

1 Statement of the problem

The National Institute of Standards and Technology (NIST) Thermodynamics Research Center (TRC) is a premier organization in evaluating thermodynamic data that are necessary for various engineering purposes. In this regard, molecular simulation has become an efficient method for predicting thermophysical properties. The accuracy of such predictions is based upon the assumptions made in the development of the molecular model and the parameterization of the force field, where the intermolecular parameters are typically of utmost importance. For example, different assumptions are too restrictive to allow accurate prediction of several different properties. After selecting the model form, the force field parameters are usually obtained to reproduce liquid density (ρ_L), vapor pressure (P_v), and/or heat of vaporization (ΔH_v) data. However, due to the high computational cost for molecular simulations these parameters are usually optimized in a brute-force manner where the entire parameter space is rarely considered. This type of optimization is limited in its ability to assure that a global minimum is obtained, assess the degree of nonlinear coupling, and quantify the uncertainty in the optimal parameters. An alternative optimization approach is necessary to overcome these deficiencies. In this study, we utilize a novel approach to force-field parameterization that facilitates a simultaneous optimization of several parameters to large amounts of experimental data. Our goal is to utilize this approach with the extensive, experimental data archives available at NIST/TRC to develop more self-consistent, predictive force-field models.

2 Background and relevance to previous work

The traditional approach of developing a family of force fields is to begin with compounds that contain the fewest number of site types. For example, it is common to start with the n -alkane family, in particular ethane and the smaller chain-lengths, and then introduce

additional functional groups by optimizing force field parameters for small compounds where previously optimized parameters are transferred when applicable. However, the assumption of transferability can constrain the new parameter set to overcompensate when a sequential optimization is implemented.

The Transferable Potentials for Phase Equilibria (TraPPE) family of force fields is one of the most inclusive databases of transferable potential parameters for industrially relevant organic functional groups. The united-atom version (TraPPE-UA) currently has more than 30 different UA Lennard-Jones 12-6 (LJ 12-6) site types. However, many studies have demonstrated that a UA LJ 12-6 model is not flexible enough to accurately describe ρ_L , P_v , and ΔH_v , which are typically the properties of interest when optimizing a force field model [1–3]. There are essentially three different ways to overcome this shortcoming: explicit hydrogen (EH), anisotropic united-atom (AUA), and a more flexible mathematical model such as the Exponential-6 (Exp-6) or Mie λ -6 model.

Chen et al. demonstrated that an EH model can improve the agreement in the normal boiling point (T_b) for *n*-alkanes from 4% for the UA model [4] to less than 0.5% without sacrificing the performance for ρ_L [1]. The disadvantage to the EH approach is the significant increase in computational cost, especially for larger compounds. For this reason, the TraPPE-UA version is still quite popular compared to TraPPE-EH. Alternatively, Ungerer et al. have observed excellent agreement between P_v , ΔH_v , and ρ_L for the AUA approach [5,6]. The AUA method has the disadvantage that it is not supported by all software packages due to a slight increase in complexity. Errington et al. [7] and Potoff et al. [2] have demonstrated that the Exp-6 and Mie λ -6 models provide considerable improvement over the LJ 12-6 model as they can simultaneously represent P_v , ΔH_v , and ρ_L . The advantage of modifying the mathematical model rather than the EH and AUA approaches is that the computational cost and complexity is relatively unaffected.

From an optimization standpoint, the EH approach significantly increases the number of

parameters. The AUA approach introduces a single additional parameter for each UA site. However, one advantage of the AUA approach is that the shifting parameter (δ) has physical meaning and can be approximated *a priori*. Although the Mie λ -6 model also requires an additional parameter when compared to the LJ 12-6 model, this model appears to be more popular than the AUA approach due to its simple implementation. For these reasons, in this study we utilize the Mie λ -6 potential with the UA assumption.

