inner approximation described here will provide an upper bound to points that lie in the convex hull of the feasible region.

As every weight vector provides a Pareto-optimal solution to (WSP($\boldsymbol{w}$)), one may naively choose to run several optimizations with many weight vectors to produce the Pareto front, e.g., by using a grid of equally distributed weights. However, it is well known that an even distribution of weights does not typically produce an evenly distributed set of points on the Pareto front (Das and Dennis, 1997), i.e., points become clustered towards certain regions. A grid search may become numerically intractable, particularly as the number of dimensions increases.

Sandwich algorithms provide an efficient way to approximate the Pareto front. The output is a finite number of non-dominated points, and a polytope (or collection of facets) that approximates the true Pareto front to within a known degree of accuracy. In essence, sandwich algorithms provide increasingly tighter inner and outer approximations to the Pareto front by iteratively solving a sequence of weighted-sum scalarizations of the MOO, and they terminate once a user-defined convergence criterion is met. Assuming convexity of the MOO problem, the true Pareto front is 'sandwiched' between the inner and outer

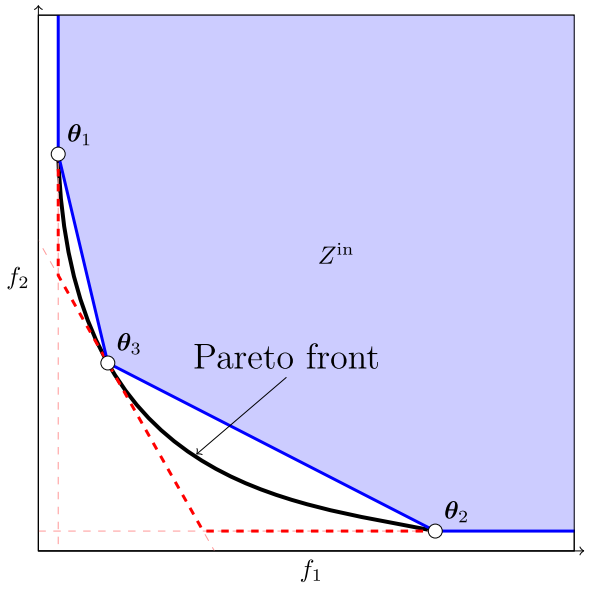


Fig. 2. Schematic illustrating the region described by the inner and outer approximations, Z^{in} and Z^{out} respectively, for three Pareto points, θ_1, θ_2, and θ_3 (white circles). The continuous black curve is the Pareto front which is convex. Z^{in} is defined by Eq. (7) . The lower edge of Z^{in} is shown by the straight blue lines, some of which connect the Pareto points, the others lie on the extreme ray of C that originates from the two extreme Pareto points θ_2 and θ_3. The blue shaded region is the convex hull of Z^{in}. The thick red dashed lines represent the lower region of Z^{out}, defined by Eq. (6), the thin red dashed lines represent redundant inequality constraints. The whole region defined by Z^{out} lies on or above the red lines. The Pareto front is 'sandwiched' in the region $Z^{\text{out}} \setminus Z^{\text{in}}$.

approximations. At each iteration of the algorithm, a new weight vector is selected so as to formulate the weighted-sum problem that leads to the largest improvement in the approximation quality. In this way, the algorithm is designed to obtain the desired approximation quality in as few weighted-sum optimizations as possible. In Fig. 3 we show the general idea behind the sandwich algorithm for two objectives, where

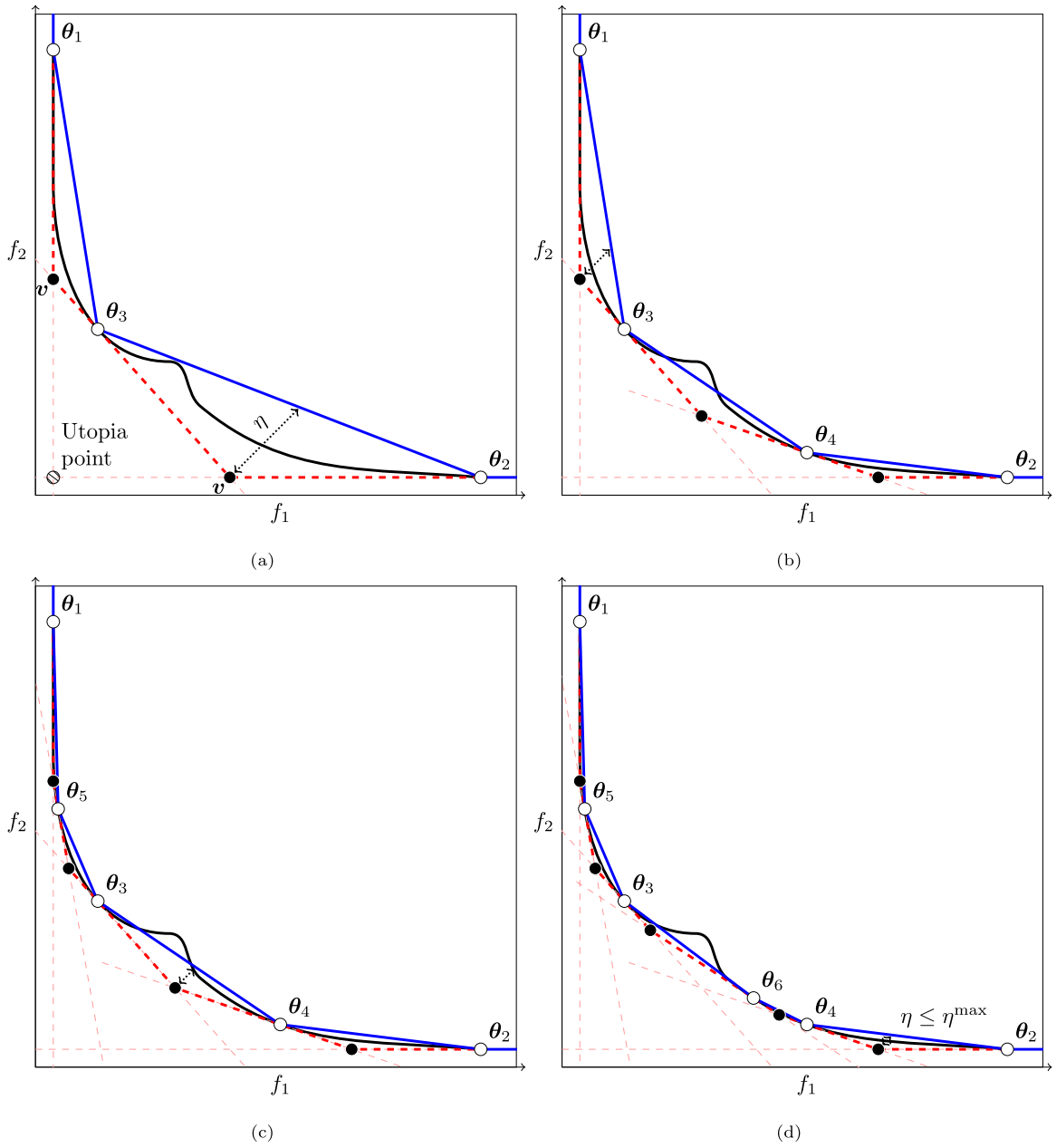


Fig. 3. Geometric interpretation of the sandwich algorithm for objectives f_1 and f_2 when there are nonconvex regions within the Pareto front. The continuous black curves represent the true Pareto front, discrete Pareto points are shown as white circles, the dashed red lines represent the outer approximation, the blue lines represent the inner approximation of the convex regions of the Pareto front, the black circles are the extreme vertices of the outer approximation. In (a), we start with an initial selection of Pareto points by solving the weighted-sum problem with three weight vectors and construct the inner and outer approximations. The Utopia point, a vector with the minimum value of each objective in the set of non-dominated solutions, is shown here. The area with the largest error between the inner and outer approximation (represented by the dotted line with double-headed arrows) is chosen and the normal vector to the facet of the inner approximation in this area is chosen as the scalarization vector for the next weighted-sum optimization. The error measure η is some scalar quantity that characterizes the distance between the inner and outer approximations, and the specific error measure (the Hausdorff distance) used in our work is defined in step 3 of the sandwich algorithm described in Section 2.3. This creates a new Pareto point (θ_4) in (b). Note that in (b) parts of the Pareto front that lie outside of the 'inner' approximation. In (c) a new Pareto point, θ_5 is created in the convex region of the Pareto front. The next largest error measure is then associated with the inner and outer approximations close to the non-convex region. In (d) the algorithm terminates since $\eta \leq \eta^{\max}$.

parts of the Pareto front are non-convex. Although the Pareto front is non-convex, the sandwich algorithm still converges.

2.3. Description of the sandwich algorithm used in our current work

In this work we use a sandwich algorithm that closely follows that of Bokrantz and Forsgren (2013). The reader is referred to section 4.2.4 of Graham (2020) for a comparison between different available sandwich algorithms and the reason we choose this particular algorithm. In Algorithm 1 we summarize the essential steps of the sandwich algorithm:

1. *Determining initial Pareto points*
   The first step is to compute an initial set of Θ Pareto points from which we can then construct Z^{in} and Z^{out}. The first N_{dim} Pareto points are found by solving a weighted-sum optimization with weight vectors equal to the extreme rays of C^* (the rows of T), with $\sum_{i=1}^{N_{dim}} w_i = 1$. We next compute a single Pareto point with

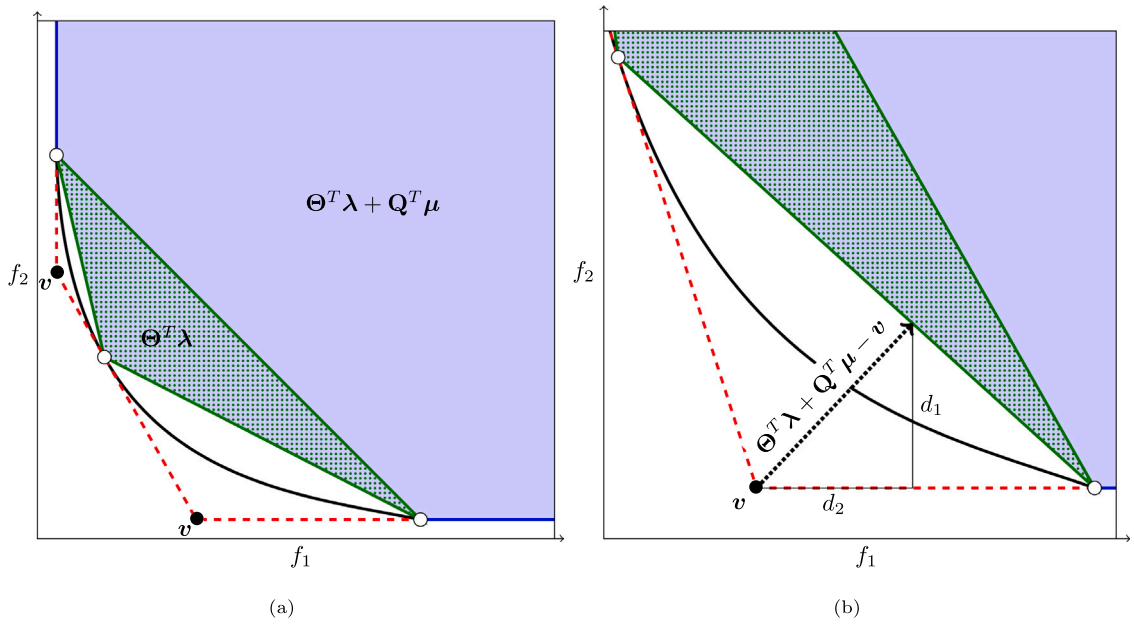


Fig. 4. Geometric interpretation of the minimization of the error measure in two dimensions. The white circles are the computed Pareto points; the region patterned with green dots is the convex hull of the Pareto points, $H = \{\mathbf{\Theta}^T\lambda : \lambda \geq 0, e^T\lambda = 1\}$; the region shaded in blue is the obtainable region when adding the ordering cone to each point in the convex hull. In (b) we zoom in on an extreme vertex of the outer approximation, $\boldsymbol{v}$, to show the geometric interpretation of the error measure, η, where $\eta = \max(d_1, d_2)$.

Algorithm 1 Sandwich algorithm, based on that of Bokrantz and Forsgren (2013)

input:
- A multi-objective optimization problem.
- A quality threshold, $\eta^{\max}$ (default 0.01).
- Reciprocal of the maximum admissible trade-offs between objective pairs, t (default 0.01).

output:
- A set of non-dominated points, $\mathbf{\Theta}$.
- A polyhedron that approximates the Pareto front.

begin

1. Generate initial Pareto points:
   Solve (WSP($\boldsymbol{w}$)) with $\boldsymbol{w}$ equal to the extreme rays of C^*.
   Solve (WSP($\boldsymbol{w}$)) with an equal weight vector.
   Initialize Z^{in} and Z^{out}.
   Set $\eta = \infty$.
2. Main loop:

while $\eta > \eta^{\max}$ **do** ▷ Terminate if quality criteria is met

3. Compute the extreme vertices, $\mathbf{V}$, of the outer approximation. For each $\boldsymbol{v}$ in $\mathbf{V}$, compute η' by solving (PLP($\boldsymbol{v}$)) and set $\eta = \max(\eta, \eta')$.
4. For the solution of (PLP($\boldsymbol{v}$)) corresponding to η, set the next weight vector, $\boldsymbol{w}^*$, to be normal to Z^{in} at the optimal solution.
5. Solve (WSP($\boldsymbol{w}$)) with $\boldsymbol{w}^*$ and update $\mathbf{\Theta}$ and Z^{out}.
6. Remove Pareto points from $\mathbf{\Theta}$ and inequality constraints from Z^{out} that correspond to sub-optimal solutions to (WSP($\boldsymbol{w}$)).

end while

end

equal weights and $\sum_{i=1}^{N_{dim}} w_i = 1$. The $N_{dim} + 1$ Pareto points and their corresponding weights are appended to matrices $\mathbf{\Theta}$ and $\mathbf{W}$ respectively, and $\boldsymbol{b}$ is initialized for the outer approximation.

2. *Main loop*

   The algorithm enters the main iteration loop, which runs until the desired error tolerance $\eta^{\max}$ is achieved.

3. *Determining the error measure between Z^{out} and Z^{in}*

The error criterion η that defines the deviation between Z^{in} and Z^{out} is the smallest value of ϵ such that $Z^{\text{out}} \subseteq (Z^{\text{in}} - \epsilon\boldsymbol{e})$, where $\boldsymbol{e}$ is the vector of ones. This error measure is equivalent to the Hausdorff distance. Since $Z^{\text{out}} \supset Z^{\text{in}}$, the Hausdorff distance is given by Bokrantz and Forsgren (2013)

$$d_{\text{Hausdorff}} = \max_{z \in Z^{\text{out}}} \min_{z' \in Z^{\text{in}}} d(\boldsymbol{z}, \boldsymbol{z}'), \tag{8}$$

where d is the distance function representing the maximum positive distance between two points in any dimension of the objective space,

$$d(\boldsymbol{z}, \boldsymbol{z}') = \max_{i=1,\ldots,N_{dim}} (0, z'_i - z_i). \tag{9}$$

One should note that the norm used in Eq. (9) is a design choice and may be replaced with a different norm (e.g., the Euclidean norm). η can then be determined by solving the bi-level optimization problem (Bokrantz and Forsgren, 2013),

$$\begin{aligned} \max_{z \in Z^{\text{out}}} \min_{\eta,\lambda,\boldsymbol{\mu}} \quad & \eta \\ \text{s.t.} \quad & \eta\boldsymbol{e} \geq \mathbf{\Theta}^T\lambda + \mathbf{Q}^T\boldsymbol{\mu} - \boldsymbol{z} \\ & \boldsymbol{e}^T\lambda = 1 \\ & \eta, \lambda, \boldsymbol{\mu} \geq 0, \end{aligned} \tag{BLP}$$

where $\boldsymbol{e}$ is a row vector of ones with length defined by its context, λ is a column vector corresponding to the weights used in a convex combination of the Pareto points, $\boldsymbol{\mu}$ is a column vector corresponding to the weights used in a conic combination of the extreme rays of the ordering cone $\mathbf{Q}$. Here, we solve (BLP) by enumerating candidate values for $\boldsymbol{z}$ in the outer problem. As Z^{out} is represented as a set of linear constraints, it can be shown (Bokrantz and Forsgren, 2013) that the values of $\boldsymbol{z}$ that are optimal in (BLP) lie on the extreme vertices of the outer approximation, i.e., points where the hyperplanes that define the outer approximation intersect. Denoting an extreme vertex by $\boldsymbol{v}$, the following linear programming problem is solved for each $\boldsymbol{v}$:

$$\begin{aligned} \min_{\eta,\lambda,\boldsymbol{\mu}} \quad & \eta \\ \text{s.t.} \quad & \eta\boldsymbol{e} \geq \mathbf{\Theta}^T\lambda + \mathbf{Q}^T\boldsymbol{\mu} - \boldsymbol{v} \\ & \boldsymbol{e}^T\lambda = 1 \\ & \eta, \lambda, \boldsymbol{\mu} \geq 0. \end{aligned} \tag{PLP($\boldsymbol{v}$)}$$

The extreme vertices are calculated using the "CON2VERT" function available on MATLAB file exchange (Kleder, 2005) , which returns a set of vertices given a system of inequalities (in our case, $\mathbf{W}^T\boldsymbol{z} \geq \boldsymbol{b}$). The function employs a primal–dual polytope method, which requires that the constraints are bounded. To ensure this, we provide a lower bound $\boldsymbol{z} \geq 0$, and a large upper bound, $\boldsymbol{z} \leq 10^{10}$. The extreme vertices that lie at the upper bound are removed from the function output.

A geometric interpretation of (PLP($\boldsymbol{v}$)) for a particular $\boldsymbol{v}$ is shown in Fig. 4. In Fig. 4a, we show the region (in blue) defined by the convex hull of the Pareto points, $H = \{\boldsymbol{\Theta}^T\lambda : \lambda \geq 0, \boldsymbol{e}^T\lambda = 1$ and the same region augmented by the ordering cone defined by $C = \{\mathbf{Q}^T\boldsymbol{\mu} : \boldsymbol{\mu} \geq 0\}$. In Fig. 4b, we show a geometric interpretation of the right-hand side of the first set of inequality constraints in (PLP($\boldsymbol{v}$)), where d_1 and d_2 are the z_1 and z_2 components of the vector $\boldsymbol{\Theta}^T\lambda + \mathbf{Q}^T\boldsymbol{\mu} - \boldsymbol{v}$. It is clear that because η is minimized, the optimal solution will lie on an inner edge of $\boldsymbol{\Theta}^T\lambda + \mathbf{Q}^T\boldsymbol{\mu}$, and $\eta = \max(d_1, d_2)$. Additionally, $\eta \geq 0$, so both d_1 and d_2 vary between 0 and some positive value, so there is a solution where $d_1 = d_2$. At this point of equality of the two distance metrics, any reduction in either d_1 or d_2 leads to an increase in the other distance, hence, at the solution, the first set of inequality constraints in (PLP($\boldsymbol{v}$)) is active and $\eta = d_1 = d_2$. In general, the first set of inequality constraints will always be active, as the supporting hyperplane of Z^{in} at the optimal solution will have a normal vector with only positive elements by construction. Hence, at a point on a hyperplane where all d_i are equal, it is not possible to choose another point on the hyperplane without increasing at least one of the d_is.

Let $\boldsymbol{v}^*$ be the extreme vertex that gives the largest value of η in (PLP($\boldsymbol{v}$)). If this value is below a user-defined tolerance $\eta^{\max}$, then the algorithm terminates here, as the inner and outer approximations are suitably close. Otherwise we continue to step 4.

4. *Choosing the next weight vector to run*

   After finding $\boldsymbol{v}^*$, the next step is to find the weight vector that is normal to Z^{in} at $\boldsymbol{\Theta}^T\lambda^* + \mathbf{Q}^T\boldsymbol{\mu}^*$, where λ^* and $\boldsymbol{\mu}^*$ are the optimal solutions to PLP($\boldsymbol{v}^*$), to use in the next weighted-sum problem. In Bokrantz and Forsgren (2013), the dual problem to (PLP($\boldsymbol{v}$)) is solved in order to find the next weight vector. Instead, we use the property that the first set of inequality constraints in (PLP($\boldsymbol{v}$)) are active, and therefore the vector of Lagrange multipliers (or dual variables) for these constraints, provided by the linear optimization programme, is precisely the normal vector needed. This avoids having to solve a separate linear programming problem. We denote the next weight vector to be run by $\boldsymbol{w}^*$.

5. *Solving* (WSP($\boldsymbol{w}$))

   If (WSP($\boldsymbol{w}$)) is a non-convex problem the output of the chosen optimization routine can be highly dependent on the initial parameter guess, as one may get trapped in minima that are locally but not globally optimal. To address this issue, we use a multi-start approach whereby initial guesses for $\boldsymbol{x}$ are chosen, which lie within the parameter bounds $\boldsymbol{x}^{\text{LB}}$ and $\boldsymbol{x}^{\text{UB}}$, based on a Sobol' sequence (Sobol', 1967). A Sobol' sequence is chosen as opposed to random sampling as its use leads to a more even coverage of the domain of the parameter space for a given number of generated points. Lee et al. (2020) showed that generating initial parameter guesses using a Sobol' sequence is an effective way to solve the weighted sum SOO to global optimality, thus generating points on the Pareto front. We choose to use powers of 2 for the number of Sobol' points due to local minima observed in the "discrepancy", a measure of how uniformly the parameter space is sampled (Morokoff and Caflisch, 1994). The default value of Sobol' points is 2048.

   Once an initial number, N_s (default 40), of Pareto points and their corresponding parameters are determined with a large number of Sobol' points, subsequent optimizations are run by using parameters already obtained from previous Pareto points (nearby solutions in objective space) and their convex combinations. This implicitly assumes that nearby points on the Pareto front will have similar parameters. The point $\boldsymbol{\Theta}^T\lambda^* + \mathbf{Q}^T\boldsymbol{\mu}^*$ lies on a hyperplane that connects the points $\boldsymbol{\theta}_k$, where k corresponds to the indices of the non-zero elements of λ^*. We expect that this point will lie close to the true Pareto point obtained in the next weighted-sum optimization. Hence, sensible initial guesses to solve (WSP($\boldsymbol{w}$)) are $\boldsymbol{x}_k$ and the convex combination: $\boldsymbol{\lambda}_{\boldsymbol{k}}^{*T}\boldsymbol{x}_k$. Note that this technique may lead to sub-optimal points being identified if there are discontinuities in the optimal set of Pareto-optimal parameters. The likelihood of this issue arising can be mitigated by increasing N_s.

   After solving (WSP($\boldsymbol{w}$)), the new Pareto point is appended to $\boldsymbol{\Theta}$ and $\boldsymbol{w}^*$ is appended to $\mathbf{W}$.

6. *Removing Pareto points*

   At each addition of a new Pareto point, we perform checks to ensure that all of the Pareto points in $\boldsymbol{\Theta}$ lie within Z^{out}. If this is not the case, then some of solutions to (WSP($\boldsymbol{w}$)) obtained in previous iterations are not globally optimal, or at least optimal with respect to the other Pareto points. For each weight vector $\boldsymbol{w}_k$ in $\mathbf{W}$ we compute the scalar product of this weight vector and all of the Pareto points. If the minimum scalar product does not correspond to $\boldsymbol{w}_k^T\boldsymbol{\theta}_k$, i.e., another Pareto point lies below the hyperplane associated with $\boldsymbol{w}_k$ in the outer approximation, then we remove the kth row from $\boldsymbol{\Theta}$ and $\mathbf{W}$. Note that this check will also remove any dominated points from $\boldsymbol{\Theta}$, and will help with the convergence of the sandwich algorithm. At this point we return to step 2 of the algorithm.

The following bullet points summarize how our algorithm differs from that of Bokrantz and Forsgren (2013):

- In step 4, a different method is used in order to determine the weight vector to be used in (WSP($\boldsymbol{w}$)), making use of the properties of the solutions to (PLP($\boldsymbol{v}$)).
- A global optimization routine is used to solve (WSP($\boldsymbol{w}$)) since the objective function may be highly non-convex, and a strategy is developed in order to provide good initial guesses to the optimization and reduce the computational effort in finding the optimal solution to (WSP($\boldsymbol{w}$)). These points are addressed in step 5. of the algorithm.
- After solving (WSP($\boldsymbol{w}$)) and appending the new point to $\boldsymbol{\Theta}$, we check if all Pareto points lie within Z^{out}. If this is not the case, we remove Pareto points $\boldsymbol{\theta}_k$ for which their lower-bounding half-space $\{\boldsymbol{z}|\boldsymbol{w}_{\boldsymbol{k}}^T\boldsymbol{z} \geq b_k\}$ lies above any of the other Pareto points. This is to ensure that solutions to (WSP($\boldsymbol{w}$)) are globally optimal, or at least optimal with respect to the other Pareto points generated. This is detailed in step 6 of the algorithm.

