Multicriteria optimization of molecular force fields by Pareto approach

K. Stöbener [a], P. Klein [a], S. Reiser [b], M. Horsch [b,*], K.-H. Küfer [a], H. Hasse [b]

[a] *Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany*
[b] *Laboratory of Engineering Thermodynamics, University of Kaiserslautern, Germany*

ARTICLE INFO

ABSTRACT

Molecular force fields are widely used for simulating thermodynamic properties of fluids. In developing such force fields, usually some of their parameters are adjusted to experimental data sets, which are often of different type. The adjustment is commonly carried out by minimizing a single objective function which represents the deviations between the model and the data. In the present work, a different approach is explored. Individual objective functions are defined for each data set and a multicriteria optimization task is solved. It is explicitly acknowledged that the different objectives are usually conflicting. The multicriteria optimization problem is solved by determining the Pareto set. By definition this set includes all solutions for which no further improvement in one objective can be achieved without having to accept a decline in at least one other objective and, hence, contains best compromises. The user can then choose out of these solutions one which is particularly suited for his application. The procedure is illustrated using the parameterization of the Lennard-Jones model for argon and methane as examples. Six different objective functions are included in the optimization. They represent the deviations between the model and the following properties at boiling conditions over a wide temperature range: (a) liquid density, (b) vapor pressure, (c) enthalpy of vaporization, (d) liquid shear viscosity, (e) liquid thermal conductivity, and (f) surface tension. First single objective fits are carried out for all properties. Then Pareto sets are determined for two triples of objectives namely, (a, b, c) on one side and (d, e, f) on the other side. An unexpected topology of the Pareto set is observed and explained. Then the full Pareto set for all six properties is determined and all results are compared. They show that good results can be achieved with the simple Lennard-Jones model for the two studied fluids, even when the goal is to simultaneously describe many different thermodynamic properties. The work also illustrates the benefits of using Pareto optimization for developing force fields, and, more generally thermodynamic models.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Molecular simulations with atomistic force fields are widely used for solving problems in physics, biology, chemistry, and engineering. They rely on the availability of suitable force fields. Usually, the functional form of the force field is known based on physical grounds, but the model parameters still need to be determined. This is often done using a combination of setting parameters based on quantum chemical data and adjusting the remaining parameters to experimental data [1]. The resulting optimization problem is usually solved by minimizing a single objective function, which contains the information on the deviations between the simulation results and the experimental data. Different solver strategies can be employed to find the minimum. Faller et al. [2] and Reith et al. [3] each presented an automatic scheme based on simplex algorithms. Wang and Kollman [4] introduced an automatic engine based on systematic search as well as a genetic algorithm. Bourasseau et al. [5] used a Gauss-Legendre least-squares estimator to find the minimum of the objective function. Hülsmann et al. [6,7] compared the performance of several algorithms. Deublein et al. [8] also present an automated method for the development of force fields based on a gradient method.

By finding the minimum of the objective function one specific parameter set for the molecular model is identified, which is optimal for the chosen objective function.

The multicriteria optimization approach used in this work considers several objective functions. These can e.g. stem from using experimental data sets of different type (e.g. densities or vapor pressures) or at different conditions (e.g. liquid or gas phase). In general, the different objective functions are conflicting, i.e. they

* Corresponding author. Tel.: +49 631 205 3227; fax: +49 631 205 3835.
E-mail address: martin.horsch@mv.uni-kl.de (M. Horsch).

N	number of temperatures
O	thermodynamic property
k	Boltzmann's constant
f	objective function
p	dimension objective space
p_C	critical pressure
p^S	saturated vapor pressure
q	dimension design space
u	Lennard-Jones potential
x	decision vector
T	temperature
T_C	critical temperature
T_{TP}	triple point temperature
Δh_V	enthalpy of vaporization
δ	mean relative deviation
ϵ	Lennard-Jones energy parameter
γ	surface tension
λ'	liquid thermal conductivity
η'	liquid shear viscosity
ρ'	saturated liquid density
ρ_C	critical liquid denisty
σ	Lennard-Jones size parameter

cannot be minimized simultaneously. In the present work the Pareto approach is used for solving this multicriteria optimization problem. It relies on identifying the set of Pareto optimal solutions (Pareto set). The Pareto set represents those solutions, for which one objective function can only be improved by having to accept a decline in at least one other objective function. Hence, Pareto sets represent best compromises. Once the Pareto set is determined the user gets an overview of what can be achieved with a certain model. Based on that knowledge he can then choose from the Pareto set the model he considers to be most attractive for his application.

Pareto optimization was previously employed in the context of molecular models. Mostaghim et al. [9] describe an optimization of bond terms for primary alcohols in which three objective functions based on the reproduction of ab initio information were designed and five different force field parameters considered. To solve the problem multi-objective evolutionary algorithms and particularly multi-objective Particle Swarm Optimization were applied to determine a Pareto set.

This work is, to the best of our knowledge, the first to use Pareto optimization for determining parameters of intermolecular interaction potentials. By a brute force enumeration, we systematically study the application of Pareto-optimization for developing atomistic force fields using the one center Lennard-Jones 6–12 potential as an example. The performance of that simple model for representing thermo-physical properties of two fluids, argon and methane, is investigated. For each of these fluids six objective functions are defined which represent the deviation of the model and the experimental data sets for the following properties: liquid density, vapor pressure, enthalpy of vaporization, liquid shear viscosity, liquid thermal conductivity, and surface tension.

In a first step two three dimensional objective spaces, each considering three properties are explored. An interesting and unexpected topology is observed and explained. To complete the analysis, the full six dimensional problem is addressed. The results illustrate the benefits from using Pareto optimization for parameterizing atomistic force fields. They also show that the Lennard-Jones model performs astonishingly well for describing fluid properties of simple substances.

2. Multicriteria optimization

A multicriteria optimization problem is characterized by several objective functions $f_i(x)$ which have to be minimized simultaneously:

$$\min f(x) = (f_1(x), \ldots, f_p(x)) \in \mathbb{R}^p \tag{1}$$

They span the objective space $\mathbb{R}^p$ and depend on the decision vector $x \in \mathbb{R}^q$ where $\mathbb{R}^q$ is the design space. The solution to such a problem is a set of best compromises: For any improvement in a single objective $f_i(x)$, a decline in at least one other objective $f_j(x)$, $i \neq j$ has to be accepted. The set of all best compromises in the objective space is called the Pareto frontier. The corresponding solutions in the design space are called the Pareto set. Once they are identified, a trade-off discussion of the individual objectives is possible. Based on this knowledge a decision vector which is particularly suitable for the studied application can be chosen.

The Pareto frontier is a subset of all feasible points in the objective space and needs to be approximated by a suitable numerical strategy. The most obvious approach for the construction of the Pareto frontier starts from a construction of the set of all feasible points in the objective space and a subsequent identification of the Pareto frontier by brute force comparison of the different objectives. This strategy is employed in the present work. However for high dimensions of the design space and for computationally intensive calculations of the objective functions, a full sampling of all feasible solutions is not possible. Then multicriteria optimization algorithms, available in the literature, have to be applied to identify the Pareto set (see e.g. [10,11]). For more information on multicriteria optimization (see e.g. [12–14]).

