Single-site Lennard-Jones models via polynomial chaos surrogates of Monte Carlo molecular simulation

Ahmad Kadoura, Adil Siripatana, Shuyu Sun, Omar Knio, and Ibrahim Hoteit

Citation: *The Journal of Chemical Physics* **144**, 214301 (2016); doi: 10.1063/1.4952976

View online: <http://dx.doi.org/10.1063/1.4952976>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/jcp/144/21?ver=pdfcov>

Published by the [AIP Publishing](#)

Articles you may be interested in

[Coexistence and interfacial properties of a triangle-well mimicking the Lennard-Jones fluid and a comparison with noble gases](#)

J. Chem. Phys. **142**, 074706 (2015); 10.1063/1.4909548

[Melting of rare-gas crystals: Monte Carlo simulation versus experiments](#)

J. Chem. Phys. **138**, 104122 (2013); 10.1063/1.4794916

[Equilibrium sizes and formation energies of small and large Lennard-Jones clusters from molecular dynamics: A consistent comparison to Monte Carlo simulations and density functional theories](#)

J. Chem. Phys. **129**, 234506 (2008); 10.1063/1.3040245

[Estimation of bulk liquid properties from Monte Carlo simulations of Lennard-Jones clusters](#)

J. Chem. Phys. **128**, 086101 (2008); 10.1063/1.2830721

[Hydration free energy of a Model Lennard-Jones solute particle: Microscopic Monte Carlo simulation studies, and interpretation based on mesoscopic models](#)

J. Chem. Phys. **128**, 064503 (2008); 10.1063/1.2828192

The cover of the journal 'AIP Applied Physics Reviews' is shown on the left. It features a blue and orange design with a molecular structure illustration. The main text on the right reads 'NEW Special Topic Sections' in large white letters. Below this, it says 'NOW ONLINE' in orange, followed by 'Lithium Niobate Properties and Applications: Reviews of Emerging Trends' in white. The AIP Applied Physics Reviews logo is in the bottom right corner.

NEW Special Topic Sections

NOW ONLINE
Lithium Niobate Properties and Applications:
Reviews of Emerging Trends

AIP Applied Physics Reviews

Single-site Lennard-Jones models via polynomial chaos surrogates of Monte Carlo molecular simulation

Ahmad Kadoura,^{1,a)} Adil Siripatana,^{2,a)} Shuyu Sun,^{1,a)} Omar Knio,^{3,a)} and Ibrahim Hoteit^{2,b)}

¹Computational Transport Phenomena Laboratory, The Earth Sciences and Engineering Department, The Physical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia

²Earth Fluid Modeling and Predicting Group, The Earth Sciences and Engineering Department, The Physical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia

³Uncertainty Quantification Center, The Applied Mathematics and Computational Science Department, The Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia

(Received 29 December 2015; accepted 16 May 2016; published online 1 June 2016)

In this work, two Polynomial Chaos (PC) surrogates were generated to reproduce Monte Carlo (MC) molecular simulation results of the canonical (single-phase) and the NVT-Gibbs (two-phase) ensembles for a system of normalized structureless Lennard-Jones (LJ) particles. The main advantage of such surrogates, once generated, is the capability of accurately computing the needed thermodynamic quantities in a few seconds, thus efficiently replacing the computationally expensive MC molecular simulations. Benefiting from the tremendous computational time reduction, the PC surrogates were used to conduct large-scale optimization in order to propose single-site LJ models for several simple molecules. Experimental data, a set of supercritical isotherms, and part of the two-phase envelope, of several pure components were used for tuning the LJ parameters (ϵ , σ). Based on the conducted optimization, excellent fit was obtained for different noble gases (Ar, Kr, and Xe) and other small molecules (CH_4 , N_2 , and CO). On the other hand, due to the simplicity of the LJ model used, dramatic deviations between simulation and experimental data were observed, especially in the two-phase region, for more complex molecules such as CO_2 and C_2H_6 . *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4952976>]

I. INTRODUCTION

Robust evaluation of fluid thermo-physical properties is essential for simulating transport in porous media. Methods emerging from the continuum based approach, such as equations of state and correlations, have been extensively used for describing the thermodynamics of subsurface reservoir fluids.^{1,2} Due to advancements in algorithms and computational power, Monte Carlo (MC) molecular simulation^{3–6} has been recently established as a powerful tool to describe phase behavior of reservoir fluids.^{7,8} However, the main obstacle hindering MC simulations from being coupled with reservoir flow simulators is the high computational cost needed.