3. Application of the MOO approach to parameter estimation of equation-of-state models

In this section we first present an appropriate form of the objective functions f_i, and then apply the sandwich algorithm to develop accurate SAFT-VR Mie models for water in order to illustrate the benefits of the MOO approach for this challenging task.

3.1. Definition of f_i

The individual objective functions are defined as follows:

$$f_i(\boldsymbol{x}) = \frac{10^4}{N_i}\sum_{j=1}^{N_i}\left[\frac{Y_{i,j}^{\text{exp.}}(\boldsymbol{k}) - Y_{i,j}^{\text{model.}}(\boldsymbol{k},\boldsymbol{x})}{Y_{i,j}^{\text{exp.}}(\boldsymbol{k})}\right]^2, \tag{10}$$

where the notation is defined as in Eq. (1). For each experiment, an equal weight is given to each data point. In our work this is a valid assumption since a constant relative error is quoted for the data considered. The prefactor of 10^4 is used in this instance so that the values of f_i on the Pareto surface are of a suitable order of magnitude. It is common to normalize the objectives after determining the initial set of Pareto points in step 1 of the algorithm, but we decide not to normalize here because the f_i, described in Eq. (10), are non-dimensional and we want to avoid any bias towards particular objectives. This least-squares type of objective function is chosen as it is mathematically well-behaved and it is typically used in the estimation of EoS model parameters. We also report the absolute average deviation (AAD(i)%) for property type i to describe the accuracy of the fit to experimental data since it provides a more intuitive indication of the deviation, although we stress that this is not used within the objective function:

$$\mathrm{AAD}(i)\%(\boldsymbol{x}) = \frac{100}{N_i}\sum_{j=1}^{N_i}\left|\frac{Y_{i,j}^{\mathrm{exp.}} - Y_{i,j}^{\mathrm{model.}}(\boldsymbol{x})}{Y_{i,j}^{\mathrm{exp.}}}\right|. \tag{11}$$

3.2. Application of the MOO approach to SAFT-VR Mie models for water

We will now apply the proposed algorithm to develop a water model for use within SAFT-VR Mie (or SAFT-γ Mie) EoS using up to *three* properties. In the introductory Section 1 we outlined the benefits of adding property types in the parameter estimation, and explained why the heat capacity C_P is a suitable property to add above the standard properties such as the saturated-liquid density ρ_L^{sat} and vapour pressure P^{vap}. We will therefore develop models using these three property types within the MOP.

3.2.1. Review of SAFT models for water

In this section we provide a brief overview of how SAFT-type equations of state can be used to model water, the archetypal associating fluid. We will review some of the proposed models and their development. This will help decide which particular set of SAFT-VR Mie parameters can be defined before the parameter estimation, in particular the parameters that define the association scheme and the sphericity of the model. A review of the existing models will also give us an idea of the values the parameters will take, and will help us specify suitable parameter bounds $\boldsymbol{x}^{\mathrm{LB}}$ and $\boldsymbol{x}^{\mathrm{UB}}$.

Within SAFT approaches, molecules are modelled as chains of spherical segments with repulsive cores, which may be either hard (e.g., the square-well potential), or soft (e.g., the Mie or Lennard-Jones potential). Association is mediated by adding off-centre, spherically symmetrical square-well bonding sites with short-range attractive interactions.

In the case of water, these association sites are chosen to represent the directional, short-range interactions representative of a hydrogen bond. It is widely known that water can form up to four hydrogen bonds, as seen in hexagonal ice, and water is therefore typically modelled using a four-site association scheme (4C in the notation proposed by Huang and Radosz (1990)). With a 4C model, two association sites of type e and two association sites of type H are used, which correspond to the two lone pairs of electrons on the oxygen atom and the two hydrogen atoms, respectively. In such a model, only e–H interactions are allowed to represent the hydrogen bonding. Due to the symmetry of the association scheme, all sites have the same fraction of nonbonded sites ($X_{\mathrm{e}} = X_{\mathrm{H}}$). This represents the statistical consideration that all of the sites are equally likely to participate in hydrogen bonding, and thus effects such as bond cooperativity (Sear and Jackson, 1996) are typically neglected in SAFT models for water. Other association schemes have been proposed to model water with SAFT using the 2B (one e and one H) or 3B (one e and two H) association schemes (Clark et al., 2006). The 4C scheme is generally chosen, however, since it is verified by molecular orbital calculations (Wolbach and Sandler, 1997), it is successful in representing pure component VLE properties (see e.g., Clark et al., 2006), and transferring the parameters for this model to aqueous mixtures generally provides a better representation than the other association schemes (Kontogeorgis and Folas, 2010). Furthermore, it is apparent from the work of Forte et al. (2017) that the non-polar 4C association scheme outperforms the 2B and 3B association schemes for the majority of the Pareto front, with objective functions defined by AAD(ρ_L^{sat})% and AAD(P^{vap})%.

Another important parameter to consider when defining a water model is m, the number of fused spherical segments used to represent the non-sphericity of the molecule. In Fig. 5 we show a spherical and non-spherical SAFT model for water using a 4C association scheme.

In Table 1 we summarize some of the parameters reported for SAFT water models, extending on the review by Kontogeorgis and Folas (2010) by including some of the more recently developed models of relevance to our current work. When the value of m is fixed, it is typically set to $m = 1$, representing a spherical model. In such cases, the chain contribution to the SAFT free energy (used to describe non-spherical molecules) is omitted. This model type is generally chosen due to its simplicity and due to its rigorously defined geometry. The other models are non-spherical and have values of m that vary quite broadly $0.98 \leq m \leq 3.792$. Quantum mechanical calculations can provide an approximate indication of the value m should take: Sheldon et al. (2006) derived values of m for water by mapping the shapes of the electron orbitals obtained from Hartree–Fock calculations onto a sphero-cylinder, and estimated a value of $m = 1.0968$ for the SAFT-VR EoS.

In Table 1 we report the values of the dispersion energy and association energy for the various water models. There is a large variation in the values of the dispersion energy (between 42.8 K and 740.5 K), and a large variation in association energy (between 825 K and 2507 K). Although one cannot directly compare the parameter values between the different SAFT theories, these ranges provide a rough guide of the order of magnitude the parameters should take.

An additional choice when defining a SAFT model for water is whether or not to explicitly include the long-range dipolar interactions present in water. Extensions to the theory to include this type of interaction have been proposed (see e.g., Müller and Gubbins, 1995; Gross and Vrabec, 2006). In earlier work, Müller and Gubbins (1995) extended the original SAFT EOS (Huang and Radosz, 1990) to account for multipolar interactions, and developed a model for water. In the recent work of Forte et al. (2017), a variety of water models were obtained for PC-SAFT (Gross and Sadowski, 2000, 2001) and the polar version PPC-SAFT (Gross and Vrabec, 2006) to account for the dipolar interaction. They developed non-polar water models for various association schemes (2B, 3B and 4C) and developed polar water models for the 2B and 4C schemes. The authors showed that using the experimental value of the dipole moment of water in the gas phase ($\mu = 1.86$ D), the polar 2B model outperforms the non-polar 2B model Pareto surface with objective functions AAD(ρ_L^{sat})% and AAD(P^{vap})% at all points. However, they found that there is a region of the Pareto front (close to the Pareto-knee where there is a large curvature in the Pareto front), where the two Pareto fronts essentially coincide, with AAD(ρ_L^{sat})% and AAD(P^{vap})% ~ 0.5. The authors found that by adjusting μ to ρ_L^{sat} and P^{vap} for the polar 2B and 4C models they were able to capture a good agreement with respect to both properties (AAD(ρ_L^{sat})% and AAD(P^{vap})% < 0.1). In our work we will not consider the explicit treatment of the dipolar interaction; the orientational and average energetic features of a dipolar fluid (which give rise to a Keesom–Lennard-Jones like interaction) will be accounted for through the regression of the typical SAFT parameters to the experimental data.

Following the arguments made in this brief review and in Section 1, we consider the development of water models for the SAFT-VR Mie EoS employing the 4C association scheme. For the specific version of the SAFT-VR Mie EoS that we consider, there is only one existing model in the literature, where water is treated as a spherical molecule ($m = 1$) (Dufal et al., 2015). In our current work, we also consider models where $m \geq 1$ by treating m as an adjustable parameter, supported by



Fig. 5. Schematic for two possible SAFT models for water. The molecule is represented by a number of spherical segments (m), which interact via a potential (in SAFT-VR Mie this is the Mie potential) with a repulsive core diameter σ. The Mie potential is characterized by a repulsive exponent (λ^r), an attractive exponent (λ^a), and a well depth ϵ. These segments are indicated by the spheres shaded red. In (a) the molecule is assumed to be spherical with just one segment ($m = 1$). In (b) the molecule is non-spherical and $1 < m < 2$, represented by two fused spheres. In (a) and (b) the hydrogen bonding is represented by a 4-site association scheme (2 e sites and 2 H sites), corresponding to the two lone pairs of electrons on the oxygen atom and the two hydrogen atoms, respectively. These association sites interact via a square-well potential with an association interaction $-\epsilon^{HB}_{e,H}$ and bonding volume $K^{HB}_{e,H}$. Only e–H interactions are allowed.

Table 1
Selected parameters reported for various SAFT models of water.
Source: The table is adapted from Kontogeorgis and Folas (2010) and extended to include the more recent models.

SAFT Variant	Reference	Non-sphericity m	Dispersion energy/ K	Association energy/ K
Original SAFT 4c (+ polar)	Müller and Gubbins (1995)	1	224.0	1415.3
Original SAFT 4c	Müller et al. (1996)	1	90	3600
SRK-CPA	Kontogeorgis et al. (1999)	–	–	2003
CK-SAFT	Huang and Radosz (1990)	1.179	528.2	1809
SAFT 3 site	Economou and Tsonopoulos (1997)	1.179	528.2	1809
SAFT 4 site	Economou and Tsonopoulos (1997)	1.236	431.7	1368
Simplified SAFT	Fu and Sandler (1995)	2	188.2	826
CK-SAFT	Button and Gubbins (1999)	1.047	504.4	1365
Original SAFT, 4 site	Li and Englezos (2004)	0.980	433.9	1195
PSAFT	Karakatsani et al. (2005)	1	52.1	1982
PC-PSAFT	Karakatsani et al. (2005)	1	42.8	1973
CK-SAFT	Boulougouris et al. (2001)	2.850	167.0	1634
PR-CPA	Wu and Prausnitz (1998)	–	–	1477
SAFT-VR	Patel et al. (2003)	1	253.0	1366
CK-SAFT 3 site HF	Wolbach and Sandler (1997)	1.278	385.1	2286
CK-SAFT 4 site HF	Wolbach and Sandler (1997)	1.406	212.9	1809
CK-SAFT 3 site DFT	Wolbach and Sandler (1997)	1	615.9	1627
CK-SAFT 4 site DFT	Wolbach and Sandler (1997)	1	546.6	1237
APACT 2 site	Economou and Donohue (1992)			2418
APACT 3 site	Economou and Donohue (1992)			2618
Original SAFT 4C	Li and Englezos (2003)	0.982	433.9	1195
SRK-CPA	Voutsas et al. (2000)	–		1794
CK-SAFT	Voutsas et al. (2000)	2.853	167.1	1635
PC-PSAFT	Karakatsani and Economou (2006)	1.750	169.5	1131
tPC-PCAFT	Karakatsani and Economou (2006)	1.600	58.1	1640
tPC-PSAFFT 4C	Karakatsani et al. (2006)	2.815	150.7	1575
PC-SAFT 2B	Gross and Sadowski (2002)	1.066	366.5	2501
sPC-SAFT 4C	Grenner et al. (2006)	1.500	180.3	1804
sPC-SAFT 4C	Grenner et al. (2007)	2.610	140.4	1695
PC-SAFT 3B	Kleiner and Sadowski (2007)	3.254	196.2	1801
PC-SAFT 4C	Kleiner and Sadowski (2007)	3.792	138.6	1718
PC-SAFT 4C	Forte et al. (2017)	2.500	147.9	1582
PC-SAFT 2B	Forte et al. (2017)	1.785	233.8	2507
PCP-SAFT 2B	Forte et al. (2017)	2.216	214.5	1544
SAFT-VR Mie (LJ Kernel) 4C	Dufal et al. (2015)	1	266.7	1985
SAFT-VR Mie (Mie Kernel) 4C	Dufal et al. (2015)	1	418.0	1600
SAFT-VR SW 4C	Clark et al. (2006)	1	250.0	1400
SAFT-VR SW 4C	Sheldon et al. (2006)	1.0968	740.5	367

evidence from quantum mechanics (Sheldon et al., 2006) and from the majority of reported models which have a value of m different to 1. Through application of the proposed MOO approach and analysis of the Pareto fronts this provides a rigorous comparison between the two model types, and will make clear if it is beneficial to choose a non-spherical model over a (simpler) spherical model.

3.3. The SAFT-VR Mie EoS

For a comprehensive description of the SAFT-VR Mie EoS, the reader is referred to Lafitte et al. (2013), and the modifications to the association term described in Dufal et al. (2015, 2018). Specifically, the equations used to model the association contribution are those

Table 2
Description of the parameters used to characterize water.

	Units	Description
m	–	Number of spherical segments
σ	Å	Size of each spherical segment
λ^r	–	Repulsive exponent of the Mie potential
λ^a	–	Attractive exponent of the Mie potential
ϵ/k_B	K	Depth of the Mie potential
$\epsilon^{HB}_{e,H}/k_B$	K	Depth of association potential between sites e and H
$K^{HB}_{e,H}$	Å^3	Bonding volume between sites e and H
NST_e	–	Number of site types of type e
NST_H	–	Number of site types of type H

Table 3
Upper and lower bounds on parameters.

	m	σ/Å	λ^r	λ^a	(ϵ/k_B)/K	$(\epsilon^{HB}_{e,H}/k_B)$/K	$K^{HB}_{e,H}$/Å^3
Lower bound	1	2.5	8	6	100	1000	0.1
Upper bound	2	3.5	40	6	500	2500	500

which use the Lennard-Jones reference fluid as the basis for the free energy perturbation, as opposed to using the generalized Mie fluid as in Dufal et al. (2015). The representation of molecules with the SAFT-VR Mie EoS is identical to the group-contribution counterpart, SAFT-γ Mie (Papaioannou et al., 2014), when the molecules consist of identical functional groups. The ideal gas contribution to the heat capacity is calculated using the third order polynomial of Poling et al. (2001).

3.4. The choice of water model

The SAFT parameters that characterize the water model, along with their physical descriptions, are summarized in Table 2.

The attractive exponent of the Mie potential, λ^a, is fixed to 6, corresponding to the attractive range of the London dispersion force as expected from a simple quantum-mechanical description of the dispersion forces. Furthermore, it has been shown that a conformal description of the thermodynamics can be achieved with an interrelationship between λ^r and λ^a (Ramrattan et al., 2015), negating the need to consider the attractive and repulsive ranges independently. Following the arguments made in Section 3.2.1, we choose to use a four-site association scheme to model water, where $NST_e = 2$ and $NST_H = 2$. A schematic for this model type is shown in Fig. 5. For spherical models of water, the value of m is fixed to 1. In Table 3 we define the lower and upper bounds for each parameter.

4. Results

The multi-objective optimization technique is applied to two types of water model: 'spherical' models, where $m = 1$, and 'non-spherical' models where $m \geq 1$ following the arguments made in Section 3.2.1. These two model types are considered as one may prefer to choose a simpler spherical model over a non-spherical model. Here we show in detail the level of improvement that one may obtain by treating m as adjustable.

We apply the MOO technique to the development of water models where three different experimental property types are considered as competing objectives: saturated-liquid density (ρ_L^{sat}), saturated vapour pressure (P^{vap}), and isobaric heat capacity (C_P) data.

In the first scenario, we analyse the two-dimensional Pareto fronts for spherical and non-spherical models of water when the two typically considered experimental properties, saturated-liquid density (ρ_L^{sat}) and vapour pressure (P^{vap}), are used as objective functions in the MOP. These are the most common data types used to regress pure component EoS parameters (Kontogeorgis and Folas, 2010), and therefore provide some insight into the objective space and the models typically obtained when solving (WSP($\boldsymbol{w}$)).

In the second scenario, we analyse the three-dimensional Pareto fronts where C_P is included in the MOP. This will highlight how the addition of a second-derivative caloric property affects the objective functions and the corresponding water models.

Finally, some preferable Pareto-optimal water models are chosen from the Pareto front and the level of agreement with respect to the experimental data is analysed, including the prediction of thermodynamic properties not included in the objective function.

4.1. Calculation details

The experimental data for water is taken from the National Institute of Standards and Technology (NIST) database (Lemmon et al., 2018). Saturation property data are taken for temperatures between the triple point of water, 273.16 K, and 613.16 K, to ensure that temperatures do not exceed 95% of the critical point. This is the same range of temperatures considered by Dufal et al. (2015) and therefore the models obtained in our current work provide a useful comparison with the previously reported SAFT-VR Mie model. We choose data points with temperature intervals of 10 K following Forte et al. (2017). The experimental values for the isobaric liquid-phase heat capacity are taken at 1 atm, between the triple point and 10 K below the saturation temperature, i.e., 273.16 K–363.16 K. We choose only one isobar at 1 atm since the liquid-phase isobaric heat capacity does not vary significantly with pressure over this temperature range. As the heat capacity function used here requires temperature and pressure as inputs, temperatures close to the saturation temperature are not included in the estimation to limit the risk of evaluating a gas-phase heat capacity. This prevents having large discontinuities in the objective function. The uncertainties in the experimental values of the thermodynamic properties examined are expected to be below 0.1% within the temperature and pressure ranges considered (Lemmon et al., 2018).

The sandwich algorithm is implemented in MATLAB R2018a, whilst the optimization problems are solved externally using a Levenberg–Marquardt algorithm as implemented in the 'lmfit' module in Python (Newville et al., 2016), with the default options. The thermodynamic properties within the optimization routine are evaluated with gSAFT (Lafitte et al., 2017), which has an in-built flash algorithm to solve for phase equilibria for the SAFT-VR Mie EoS. To determine the convex hull of the set of Pareto points and the normal vectors to the facets, the C++ implementation of Qhull (Barber et al., 1996) is used.

In all cases, we set the reciprocal of the maximum admissible trade-off between objective pairs to be $t = 0.01$. This represents a relatively narrow ordering cone, but limits the non-dominated solutions to have reasonable trade-offs, and limits the f_i such that they do not take on extremely large values that would not be acceptable for thermodynamic modelling. Furthermore, fixing $t > 0$ as opposed to $t = 0$ avoids generating weight vectors with zero elements which can lead to weakly dominated solutions.

The maximum admissible error is set to $\eta^{max} = 0.1$ in all cases, meaning that any point on the inner edge of Z^{in} will not be further than 0.1 from the true Pareto front in any dimension of the objective space, assuming that the Pareto front is convex. To solve (WSP($\boldsymbol{w}$)), a total of 2048 Sobol' points are used for the first 40 Pareto points calculated ($N_s = 40$ in step 5 of the algorithm). This number of Sobol' points has been found to be sufficiently large to provide a high degree of confidence that a global solution to (WSP($\boldsymbol{w}$)) is found. Specifically, we have solved (WSP($\boldsymbol{w}$)) with an equal weight given to ρ_L^{sat}, P^{vap}, and C_P, i.e., $\boldsymbol{w} = [1/3, 1/3, 1/3]^T$, and allowing m to vary, using 2048 Sobol' points. Of these, 712 of the solutions have an objective function that is within 0.01% of the best known objective function. The first model within 0.01% of the best known objective function value is found using the third initial guess generated by the Sobol' sequence. This provides confidence that the first 40 Pareto points are globally optimal. For the remaining 1336 Sobol' points, a number result in failure to converge as the corresponding initial parameter values yield a phase

envelope that is very far from the experimental data (specifically with many temperatures at which a single phase is predicted instead of the experimentally-observed vapour–liquid equilibrium); other points result in convergence to higher objective function values, typically corresponding to a bound-constrained solution.

Further optimizations are run using the initial guesses for the parameters of the previous Pareto points and their convex combinations as described in step 5.

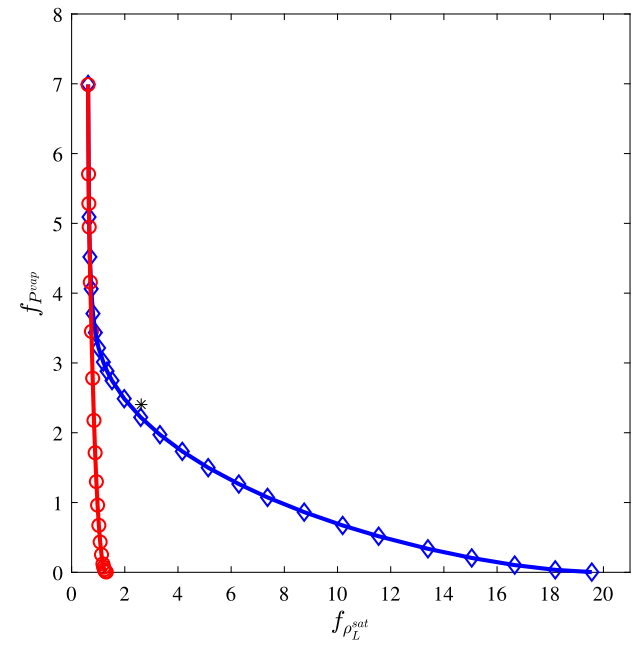
4.2. Scenario 1: Two-dimensional MOP with data types ρ_L^{sat} and P^{vap}

In Fig. 6a, we display the inner edge of Z^{in} for the two Pareto fronts and in Fig. 6b we display the corresponding AAD% for the saturated-liquid density and vapour-pressure objectives. In Appendix A.1 we provide information attributed to each Pareto point in Table A.6 (spherical) and Table A.7 (non-spherical). It is clear from Fig. 6 that the proposed sandwich algorithm works efficiently as the Pareto points are equally distributed along the Pareto front. This equal distribution is achieved with a sequence of weight vectors generated by the sandwich algorithm, and these weight vectors are not evenly distributed (particularly in the case of the non-spherical models, c.f. Table A.7). This suggests that the sandwich algorithm is more efficient than a brute-force approach (e.g., solving (WSP($\boldsymbol{w}$)) with uniformly distributed weight vectors). The true Pareto front appears to be convex in both cases as there are no obvious areas with large gaps in the Pareto front. We are therefore confident that we have efficiently captured a good approximation of the full set of non-dominated solutions for the two dimensional MOP of both of the model types, and any model that lies on the Pareto front will be acceptable. By visual inspection of the Pareto fronts, it is possible to make rigorous comparisons between the two model types. It is clear that by relaxing the constraint on m, we are able to obtain significantly lower values of $f_{\rho_L^{sat}}$ where the deviation in vapour pressure is also small. For example, if we desire a water model with a description of the vapour pressure within AAD(P^{vap})% $\leq$ 0.5% then the non-spherical model is able to capture the liquid density within AAD(ρ_L^{sat})% $\leq$ 1%, whereas for the spherical model the corresponding description is AAD(ρ_L^{sat})% $\geq$ 3%. The Pareto fronts coincide at the lowest values of $f_{\rho_L^{sat}}$ because m hits the lower bound of 1 where the two model types are identical.

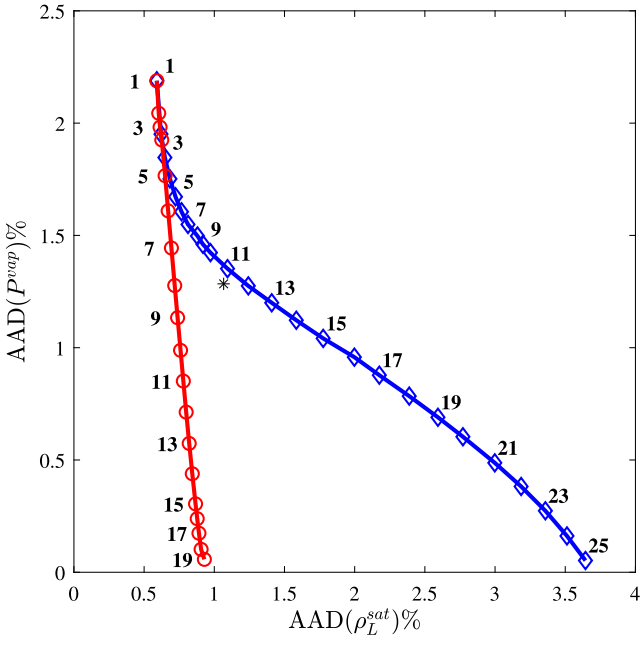
In Figs. 7 and 8, we show the predictions of ρ_L^{sat} and P^{vap} in addition to the isobaric heat capacity C_P, for the non-dominated models that only consider ρ_L^{sat} and P^{vap} in the objective function. These figures correspond to the Pareto points as labelled in Fig. 6b.