The standard optimization procedure is to find the parameter set that minimizes the objective function:

$$F(b) = \sum_i \left(\frac{\hat{f}(b; T_i) - f_{\text{exp},i}}{\delta f_i} \right)^2 \quad (1)$$

where b is the set of parameters, $\hat{f}(b; T_i)$ is the predicted value for property f at T_i which corresponds to $f_{\text{exp},i}$, and δf_i is the combined uncertainty in the predicted and experimental values. The unique difficulty when optimizing force field parameters is that the evaluation of \hat{f}_i is not straightforward. On the contrary, evaluating \hat{f}_i may require hours or days of computer time since this is typically achieved by performing molecular simulations at each T_i . Due to this high computational cost most studies do not perform a complete scan of the parameter space. However, recently it was demonstrated that GOMC enables a full parameter scan to verify that a global minimum is reached and to quantify the uncertainty in the parameters [8]. However, this approach is still limited to the case where only a few parameters are simultaneously optimized. An additional complication is that \hat{f} has some inherent uncertainty when obtained from molecular simulation. This renders gradient-based optimization and uncertainty quantification difficult because both of these methods rely upon approximating derivatives in \hat{f} with respect to b . In some cases, the uncertainty associated with the molecular simulation is too large to allow for an accurate assessment of these derivatives.

For these reasons, some researchers have resorted to using molecular based equations of

state as a surrogate model to approximate \hat{f} , e.g. PC-SAFT [9], SAFT- γ [10–13]. In fact, some attempts have been made at automating this process with a corresponding state SAFT approach [14]. The results from Dufal et al. are significant since they demonstrate that very accurate results are possible when several compounds are optimized simultaneously to large amounts of data [10]. However, use of a surrogate model is not as rigorous as molecular simulation and the development is not straightforward for more complex molecular models [15]. Therefore, it would be ideal if one could achieve the rigor and robustness of molecular simulation while not sacrificing the speed and precision of the surrogate model approach. We propose a method titled “pair correlation function post-simulation optimization” (PCF-PSO) that attempts to find a compromise between the surrogate model and molecular simulation methods for parameter optimization.

3 General Methodology

For this project we propose a systematic approach (PCF-PSO) to converting all of the Lennard-Jones 12-6 parameters from the TraPPE-UA force field to Mie λ -6 parameters. This approach assumes that the Mie λ -6 potential is capable of accurately representing P_v , ΔH_v , and ρ_L for the functional groups found in the TraPPE-UA database. The PCF-PSO method utilizes a perturbation theory approach similar in spirit to that used to develop the SPEADMD force field parameters [16]. The general assumption is that small perturbations in the intermolecular parameters can be accurately accounted for explicitly, i.e. without requiring additional simulations. Specifically, the TraPPE-UA model is taken as the unperturbed system and the perturbation is converting a LJ 12-6 model to a Mie λ -6 model.

The key advantages to the PCF-PSO approach is that approximations for P_v , ΔH_v , and ρ_L are obtained for any parameter set without additional simulations. Therefore, it is possible to perform more advanced multidimensional optimization algorithms, such as simulated

annealing. Similarly, the assumption of transferability can be relaxed by increasing the number of parameters and compounds that are simultaneously optimized. This allows for a more extensive use of the evaluated experimental data that are available at NIST/TRC. In addition, it is possible to assess the degree of nonlinear coupling between parameters for an uncertainty analysis. In summary, PCF-PSO renders multidimensional parameterization feasible with regards to nonlinear coupling analysis, global minimum searches, and uncertainty quantification assessments.

4 New or unusual methods

To our knowledge, the pair correlation function post-simulation optimization (PCF-PSO) method proposed in this document has never been implemented in the literature. In this section we discuss the implementation of the PCF-PSO method since the performance of this method is pivotal to achieving the desired goals of this research.

A force field model generally has the following form:

$$u = u_{intra} + u_{inter} \tag{2}$$

where the force field has been separated into an intramolecular term (u_{intra}) and an intermolecular term (u_{inter}). Furthermore, the intermolecular portion can be divided between short-range interactions (u_{inter}^{sr}) and long-range interactions (u_{inter}^{lr}) to yield:

$$u = u_{intra} + u_{inter}^{sr} + u_{inter}^{lr} \tag{3}$$

In this study we adopt the intramolecular and Coulombic parameters from the literature (the TraPPE-UA model) and, therefore, we are only concerned with optimizing the parameters for u_{inter}^{sr} . The importance for the distinction between the different terms in Equation 3 will

become clear in the following discussion.