In this regard, researchers have invested great effort in order to speed up MC molecular simulations. Various methods have been developed for this purpose. These include umbrella sampling in various ensembles,^{9–16} bond formation early rejection scheme,¹⁷ histogram reweighting,^{18,19} expanded ensembles method,^{20,21} MC tempering,²² MC Markov chain annealing,²³ and sampling of rejected states.²⁴ Recently, we have also proposed a set of early rejection schemes^{25,26} in

addition to the Markov chain reweighting and reconstruction technique.^{27–29}

In this work, the time consuming MC molecular simulations are replaced by fast surrogate models created via Polynomial Chaos (PC) expansions. PC expansions provide an approximate representation of the solution of the forward problem which can be used to reduce the computational cost of repetitive model evaluations. PC works by representing the model variables and parameters in terms of a spectral expansion in an orthogonal polynomial basis according to their probabilistic distributions. The resulting cost effective surrogate models can then be used to efficiently reproduce the solutions and the statistical properties of Quantities of Interest (QoIs).^{30–33} The PC-based model provides a complete probabilistic representation of the outputs in terms of the random inputs. PC may suffer from “the curse of dimensionality”³⁴ which limits the application of PC to only a moderate number of stochastic parameters for surrogate model construction. Nonetheless, PC methods have become one of the standard approaches for solving stochastic problems, to propagate and quantify uncertainties in various disciplines including physical,^{35–37} chemical, and geophysical systems.^{38–41}

So far, only few studies have applied PC expansions to molecular simulations. In fact, those studies have worked with Molecular Dynamics (MD) instead of MC molecular simulations. For instance, Rizzi *et al.* have demonstrated

^{a)}Email addresses: ahmad.kadoura@kaust.edu.sa; adil.siripatana@kaust.edu.sa; shuyu.sun@kaust.edu.sa; and omar.knio@kaust.edu.sa

^{b)}Author to whom correspondence should be addressed. Electronic mail: ibrahim.hoteit@kaust.edu.sa

a successful coupling between PC and MD in studying force field parameters of water molecules^{42,43} in addition to concentration driven ionic flow in nano-pores.^{44,45} Later on, similar coupling is adopted in order to investigate flow at nanoscale⁴⁶ and quantify parametric uncertainty in multi-scale simulations.⁴⁷

In this study, a novel combination between PC expansions and MC molecular simulations is proposed. First, two independent PC surrogate models are constructed to replace the canonical and the NVT-Gibbs ensembles, respectively. These surrogates are capable of accurately reproducing the normalized supercritical isotherms and the two-phase envelope of the Lennard-Jones (LJ) fluid. The surrogate model estimations match the MC results within a margin of $\pm 3\%$ deviation in the whole thermodynamic range studied, while requiring a tiny fraction of the computational time needed by the MC molecular simulations. In the second part of the paper, a large-scale optimization for forcefield parameters of different molecules is conducted via the pre-constructed PC surrogates. The optimization outcome leads to a recommended set of single-site LJ models for a group of molecules. An excellent fit is achieved for argon (Ar), krypton (Kr), xenon (Xe), methane (CH_4), nitrogen (N_2), and carbon monoxide (CO). Nonetheless, and not surprisingly, other bigger and/or polar molecules show larger misfits due to the simplicity of the LJ model used.

It is worthy to mention that we rely on the Non-Intrusive Spectral Projection (NISP) methodology, as will be outlined in Section II, to determine the PC coefficients. Fortunately, employing the NISP method does not require any modifications inside the model code; therefore, the model is called as it is in order to evaluate the PC coefficients of the surrogate model. Note that application of NISP to represent MD-based QoIs has been hindered by the fact that the MD estimates are typically affected by noise due to finite size sampling. This generally leads to large NISP errors, and consequently a poor representation. Various approaches have been developed in order to overcome this issue.^{42–45,48} In the present work, application of specific techniques to filter out noise proved unnecessary, because of (i) our ability to perform a very large number of MC samples and (ii) to consider multiple replica. As a result, the obtained estimates exhibit sufficient smoothness, which justifies the application of direct projection technique.

The succeeding sections of the paper consist of three main parts. In the first part, a detailed description of the simulation methods used in this work (MC and PC) is given. In Sec. III, results of PC surrogates reproducing the MC simulations for normalized LJ fluids are presented. In addition, the LJ parameter optimization scheme is discussed and the section ends by reporting the recommended LJ parameters for the studied molecules. Concluding remarks and future work plans are finally offered.