For spherical models (Fig. 7), the prediction of the isobaric liquid-phase heat capacity is relatively poor, with f_{C_P} and AAD(C_P)% exceeding 110 and 10%, respectively, at points towards the centre of the Pareto front. If we consider the 'standard' model that is obtained by weighting the two objectives equally (point 8), we obtain a model with an AAD% vector [AAD(ρ_L^{sat})%, AAD(P^{vap})%, AAD(C_P)%] = [0.88,1.50,10.45]%. This corresponds to a fairly significant deviation of C_P. At point 24, corresponding to a weight vector of [$w_{\rho_L^{sat}}$, $w_{P^{vap}}$] = [0.0335,0.9665], where most of the weight is given to the vapour pressure, we obtain the AAD% vector [3.51,0.16,0.45]%. Comparing point 24 with point 8, point 24 comes with a small sacrifice in AAD(ρ_L^{sat})% (an increase of 2.63 percentage points) against a small improvement in AAD(P^{vap})%, with an increase of 1.34 percentage points and a vast improvement in AAD(C_P)% of 10 percentage points. This illustrates that a model that is potentially preferable may be obtained by assigning an extreme weight vector in (WSP($\boldsymbol{w}$)) (i.e., a weight vector that would generally not be chosen *a priori*). Numerous Pareto points obtained using the proposed technique may be evaluated against other properties not included in the regression (e.g., property types that are practically difficult to include within the objective function).

For the Pareto-optimal non-spherical models (Fig. 8) we find that the prediction in C_P (AAD(C_P)%) is between 0.4% and 7.8%. The point obtained with equal weighting of the two objective functions is point 17 with [AAD(ρ_L^{sat})%, AAD(P^{vap})%, AAD(C_P)%] = [0.89,0.17,3.68]%. Note that AAD(P^{vap})% is fairly close to the expected experimental uncertainty (relative error 0.1%) reported by NIST (Lemmon et al., 2018). Depending on the application, model 12 is potentially preferable with an AAD% vector [0.80,0.71,0.41]%, and is obtained with [$w_{\rho_L^{sat}}$, $w_{P^{vap}}$] = [0.85,0.16].



(a)



(b)

Fig. 6. (a) Pareto frontiers for the two objectives f_P and $f_{\rho_L^{sat}}$. The blue diamonds are the calculated Pareto points for spherical models of water where $m = 1$, the blue curve is the inner approximation of the Pareto front. The red circles are the calculated Pareto points for non-spherical models for water where $m \geq 1$ and the red curve is the inner approximation of the Pareto front. (b) The corresponding AAD% for each point is shown, and the curves are a guide to the eye joining the Pareto points. The black star is the result obtained when using the parameters for the spherical model reported by Dufal et al. (2015).

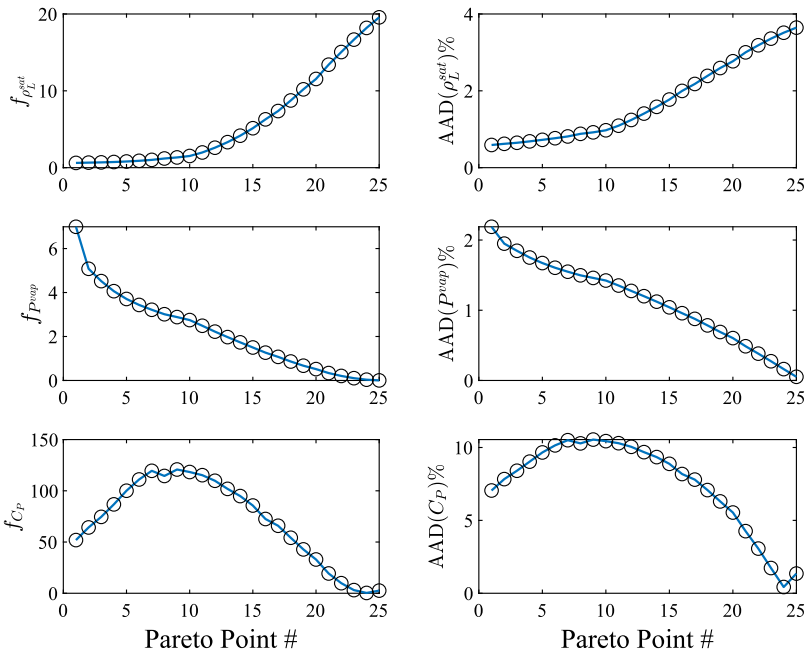


Fig. 7. The individual objectives f_i and the corresponding AAD(i)% for saturated-liquid density ρ_L^{sat} and vapour pressure P^{vap} of the non-dominated models of spherical water ($m = 1$) where only $f_{\rho_L^{sat}}$ and $f_{P^{vap}}$ are considered in the MOP. The deviation of the prediction of the isobaric heat capacity C_P is also shown.

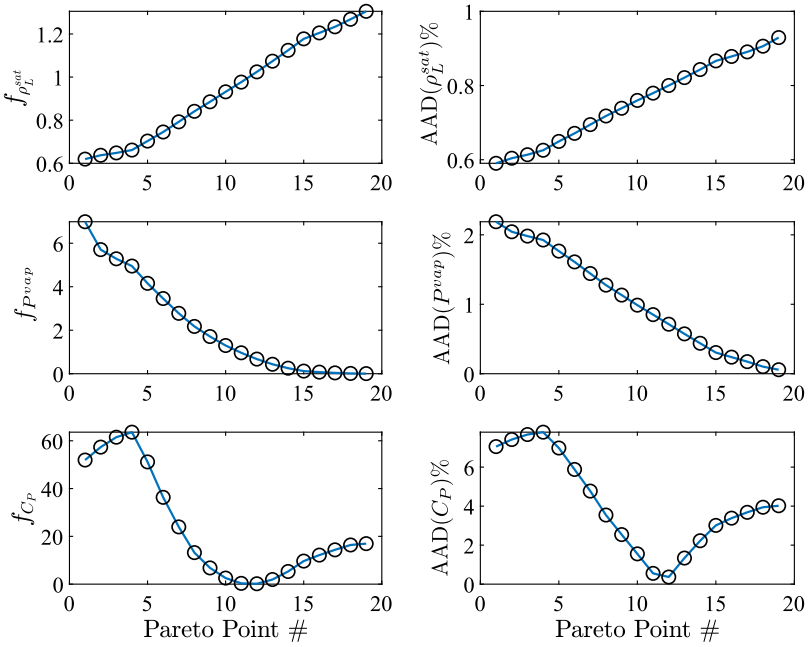


Fig. 8. The individual objectives f_i and the corresponding AAD(i)% for saturated-liquid density ρ_L^{sat} and vapour pressure P^{vap} of the non-dominated models of non-spherical water ($m \geq 1$) where only $f_{\rho_L^{sat}}$ and $f_{P^{vap}}$ are considered in the MOP. The deviation of the prediction of the isobaric heat capacity C_P is also shown.

From the analysis of the Pareto fronts obtained when ρ_L^{sat} and P^{vap} are used in (WSP($\boldsymbol{w}$)), C_P is predicted accurately for some Pareto points but other points present significant deviations. This may serve as an indication that VLE data alone does not provide sufficient information in predicting this second-derivative caloric property, and that there is some degeneracy in the parameter space.

In Fig. 9 we show the trends in parameters versus the Pareto point number, with indices indicated in Fig. 6b. Along the Pareto front for spherical models, the values of the dispersion energy ϵ/k_B vary between ~270 K and ~440 K and those of the association energy $\epsilon_{e,H}^{HB}/k_B$ between ~1550 K and ~1980 K. These represent relatively large ranges for the two energetic parameters, and there is a clear inverse relationship between the two along the Pareto front, which is again indicative of degeneracy. The value of the segment size parameter σ remains fairly constant at just over 3 Å. The size σ and energy ϵ are characterized by an almost identical curvature, and similarly with the repulsive range λ^r and bonding-volume $K_{e,H}^{HB}$ parameters, indicating a close relationship between these two pairs of parameters and how they affect the individual objectives.

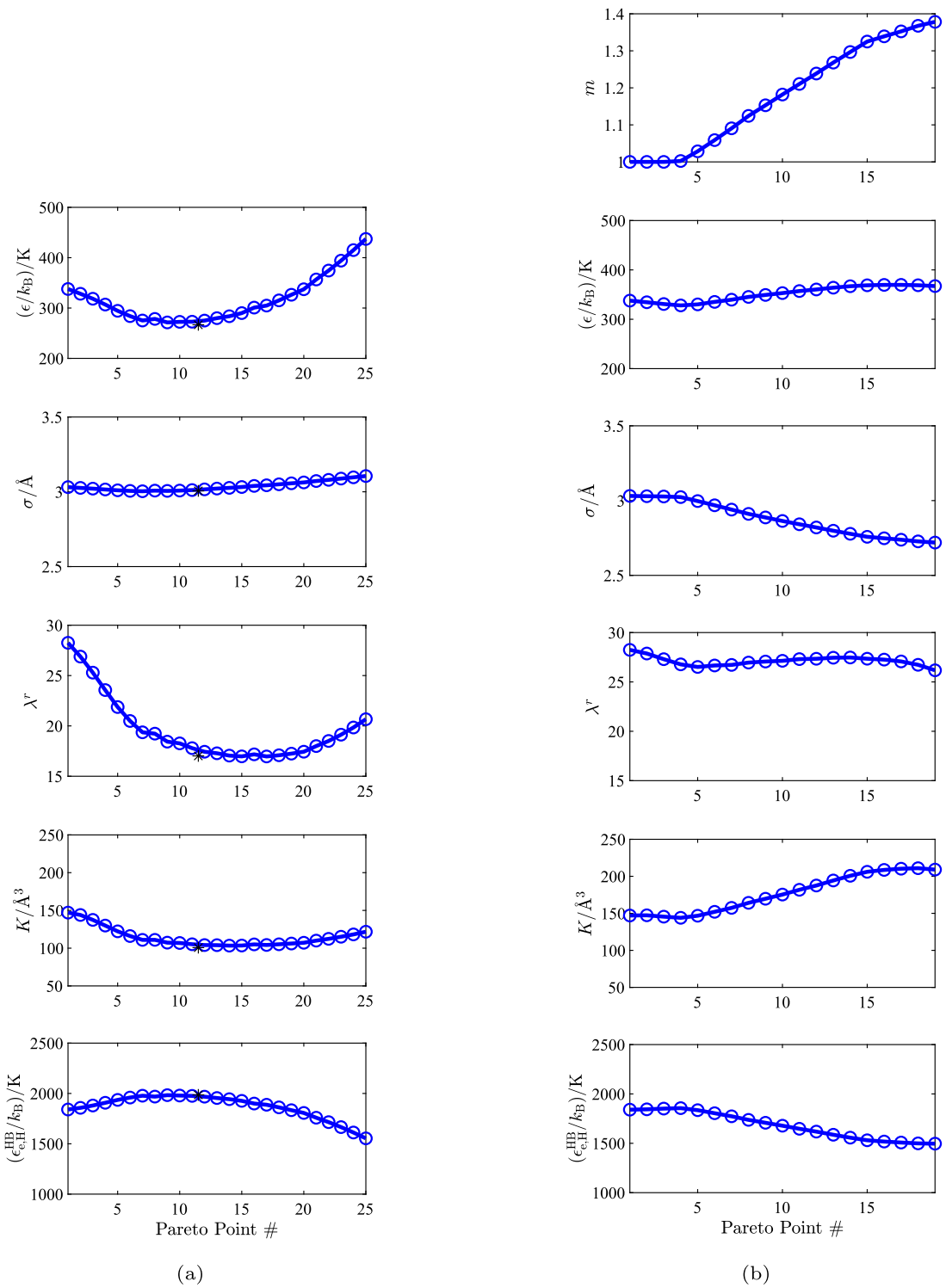


Fig. 9. Pareto fronts in parameter space. m is the non-sphericity, ϵ is the dispersion energy, σ is the size, λ^r is the repulsive exponent, K is the bonding volume, and $\epsilon^{HB}_{c,H}$ is the association energy. The numbers on the x axis correspond to the Pareto points labelled in Fig. 6b. (a) Pareto-optimal parameters for the spherical water models where $m = 1$. The black stars correspond to the set of parameters in Dufal et al. (2015). (b)Pareto-optimal parameters for the non-spherical water models where $m \geq 1$.

The model of Dufal et al. (2015) lies close to the calculated Pareto curve for spherical models of water, and these models are close to the Pareto knee. We obtain very similar parameters to those of Dufal et al. (2015) in this region. The parameters for models of water where $m \geq 1$ are shown in Fig. 9. There is a clear correlation between the non-sphericity parameter m and the segment size σ. The value of the repulsive exponent λ^r remains fairly constant and exhibits both a minimum and a maximum along the Pareto front.

4.3. Scenario 2: Three-dimensional MOP with data types ρ_L^{sat}, P^{vap}, and C_P

In the second scenario, we analyse the Pareto fronts when we include the liquid-phase isobaric heat capacity C_P as an additional property in the MOP.

Before proceeding with the MOP, the effect of adding C_P on a single weighted-sum objective function compared to the two-dimensional case

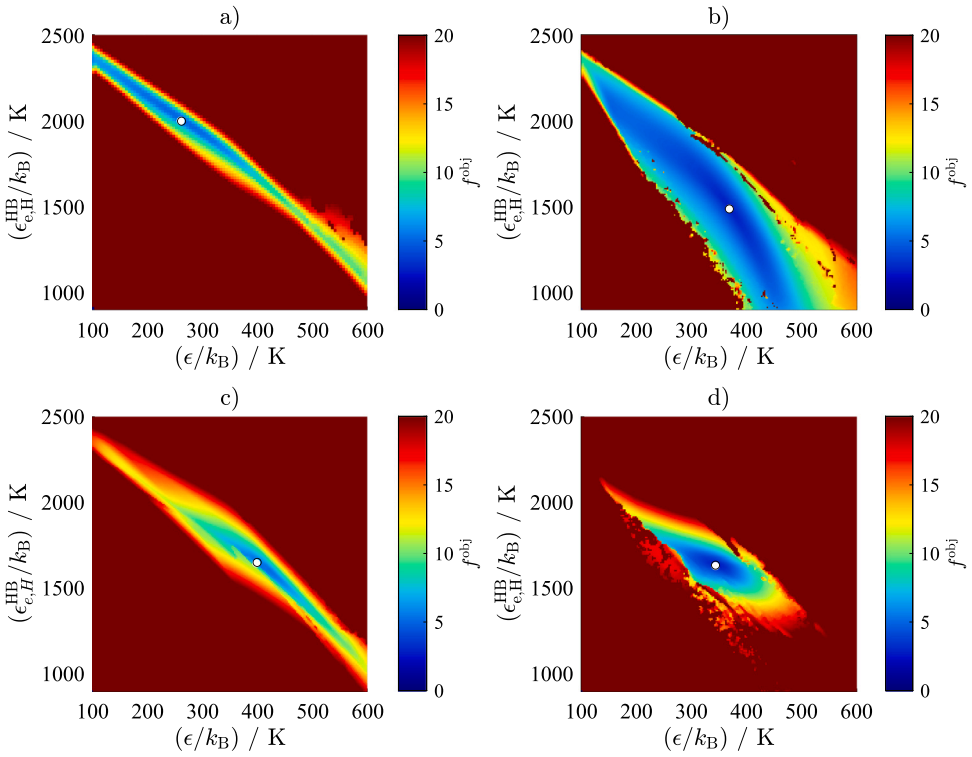


Fig. 10. The weighted-sum objective function for water models where the dispersion and association energy parameters ϵ and $\epsilon^{HB}_{e,H}$ are discretized in a 200 x 200 grid with each parameter varying between the upper and lower bounds defined in Table 3. Each point represents an individual optimization where ϵ and $\epsilon^{HB}_{e,H}$, are fixed and the other parameters are optimized. The weight vectors are normalized to sum to 1 and an equal weighting is given to each property type. The value of the objective function at the optimal solution, $f^{obj} = \boldsymbol{w}^T \boldsymbol{f}(\boldsymbol{x}^*)$, is indicated by colour. The global minimum is indicated by a white circle. The four sub-figures are different in the type of water model and the types of experimental data used in the optimization: (a) Spherical water with ρ_L^{sat} and P^{vap} used as experimental data types. (b) Non-spherical water with ρ_L^{sat} and P^{vap} used as experimental data types. (c) Spherical water with ρ_L^{sat}, P^{vap}, and C_P used as experimental data types. d) Non-spherical water with ρ_L^{sat}, P^{vap} and C_P used as experimental data types.

is investigated. In Fig. 10 we show a contour plot of the weighted-sum objective function (with equal weighting given to ρ_L^{sat}, P^{vap}, and C_P if included) versus discretized pairs of the dispersion and association energy parameters ϵ and $\epsilon^{HB}_{e,H}$. A similar analysis was conducted by Clark et al. (2006) and Dufal et al. (2015). It is apparent that the objective function surface consisting of only saturated-liquid density and vapour pressure has a large shallow region of relatively similar values. The addition of C_P data in panels b and d reduces the size of the region with similar values, and reduces the degeneracy between ϵ and $\epsilon^{HB}_{e,H}$. We also observe that treating the non-sphericity m as an adjustable parameter reduces the parameter degeneracy. The combination of adding C_P as an objective and treating m as adjustable has the most dramatic effect on reducing the size of the region with low values of the weighted sum.

In Fig. 11, we show the Pareto front as a contour plot, along with the corresponding AAD% for each parameter. When the Pareto front is plotted in 3 dimensions, the inner edge of Z^{in} is shown without the addition of the ordering cone, which is equivalent to plotting the facets of the convex hull of $\boldsymbol{\Theta}$ that consist of only positive inward-facing normals. This is to ensure that models in-between the Pareto points can be determined from the convex combinations of the parameters at the facet vertices. Information on each Pareto point is tabulated in Table A.8. We first note that the Pareto points are evenly distributed throughout the Pareto front and that the Pareto front appears to be convex since there are no obvious regions with large gaps. Thus, the sandwich algorithm allows one to efficiently identify a series of weight vectors in (WSP($\boldsymbol{w}$)). The Pareto front highlights the trade-offs that are involved in choosing a particular model, and it is clear that the three objectives are indeed conflicting. The Pareto-optimal solutions of the two dimensional MOP when the heat capacity is not considered in the objective function is also shown. These models are not found on the 3D front because the gradients between objective functions are extremely steep within this area, and hence the models become dominated with respect to the specified ordering cone.

In Fig. 12 we show the Pareto front for the non-spherical water models for the three objectives, and display information for each Pareto point in Table A.9. The surface is extremely steep at the inner edge, so the same Pareto front by swapping the f_{C_P} and $f_{P^{vap}}$ coordinate axes is shown in Fig. 13. This steepness provides a good indication that liquid-phase isobaric heat capacity as a data type is highly suited to reducing the parameter degeneracy, since providing a very small weight to C_P leads to a highly accurate prediction of this property with a negligible decrease in the quality of the representation of the vapour pressure and saturated-liquid density. We also note that all of the parameters are within a relatively small range. The non-spherical models can be used to predict all three properties extremely well (all properties to within AAD%=0.8), and the deviations are well below those attainable with spherical water models. It is apparent that in some regions of the Pareto front, the heat capacity is not conflicting with the other two objectives. For example, in the region where $f_{\rho_L^{sat}} \sim 1$, the 3D Pareto front essentially coincides with the 2D Pareto front where C_P is not considered in the objective function. If we consider the 2D Pareto point number 7 in Table A.7 with an AAD(P^{vap})% of 0.7% in vapour pressure, the corresponding AAD(C_P)% is ~5. We can choose a model that does equally well in terms of vapour pressure and saturated-liquid density but with an AAD(C_P)% = 0.2 with a small weight given to C_P of 0.07 (Pareto point 9 in Table A.9). In Fig. 13, this region where C_P does not conflict with the other two objectives is visible where $f_{\rho_L^{sat}}$ is approximately 1 due to the white area above this value. Points in this region of the Pareto front are dominated due to the limit of steepness defined in the ordering cone C, rather than the surface being non-convex in this area. Although not shown here, we are able to obtain Pareto-optimal points within this region by specifying a very small value for w_{C_P}.

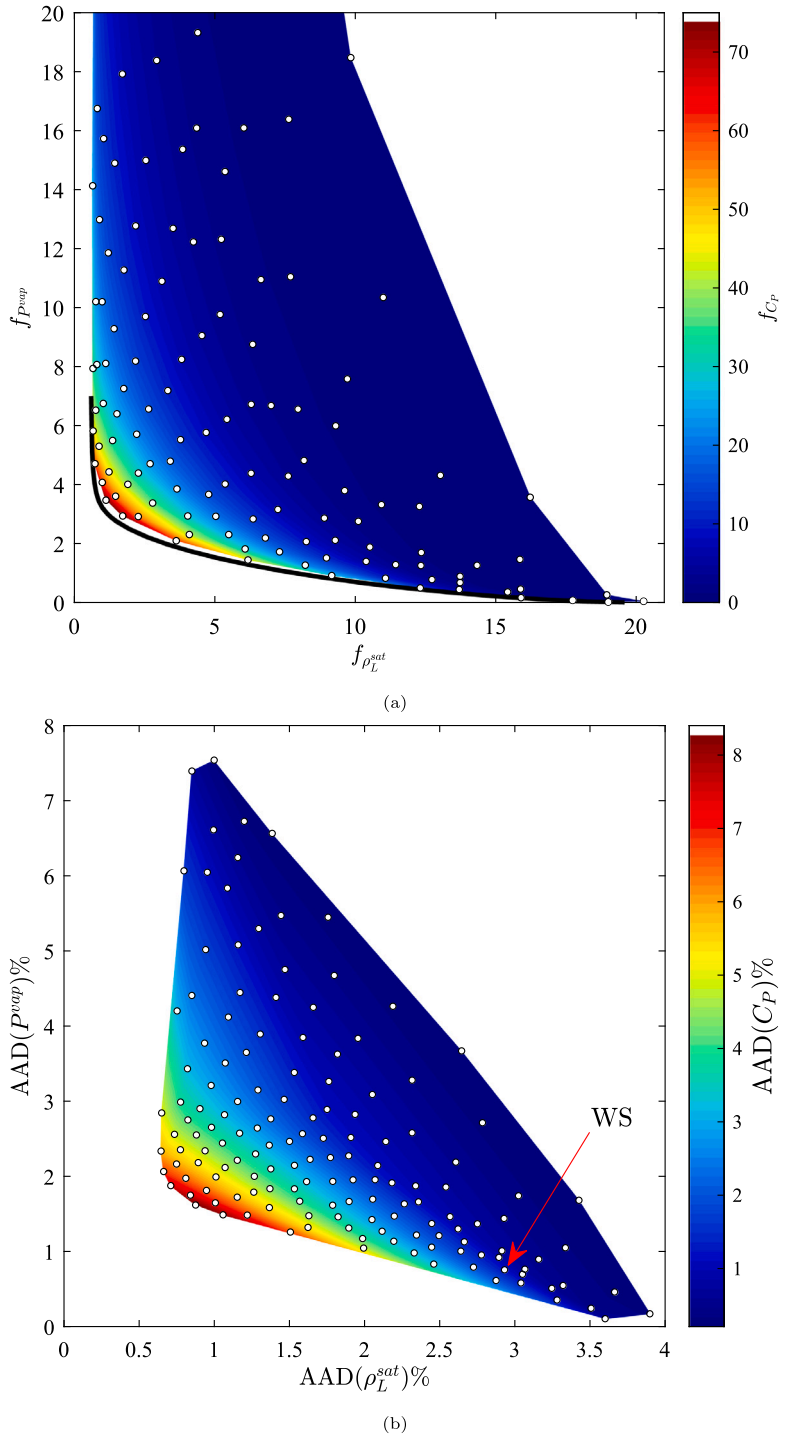


Fig. 11. Pareto front for three objectives, the vapour pressure $f_{P^{vap}}$, the saturated-liquid density $f_{\rho_L^{sat}}$, and the liquid-phase isobaric heat capacity f_{C_P}, for spherical models of water. The white circles are the individual Pareto points obtained with the sandwich algorithm. The black curve corresponds to the Pareto front when only $f_{P^{vap}}$ and $f_{\rho_L^{sat}}$ are considered, i.e., the blue curve in Fig. 6a. The Pareto front is indicated by the coloured region. The white areas indicate either that there are no models with such combinations of objective function values or that such models are dominated. The Pareto point labelled WS is the model that we will analyse further as described in Section 4.3.1. In (a) we plot the Pareto front for objectives f_i while in b) we plot the corresponding AAD% for each property type.

It is apparent from Table A.9 that the values of m vary only slightly above unity (between 1.12 and 1.33) and close to what one would expect from the quantum-mechanical predictions ($m = 1.1$) of Sheldon et al. (2006).