In the context of force field parameterization the design space is spanned by the parameters describing the model. Thus we refer to it as parameter space in the subsequent text. A point in the parameter space, hence, corresponds to a certain force field parameter set and would commonly be called a model of the substance. Mapped to each point in the parameter space is one point in the objective space. The objective functions contain the information on the quality of the force field model, e.g. as represented by the mean relative deviations between the simulation results and the chosen experimental data sets. Pareto optimization crucially depends on a suitable choice of the objective function. Different approaches are possible and can individually be adapted to the thermodynamic properties, whose quality needs to be assessed.

Once the Pareto frontier is identified, full insight over the achievable model performance is gained. It is then possible to navigate on the Pareto frontier, i.e. to carry out the trade-off discussion of the individual objectives, and to finally choose the most attractive of all Pareto optimal solutions.

3. Studied systems

In this work the parameterization of the Lennard-Jones (LJ) potential

$$u(r) = -4\epsilon \left[\left(\frac{\sigma}{r} \right)^6 - \left(\frac{\sigma}{r} \right)^{12} \right] \tag{2}$$

for argon and methane is studied as an example. The studied properties are all related to vapor–liquid equilibria: saturated liquid density, saturated vapor pressure, enthalpy of vaporization, saturated liquid shear viscosity, saturated liquid thermal conductivity, and surface tension. The two fluids are studied for technical reasons: A large variety of data and many literature models are available to compare the findings of this work to.

Table 1
Transformations reduced units to physical units.

Temperature	$T = \frac{\epsilon}{k} T^*$
Liquid density	$\rho' = \sigma^3 \rho^*$
Vapor pressure	$p^S = \frac{\epsilon}{\sigma^3} P^*$
Enthalpy of vaporization	$\Delta h_V = \epsilon \Delta h^*$
Liquid thermal conductivity	$\lambda' = \frac{k}{\sigma^2} \sqrt{\frac{\epsilon}{m}} \lambda^*$
Liquid shear viscosity	$\eta' = \frac{\sqrt{m\epsilon}}{\sigma^2} \eta^*$
Surface tension	$\gamma = \frac{\epsilon}{\sigma^2} \gamma^*$

3.1. Data basis

For this work no experiments or simulations were carried out. All the data sets are taken from the literature.

For the experimental data temperature dependent correlations from the DIPPR data bank [15] are used. The errors reported for both argon and methane coincide: for the liquid density, vapor pressure and enthalpy of vaporization they are <1%, for the liquid viscosity and surface tension <3% and for the liquid thermal conductivity <5% [15].

The simulation data are collected from several publications. The information for the six thermo-physical properties is gathered in reduced units along the vapor–liquid coexistence curve. Correlation functions for the saturated liquid density, saturated vapor pressure, and the enthalpy of vaporization are taken from Lotfi et al. [16]. The statistical errors for the simulation data are reported as <1%, <5% and <8%, respectively, the deviations of the correlation functions corresponding to the simulation data are reported to be <1 %, <1% and <14% respectively. The data set for the correlation of the liquid shear viscosity is taken from Rowley and Painter [17]. The average absolute deviation between the simulation data and the correlation is reported to be <4.5%. The liquid thermal conductivity is taken from Fernández et al. [18]. They report only data for specific substances, which, however, can be used to retrieve the information on the liquid thermal conductivity in reduced units. On that basis a correlation was established. The root mean square error (RMSE) of that correlation and the simulation data is <20 % and large compared to the other properties. The surface tension is from Grosfils and Lutsko [19]. The statistical error is <6%. Again the data set is correlated, and the RMSE is <1%. More information on the correlation functions for the liquid thermal conductivity and the surface tension and their quality is given in the Supplementary Material.

3.2. Methodology

To evaluate the quality of a specific LJ model the correlations of the simulation data in reduced units are transferred to physical units with the corresponding values for the energy parameter ϵ and the length parameter σ. The transformations are given in Table 1. For each model a mean relative deviation function δO for each property O over different temperatures T_j, $j = 1, \ldots, N$ is evaluated. The objective space is spanned by the deviation functions for the six thermophysical properties. In this work, the objective function f_i for one property O has the following form:

$$f_i = \delta O = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left(\frac{O^{exp}(T_j) - O^{sim}(T_j)}{O^{exp}(T_j)} \right)^2} \qquad (3)$$

In this work 13 equidistant reduced temperatures between $T^* = 0.7$ and $T^* = 1.2$ are considered. For a given energy parameter ϵ, it may not be possible to evaluate O^{exp} for all temperatures corresponding to these reduced temperatures. For high T^* and ϵ the temperature $T = (T^* \epsilon / k)$ may be above the critical temperature T_C ($T_C^{Ar} = 150.86$ K, $T_C^{Me} = 190.69$ K [15]), for low $T = (T^* \epsilon / k)$

the temperature may be below the temperature at the triple point T_{TP} ($T_{TP}^{Ar} = 83.78$ K, $T_{TP}^{Me} = 90.69$ K [15]). The corresponding reduced temperatures T^* were not considered.

Each solution in objective space holds for a specific force field model, represented by a point in the parameter space which is spanned by the energy parameter ϵ and the length parameter σ.

3.3. Numerics

The parameter space is explored in a restricted region. The boundaries σ_{min} / σ_{max} and ϵ_{min} / ϵ_{max} are choosen such that, depending on the scenario, the whole Pareto set is found. The sampled grid is 200×200 points.

Because correlations of the simulation data sets in reduced units are available, all relative mean deviations $\delta O(\epsilon, \sigma)$ for any given model can be calculated straightforwardly. Hence a complete sampling of the design space spanned by the two model parameters is possible. Thus, the LJ models leading to Pareto optimal solutions can be identified simply by comparing points in the objective space to each other. Reduced units were previously utilized by Merker et al. [20] to optimize force field models.

For the employed algorithm, every point in the objective space is tagged as 'Pareto optimal'. Then the solutions are sorted in descending order for the first objective function. Starting by the first point, every point is compared to the successive solutions, to find one, for which the remaining objective functions can all be improved. If the search was successful, it is aborted and the considered point is tagged as 'not Pareto optimal'. If the search was not successful, the point is Pareto optimal. Then the next point is considered and the loop continues until all points are checked for Pareto optimality. Therefore, the Pareto set is easily isolated.

3.4. Scenarios

The thermo-physical properties are first considered individually. For each property, a single objective optimization is carried out, yielding an optimal parameter set. Furthermore the topology of the relative mean deviations as a function of the model parameters is explored.