II. SIMULATION METHODS

A. Monte Carlo molecular simulation

The essence of MC molecular simulation lies in predicting macroscopic properties of fluids by simulating

the interactions among a large group of molecules (statistical ensemble) representing the fluid. Each statistical ensemble is characterized by the different physical constraints imposed in order to capture the thermo-physical conditions at which these fluids exist. Fortunately, nowadays a wide range of ensembles are available to be used in MC molecular simulations. The decision of choosing the suitable ensemble to work with mainly relies on the simulation output needed. For instance, in this work, pressures under supercritical conditions and two-phase saturated densities were needed to search for adequate LJ parameters for different molecules. Therefore, the canonical (NVT) ensemble was used to predict single-phase pressures while the NVT-Gibbs ensemble was adopted to compute the two-phase saturated densities.

In the canonical ensemble, a certain number of molecules (N) is trapped in a fixed hypothetical volume (V), leading to a constant system density ($\rho = N/V$) throughout the simulation. In addition, the system's temperature (T) is fixed as well. As a result, the fluid's pressure (P), under the prescribed ρ and T , can be predicted. The NVT-Gibbs ensemble^{49–52} is used to study phase equilibria between two phases. Each of these phases is initially represented by a separate simulation box, each containing a certain number of molecules. Similar to the canonical ensemble, the temperature is kept fixed throughout the simulation. In order to reach equilibrium between the two phases, molecules are randomly transferred between the two boxes (chemical equilibrium) while the individual volumes of each box are allowed to change (mechanical equilibrium) such that the total volume is conserved. Finally thermal equilibrium is reached by performing translational and configurational MC trials on individual molecules. For more details about the Gibbs ensemble and the conditions used to decide on accepting new configurations, see Ref. 6.

The statistical nature of the MC molecular simulation arises from the random MC trials performed on the simulated ensembles. MC trials (e.g., molecule displacement, molecule transfer between simulation boxes, and box volume change) introduce elemental changes to the system to create a large number of different molecular configurations. These configurations are sampled to construct the so-called MC Markov chain. The decision of accepting these elemental changes or rejecting them while creating the new configurations is related to the interactions among the simulated molecules. Thus, a key component of any MC molecular simulation is the potential model describing molecular interactions.

In this paper, the LJ model was used to represent the molecular interactions, in specific, the van der Waals long-range attractive force, and the repulsive force resulting from the Pauli exclusion principle that prevents the collapse of molecules. The LJ model relates the energy between any pair of particles to their separating distance as follows:

$$u_{ij} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (1)$$

where u_{ij} stands for potential energy due to dispersion-repulsion interaction between particle i and particle j with a separating distance r_{ij} . The LJ parameters ϵ_{ij} and σ_{ij} are the

potential well depth and the “hard sphere” radius, respectively. This work focuses only on pure component single-phase and two-phase systems, thus all particles are identical and the LJ parameters can be reduced to ε and σ . For convenience, both u_{ij} and r_{ij} are normalized by ε and σ such that $u_{ij}^* = u_{ij}/\varepsilon$ and $r_{ij}^* = r_{ij}/\sigma$, leading to the dimensionless LJ model form

$$u_{ij}^* = 4 \left[\left(\frac{1}{r_{ij}^*} \right)^{12} - \left(\frac{1}{r_{ij}^*} \right)^6 \right]. \quad (2)$$

B. The polynomial chaos framework

The general framework of PC technique works by treating input variable uncertainties, parameters, and model state as stochastic variables. Hence, the system state can be represented by a random vector, $\mathbf{x}(t)$, and its time evolution is expressed through the following differential equation:

$$\dot{\mathbf{x}}(t, \boldsymbol{\Theta}) = \mathbf{f}(t, \boldsymbol{\Theta}, \mathbf{x}, \mathbf{u}), \quad \mathbf{x}(t_0) = \mathbf{x}_0. \quad (3)$$

In the equation above, $\boldsymbol{\Theta}$ represents the uncertain parameter vector and $\mathbf{u}$ the assumed deterministic forcing terms. The initial state estimate, $\mathbf{x}_0$, may also be uncertain but here it is assumed known. The polynomial chaos method enables to compute the full pdf characterizing the space-time evolution of x_k , allowing to quantify the uncertainty of the system outputs with respect to the parameters.

Recently, new uncertainty propagation methods have been developed based on the polynomial chaos theory. Generalized polynomial chaos (gPC) is an extension of the homogeneous chaos idea of Wiener.⁵³ The idea is based on separating the random variables from deterministic ones while solving the stochastic differential equation. Then the random variables are expanded using a suitable polynomial expansion.