4.3.1. Choice of Pareto point

The solution of MOP provides a plethora of models and one model must be chosen for a given application. There is no rigorous way of defining the process by which a Pareto point should be chosen. In part this is because each Pareto point is equally optimal to the MOP from a mathematical perspective. However, in practical applications the decision maker tends to choose points 'in the middle and near the bulge' of the Pareto front (Das, 1999). Among other methods, a simple yet effective metric to characterize a point on this 'bulge' is to find the Pareto point that minimizes the Euclidean distance between the Utopia point (a vector with the minimum value of each objective in the set

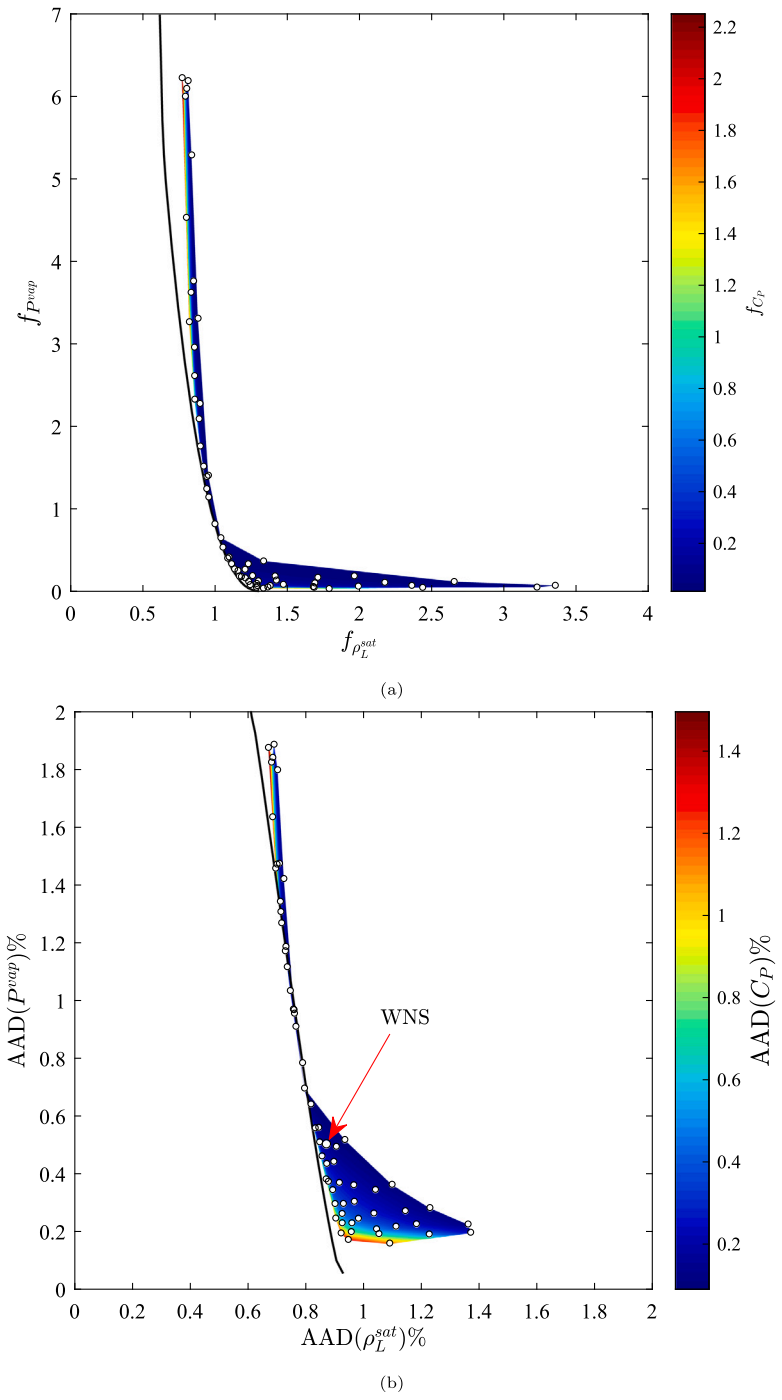


Fig. 12. Pareto front for three objectives, the vapour pressure $f_{P^{vap}}$, the saturated-liquid density $f_{\rho_L^{sat}}$, and the liquid-phase isobaric heat capacity f_{C_p}, for non-spherical models of water. The white circles are the individual Pareto points obtained with the sandwich algorithm. The black curve corresponds to the Pareto front when only $f_{P^{vap}}$ and $f_{\rho_L^{sat}}$ are considered, i.e., the red curve in Fig. 6a. The Pareto front is indicated by the coloured region. The white areas indicate either that there are no models with such combinations of objective function values or that such models are dominated. The Pareto point labelled WNS is the model that we analyse further as described in Section 4.3.1. In (a) we plot the Pareto front for objectives f_i while in (b) we plot the corresponding AAD% for each property type.

of non-dominated solutions, as shown in Fig. 3) (Cheikh et al., 2010) and the Pareto front. We therefore assign a distance measure d_k to each Pareto point $\boldsymbol{\theta}_k$ according to

$$d_k = \sqrt{\sum_{i=1}^{N_{dim}} (\theta_{ki} - f_i^*)^2}, \tag{12}$$

where f_i^* is the minimum value of objective i in the non-dominated set and θ_{ki} represents the ith index of Pareto point $\boldsymbol{\theta}_k$. d_k represents the distance between each Pareto point $\boldsymbol{\theta}_k$ and the Utopia point $\boldsymbol{f}^*$. Using this metric we can rank the Pareto points.

Other considerations may be taken into account when choosing suitable Pareto points, considering aspects not included in the MOP. SAFT-type equations of state may exhibit unrealistic or non-physical predictions at certain thermodynamic conditions, due to the empirical functional forms required to approximate statistical mechanical theories with no exact solutions. Non-physical predictions include negative heat capacities at very high pressures and multiple pure component critical points (Kalikhman et al., 2010; Polishuk, 2010). More than

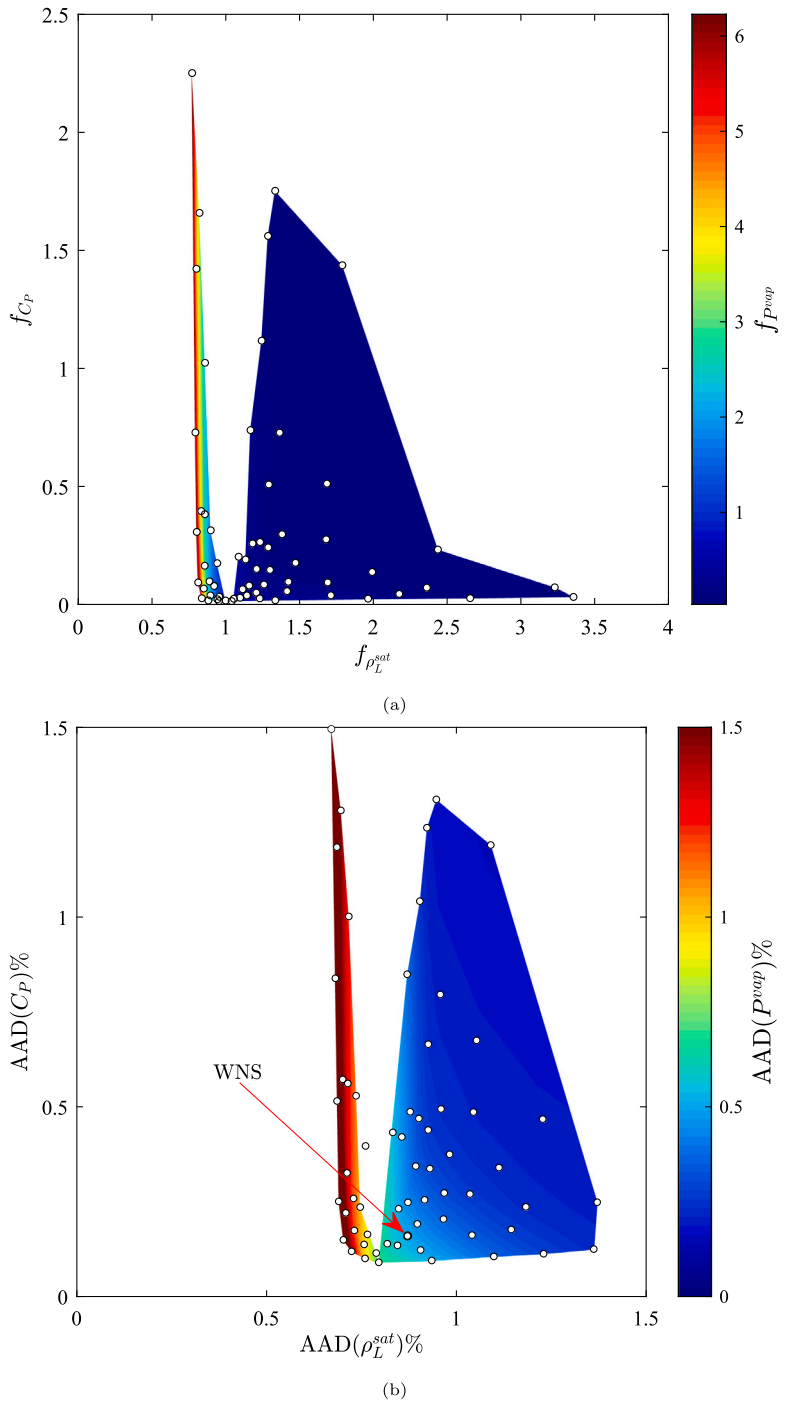


Fig. 13. Pareto front for three objectives, the vapour pressure $f_{P^{vap}}$, the saturated-liquid density $f_{\rho_L^{sat}}$, and the liquid-phase isobaric heat capacity f_{C_P}, for non-spherical models of water. The white circles are the individual Pareto points obtained with the sandwich algorithm. The black curve corresponds to the Pareto front when only f_P and $f_{\rho_L^{sat}}$ are considered, i.e., the red curve in Fig. 6a. The Pareto front is indicated by the coloured region. The white areas indicate either that there are no models with such combinations of objective function values or that such models are dominated. The red arrows indicates model WNS as detailed in Section 4.3.1. This figure differs from Fig. 12 in that C_P and P^{vap} have been swapped.

three volume roots can be present, requiring sophisticated algorithms to solve for phase equilibria (Privat et al., 2010; Alsaifi et al., 2019). After determining the set of Pareto points, models that exhibit unfavourable behaviour can be discarded. In the case of the water models identified here, negative values of the residual isochoric heat capacity are predicted at pressures above 30,000 bar and a second critical point is observed at $T_c \approx 4500$ K and $P_c \approx 1000$ bar for all models developed. These non-physical predictions are not an issue in our work since the thermodynamic conditions at which they occur are not relevant for current practical application. We do however observe models that exhibit an anomaly in the saturated-density envelope, in the form of a shallow nonconvex region (c.f., section Appendix A.2). For a given modelling strategy j, corresponding to a choice of $m = 1$ or $m \geq 1$, and 2 or 3 objective functions, we therefore select Pareto points from the subset $\boldsymbol{\theta}_{p,j}$ of physical models from the corresponding Pareto front. The models that we recommend in our current work, $\boldsymbol{\theta}_{\text{ref},j}$, are

Table 4
Promising water models chosen for further analysis. The models are labelled with WD representing the literature model (Dufal et al., 2015), WS representing spherical model picked from the three-dimensional Pareto front, and WNS representing the non-spherical model picked from the three-dimensional Pareto front. The number in brackets indicates the model number as defined in Tables A.6 and A.7. The temperature range used for the saturated-liquid density ρ_L^{sat} and vapour pressure P^{vap} is 273.16 K–613.16 K (from the triple point to 95% of the critical point). The temperature range used for the isobaric liquid heat capacity at 1 atm (C_P) is 271.16 K–363.16 K.

Model	w_i			f_i			AAD(i)%			T_c/K	AAD(T_c)%	Parameters, $\boldsymbol{x}$					
	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P			m	(ϵ/k_B)/K	σ/Å	λ^r	$K_{e,H}^{HB}$/Å^3	$(\epsilon_{e,H}^{HB}/k_B)$/K
WS (136)	0.23	0.70	0.08	12.73	0.78	3.07	2.93	0.76	1.71	668.4	3.31	1.0000	406.56	3.0912	23.097	137.87	1640.1
WNS (29)	0.19	0.22	0.59	1.14	0.34	0.04	0.87	0.50	0.16	653.5	1.00	1.2566	351.23	2.8024	25.126	177.62	1630.6
WD	–	–	–	2.62	2.40	120.52	1.06	1.28	10.52	673.9	4.16	1.0000	266.68	3.0063	17.020	101.69	1985.4

Table 5
Deviations of the chosen water models for other properties not included in the objective function. Saturated-liquid isochoric heat capacity (C_V), heat of vaporization (ΔH^{vap}), saturated-liquid speed of sound (u), and vapour–liquid interfacial tension (γ). The latter property is obtained with the SAFT-VR Mie mean-field density functional theory (DFT) developed by Graham (2020). All deviations are computed with respect to correlated experimental data from NIST (Lemmon et al., 2018).

Property i	C_V	ΔH^{vap}	u	γ
Temperature range (K)	273.16–613.16	273.16–613.16	273.16–613.16	453.9–613.16
Model	AAD(i)%			
WS (136)	7.94	3.31	18.23	17.29
WNS (29)	17.57	2.20	49.36	3.27
WD	8.23	3.01	37.53	10.06

thus determined by finding k that corresponds to the solution of the optimization:

$$k = \underset{k}{\text{argmin}} \quad d_k \qquad \text{s.t.} \quad \boldsymbol{\theta}_k \in \boldsymbol{\theta}_{p,j}. \tag{PS$_j$}$$

It is important to note that the optimum value of this problem is sensitive to the scaling of the objectives, i.e., objectives spanning a larger range on the Pareto front will be more important than others. We now select some interesting models for further analysis. Information on these Pareto points can be found in Table 4.

1. Model WS: A model chosen from the three-dimensional Pareto front for spherical water models by solving (PS$_j$).
2. Model WNS: A model chosen from the three-dimensional Pareto front for non-spherical water models by solving (PS$_j$).
3. Model WD: The existing SAFT-VR Mie model for water reported by Dufal et al. (2015).

The deviations for other thermodynamic properties not considered in the objective function for each Pareto point are summarized in Table 5. Models for the spherical WS and non-spherical WNS representations of water are indicated with arrows on the Pareto front in Figs. 11 and 13 respectively. By visual inspection, the reference points offer a good compromise between the three objectives.

We note that the parameters obtained in our current study are somewhat different to those determined by Dufal et al. (2015). Nevertheless, they are physically reasonable and are well within the range of values for other reported SAFT models for water. In comparison with model WD, our selected models have a significantly higher value of the dispersion energy ϵ: (406 K for model WS) and 351 K for model WNS, versus 267 K for model WD. The association energy $\epsilon_{e,H}^{HB}$ is noticeably lower, with the models developed in our work all $\sim$ 1600 K compared with 1985 K for model WD. This may indicate that the consideration of heat-capacity data provides suitable information on how the dispersive and associative attractive energies are partitioned. The repulsive range parameter λ^r is higher for the models developed here compared to model WD (values above 23 compared with 17 for model WD). The values of the segment diameter σ are similar for the spherical models, all slightly above 3 Å, but the non-spherical model has a slightly lower value of about 2.80 Å, consistent with the larger value of the non-sphericity parameter m and yielding very similar molecular volumes.

We analyse these models of water by comparing the description of various thermodynamic properties, including the prediction of properties that are not considered in the parameter estimation. The level of agreement between the models and the experimental data for the saturated-liquid density and saturated-vapour densities is shown in Fig. 14a. Model WS is seen to under predict the saturated-liquid density at temperatures between the triple point and 500 K, with a maximum relative deviation of 3% with respect to the experimental data at 350 K. Models WNS and WD are in better agreement with the experimental data over this temperature range, with a maximum relative deviation of less than 0.5% for both models. We also note that all models give rise to a similar curvature close to the experimental triple point, and that none of these models captures the experimentally observed maximum in the saturated-liquid density at 273 K.

The prediction of the experimental critical temperature with the SAFT-VR Mie EoS for models WS, WNS and WD is within to 3.2%, 0.9%, and 4.0%, respectively. Model WNS allows one to capture the near-critical region much more accurately than the other models providing the correct curvature of both the saturated-liquid and saturated-vapour densities, and is quantitatively accurate between 500 K and 600 K with a maximum deviation of 0.17% at 600 K.

The corresponding description of the vapour pressure is shown in Fig. 14b. All models provide fairly good agreement with the experimental data; the models are indistinguishable from the experimental data in the Clausius–Clapeyron representation. The only noticeable deviation is seen with model WS in the near-critical region with an overprediction of the critical pressure of 24.7%.

The description of the isobaric-liquid heat capacity along a single isobar (1 atm) obtained from the SAFT-VR Mie EoS with models WS, WNS, and WD is shown in Fig. 15a; the corresponding description of the saturated-liquid heat capacity between the experimental triple point and temperatures close to the experimental critical point is given in Fig. 15b. The use of model WD leads to an underprediction of the heat capacity by an average of 10.5% across the isobar, with the highest deviation of 15% at 273 K. Models WS and WNS provide a very good description of the experimental heat capacity at 1 atm; the lowest average deviation of AAD(C_P)% = 0.16% is found for model WNS. A low-temperature minimum in the saturated-liquid heat capacity is predicted with models WS and WNS but not with model WD; the temperature at which this minimum occurs is predicted with remarkable accuracy with model WNS (311 K compared with the experimental value of 310 K).

In Fig. 16, we show the predictions obtained with the SAFT-VR Mie EoS for the heat of vaporization ΔH^{vap}; this property was not included in the parameter estimation procedure. All models are seen to provide an accurate representation of the property away from the

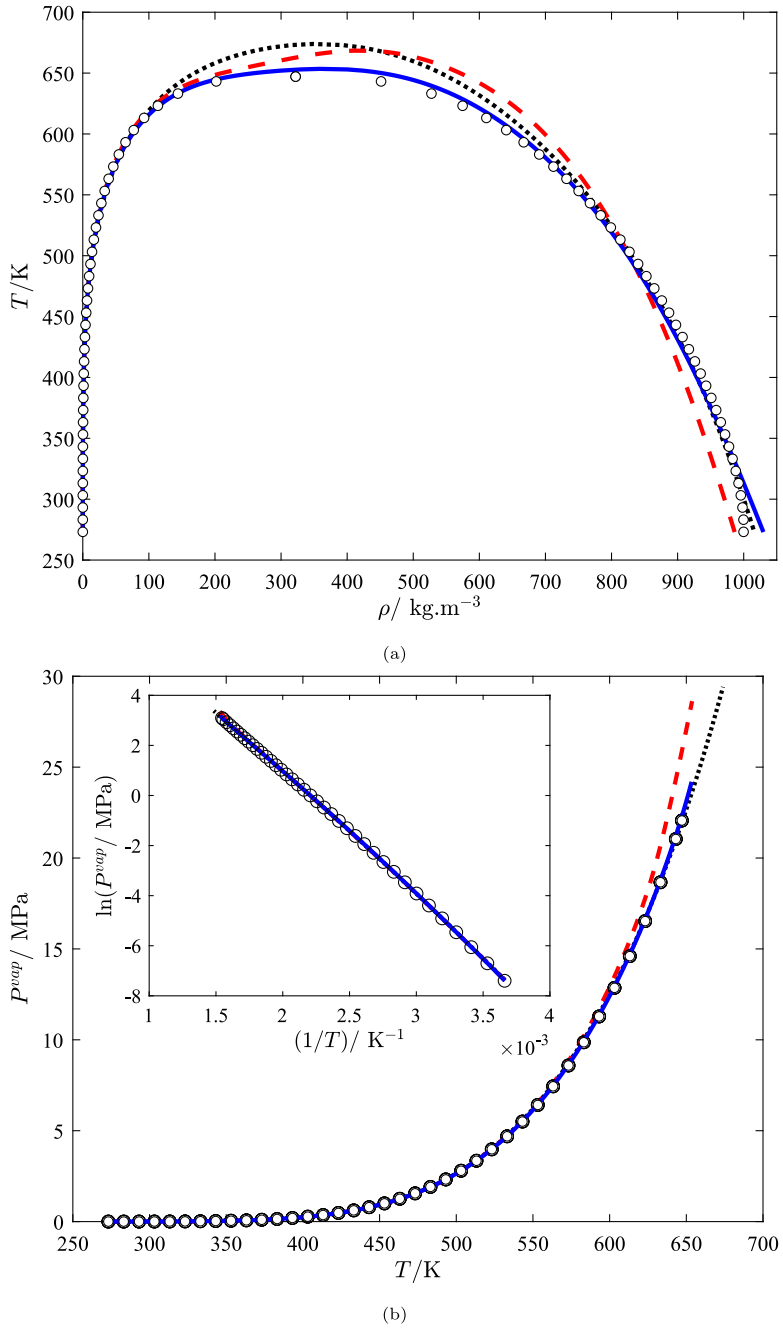


Fig. 14. (a) The saturated vapour- and liquid-phase densities, and (b) the vapour pressure (including the Clausius–Clapeyron representation) for water compared with the description obtained with the SAFT-VR Mie EoS with models WS, WNS, and WD. The theoretical description obtained with model WS is represented by the dashed red curve, that with model WNS by the continuous blue curve, and that with model WD by the dotted black curve. The white circles are the correlated experimental data from NIST (Lemmon et al., 2018). All these data points are used in the parameter estimation, apart from those above 95% of the experimental critical temperature.

critical region, with a better near-critical description seen for model WNS. The values for AAD(ΔH^{vap})% are shown in Table 5.

In Fig. 17, we show the predictions of the different models for the vapour–liquid interfacial tension, obtained with the SAFT-VR Mie mean-field density functional theory (DFT) developed by Graham (2020). This provides a stringent test of the models in terms of the partitioning of the dispersion (long-range) and association (short-range) attractive energies. Model WS provides a good prediction of the near-critical interfacial tensions well, but it leads to an overprediction of the interfacial tension at lower temperatures. Model WD leads to a poor overall description of the temperature dependence of the interfacial tension, with moderate (0.05 N m^{-1}) under-predictions at the lowest temperatures considered. The most accurate overall prediction of the interfacial tension is obtained for model WNS over the entire range of temperatures considered. The values for AAD(γ)% are shown in Table 5.

Owing to the superior performance of model WNS, we assess it further by evaluating some other thermodynamic properties not considered in the parameter estimation, to ensure that the parameters are physically sound. In Figs. 18 and 19, we show the SAFT-VR Mie predictions obtained with model WNS for the density, isobaric heat capacity, speed of sound, and isochoric heat capacity over pressures varying by several orders of magnitude. The single-phase densities are predicted accurately apart from the curvature seen experimentally in the low-temperature region. The heat capacity in the liquid and vapour phases is predicted very well over the full pressure range (varying by several

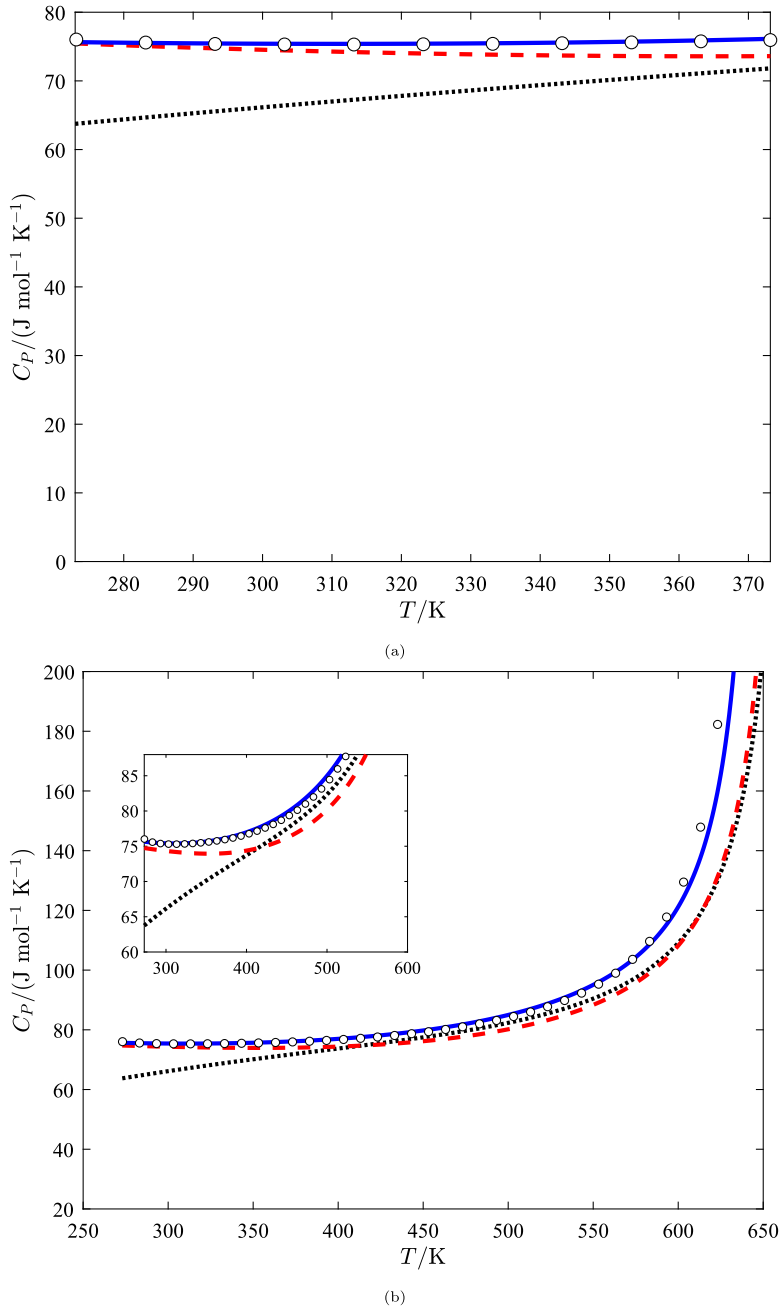


Fig. 15. (a) The isobaric liquid-phase heat capacity at 1 atm and (b) the saturated-liquid heat capacity for water compared with the description obtained with the SAFT-VR Mie EoS with models WS, WNS, and WD. The theoretical description obtained with model WS is represented by the dashed red curve, that with model WNS by the continuous blue curve, and that with model WD by the dotted black curve. The white circles are the correlated experimental data from NIST (Lemmon et al., 2018). The data points used in the parameter estimation are the experimental data shown in (a).

orders of magnitude), and the model is able to capture the maximum in C_P, which is more pronounced at near-critical/ super-critical pressures. Though the speed of sound in the gas phase is accurately represented, that in the liquid phase is not predicted well (e.g., it is over-predicted by approximately 85% at 273 K). In Table 5, we show that using model WS the speed of sound is predicted more accurately than with the other models, with an AAD(u)% of 18.23 compared to 49.36 for model WNS and 37.53 for model WD. The experimentally observed maximum is however not captured by any of the models. The relatively poor agreement for all models is due to the inability of the model to capture the anomalous curvature of the saturated-liquid density of water at low temperatures. The isochoric heat capacity is predicted moderately well with model WNS, with an average AAD of 17.57% across the full temperature range. This is slightly worse than models WS and WD, which give rise to AADs of 7.49% and 8.23% respectively, as shown in Table 5.