In a next step the six properties are split into two groups. The first group contains three static properties: saturated liquid density, saturated vapor pressure, and enthalpy of vaporization. The second group consists of the liquid thermal conductivity, shear viscosity, and surface tension. The liquid thermal conductivity and the shear viscosity additionally depend on the mass, which for the present analysis is not considered as a free parameter ($m_{Ar} = 39.948$ u, $m_{CH_4} = 16.04246$ u [15]). For argon and methane, the two triples of properties are analyzed and the corresponding objective spaces are explored. Finally, the full six dimensional objective space is investigated. The parameter sets leading to Pareto optimal solutions for those three tasks are identified and presented in the next section.

4. Results and discussion

For the different scenarios only the results of argon are given in the figures. For methane central results are discussed but the respective figures are given in the Supplementary Material.

4.1. Single objectives

Fig. 1 shows the dependence of the mean relative deviation from the LJ model parameters ϵ and σ for each of the six studied properties for argon. In the contour plots the color indicates the value of the mean relative deviations. As the parameters ϵ and σ
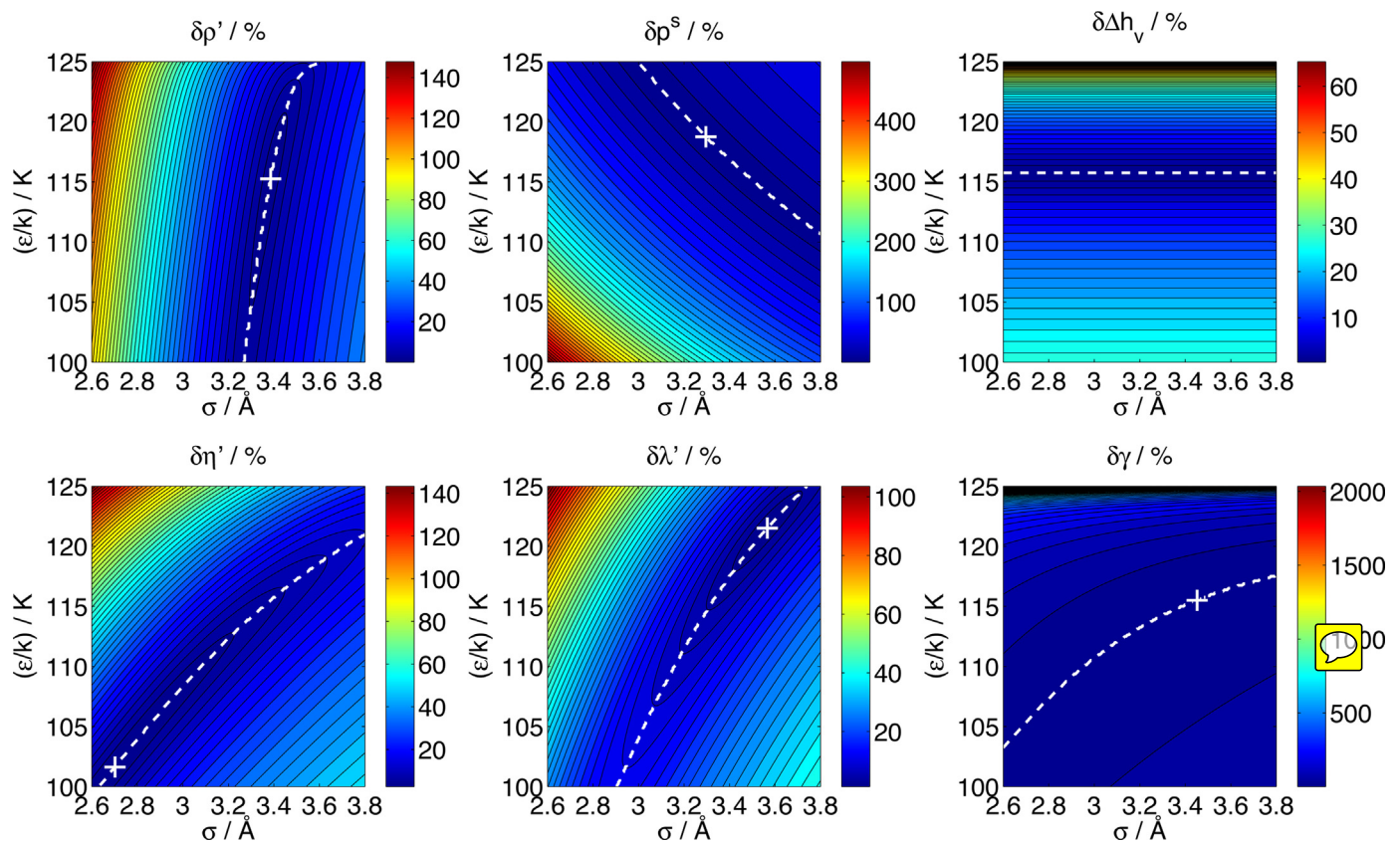
Fig. 1. Dependence of the mean relative deviations for six thermo-physical properties of argon on the Lennard-Jones parameters σ and ϵ/k (k is Boltzmann's constant): liquid density ρ', vapor pressure p^S, enthalpy of vaporization Δh_V, liquid viscosity η', liquid thermal conductivity λ' and surface tension γ.

are varied, the mean relative deviation for each property forms a valley around the white dashed line. It indicates the path of lowest ascent/descent through the valley. Additionally the parameter combinations obtained by a single criterion optimization of the corresponding properties are indicated in the diagrams: This is a single point marked by a white cross for the liquid density, vapor pressure, enthalpy of vaporization, liquid viscosity, liquid thermal conductivity and surface tension. The objective function for the enthalpy of vaporization however is independent of the variation of σ. Hence, all the parameter combinations along the white dashed line lead to an optimal mean relative deviation.

The range in which the parameters are varied (about 25% in ϵ and 40% in σ) is the same for all studied properties. It is also interesting to compare the results for the maximal deviations of the six different properties shown in Fig. 1. They can be used as indicators of the sensitivity of the different properties on the model parameters and turn out to be very different. The lowest number is found for the enthalpy of vaporization (about 20%), the highest for the vapor pressure (400%) and the surface tension (500%). The results for the liquid density (150%), liquid viscosity (150%) and liquid thermal conductivity (100%) lie in between.

Looking at every property separately, the simulations with the LJ potential can reproduce the experimental results within the experimental and statistical uncertainties. The mean relative deviations achievable for argon are for the saturated liquid density < 0.25%, for the saturated vapor pressure < 0.5%, for the heat of vaporization < 1%, for the liquid viscosity < 3%, for the liquid heat conductivity <2% and for the surface tension < 1%. For methane the corresponding mean relative deviations are <0.5%, <1%, <2%, <2%, <3% and <2% respectively. But to get acceptable deviations in all properties trade-offs have to be taken into account.