We will consider two different force field models (β and α) that only differ in u_{inter}^{sr} . In other words, the intramolecular and Coulombic parameters are the same while the short-range intermolecular parameters are different. For the remainder of the discussion the subscript will be omitted and it is implied that u^{sr} refers to the intermolecular short-range interactions. The molar internal energy (U) for the force field model with parameters β can be expressed as:

$$\langle U(\beta) \rangle = \langle U(\alpha) \rangle + U_{inter}^{sr}(\beta) - U_{inter}^{sr}(\alpha) \quad (4)$$

where $\langle \dots \rangle$ denotes an ensemble average, $U(\beta)$ and $U(\alpha)$ are the internal energies for the force field model with parameters β and α , respectively, and $U_{inter}^{sr}(\beta)$ and $U_{inter}^{sr}(\alpha)$ are the contributions to U from the short-range intermolecular interactions (u^{sr}) for the β and α force field parameters, respectively. The assumption made in deriving Equation 4 is that the intramolecular and long-range contributions are the same for the two different parameter sets, β and α . We expect that this assumption is valid within the statistical uncertainties associated with the ensemble averages. Subsequently, pair-wise additivity of the two-body interactions is assumed in order to obtain the short-range intermolecular portion (U_{inter}^{sr}) for a polyatomic molecule:

$$U_{inter}^{sr}(b) = 2\pi\rho R_g \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \int_{r_{i,j}=0}^{\infty} u_{i,j}^{sr}(b; r_{i,j}) g_{i,j}(b; r_{i,j}) r_{i,j}^2 dr_{i,j} \quad (5)$$

where ρ is the number density, R_g is the universal gas constant, N_S is the number of sites in a molecule, $u_{i,j}^{sr}(b; r_{i,j})$ is the short-range potential between sites i and j with intermolecular parameters b , and $g_{i,j}(b; r_{i,j})$ is the site-site radial distribution function between sites i and j obtained from intermolecular parameters b . Substitution of Equation 5 into Equation 4 and

combining the two integrals for U_{inter}^{sr} gives:

$$\langle U(\beta) \rangle = \langle U(\alpha) \rangle + 2\pi\rho R_g \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \int_{r_{i,j}=0}^{\infty} [u_{i,j}^{sr}(\beta; r_{i,j}) g_{i,j}(\beta; r_{i,j}) - u_{i,j}^{sr}(\alpha; r_{i,j}) g_{i,j}(\alpha; r_{i,j})] r_{i,j}^2 dr_{i,j} \quad (6)$$

The next step involves the fundamental assumption in the PCF-PSO derivation which is that the radial distribution functions for the two force fields are approximately equal:

$$g_{i,j}(\beta; r_{i,j}) \approx g_{i,j}(\alpha; r_{i,j}) \quad (7)$$

For small differences between β and α this assumption is valid. However, even for large deviations, this assumption is useful when optimizing a force field model because it allows one to predict $\langle U(\beta) \rangle$ from $\langle U(\alpha) \rangle$ without requiring any additional simulations. Applying Equation 7 to Equation 6 yields the final expression:

$$\langle U(\beta) \rangle \approx \langle U(\alpha) \rangle + 2\pi\rho R_g \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \int_{r_{i,j}=0}^{\infty} [u_{i,j}^{sr}(\beta; r_{i,j}) - u_{i,j}^{sr}(\alpha; r_{i,j})] g_{i,j}(\alpha; r_{i,j}) r_{i,j}^2 dr_{i,j} \quad (8)$$