As discussed in Section 1, the degree of hydrogen bonding is frequently analysed with respect to the values estimated from spectroscopic data when developing SAFT models for water (Clark et al., 2006; Dufal et al., 2015). The fraction of total possible O–H hydrogen bonds that are free $f_{\mathrm{free}}^{\mathrm{OH}}$ is shown in Fig. 20. This is calculated from the fraction of association sites of type e and H not bonded, which can be determined implicitly with the SAFT EoS:

$$f_{\mathrm{free}}^{\mathrm{OH}} = X_{\mathrm{H}} = X_{\mathrm{e}}, \tag{13}$$

where X_{H} and X_{e} are the fraction of sites H and e not associated to other sites. The reader is referred to the work of Clark et al. (2006) for a more

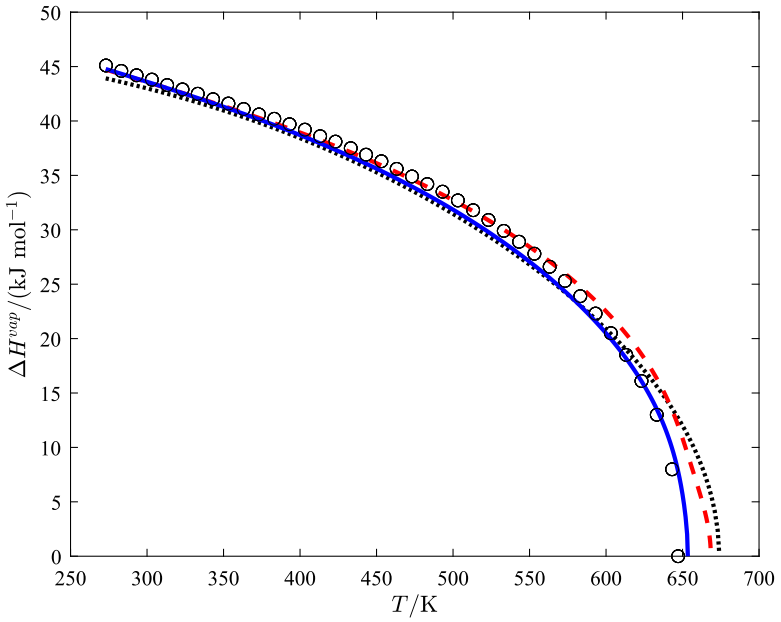


Fig. 16. The heat of vaporization for water compared with the description obtained with the SAFT-VR Mie EoS with models WS, WNS, and WD. The theoretical description obtained with model WS is represented by the dashed red curve, that with model WNS by the continuous blue curve, and that with model WD by the dotted black curve. The white circles are the correlated experimental data from NIST (Lemmon et al., 2018).

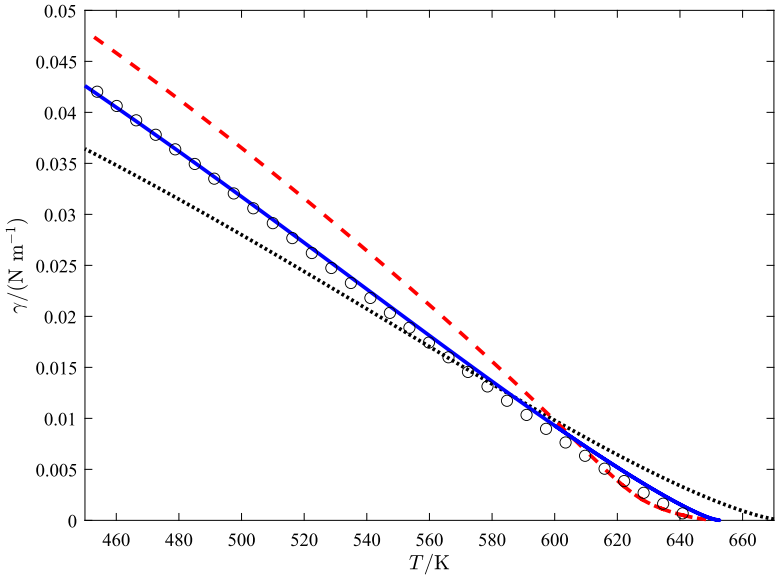
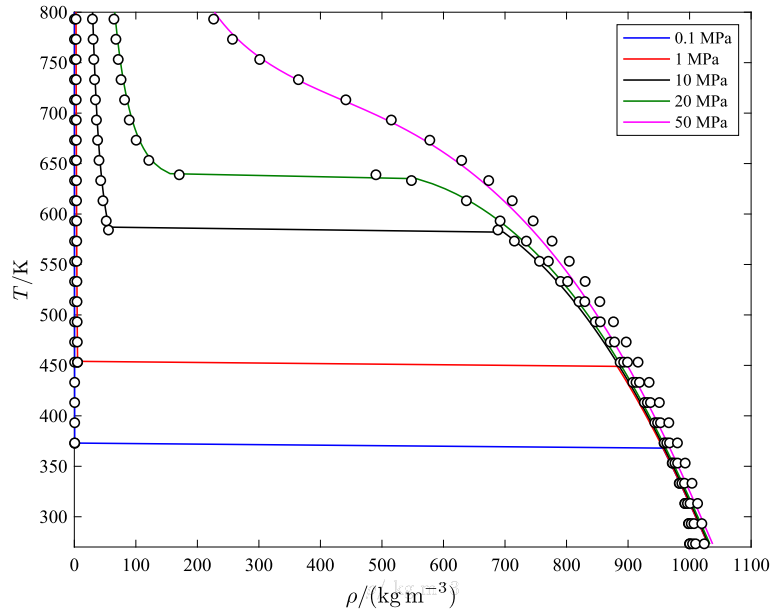
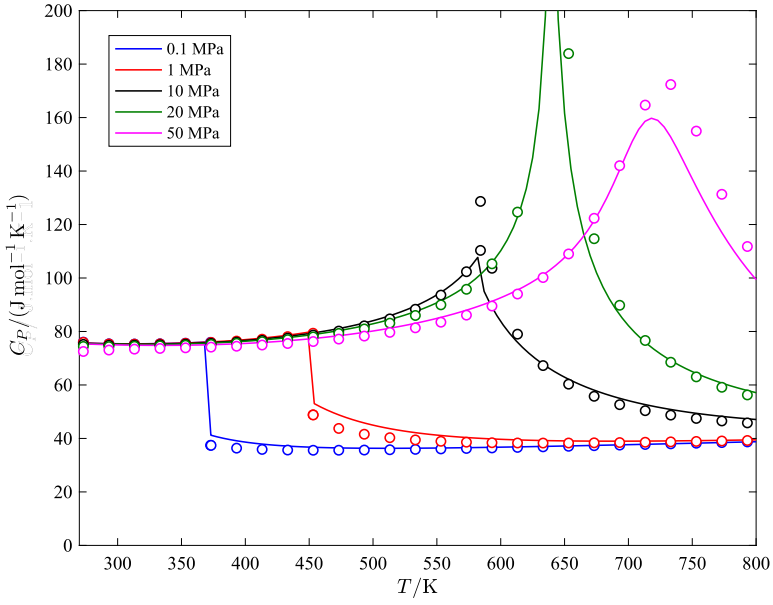


Fig. 17. The vapour–liquid interfacial tension of water compared with the description obtained with the SAFT-VR Mie mean-field DFT developed by Graham (2020) with models WS, WNS, and WD. The theoretical description obtained with model WS is represented by the dashed red curve, that with model WNS by the continuous blue curve, and that with model WD by the dotted black curve. The white circles are the correlated experimental data from NIST (Lemmon et al., 2018).

detailed discussion. Model WNS provides a description of the extent of association which is in good agreement with the data obtained from molecular simulations of the TIP4P/2005 distributed charge model of water (Dufal et al., 2015). This provides confidence that the balance between dispersive and association forces is physically sound, as the degree of association is directly related to the hydrogen-bonding energy and volume. Several authors have compared the association fractions of SAFT water models to the experimental data of Luck (Müller and Gubbins, 1995; Aparicio-Martínez and Hall, 2007; Clark et al., 2006; von Solms et al., 2006; Grenner et al., 2007; Kontogeorgis and Folas, 2010; Tsivintzelis et al., 2014; Liang et al., 2014; Dufal et al., 2015). In particular, Kontogeorgis et al. (2010) used the monomer fraction data directly in the parameter regression. Whilst some models predict the experimental data well, such agreement requires a compromise on the accuracy of the description of other thermodynamic properties, e.g., the phase behaviour of mixtures (Liang et al., 2014). For this reason and due to the non-trivial analysis of the spectroscopic data of Luck and the debate around the accuracy of such data (Dufal et al., 2015), we choose not to use these data to discriminate further between the Pareto-optimal water models. While the description of the IR data of Luck (1980) with the models in the current work is relatively poor, with an average relative deviation of 23.5% for temperatures between the triple point and 95% of the critical point, there is good agreement with respect to the molecular simulation results, where the evaluation



(a)



(b)

Fig. 18. (a) The density at and (b) the heat capacity of water compared with the description obtained with the SAFT-VR Mie EoS with model WNS. The theoretical description obtained with model WNS is represented by the continuous curves at different pressures with colours shown in the legend. The white circles are the correlated experimental data from NIST (Lemmon et al., 2018).

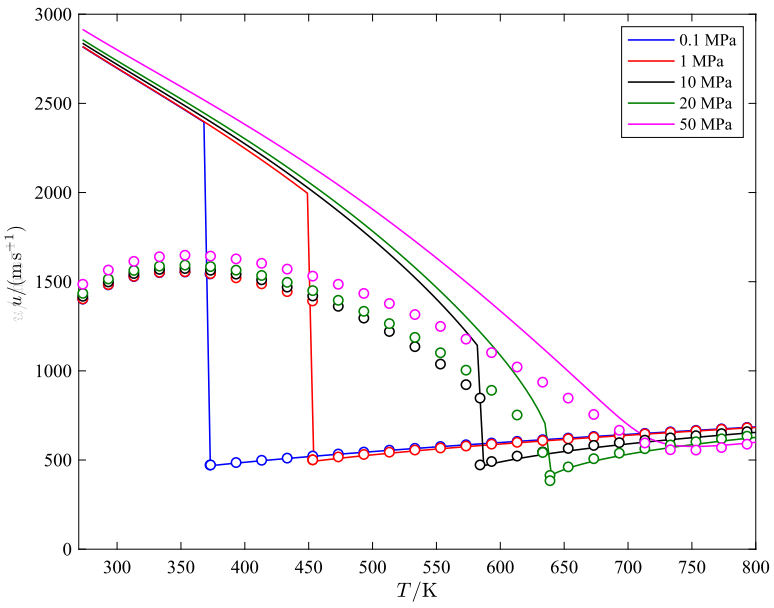
of association fractions is straightforward. This indicates the proposed models are consistent with previous models for this property.

5. Conclusion

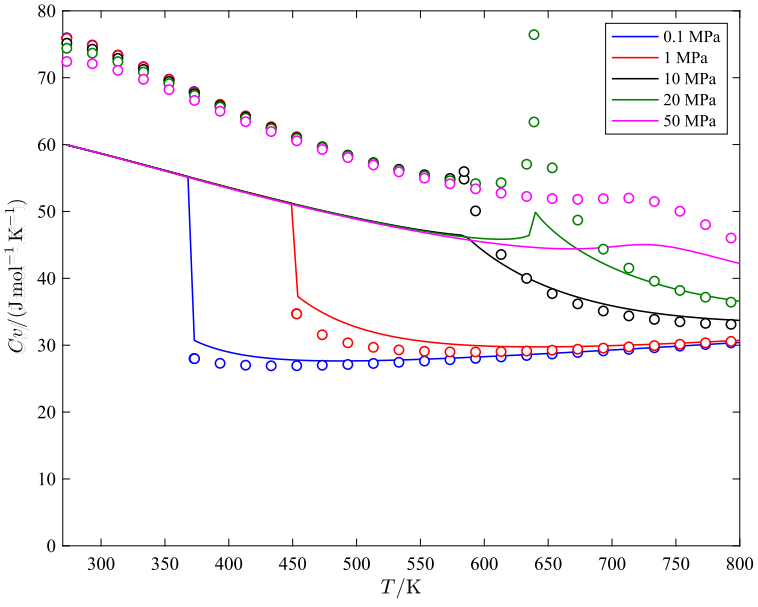
We have proposed and implemented a multi-objective optimization algorithm for parameter estimation of thermodynamic models using any number of property types. As shown by Forte et al. (2017), we find that the multi-objective approach is beneficial over the common approaches used to parameterize the models employed in SAFT EoSs.

Careful attention is paid to minimizing the computational effort involved and providing a framework that can be retrofitted to existing parameter estimation methods. We have applied the method to two- and three-dimensional problems, but the proposed algorithm is generally applicable to any number of dimensions. This allows for an efficient estimation of the convex parts of a Pareto front where numerous property types may be considered as competing objectives. We have demonstrated that this technique can be used as a platform for the rigorous comparison between different model types.

The MOO technique is applied to the development of models for water for use with the SAFT-VR Mie EoS: we consider models where water is treated as a spherical molecule ($m = 1$), or non-spherical by treating the number of segments m as an adjustable non-sphericity parameter. Numerous Pareto-optimal models for water are obtained in this manner. By analysing the Pareto front and predictions of various thermodynamic properties, we have identified that the most preferable



(a)



(b)

Fig. 19. (a) The speed of sound, u, and (b) the isochoric heat capacity, C_v, of water compared with the description obtained with the SAFT-VR Mie EoS with model WNS. The theoretical description obtained with model WNS is represented by the continuous curves at different pressures with colours shown in the legend. The white circles are the correlated experimental data from NIST (Lemmon et al., 2018).

non-dominated models can be found using a weight vector that would not typically be chosen *a priori* (i.e., not an equal weight vector). For the MOP we consider three property types as competing objectives: the saturated-liquid density (ρ_L^{sat}), the vapour pressure (P^{vap}), and the liquid isobaric heat capacity (C_P). It is shown that if only ρ_L^{sat} and P^{vap} are considered in the optimization problem (as is typically the case in regressing EoS model parameters), a variety of non-dominated models are obtained that lead to predictions of C_P with varying degrees of accuracy (some which are quite poor). This exemplifies that the determination of a Pareto front using an MOO approach is a useful way of generating candidate models which can be tested against other properties not considered in the parameter regression.

We have also considered the three-dimensional MOP including the three properties ρ_L^{sat}, P^{vap}, and C_P of water as objective functions. We have shown that C_P is a particularly useful property to reduce the parameter degeneracy in this case, and that one is able to achieve a good description of C_P with a negligible compromise in the quality of the description of ρ_L^{sat} and P^{vap}. The three-dimensional Pareto fronts are computed and preferable models are chosen through analysis of these fronts. We find that the non-spherical water model (WNS) outperforms the spherical model developed in our current work (model WS) for all of the thermodynamic properties considered. The models developed when including C_P are able to capture the key thermodynamic properties with better accuracy than the previously reported spherical model WD

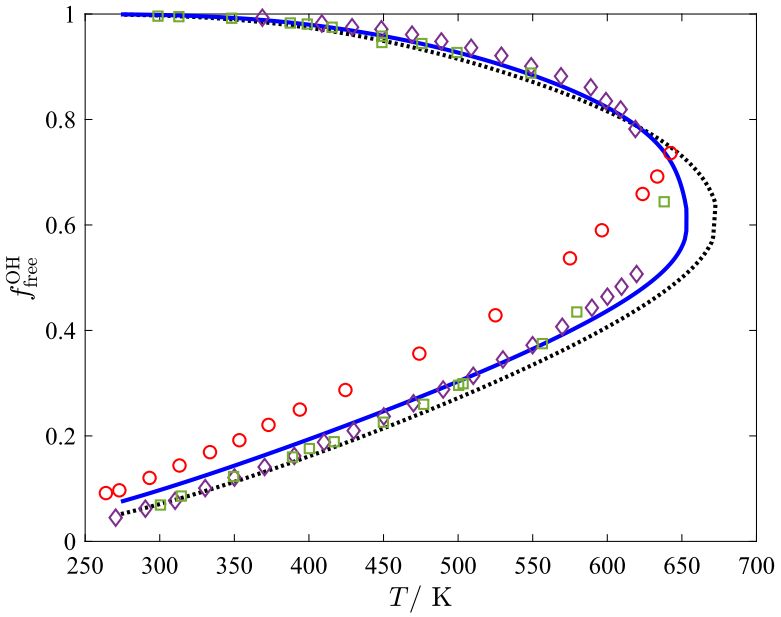


Fig. 20. The fraction of free OH hydrogen bonds, $f_{\text{free}}^{\text{OH}}$ for the saturated-vapour and saturated-liquid states of water compared with the description obtained with the SAFT-VR Mie EoS with models WNS and WD. The theoretical description obtained with model WNS is represented by the continuous blue curve, and that with model WD by the dotted black curve. The open red circles are the spectroscopic data provided by Luck (1980). The open green squares are the SPC/E simulation results of Dufal et al. (2015) and the open purple diamonds are the TIP4P/2005 simulation results of Dufal et al. (2015).

((Dufal et al., 2015), which describes C_P relatively inaccurately), with a negligible effect on the description of ρ_L^{sat} and P^{vap}.

In further work, it would be useful to explore the effect of considering further thermodynamic properties in the parameter estimation. It would be particularly useful to investigate the effect of including different second derivative properties, e.g., the speed of sound or isochoric liquid heat capacity in the MOP. Since the proposed MOO technique is applicable to dimensions higher than three, it will be useful to investigate the Pareto fronts obtained when a larger number of properties (or objective functions) are included. Furthermore, the proposed approach is readily extended to multicomponent mixtures (Graham, 2020) and it would be useful to explore the effect of including mixture properties at an earlier stage of the parameter development.

CRediT authorship contribution statement

Edward J. Graham: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. **Esther Forte:** Conceptualization, Methodology, Validation, Supervision, Writing – review & editing. **Jakob Burger:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Amparo Galindo:** Conceptualization, Methodology, Supervision, Writing – review & editing, Funding acquisition. **George Jackson:** Conceptualization, Methodology, Supervision, Writing – review & editing, Funding acquisition. **Claire S. Adjiman:** Conceptualization, Methodology, Supervision, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Edward Graham reports financial support was provided by Engineering and Physical Sciences Research Council. George Jackson reports financial support was provided by Engineering and Physical Sciences Research Council. Amparo Galindo reports was provided by Engineering and Physical Sciences Research Council. Claire Adjiman reports financial support was provided by Engineering and Physical Sciences Research Council. Edward Graham reports financial support was provided by European Commission. George Jackson reports financial support was provided by European Commission. Amparo Galindo reports financial support was provided by European Commission. Claire Adjiman reports financial support was provided by European Commission.

Data availability

We have made the data available on zenodo.org.

Acknowledgements

The authors gratefully acknowledge financial support from the Engineering and Physical Sciences Research Council (EPSRC) of the UK (grants EP/M507878/1, EP/J003840, EP/E016340, and EP/J014958), and funding from the European Union's Horizon 2020 research and innovation program (grant agreement 727503 - ROLINCAP – H2020-LCE-2016-2017/H2020-LCE-2016-RES- CCS-RIA).

Appendix

A.1. Tabulated Pareto points

See Tables A.6–A.9.

A.2. Anomalous density profiles

Some of the Pareto-optimal spherical water models found during the optimization produced saturated-density phase diagrams appear unphysical at near-critical temperatures. This behaviour of the near-critical region contrasts with the well-known law of the rectilinear diameter, which states that the average of the densities ρ_L^{sat} and ρ_G^{sat} is a linear function of temperature (Zollweg and Mulholland, 1972) . After a detailed parametric investigation on the near-critical region of the $\rho-T$ phase diagram, we find that these shapes, which we may define as a non-convex vapour–liquid envelope, occur for relatively high values of λ^r. The A_1 term in SAFT-VR Mie is parametrized for two Sutherland potentials with exponents ranging between $\lambda = 5$ and $\lambda = 100$ (where λ

Table A.6
Pareto-optimal spherical models of water ($m = 1$), when only $f_{\rho_L^{sat}}$ and $f_{P^{vap}}$ are treated as objectives. We also provide the deviations in C_P. The models are sorted with $f_{\rho_L^{sat}}$ in increasing order.

θ_k	w_i			f_i			AAD(i)%			Parameters, $\boldsymbol{x}$					
	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	m	(ϵ/k_B)/K	σ/Å	λ^r	$K_{c,H}^{HB}$/Å^3	$(\epsilon_{c,H}^{HB}/k_B)$/K
1	0.9901	0.0099	0.0000	0.6194	6.9906	52.4530	0.5905	2.1887	7.0610	1	337.80287	3.03114	28.25278	147.08527	1840.58221
2	0.9616	0.0384	0.0000	0.6550	5.0907	64.7120	0.6204	1.9510	7.8490	1	328.31646	3.02607	26.88362	143.95944	1857.10720
3	0.9267	0.0733	0.0000	0.6876	4.5189	74.9030	0.6480	1.8467	8.4340	1	318.42870	3.02106	25.29042	137.44404	1880.16773
4	0.8741	0.1259	0.0000	0.7364	4.0619	87.3960	0.6840	1.7517	9.0220	1	306.88405	3.01552	23.56082	129.91951	1907.25622
5	0.8004	0.1996	0.0000	0.8033	3.7071	100.3040	0.7237	1.6721	9.6610	1	294.44697	3.00995	21.87371	122.29255	1935.95367
6	0.6971	0.3029	0.0000	0.8927	3.4351	111.0290	0.7665	1.6054	10.1700	1	283.92050	3.00609	20.49043	116.05522	1959.13400
7	0.5852	0.4148	0.0000	1.0140	3.2155	119.1040	0.8127	1.5488	10.5090	1	275.18560	3.00363	19.36393	110.97833	1977.48461
8	0.5000	0.5000	0.0000	1.1923	3.0124	114.2610	0.8802	1.4975	10.2860	1	278.38985	3.00784	19.23058	111.01864	1968.60269
9	0.4471	0.5529	0.0000	1.3328	2.8831	121.1450	0.9191	1.4604	10.5620	1	271.38001	3.00598	18.43544	107.29818	1983.02618
10	0.4002	0.5998	0.0000	1.5199	2.7475	118.8330	0.9725	1.4233	10.4400	1	272.52796	3.00816	18.26557	106.86603	1979.14882
11	0.3282	0.6718	0.0000	1.9784	2.4878	115.2320	1.0941	1.3520	10.3090	1	272.82700	3.01162	17.79038	105.19051	1975.79311
12	0.2774	0.7226	0.0000	2.5976	2.2211	109.5830	1.2423	1.2752	10.0370	1	275.05570	3.01619	17.43385	104.10442	1967.92583
13	0.2389	0.7611	0.0000	3.3213	1.9723	102.0180	1.4091	1.2000	9.7090	1	279.87340	3.02155	17.28108	103.97918	1954.51768
14	0.2067	0.7933	0.0000	4.1621	1.7331	94.8840	1.5841	1.1223	9.3810	1	283.68725	3.02633	17.05033	103.35989	1943.15761
15	0.1817	0.8183	0.0000	5.1357	1.4994	85.7140	1.7757	1.0413	8.8730	1	289.93519	3.03195	16.97441	103.48767	1926.46549
16	0.1602	0.8398	0.0000	6.2878	1.2653	72.6770	1.9977	0.9579	8.1860	1	300.83182	3.03935	17.17551	104.99422	1899.64923
17	0.1407	0.8593	0.0000	7.3675	1.0736	66.5060	2.1755	0.8776	7.8200	1	304.85362	3.04354	16.96611	104.23066	1887.73446
18	0.1246	0.8754	0.0000	8.7460	0.8624	54.7030	2.3886	0.7844	7.0530	1	315.19019	3.05012	17.08512	105.13793	1861.65890
19	0.1088	0.8912	0.0000	10.1938	0.6715	42.8890	2.5931	0.6901	6.3230	1	326.33165	3.05670	17.24536	106.13407	1833.72872
20	0.0945	0.9055	0.0000	11.5457	0.5170	32.8000	2.7710	0.6034	5.5560	1	337.28783	3.06262	17.43693	107.16714	1806.39249
21	0.0807	0.9193	0.0000	13.3992	0.3371	19.2640	2.9982	0.4872	4.2440	1	356.71533	3.07200	17.99663	109.97873	1758.92134
22	0.0667	0.9333	0.0000	15.0471	0.2059	9.5340	3.1857	0.3820	3.0430	1	374.43513	3.07992	18.50414	112.34087	1714.95475
23	0.0514	0.9486	0.0000	16.6635	0.1038	3.5050	3.3585	0.2737	1.7210	1	394.07641	3.08829	19.12669	115.12702	1665.66220
24	0.0335	0.9665	0.0000	18.1875	0.0352	1.1050	3.5123	0.1618	0.4500	1	415.04273	3.09678	19.84072	118.21748	1612.01748
25	0.0099	0.9901	0.0000	19.5617	0.0035	2.9880	3.6441	0.0523	1.3710	1	437.19481	3.10538	20.66490	121.75049	1553.92136

Table A.7
Pareto-optimal non-spherical models of water ($m \geq 1$), when only $f_{\rho_L^{sat}}$ and $f_{P^{vap}}$ are treated as objectives. We also provide the deviations in C_P. The models are sorted with $f_{\rho_L^{sat}}$ in increasing order.