The results obtained from a single criterion optimization of the studied properties (white crosses for liquid density, vapor pressure, liquid viscosity, liquid thermal conductivity, and surface tension, white dashed line for the enthalpy of vaporization) lie in the upper right quadrant of the studied parameter range for the liquid density, vapor pressure, liquid thermal conductivity, and surface tension, but in the lower right quadrant for the liquid viscosity. The distance between these optimal solutions in the parameter space alone is, however, not a good indicator for the difficulty of finding good compromises. It merely underlines the necessity of multi-criteria optimization.

A comparison of the results for argon (Fig. 1) and methane (Supplementary Materials), shows strong similarities. They include the course of the valleys in the σ, ϵ plane and the ranking of the sensitivity of the results for the different properties.

4.2. Three objectives

The results for the two three dimensional Pareto optimization scenarios for argon are presented in Figs. 2 and 3.

Fig. 2 shows the results obtained for the Pareto set of argon with three objective functions: liquid density, vapor pressure and enthalpy of vaporization. On the left hand side the parameter space is shown, on the right hand side the objective space. The mapping between both spaces is indicated by colors. The upper row (a) shows results from a brute force variation of the parameters within their bounds and the corresponding results in the objective space. The Pareto optimal points found by enumeration of these results are marked black. For clarity the lower row (b) shows only the isolated Pareto set in the parameter space and the corresponding Pareto frontier in the objective space.
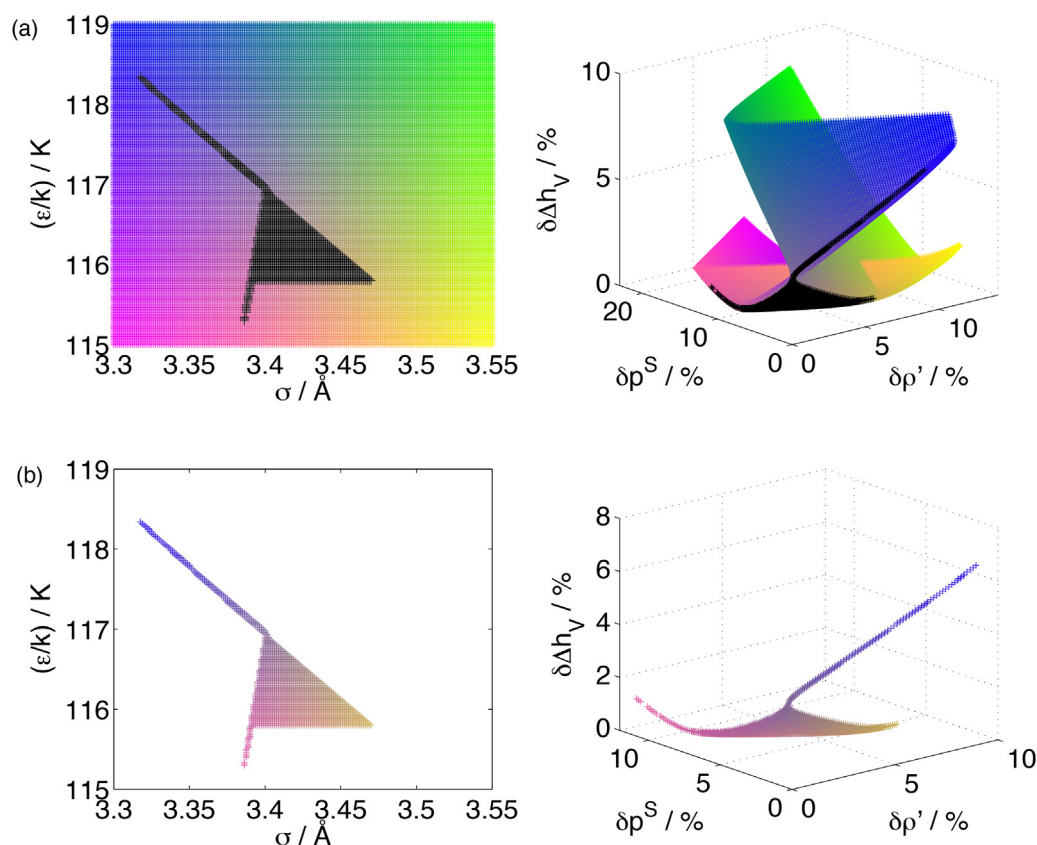
Fig. 2. Results for Pareto optimization of argon with three objective functions: mean relative deviation of liquid density $\delta\rho'$, vapor pressure δp^S and enthalpy of vaporization $\delta\Delta h_V$. Left side: parameter space, right side: objective space. Mapping between both spaces indicated by colours. (a) Brute force variation of the parameters and corresponding results. The Pareto set is marked black. (b) Isolated Pareto set.

In the objective space, an interesting and unexpected topology is observed. It self-intersects, meaning, that for two different parameter sets the same mean relative deviations in all three properties may be found. Three lines of self-intersections are observed. Only part of the surface in the objective space, found by brute force enumeration of the parameters, is Pareto optimal. Its main structure is triangular, with tails stretching out from the corners of the triangle both in the parameter and objective space. The corresponding triangular Pareto set in the parameter space has almost linear edges and tails.

The results obtained for the Pareto frontier show that compromises between all three objectives can be found such that in all objectives the observed mean relative deviations are of the order of only a few percent and, hence, not significantly above the cumulated errors of the experiment and simulation. Within the triangular structure, reasonable compromises between all three objectives are obtained. The tails however basically correspond to compromises between only two objectives whereas for the third objective high deviations are accepted. The triangular Pareto set in the parameter space is narrow. Its width is of the order of 1 K in ϵ/k and 0.05 Å in σ.

For methane the Pareto optimal points show the same structure in both the objective and the parameter space. In the objective space the mean relative deviations for methane in the triangle are also of the order of a few percent. Thus compromises, with mean relative deviations below the cumulated errors can be found. The triangular Pareto set is again narrow. Its width is of the order of 2 K in ϵ/k and 0.1 Å in σ.

Fig. 3 shows the results obtained for the Pareto set of argon for the second scenario: the evaluation of the three dimensional optimization task concerning the liquid viscosity, liquid thermal conductivity and surface tension.

In the objective space again a topology with three, barely distinguishable, lines of self-intersections is observed. The objective function for the surface tension $\delta\gamma$ is extremely sensible on a variation of the parameters, so that the mean relative deviation for the surface tension reaches up to 2000%. Evaluated in the same parameter range, the objective functions for the liquid viscosity and the liquid thermal conductivity are of the order of a few hundred percent.

The Pareto optimal points in both the objective and parameter space again form a triangular region. Three tails, each of them spreading from one corner of this triangular structure are observed. In the parameter space the tails reach far, thus the range in which the parameter space is sampled to capture the whole Pareto set is very large (about 25% in ϵ and 40% in σ). But the width of the triangular Pareto set in the parameter space is of the order of 2 K in ϵ/k and 0.05 Å in σ. The compromises that are found for these models lead to mean relative deviations below 10% in the three considered properties liquid viscosity, liquid thermal conductivity and surface tension.