A similar derivation is possible for pressure (P) starting with the expression:

$$\langle P(\beta) \rangle = \langle P(\alpha) \rangle + P_{inter}^{sr}(\beta) - P_{inter}^{sr}(\alpha) \quad (9)$$

where the short-range intermolecular contributions to the pressure (P_{inter}^{sr}) for a polyatomic compound is (assuming pair-wise additivity):

$$P_{inter}^{sr}(b) = -\frac{2}{3}\pi\rho^2 \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \int_{r_{i,j}=0}^{\infty} \frac{du_{i,j}^{sr}(b; r_{i,j})}{dr_{i,j}} g_{i,j}(b; r_{i,j}) r_{i,j}^3 dr_{i,j} \quad (10)$$

Substitution of Equation 10 into Equation 9 and combining the two integrals for P_{inter}^{sr} gives:

$$\langle P(\beta) \rangle = \langle P(\alpha) \rangle - \frac{2}{3}\pi\rho^2 \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \int_{r_{i,j}=0}^{\infty} \left[\frac{du_{i,j}^{sr}(\beta; r_{i,j})}{dr_{i,j}} g_{i,j}(\beta; r_{i,j}) - \frac{du_{i,j}^{sr}(\alpha; r_{i,j})}{dr_{i,j}} g_{i,j}(\alpha; r_{i,j}) \right] r_{i,j}^3 dr_{i,j} \quad (11)$$

Again we assume that Equation 7 is valid, i.e. that the radial distribution function is constant with respect to a change in the parameters such that:

$$\langle P(\beta) \rangle \approx \langle P(\alpha) \rangle - \frac{2}{3}\pi\rho^2 \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \int_{r_{i,j}=0}^{\infty} \left[\frac{du_{i,j}^{sr}(\beta; r_{i,j})}{dr_{i,j}} - \frac{du_{i,j}^{sr}(\alpha; r_{i,j})}{dr_{i,j}} \right] g_{i,j}(\alpha; r_{i,j}) r_{i,j}^3 dr_{i,j} \quad (12)$$

With expressions for P and U the heat of vaporization is calculated as:

$$\Delta H_v = U_v - U_L + P_v(V_v - V_L) \quad (13)$$

where U_v and U_L are the molar internal energies for the vapor and liquid phases, respectively, P_v is the vapor pressure, and V_v and V_L are the vapor and liquid molar volumes, respectively. U_v and U_L are obtained from Equation 8 by using ρ_v and ρ_L and where $g_{i,j}(r_{i,j})$ is obtained from the vapor and liquid phases, respectively. As is typically recommended for calculating P_v from molecular simulation, P_v in Equation 13 was calculated using the vapor phase simulations. In other words, Equation 12 is evaluated using ρ_v for ρ and $g_{i,j}(r_{i,j})$ is the radial distribution function for the vapor phase.

The largest assumption made in this approach is that $g_{i,j}(r_{i,j})$ is constant with respect to the intermolecular parameters. This assumption has been tested for the single-site Mie λ -6 model. Specifically, we varied the reduced values of ϵ and σ between 0.8-1.2 while also varying λ between integer values of 8-18. We calculated P_v , U_v , and U_L with the $g(r)$ obtained from simulation for each set of ϵ , σ , and λ . We also calculated P_v , U_v , U_L for each set of ϵ , σ , and λ but with the $g(r)$ obtained from simulation of $\epsilon = 1$, $\sigma = 1$, and $\lambda = 12$, i.e. the LJ 12-6.

Our findings were that the PCF-PSO approach accurately predicts U_L (within 1%) while P_v is less accurate (deviations as large as 5%).

One disadvantage to the PCF-PSO method is that liquid density (ρ_L) is not included directly in the optimization. However, ρ_L is known to be intimately related to both σ and r_{\min} . Therefore, we recommend that the optimization procedure utilizes a constraint for σ and/or r_{\min} . We propose two different means for determining the constraints that are placed upon $\sigma_{i,j}$ and $r_{\min,i,j}$. The first approach is to allow these values to vary by the same amount as the deviation for ρ_L obtained from the TraPPE-UA model. In other words, if the TraPPE-UA model does a poor job representing ρ_L then $\sigma_{i,j}$ and $r_{\min,i,j}$ will be less constrained. Whereas if the TraPPE-UA model accurately represents ρ_L for a given species then the constraints will be tighter. Alternatively, comparing the results by Potoff et al. for ethane it can be assumed that neither value vary by more than 0.2%.