θ_k	w_i			f_i			AAD(i)%			Parameters, $\boldsymbol{x}$						
	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	m	(ϵ/k_B)/K	σ/Å	λ^r	$K_{c,H}^{HB}$/Å^3	$(\epsilon_{c,H}^{HB}/k_B)$/K
1	0.9901	0.0099	0.0000	0.6194	6.9883	52.0250	0.5904	2.1886	7.0730	1.00000	337.75987	3.03112	28.24623	147.05562	1840.69549
2	0.9798	0.0202	0.0000	0.6373	5.7063	57.4990	0.6041	2.0442	7.4160	1.00000	334.47266	3.02930	27.87635	147.35113	1844.03705
3	0.9690	0.0310	0.0000	0.6481	5.2840	61.5330	0.6136	1.9820	7.6770	1.00000	330.86469	3.02739	27.30268	145.51528	1851.44217
4	0.9560	0.0440	0.0000	0.6615	4.9475	63.6970	0.6258	1.9250	7.7770	1.00237	328.04744	3.02341	26.77826	143.97359	1856.12122
5	0.9468	0.0532	0.0000	0.7030	4.1577	51.1990	0.6494	1.7661	6.9870	1.02857	330.13712	2.99679	26.51523	146.76066	1835.48255
6	0.9392	0.0608	0.0000	0.7456	3.4518	36.4470	0.6712	1.6094	5.9060	1.05883	335.18845	2.96860	26.66534	152.06939	1804.62657
7	0.9304	0.0696	0.0000	0.7920	2.7810	24.0180	0.6946	1.4439	4.8070	1.09064	339.65932	2.93978	26.72327	157.38033	1773.53096
8	0.9189	0.0811	0.0000	0.8409	2.1768	13.4560	0.7180	1.2768	3.5790	1.12425	345.24842	2.91109	26.95946	164.17938	1737.41107
9	0.9066	0.0934	0.0000	0.8850	1.7131	6.8800	0.7385	1.1335	2.5710	1.15290	349.28501	2.88723	27.06414	169.71561	1707.85087
10	0.8914	0.1086	0.0000	0.9313	1.2993	2.6340	0.7597	0.9883	1.5860	1.18193	353.02710	2.86375	27.13743	175.37069	1678.21288
11	0.8709	0.1291	0.0000	0.9766	0.9625	0.4850	0.7794	0.8514	0.5910	1.21049	357.16610	2.84169	27.30995	181.81064	1647.03618
12	0.8450	0.1550	0.0000	1.0241	0.6731	0.2610	0.7996	0.7131	0.4050	1.23866	360.34800	2.82020	27.34463	187.57936	1618.31453
13	0.8075	0.1925	0.0000	1.0738	0.4349	2.0290	0.8213	0.5737	1.3660	1.26817	363.97545	2.79858	27.45117	194.37886	1586.67667
14	0.7523	0.2477	0.0000	1.1241	0.2535	5.1900	0.8432	0.4387	2.2370	1.29678	366.91760	2.77802	27.47369	200.71517	1557.14185
15	0.6638	0.3362	0.0000	1.1774	0.1205	9.7150	0.8664	0.3045	3.0290	1.32500	368.82805	2.75787	27.34959	206.10767	1530.64678
16	0.5960	0.4040	0.0000	1.2049	0.0732	12.3330	0.8784	0.2386	3.3660	1.33907	369.60073	2.74796	27.25961	208.64714	1517.84524
17	0.5000	0.5000	0.0000	1.2332	0.0382	14.5810	0.8907	0.1739	3.6830	1.35249	369.67963	2.73832	27.07443	210.27491	1507.73458
18	0.3192	0.6808	0.0000	1.2682	0.0126	16.5710	0.9060	0.1022	3.9700	1.36756	368.91007	2.72724	26.73373	210.97036	1499.17827
19	0.0099	0.9901	0.0000	1.3051	0.0044	16.9690	0.9292	0.0575	4.0240	1.37826	367.22287	2.71932	26.17634	209.07780	1496.33243

is the repulsive or attractive exponent), and the A_2 term is a function of $2\lambda^r$. Hence, the theory is strictly only applicable for values of $\lambda^r \leq 50$, as has been previously pointed out (Lafitte et al., 2013). However, when association is introduced, the critical temperature and the range of VLE coexistence increases, and non-convex density profiles occur for lower values of λ^r. We illustrate this point in Fig. A.21 by changing the value of λ^r while keeping other variables fixed to values that are similar to the values obtained for the water models developed in our current work. Non-convex envelopes occur at values of λ^r above 30 for this particular set of parameters.

We attribute this phenomenon to the increase in the number of density roots that appear in the near-critical region, which can lead to practically unrealistic predictions such as multiple critical points for pure compounds. Multiple density roots have been reported for SAFT-type equations of state (Koak et al., 1999; Aslam and Sunol, 2006; Privat et al., 2010; Polishuk, 2010; Alsaifi and Englezos, 2011; Alsaifi et al., 2017, 2019). In the recent work (Alsaifi et al., 2019) have shown that as many as 10 density roots may occur in the SAFT-VR Mie description, via a sophisticated use of bifurcation diagrams to analyse rigorously the metastable region of the equation of state. In our test, we found that 5 volume roots can appear at temperatures close to the critical point. Alsaifi et al. (2019) note that these non-physical predictions are inevitable in more sophisticated equations of state due to the empirical functional forms (e.g., an empirical analytical expression for the pair distribution function at contact) required to approximate statistical mechanical theories with no exact solutions. It is possible to follow a rigorous method such as that proposed by Alsaifi et al. (2019) to identify models with non-physical regions, or simply to inspect the van der Waals loop on a pressure–volume diagram at various temperatures. It is possible to include such checks as constraints within the MOP to avoid the generation of anomalous models, but such checks are not included in our current implementation as they would

Table A.8
Pareto points and parameters for water models with $m = 1$, derived using all three objective functions. The models are sorted with $f_{\rho_L^{sat}}$ in increasing order.

θ_k	w_i			f_i			AAD(i)%			Parameters, $\boldsymbol{x}$					
	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	m	(ϵ/k_B)/K	σ/Å	λ^r	$K_{c,H}^{HB}$/Å^3	$(\epsilon_{c,H}^{HB}/k_B)$/K
1	0.9804	0.0098	0.0098	0.6517	14.1300	25.2830	0.6499	2.8436	4.9834	1	367.53841	3.04818	33.56824	164.62466	1774.31730
2	0.9494	0.0399	0.0108	0.6628	7.9364	34.8796	0.6459	2.3360	5.8595	1	361.58200	3.04495	32.92614	168.30664	1775.11750
3	0.8800	0.1102	0.0098	0.6664	5.8178	48.5804	0.6631	2.0639	6.8659	1	345.59187	3.03663	29.48471	156.03468	1812.52240
4	0.7777	0.2125	0.0098	0.7337	4.7058	60.4582	0.7105	1.8761	7.6107	1	332.18496	3.03049	26.69834	144.87712	1844.95346
5	0.8154	0.1611	0.0235	0.7607	6.5204	37.6405	0.7489	2.1637	6.0801	1	358.50354	3.04517	31.35607	164.56748	1779.18382
6	0.9113	0.0591	0.0296	0.7632	10.2025	23.7610	0.7352	2.5588	4.8673	1	377.46647	3.05490	36.14391	180.85571	1734.37307
7	0.9552	0.0098	0.0350	0.7733	33.0080	9.2640	0.7525	4.2015	3.0415	1	390.45510	3.06182	38.21938	175.05650	1729.54724
8	0.8430	0.1240	0.0330	0.7983	8.0673	28.6943	0.7748	2.3537	5.3372	1	370.68664	3.05199	34.01176	174.70272	1748.80367
9	0.9291	0.0287	0.0422	0.8151	16.7503	15.0339	0.7757	2.9879	3.8759	1	388.10231	3.06104	38.41947	184.60724	1716.34650
10	0.8881	0.0098	0.1021	0.8388	65.8909	2.6647	0.7984	6.0660	1.6236	1	400.83947	3.06840	39.53367	171.16408	1722.28340
11	0.7189	0.2583	0.0228	0.8780	5.2975	44.6224	0.8106	1.9739	6.5885	1	349.24593	3.04145	28.88369	155.31290	1801.47944
12	0.8818	0.0616	0.0566	0.8949	12.9829	16.7848	0.8244	2.7504	4.0952	1	387.90028	3.06186	38.07096	187.14218	1709.88976
13	0.9003	0.0278	0.0719	0.8965	23.8938	10.0560	0.8215	3.4318	3.1648	1	394.73430	3.06555	39.34556	184.85993	1706.21829
14	0.7091	0.0098	0.2811	0.9354	92.3154	0.5325	0.8511	7.3942	0.7062	1	405.25596	3.07184	38.92119	163.73707	1725.71258
15	0.8739	0.0212	0.1048	0.9401	37.8256	5.7241	0.8506	4.4073	2.3778	1	399.82207	3.06875	39.82501	181.13267	1705.04996
16	0.8223	0.1169	0.0608	0.9875	10.1984	19.4775	0.8810	2.5505	4.4119	1	384.49597	3.06090	36.63574	184.66905	1714.14202
17	0.6215	0.3641	0.0144	0.9933	4.0708	60.9243	0.8415	1.7493	7.6289	1	330.54564	3.03250	25.38992	140.40301	1847.45812
18	0.7341	0.2238	0.0421	1.0280	6.7488	29.9651	0.8933	2.1822	5.4468	1	368.01105	3.05256	32.17302	169.24478	1753.88890
19	0.8528	0.0602	0.0870	1.0424	15.7297	12.6042	0.9046	2.9003	3.5426	1	393.75167	3.06618	38.70085	188.24054	1698.22077
20	0.8378	0.0346	0.1276	1.0990	28.6982	6.6959	0.9354	3.7725	2.5662	1	400.52266	3.07024	39.58270	184.63964	1695.03693
21	0.7976	0.0232	0.1792	1.1092	47.0912	3.1832	0.9428	5.0175	1.7517	1	404.27868	3.07239	39.63630	178.04883	1700.43663
22	0.7540	0.1879	0.0581	1.1156	8.1056	23.0267	0.9387	2.3396	4.7925	1	378.45360	3.05861	34.38438	177.83569	1727.53435
23	0.5518	0.4384	0.0098	1.1259	3.4693	74.8018	0.8759	1.6181	8.4021	1	316.10525	3.02599	23.12063	130.18223	1881.91932
24	0.7269	0.0179	0.2551	1.1299	64.5210	1.4154	0.9537	6.0471	1.1498	1	406.18339	3.07359	39.17057	171.43850	1707.73228
25	0.8242	0.0510	0.1248	1.1992	20.9022	8.6144	0.9792	3.2097	2.9148	1	399.21583	3.07009	39.17561	187.38935	1690.26536
26	0.7988	0.1096	0.0915	1.2086	11.8553	14.5645	0.9815	2.6522	3.8100	1	392.18082	3.06636	37.72349	188.41622	1695.91795
27	0.6162	0.0160	0.3678	1.2268	74.0940	0.6601	0.9954	6.6117	0.7605	1	407.42564	3.07481	38.47438	166.75764	1711.31176
28	0.5994	0.3756	0.0249	1.2321	4.4280	46.4877	0.9480	1.8118	6.7082	1	345.76636	3.04213	27.11424	148.99007	1808.91810
29	0.4028	0.0098	0.5874	1.2371	91.4259	0.1420	0.9991	7.5397	0.3229	1	407.10232	3.07478	37.36208	158.93433	1723.54118
30	0.6375	0.3206	0.0419	1.3599	5.4924	32.6586	1.0110	1.9918	5.6709	1	363.11144	3.05187	30.00203	161.72254	1765.08853
31	0.7358	0.1775	0.0867	1.4126	9.2841	16.8757	1.0543	2.4421	4.1057	1	388.35290	3.06536	35.91562	184.10770	1702.29862
32	0.7828	0.0867	0.1306	1.4386	14.8970	10.2017	1.0681	2.8197	3.1722	1	398.71468	3.07095	38.46849	189.40948	1683.48647
33	0.7751	0.0498	0.1750	1.4462	24.9968	5.8813	1.0729	3.5083	2.3852	1	403.70739	3.07370	39.16294	185.88775	1683.01538
34	0.5121	0.4700	0.0178	1.4666	3.6023	56.7316	1.0072	1.6471	7.3644	1	333.58441	3.03717	24.69590	138.56834	1838.38428
35	0.6117	0.0226	0.3657	1.4840	58.9976	1.0146	1.0876	5.8359	0.9340	1	408.70348	3.07663	38.37843	170.71434	1698.95451
36	0.7422	0.0400	0.2178	1.5059	32.8709	3.8486	1.0937	4.1194	1.9066	1	406.25100	3.07534	39.15667	182.44843	1684.14493
37	0.6557	0.2860	0.0583	1.5144	6.3977	25.0884	1.0719	2.1169	4.9940	1	374.03576	3.05832	31.94133	169.98481	1737.14733
38	0.4736	0.0195	0.5070	1.6899	64.4619	0.4800	1.1565	6.2428	0.5993	1	409.47510	3.07787	37.31564	165.54182	1702.30452
39	0.6313	0.0306	0.3381	1.7039	46.1878	1.5847	1.1591	5.0811	1.1725	1	409.46196	3.07784	38.44533	175.18116	1687.87326
40	0.7492	0.0761	0.1748	1.7112	17.9210	7.1957	1.1551	2.9960	2.6394	1	403.49265	3.07460	38.62423	188.51207	1675.25631
41	0.4284	0.5618	0.0098	1.7205	2.9346	73.4015	1.0582	1.4876	8.3131	1	315.61931	3.02955	21.98910	126.07180	1880.77022
42	0.6763	0.0384	0.2852	1.7479	36.8614	2.5259	1.1713	4.4467	1.5098	1	408.89055	3.07766	38.72825	179.63578	1681.43071
43	0.6610	0.2602	0.0788	1.7573	7.2545	19.2839	1.1531	2.2137	4.3889	1	383.34324	3.06414	33.51921	176.66211	1713.16647
44	0.7245	0.1454	0.1301	1.7662	11.2734	11.4185	1.1685	2.5738	3.3582	1	397.92591	3.07184	37.37644	189.15692	1679.45883
45	0.3090	0.0137	0.6773	1.8330	71.4861	0.1798	1.1996	6.7251	0.3397	1	408.83142	3.07813	36.04581	159.17708	1710.47521
46	0.7003	0.0543	0.2454	1.8969	26.2407	3.9673	1.2132	3.6491	1.9161	1	408.13504	3.07774	38.72168	184.51861	1673.11548
47	0.5152	0.4529	0.0319	1.9018	4.0065	39.7148	1.1526	1.7222	6.2140	1	352.34639	3.04867	26.76551	148.99702	1790.88175
48	0.6534	0.2359	0.1108	2.1778	8.1864	13.8505	1.2747	2.2979	3.7143	1	392.82544	3.07051	34.84933	182.36931	1688.78659
49	0.6964	0.1275	0.1761	2.1832	12.7704	8.1133	1.2863	2.6419	2.7979	1	404.06364	3.07644	37.78233	190.21733	1665.72376
50	0.6857	0.0737	0.2406	2.1879	20.0945	4.7282	1.2904	3.1494	2.0911	1	408.55347	3.07884	38.40002	187.15315	1664.99720
51	0.4793	0.0292	0.4915	2.1938	47.5620	0.8083	1.2959	5.2981	0.7639	1	411.57671	3.08058	37.08806	170.15071	1685.59555
52	0.5860	0.3453	0.0687	2.2169	5.7054	21.7131	1.2665	1.9999	4.6503	1	377.96409	3.06288	31.12577	168.32929	1725.62337
53	0.6317	0.0519	0.3164	2.2424	28.7606	2.6559	1.3060	3.8936	1.5167	1	411.04173	3.08034	38.23327	181.93216	1668.77334
54	0.4059	0.5780	0.0161	2.2734	2.9139	56.4887	1.2195	1.4837	7.3336	1	331.93560	3.03995	23.17861	132.78927	1840.27056
55	0.5228	0.4316	0.0455	2.2767	4.3890	30.1785	1.2628	1.7890	5.4488	1	364.40408	3.05617	28.19810	156.08461	1759.81283
56	0.6525	0.1928	0.1547	2.5311	9.6975	9.6120	1.3660	2.4143	3.0643	1	401.46561	3.07605	36.28701	187.60016	1667.70478
57	0.1426	0.0098	0.8476	2.5383	65.0754	0.0806	1.3853	6.5642	0.2362	1	408.39688	3.08018	33.82458	152.97294	1710.94382
58	0.6636	0.1061	0.2304	2.5465	14.9914	5.6002	1.3758	2.7646	2.2774	1	408.83489	3.07996	37.94123	189.48842	1657.03087
59	0.5960	0.3055	0.0985	2.6433	6.5617	15.2557	1.3760	2.0985	3.9031	1	389.05973	3.06986	32.78701	175.66967	1696.89830
60	0.5032	0.0433	0.4535	2.6757	33.9959	1.3284	1.4107	4.3810	0.9795	1	413.62013	3.08291	37.17836	175.97739	1668.70481
61	0.5273	0.4123	0.0604	2.6961	4.7051	23.4377	1.3703	1.8354	4.8225	1	374.14479	3.06231	29.36757	161.83788	1734.43563
62	0.4395	0.5279	0.0327	2.7852	3.3731	36.3826	1.3667	1.5838	5.9483	1	354.88070	3.05288	25.73817	145.74172	1782.79099
63	0.3004	0.0243	0.6753	2.7954	47.8816	0.3868	1.4442	5.4718	0.4838	1	412.13258	3.08263	35.21583	163.38949	1687.12696
64	0.3980	0.0358	0.5662	2.9203	38.0467	0.7911	1.4705	4.7531	0.6995	1	413.92632	3.08375	36.15728	170.55471	1673.10753
65	0.6051	0.0850	0.3099	2.9313	18.3792	3.4707	1.4650	3.0241	1.7178	1	413.03259	3.08314	37.76290	186.76992	1651.49463
66	0.6210	0.1649	0.2141	3.1180	10.8929	6.3751	1.5007	2.4645	2.4356	1	408.69023	3.08127	36.81349	189.41695	1650.60082
67	0.5329	0.0699	0.3972	3.2267	21.8412	2.2206	1.5325	3.3821	1.2898	1	415.41625	3.08509	37.29487	182.99613	1650.57203
68	0.5842	0.2765	0.1393	3.3259	7.1861	10.2429	1.5336	2.1460	3.1769	1	399.02243	3.07666	33.84917	180.69125	1670.91174
69	0.5148	0.4066	0.0786	3.4142	4.7917	17.4969	1.5331	1.8337	4.1784	1	382.98105	3.06860	29.93445	165.31156	1710.88355
70	0.4341	0.0546	0.5113	3.4888	26.3679	1.3101	1.5903	3.8480	0.9057	1	416.25988	3.08619	36.31751	176.76322	1655.19039
71	0.5848	0.1329	0.2823	3.5143	12.6873	4.3176	1.5867	2.5674	1.9250	1	413.55061	3.08467	37.07605	189.19244	1641.22981
72	0.2901	0.7001	0.0098	3.6259	2.0944	61.1259	1.5065	1.2584	7.5883	1	323.58414	3.04067	20.77653	122.26775	1856.82544

(*continued on next page*)

Table A.8 (*continued*).