A comparison of these findings to the results obtained for the first optimization task for argon (Fig. 2) shows, that the static properties are represented with a higher accuracy than the two transport properties and the surface tension. This is consistent with the uncertainties given for the experimental and statistical errors for the six properties. For the second scenario (Fig. 3) due to the long tails, a wider variation of the parameters ϵ and σ is needed to find the full Pareto set. The triangles in the parameter space for the two tasks, however, are of similar size and lie in close proximity.
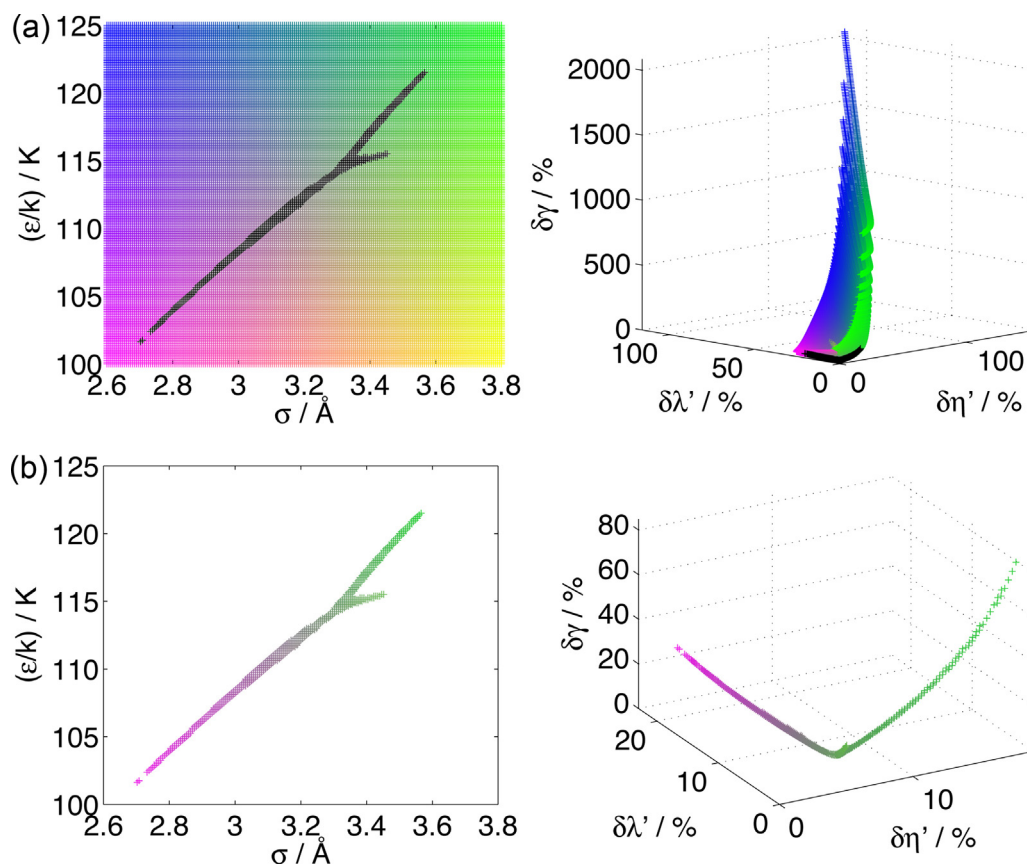
Fig. 3. Results for Pareto optimization of argon with three objective functions: mean relative deviation of liquid viscosity $\delta\eta'$, liquid thermal conductivity $\delta\lambda'$ and surface tension $\delta\gamma$. Left side: parameter space, right side: objective space. Mapping between both spaces indicated by colours. (a) Brute force variation of the parameters and corresponding results. The Pareto set is marked black. (b) Isolated Pareto set.

The results obtained for methane are again very similar to the findings discussed above for argon: The mean relative deviation for the surface tension is very sensible to variations of the parameters, whereas the mean relative deviations in the liquid viscosity and the liquid thermal conductivity are about a few hundred percent. The width of the triangular region in the parameter space is of the order of 30 K in ϵ/k and 2 Å in σ. It is considerably larger than the triangles found before.

The results for all three dimensional optimization scenarios studied here show a topology in the objective space which self intersects and leads to a triangular Pareto set. This topology is explained in more detail below (Section 4.4).

4.3. Six objectives

Fig. 4 shows part of the Pareto optimal points in the objective space for argon for six objective functions: liquid density, vapor pressure, enthalpy of vaporization, liquid viscosity, liquid thermal conductivity and surface tension. To view all six objectives simultaneously, they are plotted against a model ID. To distinguish the different objectives, different markers are used. For clarity, not the whole Pareto optimal points are shown, but only those which yield mean relative deviations below 5% for the liquid density, the vapor pressure and the enthalpy of vaporization. The numerical values for the objective functions and the parameter sets corresponding to the model IDs are given in the Supplementary Material.

It is possible to keep the mean relative deviations for the saturated liquid density, for the saturated vapor pressure and for the enthalpy of vaporization below 5% and for the liquid viscosity, liquid heat conductivity and surface tension below 10% (e.g. model ID 1–10). If however a higher mean relative deviation in the surface tension and enthalpy of vaporization are accepted (higher model IDs), lower mean relative deviations in the liquid density, vapor pressure and liquid thermal conductivity can be achieved. Taking the uncertainties of the data sets into account, a good compromise is e.g. model number 29 ($\epsilon/k = 116.6$ K, $\sigma = 3.4$ Å), with mean relative deviations of 0.6% for the liquid density, 1.7% for the vapor pressure, 2.3% for the enthalpy of vaporization, 8.4% for the liquid viscosity, 4.3% for the liquid heat conductivity, and 11.4% for the surface tension. Only the mean relative deviation of the surface tension does not lie in the range of the experimental or statistical errors (cf. Section 3.1).

For methane excellent results can be achieved simultaneously for the liquid density, vapor pressure, enthalpy of vaporization, liquid thermal conductivity and surface tension. Compromises for which the mean relative deviation for all five properties is less than 5% are found. Only for the liquid viscosity error sets rally above the cumulated uncertainties of the experiments and simulations observed (about 20–25%). The order of magnitude of that deviation does not change upon changing the model. This might indicate errors in the experimental or simulation data, which could also be systematic. But it could of course also be related to problems of the LJ model in representing that property of methane. Disregarding the performance of the model for the liquid viscosity, a good comprise is e.g. $\epsilon/k = 149$ K, $\sigma = 3.73$ Å, with mean relative deviations of 0.4% for the liquid density, 1.5% for the vapor pressure, 2.1% for the enthalpy of vaporization, 1.6% for the liquid heat conductivity, 26.5% for the liquid viscosity, and 3.2% for the surface tension.