Having discussed how P_v , ρ_L , and ΔH_v are accounted for in the PCF-PSO approach, we discuss the unique aspects of the optimization procedure. Since ΔH_v depends upon P_v we can simplify the optimization by utilizing an objective function that only depends on ΔH_v .

$$F = \sum_i^{N_T} \left(\frac{\Delta H_v(T_i) - \Delta H_{v,i}}{\delta \Delta H_{v,i}} \right)^2 \quad (14)$$

where N_T is the number of temperatures used in evaluation, $\Delta H_v(T_i)$ is the predicted ΔH_v at T_i , $\Delta H_{v,i}$ is the experimental ΔH_v at T_i , and $\delta \Delta H_{v,i}$ is the combined uncertainty in ΔH_v at T_i . That is, this accounts for both the uncertainty in the predicted and the experimental values for ΔH_v .

One advantage for using Equation 14 is that performing a single property (ΔH_v) optimization eliminates the need for weighting different terms in the objective function. Furthermore, P_v is a noisy property and is highly sensitive to $g(r)$. Therefore, the assumption that $g(r)$ is constant is worse for approximating P_v than for U_v and U_L . For these reasons,

we recommend that P_v be directly excluded from the objective function until the optimal set of parameters (\hat{b}) are obtained for Equation 14. Subsequently, a P_v term may be added since only small deviations from \hat{b} are expected at this stage in the optimization. Specifically, after obtaining \hat{b} for Equation 14 we propose an objective function of the form:

$$F = w_{\Delta H_v} \sum_i^{N_T} \left(\frac{\Delta H_v(T_i) - \Delta H_{v,i}}{\delta \Delta H_{v,i}} \right)^2 + w_{P_v} \sum_i^{N_T} \left(\frac{P_v(T_i) - P_{v,i}}{\delta P_{v,i}} \right)^2 \quad (15)$$

where $w_{\Delta H_v}$ is the weighting for the ΔH_v portion of the objective function, w_{P_v} is the weighting for the P_v portion of the objective function, $P_v(T_i)$ is the predicted P_v at T_i , $P_{v,i}$ is the experimental P_v at T_i , and $\delta P_{v,i}$ is the combined uncertainty associated with P_v at T_i . We recommend that the ratio of weighting be approximately 1:40 $w_{P_v}:w_{\Delta H_v}$.

Having discussed the details, we now present the generic PCF-PSO algorithm for a single compound:

1. Select an initial set of parameters (b_0)
2. Simulate compound in NVT ensemble at $\rho_v^*(T_j)$ and $\rho_L^*(T_j)$ over range of T
3. Calculate radial distribution function ($g_{i,j}(b_0; r_{i,j})$) for each simulation
4. Calculate P_v and ΔH_v using Equations 12 and 13, respectively
5. Minimize objective function (Equation 14) to obtain b_1 with constraints for $\sigma_{i,j}$ and $r_{\min,i,j}$
6. Repeat Steps 2-5 where in Step 2 the simulations use b_i and in Step 5 a new set of parameters b_{i+1} are obtained
7. Stop when F and b do not change (to within a predetermined tolerance) from iteration i to step $i + 1$

8. Repeat Steps 6-7 where the objective function includes P_v such as Equation 15

In Step 2 $\rho_v^*(T_j)$ and $\rho_L^*(T_j)$ correspond to the experimental saturated vapor and liquid densities at T_j . Although validation of the aforementioned algorithm for a single compound is an important task, the primary goal of this study is to obtain improved parameters for a large set of functional groups. The complete algorithm to achieve this goal is:

1. Select several compounds
2. Set initial parameters (b_0) to values found in literature (i.e. TraPPE-UA)
3. Simulate each compound in NVT ensemble at $\rho_v^*(T_j)$ and $\rho_L^*(T_j)$ over range of T
4. Calculate radial distribution function ($g_{i,j}(b_0; r_{i,j})$) for each simulation
5. Calculate P_v and ΔH_v using Equations 12 and 13, respectively
6. Minimize objective function (Equation 14) to obtain b_1 with constraints for $\sigma_{i,j}$ and $r_{\min,i,j}$
7. Repeat Steps 2-5 where in Step 2 the simulations use b_i and in Step 5 a new set of parameters b_{i+1} are obtained
8. Stop when F and b do not change (to within a predetermined tolerance) from iteration i to step $i + 1$
9. Repeat Steps 7-8 where the objective function includes P_v such as Equation 15

The compounds chosen in Step 1 are determined by several factors. The first factor is the availability of reliable experimental data, namely, ρ_v , ρ_L , P_v , and ΔH_v . For this reason, NIST/TRC is an ideal location for this project. That is, NIST/TRC can provide large amounts of evaluated data and correlations for a plethora of different compounds. The

second factor is the feasibility of simulating the compound. Specifically, larger compounds require larger simulations and, thus, more computational time. Fortunately, the first and second factor are typically cooperative since larger compounds tend to have fewer reliable experimental data. The third factor is the molecular structure of the compound. A compound with more functional groups can provide more insight and less redundant information. For example, if the selection of compounds is done appropriately, one can elucidate cross-interactions and, thereby, avoid the assumption of Lorentz-Berthelot combining rules. In fact, it is necessary to consider the molecular structures collectively of the entire set of compounds to assure that a unique solution is achieved. Since we are attempting to build upon the TraPPE-UA force field, we are inherently limited to compounds that contain UA sites that can be represented by TraPPE-UA.

5 Expected results, significance, and application

The expected results from this study are a family of Mie λ -6 force field parameters as inclusive as those for TraPPE-UA but for which P_v , ΔH_v , and ρ_L are accurately predicted. This family of force fields will provide researchers with more reliable prediction models than those existing in the literature. In addition, an uncertainty analysis will be performed to determine the nonlinear coupling between parameters. This will enable an uncertainty propagation when predicting properties for compounds and state points not used in the optimization procedure.

In the scenario that the development of the PCF-PSO method takes more time than foreseen, we may only be able to reparameterize a few of the UA site types. At the very least, we will present a proof of concept for the *n*-alkane family. This will allow future researchers to implement this method with more UA site types at a later date.

Previously we mentioned that ρ_L can be implicitly included in the optimization by con-

straining $\sigma_{i,j}$ and/or $r_{\min,i,j}$. In fact, it may be beneficial to assign constraints to each of the parameters such that a “trust region” approach assures that the new set of optimal parameters (b_{i+1}) does not deviate too much from the previous iteration set (b_i). The reason why a “trust region” might be necessary is because the most likely impediment to the success of the PCF-PSO method is the assumption of constant $g_{i,j}(r_{i,j})$ for the entire parameter space (b). The size of these constraints would be determined by assessing the range over which $\epsilon_{i,j}$, $\sigma_{i,j}$, and $\lambda_{i,j}$ yield accurate ΔH_v and P_v .

Alternatively, the PCF-PSO approach may be used in conjunction with a gradient-based optimization method (such as those proposed by Ungerer et al. [5], Bourasseau et al. [6] and Hulsmann et al. [17]). Each one of these methods requires partial derivatives of a macroscopic property with respect to the parameters. In the case of Bourasseau et al. these partial derivatives are rigorously related to the partial derivatives of P_v and U with respect to the parameters. For each of these methods the partial derivatives are approximated with crude forward-difference numerical expressions such as:

$$\frac{\partial X}{\partial b_k} = \frac{X(b_1, \dots, b_k + \delta_k, \dots, b_p) - X(b_1, \dots, b_k, \dots, b_p)}{\delta_k} \quad (16)$$

where X is the desired property, there are p parameters, and b_k is the parameter being perturbed by the small amount δ_k . The error for this method is on the order of δ_k , however, this error can be reduced by using a higher order approximation. For example, a forward-backward-difference yields an error on the order of δ_k^2 (remember that δ_k is small). This is expressed as:

$$\frac{\partial X}{\partial b_k} = \frac{X(b_1, \dots, b_k + \delta_k, \dots, b_p) - X(b_1, \dots, b_k - \delta_k, \dots, b_p)}{2\delta_k} \quad (17)$$