θ_k	w_i			f_i			AAD(i)%			Parameters, $\boldsymbol{x}$					
	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	m	(ϵ/k_B)/K	σ/Å	λ^r	$K_{c,H}^{HB}$/Å^3	$(\epsilon_{c,H}^{HB}/k_B)$/K
73	0.4630	0.4788	0.0582	3.6502	3.8511	21.9174	1.5686	1.6685	4.6615	1	374.73374	3.06499	27.89445	156.73278	1731.45219
74	0.5342	0.3561	0.1097	3.7787	5.5230	12.4230	1.6143	1.9295	3.5181	1	393.06105	3.07453	31.52018	172.38175	1684.77867
75	0.5755	0.2303	0.1941	3.8190	8.2438	6.8554	1.6376	2.2257	2.5434	1	407.29658	3.08197	34.98792	185.13888	1650.34092
76	0.3250	0.0424	0.6326	3.8228	30.1471	0.7658	1.6589	4.2495	0.6902	1	415.72941	3.08672	34.87544	169.36151	1662.56659
77	0.5240	0.1021	0.3739	3.8569	15.3673	2.7653	1.6572	2.7784	1.4379	1	416.85139	3.08712	36.85030	186.18742	1637.88463
78	0.3932	0.5687	0.0381	4.0356	2.9334	28.5506	1.6293	1.4751	5.2861	1	363.56833	3.06037	25.44993	145.83762	1759.12927
79	0.3263	0.6534	0.0203	4.0917	2.3038	41.4280	1.6237	1.3197	6.3054	1	345.95004	3.05212	22.89436	133.37934	1802.65450
80	0.5235	0.1324	0.3441	4.2403	12.2262	3.2201	1.7256	2.5061	1.5744	1	416.89497	3.08782	36.34828	187.14235	1632.60446
81	0.1163	0.0142	0.8695	4.2881	43.9780	0.1134	1.7572	5.4484	0.2740	1	409.66125	3.08467	31.12231	149.29987	1695.08949
82	0.4528	0.0916	0.4557	4.3542	16.0887	2.0049	1.7505	2.8904	1.1290	1	418.63662	3.08902	36.01962	182.46593	1635.63635
83	0.3964	0.0718	0.5318	4.3922	19.3217	1.4390	1.7624	3.2629	0.9507	1	418.40946	3.08905	35.40724	177.63445	1641.98856
84	0.1977	0.0280	0.7743	4.5192	33.7775	0.3275	1.7984	4.6738	0.4468	1	413.05510	3.08675	32.17799	157.24511	1676.18812
85	0.5351	0.1923	0.2725	4.5414	9.0513	4.3447	1.7729	2.2498	1.9221	1	414.24622	3.08691	35.30756	186.22877	1633.75992
86	0.3169	0.0557	0.6274	4.6704	22.3658	0.9336	1.8198	3.6262	0.7693	1	417.41979	3.08912	34.06994	170.36699	1650.33334
87	0.5197	0.3276	0.1527	4.6919	5.7647	8.0364	1.7884	1.9327	2.7960	1	402.72591	3.08116	32.20298	176.15816	1659.13130
88	0.4442	0.4810	0.0748	4.7806	3.6666	15.4910	1.7876	1.6152	3.9308	1	384.66083	3.07218	28.15369	159.16011	1704.65130
89	0.3937	0.5536	0.0527	5.0337	2.9222	19.9209	1.8239	1.4607	4.4410	1	375.71969	3.06825	26.21479	150.41691	1727.12604
90	0.4767	0.1593	0.3640	5.1893	9.7667	2.8932	1.8945	2.2730	1.4431	1	418.48273	3.09020	34.98466	184.30014	1625.10892
91	0.4211	0.1145	0.4643	5.2319	12.3148	2.0125	1.9088	2.5167	1.1305	1	419.92773	3.09113	34.86609	181.12136	1627.20873
92	0.3629	0.0884	0.5487	5.3646	14.6127	1.4520	1.9369	2.8244	0.9547	1	419.68797	3.09130	34.11544	175.90624	1633.16638
93	0.4581	0.4377	0.1042	5.3665	4.0181	10.6537	1.8973	1.6646	3.2563	1	394.53862	3.07809	29.30187	164.84241	1678.68728
94	0.2306	0.0434	0.7260	5.4251	23.4757	0.5486	1.9563	3.8362	0.5860	1	415.60155	3.08947	31.82485	160.61768	1658.96538
95	0.5023	0.2833	0.2144	5.4312	6.2118	5.1687	1.9232	1.9535	2.1654	1	410.77144	3.08645	32.86660	179.22706	1638.47093
96	0.3369	0.6270	0.0361	5.4955	2.3030	24.3928	1.8952	1.3115	4.8884	1	367.53146	3.06516	24.40619	142.00772	1747.23448
97	0.2781	0.0677	0.6542	6.0339	16.0922	0.9044	2.0519	3.0890	0.7613	1	418.29670	3.09165	32.09899	166.22003	1641.75722
98	0.2826	0.6941	0.0233	6.0776	1.8164	29.0682	1.9870	1.1716	5.3057	1	359.28778	3.06237	22.73142	133.95104	1766.86533
99	0.2217	0.7685	0.0098	6.1798	1.4410	44.6497	1.9936	1.0444	6.4995	1	337.20676	3.05292	20.21743	120.92786	1819.45372
100	0.4599	0.3796	0.1605	6.2922	4.3814	6.1658	2.0542	1.6955	2.4268	1	405.95453	3.08515	30.37620	170.22729	1648.60204
101	0.4537	0.2249	0.3214	6.2963	6.7194	3.0671	2.0699	1.9548	1.5176	1	417.79950	3.09132	33.00602	179.52341	1621.71492
102	0.3972	0.1499	0.4529	6.3485	8.7529	2.0318	2.0876	2.1438	1.1333	1	420.62464	3.09300	33.25622	177.95029	1619.96347
103	0.3894	0.5357	0.0749	6.3618	2.8325	12.3667	2.0493	1.4240	3.5125	1	388.80932	3.07673	26.88627	154.68731	1692.14959
104	0.3190	0.1021	0.5789	6.6430	10.9500	1.2989	2.1414	2.4628	0.9109	1	420.24395	3.09334	32.13402	170.51363	1627.68677
105	0.1057	0.0210	0.8732	6.7768	26.4505	0.1454	2.1886	4.2634	0.3044	1	410.96104	3.08901	28.15059	143.57129	1677.88097
106	0.3354	0.6142	0.0504	6.7989	2.1876	15.7803	2.1167	1.2686	3.9562	1	380.77418	3.07350	25.04272	146.05747	1712.17569
107	0.4023	0.1948	0.4029	7.0037	6.6773	2.2457	2.1823	1.9140	1.1873	1	420.26317	3.09353	32.27886	176.19869	1616.65611
108	0.4088	0.4690	0.1223	7.2438	3.1557	6.9075	2.1977	1.4702	2.6006	1	401.87891	3.08429	28.20532	161.46067	1657.64701
109	0.2877	0.6778	0.0346	7.3033	1.7180	19.2302	2.1945	1.1354	4.3457	1	373.29513	3.07072	23.46589	138.34927	1730.53036
110	0.4231	0.3285	0.2484	7.6145	4.2859	3.3481	2.2645	1.6335	1.6603	1	415.14880	3.09142	30.41234	171.01197	1624.33576
111	0.1704	0.0453	0.7844	7.6336	16.3857	0.3872	2.3162	3.2798	0.4913	1	414.94972	3.09197	28.46866	149.94940	1654.75947
112	0.2431	0.0807	0.6762	7.6943	11.0445	0.8160	2.3160	2.5798	0.7203	1	418.65737	3.09385	29.79935	159.75708	1635.00003
113	0.3284	0.1572	0.5144	7.9621	6.5567	1.5235	2.3398	1.8730	0.9782	1	421.24822	3.09523	30.73827	168.87542	1616.56438
114	0.3767	0.2423	0.3810	8.1756	4.8164	2.1108	2.3588	1.6601	1.1468	1	420.24994	3.09474	30.55253	170.72278	1613.36277
115	0.2327	0.7460	0.0213	8.2195	1.2651	22.2627	2.3301	0.9800	4.6482	1	366.19228	3.06883	21.82404	130.16754	1746.86423
116	0.3377	0.5871	0.0751	8.2553	2.0647	8.9370	2.3455	1.2185	2.9835	1	394.15444	3.08167	25.62217	149.76887	1676.37167
117	0.3805	0.4269	0.1926	8.8987	2.8610	3.4204	2.4471	1.3715	1.7286	1	412.68542	3.09138	28.16691	162.14349	1628.36055
118	0.2840	0.6662	0.0498	8.9739	1.5104	10.9567	2.4464	1.0572	3.3005	1	387.69785	3.07946	23.92676	141.47286	1692.20108
119	0.1742	0.8160	0.0098	9.1606	0.9119	27.7951	2.4605	0.8315	5.1490	1	354.49801	3.06528	19.98670	120.60527	1773.25014
120	0.3463	0.5347	0.1189	9.2860	2.1099	4.8696	2.4953	1.2105	2.1642	1	405.54359	3.08818	26.41597	154.15871	1646.01850
121	0.2531	0.1275	0.6194	9.3029	5.9892	0.9806	2.5413	1.8572	0.7888	1	420.33208	3.09617	28.39240	157.83966	1621.01754
122	0.3129	0.2288	0.4583	9.6211	3.7930	1.4954	2.5687	1.4632	0.9585	1	421.36256	3.09670	28.65216	162.51983	1610.23038
123	0.1948	0.0849	0.7203	9.7176	7.5810	0.5953	2.6071	2.1889	0.6124	1	417.96782	3.09544	27.03840	149.29199	1632.89044
124	0.0337	0.0098	0.9565	9.8384	18.4750	0.0578	2.6457	3.6694	0.2093	1	408.64571	3.09119	24.30685	128.89563	1677.43894
125	0.3333	0.3333	0.3333	10.1088	2.7515	1.8215	2.6222	1.3004	1.0667	1	419.76158	3.09607	27.89070	160.65411	1610.56455
126	0.2867	0.6360	0.0773	10.3863	1.3927	5.5845	2.6406	1.0039	2.3467	1	400.79016	3.08686	24.48329	144.83885	1657.21115
127	0.3272	0.5007	0.1721	10.5181	1.8788	2.7669	2.6638	1.1285	1.5436	1	413.13974	3.09291	26.29528	153.95257	1625.37918
128	0.2444	0.1828	0.5728	10.9319	3.3171	0.9772	2.7507	1.3683	0.7813	1	420.97197	3.09766	26.73006	152.71020	1612.98799
129	0.0919	0.0351	0.8729	10.9952	10.3433	0.1621	2.7845	2.7135	0.3244	1	412.59384	3.09383	24.25039	132.76548	1654.98893
130	0.2030	0.7688	0.0281	11.0751	0.8241	10.4258	2.7248	0.7910	3.2072	1	384.89974	3.08029	21.82214	130.80256	1696.84541
131	0.2827	0.6012	0.1162	11.4371	1.2828	3.0916	2.7775	0.9529	1.6915	1	409.43161	3.09173	24.72057	146.31240	1633.93705
132	0.1759	0.1274	0.6967	12.2828	3.2558	0.5477	2.9280	1.4396	0.5853	1	418.74971	3.09757	24.63408	140.93526	1621.68645
133	0.1344	0.8558	0.0098	12.3123	0.4930	13.8278	2.8749	0.6133	3.6545	1	373.28226	3.07676	19.85287	120.09880	1723.20404
134	0.2735	0.5039	0.2225	12.3354	1.2533	1.5628	2.8938	0.9211	1.0401	1	417.67711	3.09630	25.00930	147.71856	1612.29768
135	0.2444	0.2965	0.4591	12.3497	1.6935	1.0210	2.9137	1.0091	0.7872	1	421.21464	3.09842	25.39218	148.29559	1606.47089
136	0.2263	0.6986	0.0751	12.7278	0.7768	3.0739	2.9313	0.7557	1.7140	1	406.56050	3.09117	23.09716	137.86527	1640.14137
137	0.1173	0.0728	0.8099	13.0295	4.3111	0.2660	3.0243	1.7414	0.3983	1	415.31995	3.09643	23.16076	131.30275	1635.36247
138	0.1600	0.8093	0.0308	13.6999	0.4408	4.4079	3.0402	0.5811	2.0869	1	397.91599	3.08792	21.32724	128.10920	1661.04567
139	0.2209	0.4267	0.3525	13.7289	0.8864	0.9280	3.0664	0.7642	0.7318	1	420.43040	3.09851	23.94856	141.74233	1605.12276
140	0.2192	0.6455	0.1352	13.7294	0.6753	1.4720	3.0508	0.6966	1.0737	1	414.87535	3.09552	23.26273	138.69506	1617.90032
141	0.1584	0.1965	0.6451	14.3337	1.2598	0.4947	3.1584	0.8965	0.5507	1	418.99854	3.09865	22.93060	133.97819	1614.09426
142	0.1654	0.6162	0.2184	15.4225	0.3561	0.7142	3.2455	0.5095	0.6334	1	418.69554	3.09807	22.23318	132.60271	1606.94409
143	0.0876	0.0985	0.8139	15.8594	1.4618	0.1985	3.3361	1.0493	0.3595	1	414.12353	3.09738	20.99247	121.57431	1628.94425
144	0.1355	0.3002	0.5643	15.8827	0.4561	0.4306	3.3206	0.5473	0.5060	1	418.52538	3.09894	21.70424	128.21162	1610.43494
145	0.0868	0.9020	0.0112	15.9004	0.1628	2.7274	3.2804	0.3544	1.6324	1	399.08175	3.08974	19.97264	120.07966	1654.98762

(*continued on next page*)

Table A.8 (*continued*).

θ_k	w_i			f_i			AAD(i)%			Parameters, $\boldsymbol{x}$					
	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	m	(ϵ/k_B)/K	σ/Å	λ^r	$K_{c,H}^{HB}$/Å^3	$(\epsilon_{c,H}^{HB}/k_B)$/K
146	0.0098	0.0098	0.9804	16.2254	3.5659	0.0716	3.4262	1.6809	0.2276	1	407.27064	3.09634	19.92167	112.86677	1653.84095
147	0.0779	0.5524	0.3698	17.7383	0.0734	0.3085	3.5067	0.2443	0.4323	1	416.34854	3.09855	20.23600	120.25753	1610.89222
148	0.0373	0.1234	0.8393	18.9514	0.2569	0.1190	3.6633	0.4608	0.2962	1	408.78085	3.09743	18.82192	110.33025	1631.46323
149	0.0098	0.9804	0.0098	19.0067	0.0140	0.4702	3.6008	0.1051	0.6125	1	421.77481	3.09985	19.83148	117.56479	1594.06433
150	0.0098	0.4712	0.5189	20.2645	0.0408	0.1285	3.8982	0.1703	0.3089	1	408.65723	3.10253	18.56456	109.98060	1626.91189

Table A.9
Pareto points and parameters for water models with $m \geq 1$, derived using all three objective functions. The models are sorted with $f_{\rho_L^{sat}}$ in increasing order.

θ_k	w_i			f_i			AAD(i)%			Parameters, $\boldsymbol{x}$					
	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	ρ_L^{sat}	P^{vap}	C_P	m	(ϵ/k_B)/K	σ/Å	λ^r	$K_{c,H}^{HB}$/Å^3	$(\epsilon_{c,H}^{HB}/k_B)$/K
1	0.9804	0.0098	0.0098	0.7719	6.2275	2.2512	0.6706	1.8767	1.4957	1.11773	371.50797	2.93110	31.39086	180.15540	1680.81699
2	0.9716	0.0100	0.0184	0.7942	6.0024	0.7286	0.6813	1.8259	0.8388	1.13396	374.29928	2.91758	31.53519	184.00066	1662.19418
3	0.9738	0.0140	0.0122	0.8012	4.5335	1.4220	0.6852	1.6365	1.1844	1.13182	372.68108	2.91864	31.49293	185.44185	1663.85288
4	0.9586	0.0098	0.0316	0.8034	6.0970	0.3072	0.6856	1.8424	0.5154	1.14215	375.24596	2.91063	31.48800	185.07336	1654.66835
5	0.9061	0.0098	0.0841	0.8132	6.1915	0.0932	0.6901	1.8877	0.2508	1.15264	374.71137	2.90089	31.07680	184.53258	1650.27798
6	0.9665	0.0237	0.0098	0.8219	3.2685	1.6589	0.6956	1.4586	1.2811	1.13972	370.49508	2.91021	31.07029	186.93011	1660.53970
7	0.9555	0.0178	0.0267	0.8338	3.6252	0.3955	0.7004	1.4727	0.5724	1.15216	374.38711	2.90107	31.43029	190.19240	1643.37314
8	0.6677	0.0098	0.3225	0.8376	5.2910	0.0265	0.7023	1.7994	0.1495	1.17033	370.33900	2.88287	29.87199	181.76605	1650.75594
9	0.8271	0.0156	0.1574	0.8510	3.7632	0.0673	0.7088	1.4755	0.2207	1.17053	371.89983	2.88348	30.38047	186.85978	1641.21627
10	0.9180	0.0223	0.0597	0.8572	2.9603	0.1639	0.7120	1.3439	0.3259	1.16687	373.43135	2.88748	30.94974	191.25963	1635.00362
11	0.9468	0.0299	0.0234	0.8583	2.6154	0.3821	0.7137	1.3084	0.5616	1.16296	372.32603	2.89034	30.93058	191.54188	1638.44977
12	0.9485	0.0416	0.0098	0.8592	2.3294	1.0237	0.7165	1.2694	1.0017	1.15960	368.18918	2.89112	30.26195	187.95819	1651.19855
13	0.4141	0.0098	0.5761	0.8820	3.3107	0.0169	0.7239	1.4225	0.1193	1.18904	366.09880	2.86437	28.91049	181.88390	1646.51051
14	0.8674	0.0341	0.0986	0.8895	2.0923	0.0976	0.7291	1.1723	0.2585	1.18306	370.23201	2.87165	30.07703	190.93115	1630.99619
15	0.6954	0.0230	0.2816	0.8958	2.2775	0.0383	0.7311	1.1876	0.1743	1.19017	367.46658	2.86409	29.30879	186.55900	1637.18188
16	0.9259	0.0589	0.0152	0.8980	1.7620	0.3140	0.7358	1.1172	0.5291	1.18131	367.25982	2.87162	29.64280	189.41987	1638.23797
17	0.8747	0.0602	0.0651	0.9210	1.5172	0.0782	0.7465	1.0348	0.2357	1.19494	367.29650	2.85995	29.36376	190.49775	1629.04797
18	0.8972	0.0930	0.0098	0.9421	1.2451	0.1747	0.7607	0.9561	0.3968	1.20015	362.74795	2.85320	28.45624	186.66266	1637.88933
19	0.5412	0.0317	0.4272	0.9435	1.3911	0.0217	0.7569	0.9705	0.1373	1.20797	363.25104	2.84681	28.33412	185.77280	1634.52741
20	0.1783	0.0098	0.8119	0.9537	1.4074	0.0142	0.7593	0.9685	0.1005	1.21334	361.16260	2.84118	27.87938	183.40497	1638.35826
21	0.7810	0.0743	0.1447	0.9563	1.1438	0.0327	0.7657	0.9108	0.1637	1.21004	363.54923	2.84528	28.40757	188.02054	1629.43721
22	0.8528	0.1374	0.0098	0.9989	0.8186	0.0164	0.7888	0.7849	0.1146	1.22488	359.53340	2.83100	27.44930	185.49637	1630.24497
23	0.0098	0.0098	0.9804	1.0399	0.6499	0.0141	0.7956	0.6968	0.0902	1.23882	356.32783	2.81742	26.92255	185.37793	1628.94752
24	0.7522	0.1887	0.0591	1.0546	0.5356	0.0243	0.8179	0.6421	0.1391	1.24449	354.64271	2.81267	26.31932	182.01214	1631.30549
25	0.7606	0.2296	0.0098	1.0871	0.4029	0.2020	0.8325	0.5591	0.4325	1.25995	354.52660	2.80032	26.06227	183.18016	1621.17614
26	0.2473	0.1709	0.5818	1.0972	0.4172	0.0285	0.8448	0.5610	0.1348	1.25175	352.54771	2.80630	25.66531	179.91774	1630.85291
27	0.6383	0.2689	0.0928	1.1145	0.3352	0.0644	0.8479	0.5108	0.2313	1.26341	349.64239	2.79527	25.24397	178.31599	1633.23677
28	0.6737	0.3039	0.0224	1.1349	0.2717	0.1904	0.8562	0.4620	0.4206	1.27398	349.76269	2.78703	25.13530	179.35368	1625.95602
29	0.1940	0.2199	0.5861	1.1434	0.3379	0.0382	0.8712	0.5037	0.1596	1.25656	351.23210	2.80242	25.12615	177.6236	1630.62645
30	0.3333	0.3333	0.3333	1.1590	0.2525	0.0802	0.8724	0.4359	0.2482	1.27100	347.26919	2.78867	24.58727	175.71894	1633.94231
31	0.6273	0.3629	0.0098	1.1678	0.1845	0.7386	0.8703	0.3822	0.8494	1.29083	350.70344	2.77456	25.03419	181.71037	1611.65480
32	0.5834	0.3896	0.0270	1.1820	0.1793	0.2584	0.8782	0.3738	0.4870	1.28850	346.01617	2.77410	24.38553	176.59154	1627.40023
33	0.1408	0.2689	0.5903	1.2078	0.2671	0.0501	0.8970	0.4431	0.1913	1.26180	350.09497	2.79832	24.59792	175.39033	1629.39518
34	0.3189	0.4707	0.2104	1.2095	0.1628	0.1503	0.8930	0.3450	0.3438	1.28750	343.03392	2.77389	23.80463	172.75207	1635.85910
35	0.0663	0.1502	0.7835	1.2301	0.3345	0.0259	0.9058	0.4952	0.1227	1.24979	353.63873	2.81008	25.00890	176.81948	1625.75383
36	0.4343	0.5158	0.0499	1.2326	0.1203	0.2645	0.9012	0.2969	0.4693	1.30134	341.29706	2.76227	23.53922	172.66032	1632.77546
37	0.4401	0.5501	0.0098	1.2434	0.0812	1.1175	0.9038	0.2465	1.0419	1.31622	345.84190	2.75321	24.00988	178.64380	1609.24096
38	0.1449	0.4028	0.4523	1.2594	0.1906	0.0845	0.9162	0.3702	0.2548	1.27272	347.18776	2.78863	23.96077	172.76385	1630.55164
39	0.2569	0.7333	0.0098	1.2865	0.0495	1.5611	0.9224	0.1950	1.2355	1.33137	344.13926	2.74139	23.56392	177.92134	1604.01163
40	0.1890	0.6630	0.1480	1.2879	0.0960	0.2418	0.9257	0.2625	0.4390	1.30408	339.17107	2.76008	22.90371	169.10635	1635.57306
41	0.2379	0.7315	0.0306	1.2920	0.0702	0.5083	0.9259	0.2302	0.6650	1.32142	338.52033	2.74616	22.91348	171.28099	1627.59867
42	0.1469	0.5461	0.3070	1.2999	0.1275	0.1464	0.9301	0.2972	0.3372	1.28741	343.27997	2.77543	23.28348	170.05583	1632.65640
43	0.0510	0.9392	0.0098	1.3347	0.0386	1.7523	0.9472	0.1725	1.3100	1.33780	343.61283	2.73734	23.11968	176.17845	1599.75877
44	0.0098	0.0412	0.9490	1.3358	0.3683	0.0179	0.9352	0.5186	0.0949	1.24010	356.69676	2.81991	25.13331	176.91324	1621.50113
45	0.0503	0.9279	0.0218	1.3654	0.0521	0.7279	0.9576	0.1996	0.7961	1.32823	339.36753	2.74286	22.50439	170.11667	1618.20145
46	0.0804	0.8240	0.0956	1.3818	0.0704	0.2976	0.9595	0.2301	0.4943	1.31173	338.08926	2.75489	22.32681	166.58421	1631.96313
47	0.0778	0.3268	0.5953	1.4157	0.1869	0.0564	0.9661	0.3620	0.2043	1.26343	350.03607	2.79864	23.80916	171.41937	1625.60122
48	0.0897	0.4801	0.4302	1.4243	0.1359	0.0962	0.9680	0.3048	0.2730	1.27539	346.75094	2.78763	23.30379	169.40713	1627.82656
49	0.0758	0.7006	0.2235	1.4728	0.0848	0.1765	0.9822	0.2459	0.3749	1.29299	342.27345	2.77211	22.54882	166.38978	1629.52166
50	0.0249	0.9178	0.0573	1.6804	0.0578	0.2760	1.0449	0.2094	0.4864	1.30275	342.09295	2.76614	21.95364	164.07908	1621.24596
51	0.0168	0.9582	0.0250	1.6859	0.0488	0.5119	1.0532	0.1926	0.6752	1.30796	344.38607	2.76350	22.14716	166.41554	1610.77227
52	0.0544	0.5448	0.4007	1.6910	0.1021	0.0932	1.0357	0.2641	0.2702	1.27346	347.58920	2.79104	22.76179	166.21902	1623.83739
53	0.0393	0.2812	0.6795	1.7121	0.1701	0.0387	1.0411	0.3459	0.1619	1.25378	353.03633	2.80925	23.54388	169.24769	1620.36697
54	0.0098	0.9804	0.0098	1.7899	0.0335	1.4370	1.0905	0.1599	1.1899	1.31059	351.06132	2.76552	22.63659	171.23157	1587.38629
55	0.0161	0.1409	0.8430	1.9643	0.1871	0.0246	1.0987	0.3634	0.1056	1.24216	356.50974	2.82109	23.61552	168.80043	1616.06082
56	0.0288	0.8031	0.1682	1.9921	0.0639	0.1376	1.1123	0.2185	0.3398	1.28230	345.88046	2.78494	21.93198	162.04784	1620.55440
57	0.0284	0.3835	0.5881	2.1755	0.1092	0.0446	1.1446	0.2724	0.1768	1.25372	353.30379	2.81116	22.74299	164.61070	1616.47529
58	0.0249	0.6222	0.3529	2.3623	0.0724	0.0713	1.1831	0.2271	0.2365	1.26301	350.82493	2.80339	22.08567	161.38225	1616.34056
59	0.0098	0.9501	0.0401	2.4376	0.0482	0.2323	1.2274	0.1915	0.4677	1.27927	350.10681	2.79158	21.72572	161.10134	1606.75502
60	0.0098	0.1709	0.8193	2.6561	0.1190	0.0268	1.2297	0.2827	0.1129	1.23831	357.82540	2.82658	22.76946	163.52130	1611.06279
61	0.0098	0.8565	0.1337	3.2295	0.0514	0.0724	1.3709	0.1974	0.2486	1.25473	353.81957	2.81451	21.39144	156.39127	1608.35986
62	0.0098	0.3253	0.6649	3.3565	0.0730	0.0318	1.3620	0.2259	0.1251	1.23661	358.25215	2.83008	21.96677	158.37690	1607.70377

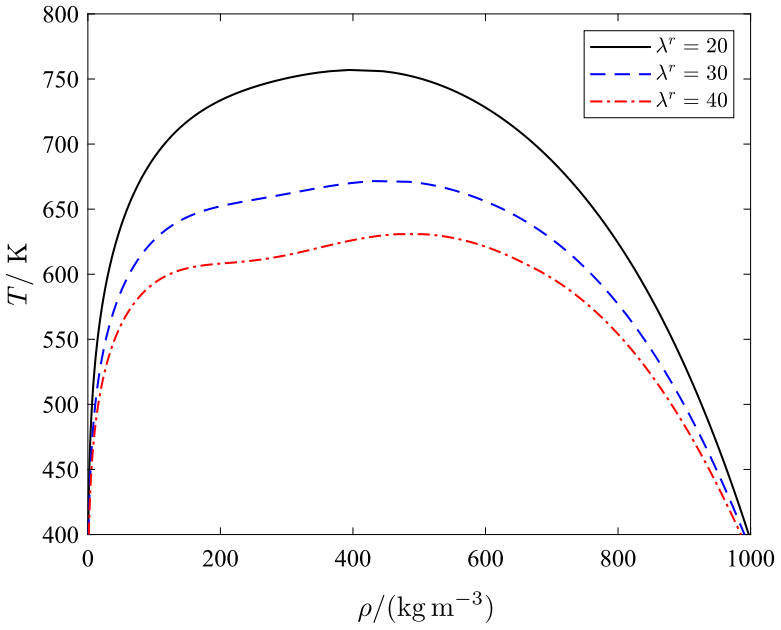


Fig. A.21. Effect of λ^r on the saturated-liquid density and saturated-vapour density obtained with the SAFT-VR Mie EoS (Lafitte et al., 2013). The values of the other parameters are fixed to the typical values for water $m = 1$, $\lambda^a = 6$, $K^{HB}_{c,H} = 180$ Å^3, $\sigma = 3$ Å, $\epsilon/k_B = 390$ K and $\epsilon^{HB}_{c,H} = 1700$ K.

require significant computational effort. From a chemical engineering perspective, the spherical models of water that we have identified during the course of the MOO solution are suitable for typical process operating conditions that are away from the critical point of water, but one must be careful when using these models to calculate properties at near-critical conditions.