Fig. 5 shows the models in the parameter space which form the Pareto set of argon for six objective functions: liquid density, vapor
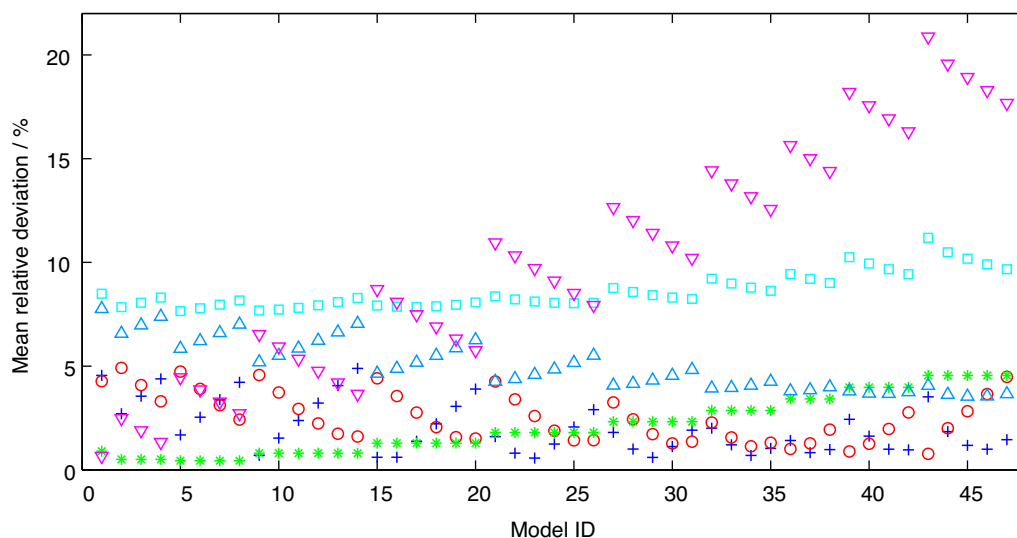
Fig. 4. Pareto optimal points of argon in objective space for six objective functions plotted against model ID: liquid density (+), vapor pressure (○), enthalpy of vaporization (*), liquid viscosity (□), liquid thermal conductivity (△) and surface tension (▽). Only models with relative mean deviations below 5% for the liquid density, vapor pressure and enthalpy of vaporization. For numerical data see Supplementary Material.

pressure, enthalpy of vaporization, liquid viscosity, liquid thermal conductivity and surface tension. The parameter sets are identified in a brute force variation of the bounded region in the parameter space shown on the left hand side. On the right hand side a zoom into the parameter space on the left hand side is presented. Additionally, models found in the literature are shown.

The resulting geometrical structure in the parameter space is an expansion of the combination of the two three dimensional optimization scenarios for argon: Three tails form. The shortest is similar to that observed in the parameter space in the optimization scenario considering the liquid density, vapor pressure and enthalpy of vaporization in Fig. 2. The two longer tails are similar to those observed in the parameter space for the second optimization task of argon considering the liquid viscosity, liquid thermal conductivity and surface tension in Fig. 3. The core of the structure, of which the tails stretch out, is shaped like a triangle, whose main part is similar to that in the first optimization scenario of argon. It is however noticeable larger and the tails are wider.

There is a large variety of LJ parameter sets for argon published in the literature. Only some are compared to the results of the Pareto optimization of the present work in the following. The model

of Bembenek and Rice [21] is derived by using time dependent parameters ϵ and σ to steer a non-equilibrium molecular dynamics simulation to a steady state with a desired bulk density. An almost identical model is found by Shukla [22], who fits simulation data for the saturated liquid density and the vapor pressure to the corresponding properties derived by the equation of state. Both models lie in close proximity to the main structure. Vrabec and Fischer [23] fit their model to best reproduce saturated volumes and enthalpies of vaporization. Fischer et al. [24] and Vrabec et al. [25] both fit simulation data of the saturated liquid density and the vapor pressure to experimental data. These three models lie in the region of models corresponding to Pareto optimal mean relative deviations.

Fischer et al. [24], Shukla [22], Vrabec and Fischer [23] and Vrabec et al. [25] not only optimized the LJ model for argon, but also published an optimized LJ model for methane. The parameter sets from Vrabec and Fischer [23] and Vrabec et al. [25] again lie within the region of models corresponding to Pareto optimal mean relative deviations, whereas the model from Fischer et al. [24] lies close to it. The model of Shukla [22] lies somewhat further off. The OPLS-UA model from Jorgensen et al. [26] was optimized to best reproduce the energy and density of liquid methane and
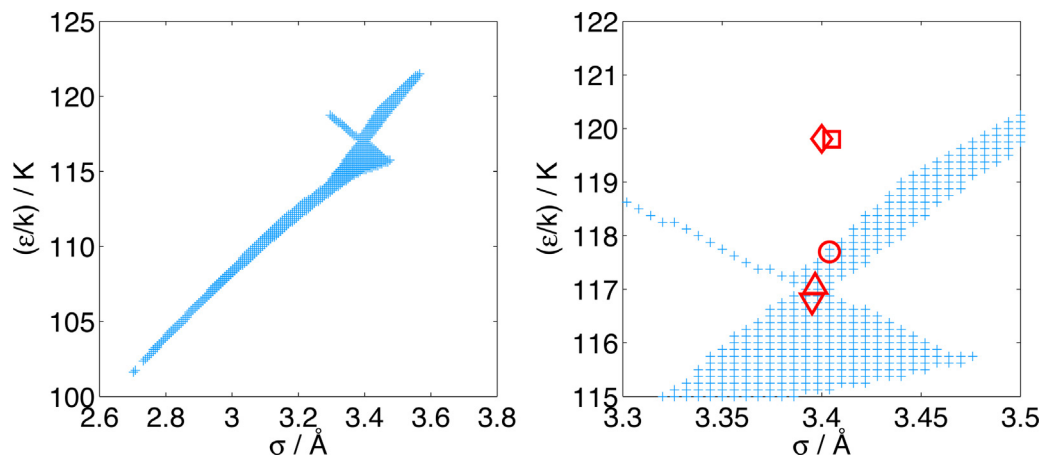

Fig. 5. Pareto set in parameter space for six objective functions of argon: liquid density, vapor pressure, enthalpy of vaporization, liquid viscosity, liquid thermal conductivity and surface tension. Left side: Full investigated range. Right side: Zoom plus models found in literature: Fischer et al. (○), Shukla (□), Bembenek and Rice (◇), Vrabec and Fischer (△), Vrabec et al. (▽).

the TraPPE model from Martin and Siepmann [27] was adjusted to fit the critical temperature and the saturated liquid density of methane. Both models are also within the region of models corresponding to Pareto optimal mean relative deviations.

The model parameters can also be compared to values derived from the principle of corresponding states. Lotfi et al. [16] obtained for the reduced temperature at the critical point $T_C^* = kT_C/\epsilon = 1.31$ and for the critical pressure $p_C^* = p_C\sigma^3/\epsilon = 0.142$. The values for the critical temperature ($T_C^{Ar} = 150.86$ K, $T_C^{Me} = 190.69$ K) and the critical pressure ($p_C^{Ar} = 4.898$ MPa, $p_C^{Me} = 4.599$ MPa) are taken from the DIPPR databank [15]. Thus calculating the parameter set for argon yields $\epsilon_{Ar}/k = 115.16$ K and $\sigma_{Ar} = 3.59$ Å, which is not within the Pareto set. For methane the parameter set is $\epsilon_{Me}/k = 145.47$ K and $\sigma_{Me} = 3.88$ Å and therefore within the Pareto set.