The obvious disadvantage when using a higher order approximation is that additional simulations are required. For example, in Equation 16 an additional simulation is required to obtain $X(b_1, \dots, b_k + \delta_k, \dots, b_p)$ while in Equation 17 two additional simulations are required,

one for $X(b_1, \dots, b_k + \delta_k, \dots, b_p)$ and one for $X(b_1, \dots, b_k - \delta_k, \dots, b_p)$. It is important to remember that an estimate for $\frac{\partial X}{\partial b_k}$ is necessary for all of the p parameters. Therefore, a method using Equation 16 requires p additional simulations while a method that uses Equation 17 requires $2p$ additional simulations. By contrast, the PCF-PSO approach allows for an approximation of P_v and U directly without additional simulations. Therefore, this allows a very accurate numerical estimate of $\frac{\partial X}{\partial b_k}$. For example, an error on the order of δ_k^4 can be achieved by using:

$$\frac{\partial X}{\partial b_k} = \frac{X(b_k - 2\delta_k) - 8X(b_k - \delta_k) + 8X(b_k + \delta_k) - X(b_k + 2\delta_k)}{12\delta_k} \quad (18)$$

where the notation has been simplified for clarity. In fact, numerical approximations are not even necessary for the PCF-PSO approach because analytic derivatives are attainable, although the integration must be performed numerically. For example, the partial derivatives for internal energy for the Mie λ -6 potential are calculated as:

$$\frac{\partial U}{\partial \epsilon_{i,j}} = 2\pi\rho R_g \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \int_{r=0}^{\infty} \frac{\partial u_{i,j}(r_{i,j})}{\partial \epsilon_{i,j}} g_{i,j}(r_{i,j}) r_{i,j}^2 dr_{i,j} \quad (19)$$

$$\frac{\partial U}{\partial \sigma_{i,j}} = 2\pi\rho R_g \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \int_{r=0}^{\infty} \frac{\partial u_{i,j}(r_{i,j})}{\partial \sigma_{i,j}} g_{i,j}(r_{i,j}) r_{i,j}^2 dr_{i,j} \quad (20)$$

$$\frac{\partial U}{\partial \lambda_{i,j}} = 2\pi\rho R_g \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} \int_{r=0}^{\infty} \frac{\partial u_{i,j}(r_{i,j})}{\partial \lambda_{i,j}} g_{i,j}(r_{i,j}) r_{i,j}^2 dr_{i,j} \quad (21)$$

Similar expressions can be determined for $\frac{\partial P_v}{\partial \epsilon_{i,j}}$, $\frac{\partial P_v}{\partial \sigma_{i,j}}$, and $\frac{\partial P_v}{\partial \lambda_{i,j}}$. These analytic forms are possible because we have assumed that $g(r)$ is constant with respect to the parameters, i.e. $\frac{\partial g(r)}{\partial \epsilon} = 0$, $\frac{\partial g(r)}{\partial \sigma} = 0$, and $\frac{\partial g(r)}{\partial \lambda} = 0$. This is a good assumption in this case since a gradient-based optimization approach performs small steps in the parameter space. The downside to a gradient-based optimization is that it is difficult to discern between local minima and to evaluate nonlinear coupling for a multidimensional parameter space. However, if the

PCF-PSO algorithm outlined previously does not work for a global optimization, we will demonstrate how it can be used as a compliment to previously developed gradient-based optimization algorithms.