References

Alsaifi, N.M., Alkhater, M., Binous, H., Aslani, I.A., Alsunni, Y., Wang, Z.-G., 2019. Non-physical behavior in several statistical mechanically based equations of state. Ind. Eng. Chem. Res. 58 (3), 1382–1395. http://dx.doi.org/10.1021/acs.iecr.8b04656.

Alsaifi, N.M., Aslani, I.A., Binous, H., Wang, Z.G., 2017. A priori determination of the region of the three physical volume root loci in the perturbed-chain SAFT EOS. Fluid Phase Equilib. 434, 152–166. http://dx.doi.org/10.1016/j.fluid.2016.11.023.

Alsaifi, N.M., Englezos, P., 2011. Prediction of multiphase equilibrium using the PC-SAFT equation of state and simultaneous testing of phase stability. Fluid Phase Equilib. 302 (1–2), 169–178. http://dx.doi.org/10.1016/j.fluid.2010.09.002.

Aparicio-Martínez, S., Hall, K.R., 2007. Phase equilibria in water containing binary systems from molecular based equations of state. Fluid Phase Equilib. 254 (1–2), 112–125. http://dx.doi.org/10.1016/j.fluid.2007.02.030.

Aslam, N., Sunol, A.K., 2006. Reliable computation of all the density roots of the statistical associating fluid theory equation of state through global fixed-point homotopy. Ind. Eng. Chem. Res. 45 (9), 3303–3310. http://dx.doi.org/10.1021/ie0489554.

Barber, C.B., Dobkin, D.P., Huhdanpaa, H., 1996. The quickhull algorithm for convex hulls. ACM Trans. Math. Software (TOMS) 22 (4), 469–483. http://dx.doi.org/10.1145/235815.235821.

Bard, Y., 1974. Nonlinear Parameter Estimation. Academic Press.

Blas, F.J., Vega, L.F., 1997. Thermodynamic behaviour of homonuclear and heteronuclear Lennard-Jones chains with association sites from simulation and theory. Mol. Phys. 92 (1), 135–150. http://dx.doi.org/10.1080/002689797170707.

Blekherman, G., Parrilo, P.A., Thomas, R.R., 2012. Semidefinite Optimization and Convex Algebraic Geometry. SIAM.

Bokrantz, R., Forsgren, A., 2013. An algorithm for approximating convex Pareto surfaces based on dual techniques. INFORMS J. Comput. 25 (2), 377–393. http://dx.doi.org/10.1287/ijoc.1120.0508.

Boulougouris, G.C., Voutsas, E.C., Economou, I.G., Theodorou, D.N., Tassios, D.P., 2001. Henry's constant analysis for water and nonpolar solvents from experimental data, macroscopic models, and molecular simulation. J. Phys. Chem. B 105 (32), 7792–7798. http://dx.doi.org/10.1021/jp010426f.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.

Button, J., Gubbins, K., 1999. SAFT prediction of vapour-liquid equilibria of mixtures containing carbon dioxide and aqueous monoethanolamine or diethanolamine. Fluid Phase Equilib. 158–160, 175–181. http://dx.doi.org/10.1016/s0378-3812(99)00150-8.

Cerdeiriña, C.A., González-Salgado, D., Romani, L., del Carmen Delgado, M., Torres, L.A., Costas, M., 2004. Towards an understanding of the heat capacity of liquids. A simple two-state model for molecular association. J. Chem. Phys. 120 (14), 6648–6659. http://dx.doi.org/10.1063/1.1667469.

Cerdeiriña, C.A., Troncoso, J., González-Salgado, D., García-Miaja, G., Hernández-Segura, G.O., Bessières, D., Medeiros, M., Romaní, L., Costas, M., 2007. Heat capacity of associated systems. Experimental data and application of a two-state model to pure liquids and mixtures. J. Phys. Chem. B 111 (5), 1119–1128. http://dx.doi.org/10.1021/jp0640272.

Cheikh, M., Jarboui, B., Loukil, T., Siarry, P., 2010. A method for selecting Pareto optimal solutions in multiobjective optimization. J. Inform. Math. Sci. 2 (1), 51–62.

Clark, G.N.I., Haslam, A.J., Galindo, A., Jackson, G., 2006. Developing optimal Wertheim-like models of water for use in statistical associating fluid theory (SAFT) and related approaches. Mol. Phys. 104 (22–24), 3561–3581. http://dx.doi.org/10.1080/00268970601081475.

Costa, A.L.H., da Silva, F.P.T., Pessoa, F.L.P., 2000. Parameter estimation of thermodynamic models for high-pressure systems employing a stochastic method of global optimization. Braz. J. Chem. Eng. 17 (3), 349–354. http://dx.doi.org/10.1590/s0104-66322000000300011.

Cripwell, J.T., Smith, S.A.M., Schwarz, C.E., Burger, A.J., 2018. SAFT-VR Mie: Application to phase equilibria of alcohols in mixtures with n-alkanes and water. Ind. Eng. Chem. Res. 57 (29), 9693–9706. http://dx.doi.org/10.1021/acs.iecr.8b01042.

Das, I., 1999. On characterizing the 'knee' of the Pareto curve based on normal-boundary intersection. Struct. Optim. 18 (2–3), 107–115. http://dx.doi.org/10.1007/bf01195985.

Das, I., Dennis, J.E., 1997. A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. Struct. Optim. 14 (1), 63–69. http://dx.doi.org/10.1007/bf01197559.

Dufal, S., Lafitte, T., Haslam, A.J., Galindo, A., Clark, G.N., Vega, C., Jackson, G., 2015. The A in SAFT: Developing the contribution of association to the Helmholtz free energy within a Wertheim TPT1 treatment of generic Mie fluids. Mol. Phys. 113 (9–10), 948–984. http://dx.doi.org/10.1080/00268976.2015.1029027.

Dufal, S., Lafitte, T., Haslam, A.J., Galindo, A., Clark, G.N., Vega, C., Jackson, G., 2018. Corrigendum: The a in SAFT: Developing the contribution of association to the Helmholtz free energy within a Wertheim TPT1 treatment of generic Mie fluids. Mol. Phys. 2, 283–285. http://dx.doi.org/10.1080/00268976.2017.1402604.

Economou, I.G., Donohue, M.D., 1992. Equation of state with multiple associating sites for water and water-hydrocarbon mixtures. Ind. Eng. Chem. Res. 31 (10), 2388–2394. http://dx.doi.org/10.1021/ie00010a019.

Economou, I.G., Tsonopoulos, C., 1997. Associating models and mixing rules in equations of state for water/hydrocarbon mixtures. Chem. Eng. Sci. 52 (4), 511–525. http://dx.doi.org/10.1016/s0009-2509(96)00441-1.

Englezos, P., Kalogerakis, N., 2000. Applied Parameter Estimation for Chemical Engineers. CRC Press.

Fawzi, H., 2017. Topics in convex optimization.

Forte, E., Burger, J., Langenbach, K., Hasse, H., Bortz, M., 2017. Multi-criteria optimization for parameterization of SAFT-type equations of state for water. AIChE Journal 64 (1), 226–237. http://dx.doi.org/10.1002/aic.15857.

Forte, E., Kulkarni, A., Burger, J., Bortz, M., Küfer, K.-H., Hasse, H., 2020. Multi-criteria optimization for parametrizing excess Gibbs energy models. Fluid Phase Equilib. 522, 112676. http://dx.doi.org/10.1016/j.fluid.2020.112676.

Fu, Y.H., Sandler, S.I., 1995. A simplified SAFT equation of state for associating compounds and mixtures. Ind. Eng. Chem. Res. 34 (5), 1897–1909. http://dx.doi.org/10.1021/ie00044a042.

Glass, M., Djelassi, H., Mitsos, A., 2018. Parameter estimation for cubic equations of state models subject to sufficient criteria for thermodynamic stability. Chem. Eng. Sci. 192, 981–992. http://dx.doi.org/10.1016/j.ces.2018.08.033.

Gloor, G.J., 2003. Modelling the Interfacial Properties of Complex Fluids with Density Functional Theory and Computer Simulation (Ph.D. thesis). Imperial College London (University of London).

Gloor, G.J., Jackson, G., Blas, F.J., del Rıo, E.M., de Miguel, E., 2004. An accurate density functional theory for the vapor-liquid interface of associating chain molecules based on the statistical associating fluid theory for potentials of variable range. J. Chem. Phys. 121 (24), 12740. http://dx.doi.org/10.1063/1.1807833.

Graham, E.J., 2020. Development of Advanced Thermodynamic Models for CO_2 Absorption: From Numerical Methods to Process Modelling (Ph.D. thesis). Imperial College London.

Grenner, A., Kontogeorgis, G.M., Michelsen, M.L., Folas, G.K., 2007. On the estimation of water pure compound parameters in association theories. Mol. Phys. 105 (13–14), 1797–1801. http://dx.doi.org/10.1080/00268970701416597.

Grenner, A., Schmelzer, J., von Solms, N., Kontogeorgis, G.M., 2006. Comparison of two association models (Elliott-Suresh-Donohue and simplified PC-SAFT) for complex phase equilibria of hydrocarbon-water and amine-containing mixtures. Ind. Eng. Chem. Res. 45 (24), 8170–8179. http://dx.doi.org/10.1021/ie0605332.

Gross, J., Sadowski, G., 2000. Application of perturbation theory to a hard-chain reference fluid: An equation of state for square-well chains. Fluid Phase Equilib. 168 (2), 183–199. http://dx.doi.org/10.1016/s0378-3812(00)00302-2.

Gross, J., Sadowski, G., 2001. Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules. Ind. Eng. Chem. Res. 40 (4), 1244–1260. http://dx.doi.org/10.1021/ie0003887.

Gross, J., Sadowski, G., 2002. Application of the perturbed-chain SAFT equation of state to associating systems. Ind. Eng. Chem. Res. 41 (22), 5510–5515. http://dx.doi.org/10.1021/ie010954d.

Gross, J., Vrabec, J., 2006. An equation-of-state contribution for polar components: Dipolar molecules. AIChE J. 52 (3), 1194–1204. http://dx.doi.org/10.1002/aic.10683.

Haimes, Y., 1971. On a bicriterion formulation of the problems of integrated system identification and system optimization. IEEE Trans. Syst. Man Cybern. 1 (3), 296–297.

Hendriks, E., Kontogeorgis, G.M., Dohrn, R., de Hemptinne, J.C., Economou, I.G., Žilnik, L.F., Vesovic, V., 2010. Industrial requirements for thermodynamics and transport properties. Ind. Eng. Chem. Res. 49 (22), 11131–11141. http://dx.doi.org/10.1021/ie101231b.

Holtzman, J.M., Halkin, H., 1966. Discretional convexity and the maximum principle for discrete systems. SIAM J. Control 4 (2), 263–275. http://dx.doi.org/10.1137/0304023.

Huang, S.H., Radosz, M., 1990. Equation of state for small, large, polydisperse, and associating molecules. Ind. Eng. Chem. Res. 29 (11), 2284–2294. http://dx.doi.org/10.1021/ie00107a014.

Kalikhman, V., Kost, D., Polishuk, I., 2010. About the physical validity of attaching the repulsive terms of analytical EOS models by temperature dependencies. Fluid Phase Equilib. 293 (2), 164–167.

Karakatsani, E.K., Economou, I.G., 2006. Perturbed chain-statistical associating fluid theory extended to dipolar and quadrupolar molecular fluids. J. Phys. Chem. B 110 (18), 9252–9261. http://dx.doi.org/10.1021/jp056957b.

Karakatsani, E.K., Kontogeorgis, G.M., Economou, I.G., 2006. Evaluation of the truncated perturbed chain-polar statistical associating fluid theory for complex mixture fluid phase equilibria. Ind. Eng. Chem. Res. 45 (17), 6063–6074. http://dx.doi.org/10.1021/ie060313o.

Karakatsani, E.K., Spyriouni, T., Economou, I.G., 2005. Extended statistical associating fluid theory (SAFT) equations of state for dipolar fluids. AIChE J. 51 (8), 2328–2342. http://dx.doi.org/10.1002/aic.10473.

Kleder, M., 2005. CON2vert - constraints to vertices. URL: https://uk.mathworks.com/matlabcentral/fileexchange/7894-con2vert-constraints-to-vertices. (Accessed 1 September 2020).

Kleiner, M., Sadowski, G., 2007. Modeling of polar systems using PCP-SAFT: An approach to account for induced-association interactions†. J. Phys. Chem. C 111 (43), 15544–15553. http://dx.doi.org/10.1021/jp072640v.

Koak, N., de Loos, T.W., Heidemann, R.A., 1999. Effect of the power series dispersion term on the pressure-volume behavior of statistical associating fluid theory. Ind. Eng. Chem. Res. 38 (4), 1718–1722. http://dx.doi.org/10.1021/ie9804069.

Kontogeorgis, G.M., Folas, G.K., 2010. Thermodynamic Models for Industrial Applications: From Classical and Advanced Mixing Rules to Association Theories. John Wiley & Sons, New York.

Kontogeorgis, G.M., Tsivintzelis, I., von Solms, N., Grenner, A., Bøgh, D., Frost, M., Knage-Rasmussen, A., Economou, I.G., 2010. Use of monomer fraction data in the parametrization of association theories. Fluid Phase Equilib. 296 (2), 219–229. http://dx.doi.org/10.1016/j.fluid.2010.05.028.

Kontogeorgis, G.M., Yakoumis, I.V., Meijer, H., Hendriks, E., Moorwood, T., 1999. Multicomponent phase equilibrium calculations for water–methanol–alkane mixtures. Fluid Phase Equilib. 158–160, 201–209. http://dx.doi.org/10.1016/s0378-3812(99)00060-6.

Kulkarni, A., Bortz, M., Küfer, K.-H., Kohns, M., Hasse, H., 2020. Multicriteria optimization of molecular models of water using a reduced units approach. J. Chem. Theory Comput. 16 (8), 5127–5138. http://dx.doi.org/10.1021/acs.jctc.0c00301.

Lafitte, T., Apostolakou, A., Avendaño, C., Galindo, A., Adjiman, C.S., Müller, E.A., Jackson, G., 2013. Accurate statistical associating fluid theory for chain molecules formed from Mie segments. J. Chem. Phys. 139 (15), 154504. http://dx.doi.org/10.1063/1.4819786.

Lafitte, T., Papaioannou, V., Dufal, S., Pantelides, C.C., 2017. gSAFT: Advanced physical property prediction for process modelling. In: Computer Aided Chemical Engineering. Elsevier, pp. 1003–1008. http://dx.doi.org/10.1016/b978-0-444-63965-3.50169-0.

Lafitte, T., Piñeiro, M.M., Daridon, J.-L., Bessières, D., 2007. A comprehensive description of chemical association effects on second derivative properties of alcohols through a SAFT-VR approach. J. Phys. Chem. B 111 (13), 3447–3461. http://dx.doi.org/10.1021/jp0682208.

Lee, Y.S., Graham, E.J., Galindo, A., Jackson, G., Adjiman, C.S., 2020. A comparative study of multi-objective optimization methodologies for molecular and process design. Comput. Chem. Eng. 136, 106802. http://dx.doi.org/10.1016/j.compchemeng.2020.106802.

Lemmon, E.W., Bell, I., Huber, M.L., McLinden, M.O., 2018. NIST standard reference database 23: Reference fluid thermodynamic and transport properties-REFPROP, version 10.0, national institute of standards and technology. http://dx.doi.org/10.18434/T4JS3C.

Li, X.S., Englezos, P., 2003. Vapor-liquid equilibrium of systems containing alcohols using the statistical associating fluid theory equation of state. Ind. Eng. Chem. Res. 42 (20), 4953–4961. http://dx.doi.org/10.1021/ie030256o.

Li, X.S., Englezos, P., 2004. Vapor–liquid equilibrium of systems containing alcohols, water, carbon dioxide and hydrocarbons using SAFT. Fluid Phase Equilib. 224 (1), 111–118. http://dx.doi.org/10.1016/j.fluid.2004.06.052.

Liang, X., Tsivintzelis, I., Kontogeorgis, G.M., 2014. Modeling water containing systems with the simplified PC-SAFT and CPA equations of state. Ind. Eng. Chem. Res. 53 (37), 14493–14507. http://dx.doi.org/10.1021/ie501993y.

Lin, J.G., 1976. Three methods for determining Pareto-optimal solutions of multiple-objective problems. In: Directions in Large-Scale Systems. Springer US, pp. 117–138. http://dx.doi.org/10.1007/978-1-4684-2259-7_9.

Luck, W.A.P., 1980. A model of hydrogen-bonded liquids. Angew. Chem. Int. Edition English 19 (1), 28–41. http://dx.doi.org/10.1002/anie.198000281.

Marler, R., Arora, J., 2004. Survey of multi-objective optimization methods for engineering. Struct. Multidiscip. Optim. 26 (6), 369–395. http://dx.doi.org/10.1007/s00158-003-0368-6.

Marler, R.T., Arora, J.S., 2009. The weighted sum method for multi-objective optimization: New insights. Struct. Multidiscip. Optim. 41 (6), 853–862. http://dx.doi.org/10.1007/s00158-009-0460-7.

Morokoff, W.J., Caflisch, R.E., 1994. Quasi-random sequences and their discrepancies. SIAM J. Sci. Comput. 15 (6), 1251–1279. http://dx.doi.org/10.1137/0915077.

Müller, E.A., Gubbins, K.E., 1995. An equation of state for water from a simplified intermolecular potential. Ind. Eng. Chem. Res. 34 (10), 3662–3673.

Müller, E.A., Rull, L.F., Vega, L.F., Gubbins, K.E., 1996. Adsorption of water on activated carbons: A molecular simulation study. J. Phys. Chem. 100 (4), 1189–1196.

Newville, M., Stensitzki, T., Allen, D.B., Rawlik, M., Ingargiola, A., Nelson, A., 2016. LMFIT: Non-linear least-square minimization and curve-fitting for Python. Astrophys. Source Code Libr..

Oliveira, M.B., Llovell, F., Coutinho, J.A.P., Vega, L.F., 2016. New procedure for enhancing the transferability of statistical associating fluid theory (SAFT) molecular parameters: The role of derivative properties. Ind. Eng. Chem. Res. 55 (37), 10011–10024. http://dx.doi.org/10.1021/acs.iecr.6b02205.

Papaioannou, V., Lafitte, T., Avendaño, C., Adjiman, C.S., Jackson, G., Müller, E.A., Galindo, A., 2014. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments. J. Chem. Phys. 140 (5), 054107. http://dx.doi.org/10.1063/1.4851455.

Pascoletti, A., Serafini, P., 1984. Scalarizing vector optimization problems. J. Optim. Theory Appl. 42 (4), 499–524. http://dx.doi.org/10.1007/bf00934564.

Patel, B.H., Paricaud, P., Galindo, A., Maitland, G.C., 2003. Prediction of the salting-out effect of strong electrolytes on water + alkane solutions. Ind. Eng. Chem. Res. 42 (16), 3809–3823. http://dx.doi.org/10.1021/ie020918u.

Poling, B.E., Prausnitz, J.M., O'connell, J.P., et al., 2001. The Properties of Gases and Liquids, Vol. 3. McGraw-Hill New York.

Polishuk, I., 2010. About the numerical pitfalls characteristic for SAFT EOS models. Fluid Phase Equilib. 298 (1), 67–74. http://dx.doi.org/10.1016/j.fluid.2010.07.003.

Privat, R., Gani, R., Jaubert, J.N., 2010. Are safe results obtained when the PC-SAFT equation of state is applied to ordinary pure chemicals? Fluid Phase Equilib. 295 (1), 76–92. http://dx.doi.org/10.1016/j.fluid.2010.03.041.

Ramrattan, N., Avendaño, C., Müller, E., Galindo, A., 2015. A corresponding-states framework for the description of the Mie family of intermolecular potentials. Mol. Phys. 113 (9–10), 932–947. http://dx.doi.org/10.1080/00268976.2015.1025112.

Rangaiah, G.P., 2009. Multi-Objective Optimization: Techniques and Applications in Chemical Engineering, Vol. 1. World Scientific.

Rehner, P., Gross, J., 2018. Predictive density gradient theory based on nonlocal density functional theory. Phys. Rev. E 98 (6), http://dx.doi.org/10.1103/physreve.98.063312.

Rehner, P., Gross, J., 2020. Multiobjective optimization of PCP-SAFT parameters for water and alcohols using surface tension data. J. Chem. Eng. Data 65 (12), 5698–5707. http://dx.doi.org/10.1021/acs.jced.0c00684.

Sear, R.P., Jackson, G., 1996. Thermodynamic perturbation theory for association with bond cooperativity. J. Chem. Phys. 105 (3), 1113–1120. http://dx.doi.org/10.1063/1.471955.

Sheldon, T., Giner, B., Adjiman, C., Galindo, A., Jackson, G., Jacquemin, D., Wathelet, V., Perpete, E., 2006. The derivation of size parameters for the SAFT–VR equation of state from quantum mechanical calculations. In: Computer Aided Chemical Engineering, vol. 22, Elsevier, pp. 143–159.

Sobol', I., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Comput. Math. Math. Phys. 7 (4), 86–112. http://dx.doi.org/10.1016/0041-5553(67)90144-9.

Stöbener, K., Klein, P., Horsch, M., Küfer, K., Hasse, H., 2016. Parametrization of two-center Lennard-Jones plus point-quadrupole force field models by multicriteria optimization. Fluid Phase Equilib. 411, 33–42. http://dx.doi.org/10.1016/j.fluid.2015.11.028.

Stöbener, K., Klein, P., Reiser, S., Horsch, M., Küfer, K.H., Hasse, H., 2014. Multicriteria optimization of molecular force fields by Pareto approach. Fluid Phase Equilib. 373, 100–108. http://dx.doi.org/10.1016/j.fluid.2014.04.009.

Tsivintzelis, I., Bøgh, D., Karakatsani, E., Kontogeorgis, G.M., 2014. The role of monomer fraction data in association theories—can we improve the performance for phase equilibrium calculations? Fluid Phase Equilib. 365, 112–122. http://dx.doi.org/10.1016/j.fluid.2013.12.013.

von Solms, N., Michelsen, M.L., Passos, C.P., Derawi, S.O., Kontogeorgis, G.M., 2006. Investigating models for associating fluids using spectroscopy. Ind. Eng. Chem. Res. 45 (15), 5368–5374. http://dx.doi.org/10.1021/ie051341u.

Voutsas, E.C., Boulougouris, G.C., Economou, I.G., Tassios, D.P., 2000. Water/hydrocarbon phase equilibria using the thermodynamic perturbation theory. Ind. Eng. Chem. Res. 39 (3), 797–804. http://dx.doi.org/10.1021/ie990559b.

Wang, Y., Biegler, L.T., Patel, M., Wassick, J., 2018. Parameters estimation and model discrimination for solid-liquid reactions in batch processes. Chem. Eng. Sci. 187, 455–469. http://dx.doi.org/10.1016/j.ces.2018.05.040.

Wolbach, J.P., Sandler, S.I., 1997. Using molecular orbital calculations to describe the phase behavior of hydrogen-bonding fluids†. Ind. Eng. Chem. Res. 36 (10), 4041–4051. http://dx.doi.org/10.1021/ie9607255.

Wu, J., Prausnitz, J.M., 1998. Phase equilibria for systems containing hydrocarbons, water, and salt: An extended Peng-Robinson equation of state. Ind. Eng. Chem. Res. 37 (5), 1634–1643. http://dx.doi.org/10.1021/ie9706370.

Zadeh, L., 1963. Optimality and non-scalar-valued performance criteria. IEEE Trans. Automat. Control 8 (1), 59–60. http://dx.doi.org/10.1109/tac.1963.1105511.

Zollweg, J.A., Mulholland, G.W., 1972. On the law of the rectilinear diameter. J. Chem. Phys. 57 (3), 1021–1025. http://dx.doi.org/10.1063/1.1678352.