4.4. Topological discussion

To explain the structure of the triangular Pareto set and its tails as shown in Figs. 2 and 3, the white dashed lines in Fig. 1 indicating the paths of lowest descent for the individual properties as well as the global minima marked with white crosses in this figure are crucial: The crosses indicate the solutions obtained for the single objective optimization when only the corresponding property is used. As expected, the optima obtained for the different objectives do not coincide. Therefore, in any optimization which takes into account more than one of the objectives, a compromise has to be made. Thus moving away from the global minimum, the least loss in the quality of the representation for one property has to be accepted if following the path of lowest descent.

Let us first consider the class of optimization problems that are obtained when two objectives are used. For the six objectives that are studied here, there are 30 such combinations for each of the two fluids. We illustrate our argument only using the liquid density and vapor pressure of argon: The global minima for the two individual properties belong to the Pareto set. E.g. starting at the global minimum for the mean relative deviation of the liquid density, the goal is to improve the quality of the mean relative deviation for the vapor pressure by a variation of the model parameters ϵ and σ. Best compromises are only achieved, if simultaneously the trade-off for the mean relative deviation in the liquid density is kept minimal. Hence the variation of the model parameters has to be along the path of lowest descent of the liquid density, up to the point, where it intersects the path of lowest descent of the vapor pressure. Then the further variation has to follow the lowest descent of the vapor pressure until the global optimum in the mean relative deviation of the vapor pressure is reached.

With the help of Fig. 6 we illustrate, how the argument is expendable to three objectives.

Fig. 6 shows the parameter space of the three dimensional optimization task of argon, considering the mean relative deviations of the liquid density, vapor pressure and enthalpy of vaporization (cf. Fig. 2). The blue markers represent the parameter combinations ϵ and σ for which Pareto optimal mean relative deviations in the three properties are achieved. Additionally the lines of lowest descent for the three individual properties, as well as the two global minima for the liquid density and the vapor pressure are plotted. For the enthalpy of vaporization no global minimum forms, as it is independent from the parameter σ (see Fig. 1).

For three properties, the triangle is bounded by the three lines of lowest descent of the three individual properties. The models yielding best compromises lie in this region. Starting from its bounds and moving inside the triangle, whichever path is taken, the distance to at least one line of lowest descent is reduced and therefore the quality in the corresponding property improves. Simultaneously, the distance to at least one other line of lowest descent increases,
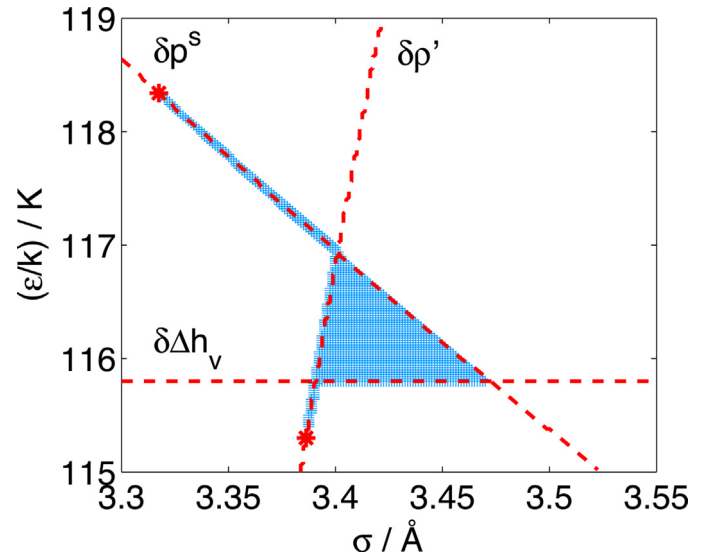

Fig. 6. Pareto set in parameter space for three objective functions of argon: liquid density ρ', vapor pressure p^S and enthalpy of vaporization Δh_V. Dashed lines: path of lowest descent for properties. Asterisks: Global minima of properties.

leading to a decline in the quality of this property. Hence, the models inside the triangle belong to the Pareto set.

A tail is formed, if the global minimum along one line of lowest descent of one property does not lie between the two points of intersection with the other lines of lowest descent. It is therefore possible, to improve the quality of the representant further until their respective global minimum is reached. The global minimum of the single objective optimizations are, hence, the end points of the tails.

In Fig. 2 three self-intersections are observed in the objective space. Each starts at one of the corners of the triangle of the Pareto set. Let us consider the intersection of the lines of lowest descent for the liquid density and the vapor pressure, as an example. A contour line of the enthalpy of vaporization spreading from this point can be drawn. There are two directions along the contour line of the enthalpy of vaporization. In both directions the mean relative
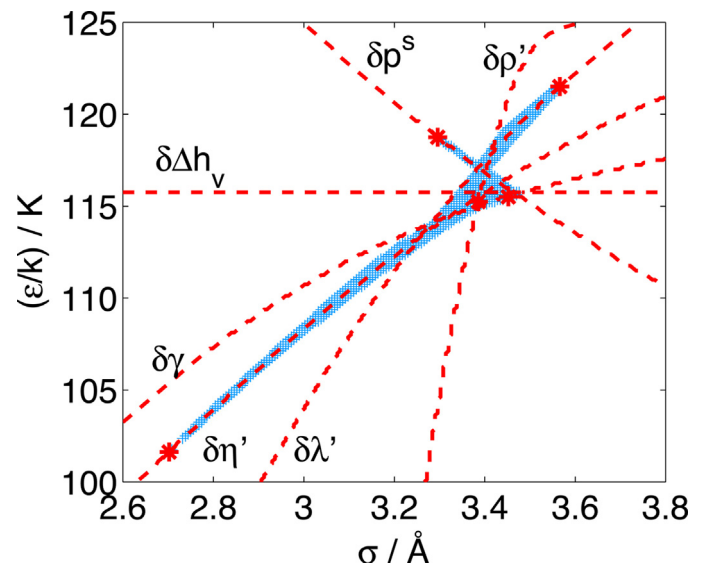

Fig. 7. Pareto set in parameter space for six objective functions of argon: liquid density ρ', vapor pressure p^S, enthalpy of vaporization Δh_V, liquid viscosity η', liquid thermal conductivity λ' and surface tension γ. Dashed lines: path of lowest descent for properties. Asterisks: Global minima of properties.

deviation in the liquid density and in the vapor pressure increases monotonically. Thus, it is possible to find two different parameter sets resulting in the same relative mean deviations for all three properties. As there are three points of intersecting lines of descent, three lines of self-intersections are found in the objective space.