References

- [1] B. Chen and J. I. Siepmann. Transferable potentials for phase equilibria. 3. explicit-hydrogen description of normal alkanes. *Journal of Physical Chemistry B*, 103(25):5370–5379, 1999.
- [2] J. J. Potoff and D. A. Bernard-Brunel. Mie potentials for phase equilibria calculations: Applications to alkanes and perfluoroalkanes. *Journal of Physical Chemistry B*, 113(44):14725–14731, 2009.
- [3] K. Stobener, P. Klein, S. Reiser, M. Horsch, K. H. Kufer, and H. Hasse. Multicriteria optimization of molecular force fields by pareto approach. *Fluid Phase Equilibria*, 373:100–108, 2014.
- [4] M. G. Martin and J. I. Siepmann. Transferable potentials for phase equilibria. 1. united-atom description of n-alkanes. *Journal of Physical Chemistry B*, 102(14):2569–2577, 1998.
- [5] P. Ungerer, C. Beauvais, J. Delhommelle, A. Boutin, B. Rousseau, and A. H. Fuchs. Optimization of the anisotropic united atoms intermolecular potential for n-alkanes. *Journal of Chemical Physics*, 112(12):5499–5510, 2000.
- [6] E. Bourasseau, M. Haboudou, A. Boutin, A. H. Fuchs, and P. Ungerer. New optimization method for intermolecular potentials: Optimization of a new anisotropic united

- atoms potential for olefins: Prediction of equilibrium properties. *Journal of Chemical Physics*, 118(7):3020–3034, 2003.
- [7] J. R. Errington and A. Z. Panagiotopoulos. A new intermolecular potential model for the n-alkane homologous series. *Journal of Physical Chemistry B*, 103(30):6314–6322, 1999.
- [8] J. R. Mick, M. S. Barhaghi, B. Jackman, K. Rushaidat, L. Schwiebert, and J. J. Potoff. Optimized mie potentials for phase equilibria: Application to noble gases and their mixtures with n-alkanes. *Journal of Chemical Physics*, 143(11), 2015.
- [9] T. van Westen, T. J. H. Vlugt, and J. Gross. Determining force field parameters using a physically based equation of state. *Journal of Physical Chemistry B*, 115(24):7872–7880, 2011.
- [10] S. Dufal, V. Papaioannou, M. Sadeqzadeh, T. Pogiatzis, A. Chremos, C. S. Adjiman, G. Jackson, and A. Galindo. Prediction of thermodynamic properties and phase behavior of fluids and mixtures with the soft- γ mie group-contribution equation of state. *Journal of Chemical and Engineering Data*, 59(10):3272–3288, 2014.
- [11] V. Papaioannou, F. Calado, T. Lafitte, S. Dufal, M. Sadeqzadeh, G. Jackson, C. S. Adjiman, and A. Galindo. Application of the soft- γ mie group contribution equation of state to fluids of relevance to the oil and gas industry. *Fluid Phase Equilibria*, 416:104–119, 2016.
- [12] T. Lafitte, C. Avendano, V. Papaioannou, A. Galindo, C. S. Adjiman, G. Jackson, and E. A. Muller. Soft- γ force field for the simulation of molecular fluids: 3. coarse-grained models of benzene and hetero-group models of *n*-decylbenzene. *Molecular Physics*, 110(11-12):1189–1203, 2012.

- [13] A. Hemmen and J. Gross. Transferable anisotropic united-atom force field based on the mie potential for phase equilibrium calculations: *n*-alkanes and *n*-olefins. *Journal of Physical Chemistry B*, 119(35):11695–11707, 2015.
- [14] A. Mejia, C. Herdes, and E. A. Muller. Force fields for coarse-grained molecular simulations from a corresponding states correlation. *Industrial & Engineering Chemistry Research*, 53(10):4131–4141, 2014.
- [15] V. Papaioannou, T. Lafitte, C. Avendano, C. S. Adjiman, G. Jackson, E. A. Muller, and A. Galindo. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from mie segments. *Journal of Chemical Physics*, 140(5), 2014.
- [16] A. Sans, A. Vahid, and J. R. Elliott. Transferable intermolecular potential models for a broad range of organic compounds. *Journal of Chemical and Engineering Data*, 59(10):3069–3079, 2014.
- [17] M. Hulsmann, T. Koddermann, J. Vrabec, and D. Reith. Grow: A gradient-based optimization workflow for the automated development of molecular models. *Computer Physics Communications*, 181(3):499–513, 2010.