Fig. 7 shows the parameter space of the six dimensional optimization task of argon, considering the mean relative deviations of the liquid density, vapor pressure, enthalpy of vaporization, liquid viscosity, liquid thermal conductivity and surface tension.

The triangular core structure is bound by the lines of lowest descent of the vapor pressure, liquid thermal conductivity and surface tension. The lines of lowest descent for the liquid density, enthalpy of vaporization and liquid viscosity run through the triangular region. The three tails spread along the lines of lowest descent for the vapor pressure, liquid viscosity and liquid thermal conductivity and end at the respective global minimum for the three properties. This again confirms the explanation given for the topological characteristics for argon for the three dimensional optimization task.

5. Conclusions

In this work we study the multicriteria optimization of the Lennard-Jones model by Pareto approach for the two fluids argon and methane. As objectives the mean relative deviations of six different properties, i.e. liquid density, vapor pressure, enthalpy of vaporization, liquid viscosity, liquid thermal conductivity, as well as the surface tension, are considered. In a first step a single objective fit shows the expected conflict: Though for each property excellent results can be achieved with the Lennard-Jones model, those results are only possible with very different parameter sets for both argon and methane. To answer the question if good compromises are achievable, we apply Pareto optimization, to identify several Pareto sets. First two three dimensional optimization tasks are considered. An unexpected topology is observed, which can be fully explained by the results of the single objective optimization problem. Additionally, the Pareto analysis shows, that good compromises are possible. Finally, all six objectives are optimized simultaneously and the results presented in the objective and parameter space. The selection of models leading to Pareto optimal relative mean deviations in the various properties is again explained by the results of the single objective optimization. Despite of the complexity of the optimization task, good compromises are possible. For argon, with the parameter set $\epsilon/k = 116.6$ K and $\sigma = 3.4$ Å mean relative deviations of 0.6% for the liquid density, 1.7% for the vapor pressure, 2.3% for the enthalpy of vaporization, 8.4% for the liquid viscosity, 4.3% for the liquid heat conductivity, and 11.4% for the surface tension are achieved. For methane, with the parameter set $\epsilon/k = 149$ K and $\sigma = 3.73$ Å mean relative deviations of 0.4% for the liquid density, 1.5% for the vapor pressure, 1.6% for the enthalpy of vaporization, 26.5% for the liquid viscosity, 2.1% for the liquid heat conductivity, and 3.2% for the surface tension are achieved. A comparison to the model calculated from the principle of corresponding states as well as to models from former optimization efforts agree well with the identified Pareto set.

Pareto optimization is attractive for developing force field models, as it allows complete assessment of what can be achieved with a given model.

A future challenge is to directly combine multicriteria optimization algorithms with molecular simulations. Therefore, suitable algorithms need to be developed for the purpose of determining the Pareto set in such cases, permitting a more efficient exploration of the Pareto frontier. Then a complete sampling of the parameter space is not necessary and the performance of more complex molecules can be studied.

Beyond the scope of this work, Pareto optimization is attractive for parameterizing other models for thermodynamic properties like e.g. equations of states and g^E-models.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fluid.2014.04.009.

References

[1] B. Eckl, J. Vrabec, H. Hasse, J. Phys. Chem. B 112 (2008) 12710–12721.
[2] R. Faller, H. Schmitz, O. Biermann, F. Müller-Plathe, J. Comput. Chem. 20 (1998) 100–109.
[3] D. Reith, H. Meyer, F. Müller-Plathe, Comput. Phys. Commun. 148 (2002) 299–313.
[4] J. Wang, P.A. Kollman, J. Comput. Chem. 22 (2001) 1219–1228.
[5] E. Bourasseau, M. Haboudou, A. Boutin, A.H. Fuchs, P. Ungerer, J. Chem. Phys. 118 (2003) 3020–3034.
[6] M. Hülsmann, T. Köddermann, J. Vrabec, D. Reith, Comput. Phys. Commun. 181 (2010) 499–513.
[7] M. Hülsmann, J. Vrabec, A. Maaß, D. Reith, Comput. Phys. Commun. 181 (2010) 887–905.
[8] S. Deublein, P. Metzler, J. Vrabec, H. Hasse, Mol. Simul. 39 (2013) 109–118.
[9] S. Mostaghim, M. Hoffmann, P.H. Konig, T. Frauenheim, J. Teich, Molecular force field parametrization using multi-objective evolutionary algorithms Congress on Evolutionary Computation (CEC), vol. 1, 2004, pp. 212–219.
[10] J.I. Hernández, Multi-objective Optimization in Mixed Integer Problems: With Application to the Beam Selection Optimization Problem in IMRT, in: International Series in Operations Research and Management Science, Mensch und Buch Verlag, Berlin, 2012.
[11] F. Logist, B. Houska, J.V. Impe, Struct. Multidiscip. Optim. 42 (2010) 591–603.
[12] K.M. Miettinen, Nonlinear Multiobjective Optimization, volume 12 of International Series in Operations Research and Management Science, Kluwer Academic Publishers, Dordrecht, 1999.
[13] C. Hillermeier, Nonlinear Multiobjective Optimization, vol. 135 of International Series of numerical mathematics, Birkhäuser Verlag, Basel-Boston-Berlin, 2001.
[14] M. Ehrgott, Multicriteria Optimization, vol. 491 of Lecture Notes in Economics and Mathematical Systems, Springer Verlag, Berlin-Heidelberg, 2005.
[15] Design Institute for Physical Properties Sponsored by AIChE, DIPPR Project 801 – Full Version, 2012.
[16] A. Lotfi, J. Vrabec, J. Fischer, Mol. Phys. 76 (1992) 1319–1333.
[17] R.L. Rowley, M.M. Painter, Int. J. Thermophys. 18 (1997) 1109–1121.
[18] G.A. Fernández, J. Vrabec, H. Hasse, Fluid Phase Equilib. 221 (2004) 157–163.
[19] P. Grosfils, J.F. Lutsko, J. Chem. Phys. 130 (2009) 054703.
[20] T. Merker, J. Vrabec, H. Hasse, Soft Mater. 10 (2012) 3–25.
[21] S.D. Bembenek, B.M. Rice, Mol. Phys. 97 (1999) 1085–1094.
[22] K. Shukla, Fluid Phase Equilib. 94 (1994) 19–49.
[23] J. Vrabec, J. Fischer, Mol. Phys. 85 (1995) 781–792.
[24] J. Fischer, R. Lustig, H. Breitenfelder-Manske, W. Lemming, Mol. Phys. 52 (1984) 485–497.
[25] J. Vrabec, J. Stoll, H. Hasse, J. Phys. Chem. B 105 (2001) 12126–12133.
[26] W.L. Jorgensen, J.D. Madura, C.J. Swenson, J. Am. Chem. Soc. 106 (1984) 6638–6646.
[27] M.G. Martin, J.I. Siepmann, J. Phys. Chem. B 102 (1998) 2569–2577.