The gPC approach assumes that the model outputs $\mathbf{X} \in L^2(\Omega^*)$ to admit a spectral expansion of the form

$$\mathbf{X} = \sum_{k=0}^{\infty} c_k \Psi_k \doteq \sum_{k=0}^P c_k \Psi_k(\boldsymbol{\xi}), \quad (4)$$

where $\xi_i \stackrel{iid}{\sim} \mathcal{U}(-1, 1)$ are canonical random variables used to parameterize random inputs, Ψ_k is the d -variate Legendre polynomial, and c_k represents the PC coefficients. The order of truncation P_{tr} depends on the stochastic dimension d and expansion order, p , as

$$P_{tr} = \frac{(d+p)!}{d!p!} - 1. \quad (5)$$

Since $\{\Psi_k\}_0^{P_{tr}}$ forms an orthogonal system, we can write the following dot product:

$$\langle X, \Psi_k \rangle = c_k \langle \Psi_k, \Psi_k \rangle, \quad (6)$$

with

$$c_k = \frac{\langle X, \Psi_k \rangle}{\langle \Psi_k^2 \rangle}. \quad (7)$$

In (7), the moments $\langle \Psi_k^2 \rangle$ of the multivariate Legendre polynomials can be analytically evaluated³⁶ while $\langle X, \Psi_k \rangle$ requires more elaborate computation, as we note that

$$\langle X, \Psi_k \rangle = \int_{\Omega^*} X(\mathbf{s}) \Psi_k(\mathbf{s}) dF_{\boldsymbol{\xi}}(\mathbf{s}), \quad k = 0, \dots, P_{tr}.$$

Hence, evaluating c_k involves computing the values of a set of $P_{tr} + 1$ integrals over $\Omega^* \subseteq \mathbb{R}^d$, which can be discretized as finite sums using an appropriate quadrature formula,

$$\int_{\Omega^*} X(\mathbf{s}) \Psi_k(\mathbf{s}) dF_{\boldsymbol{\xi}}(\mathbf{s}) \doteq \sum_{j=1}^{N_q} w_j X(\boldsymbol{\xi}_j) \Psi_k(\boldsymbol{\xi}_j). \quad (8)$$

In the equation above, $\boldsymbol{\xi}_j \in \Omega^*$ and w_j are the nodes and weights of the quadrature rule. In this approach, the main computational burden essentially amounts to the deterministic evaluation of the model over the sample set of quadrature nodes,

$$\mathcal{S} \equiv \{\boldsymbol{\xi}_j\}_{j=1}^{N_q} \subset \Omega^*. \quad (9)$$

Let $\boldsymbol{\Pi} \in \mathbb{R}^{(P_{tr}+1) \times N_q}$ be the NISP matrix,

$$\boldsymbol{\Pi}_{k,j} = \frac{w_j \Psi_k(\boldsymbol{\xi}_j)}{\langle \Psi_k^2 \rangle}, \quad k = 0, \dots, P_{tr}, \quad j = 1, \dots, N_q$$

also let $\boldsymbol{\zeta}$ be the vector with coordinates $\zeta_j = X(\boldsymbol{\xi}_j)$. Then the vector $\mathbf{c}$ can be expressed as $\boldsymbol{\Pi} \boldsymbol{\zeta}$, or in component form,

$$c_k = \sum_{j=1}^{N_q} \boldsymbol{\Pi}_{kj} \zeta_j = \sum_{j=1}^{N_q} \boldsymbol{\Pi}_{kj} X(\boldsymbol{\xi}_j), \quad k = 0, \dots, P_{tr}. \quad (10)$$

With a fully tensorized quadrature rule, the complexity of NISP scales with $N_q = n^d$, the number of nodes n , and number of uncertain parameters d . Hence, this approach is computationally limited to a moderate number of parameters. For detailed mathematical formulation of the NISP method, the reader is referred to Ref. 34.

III. RESULTS AND DISCUSSION

A. Surrogate model construction

The construction of the surrogate models through PC expansions, which are then used to produce the MC simulation results, was conducted as follows. At first, let $\mathbf{m} = (m_1, m_2, \dots, m_n)^T$ be the vector of random model inputs

FIG. 1. Normalized LJ fluid isotherms at supercritical thermodynamic conditions ($0.1 \leq \beta^* \leq 0.6$) generated by MC canonical ensemble simulations (symbols) in comparison with the results obtained by the pre-constructed single-phase PC surrogate model (lines). For this case, a polynomial of degree 6 was used with 64 total quadrature nodes.

FIG. 2. Normalized LJ fluid two-phase envelope at subcritical thermodynamic conditions ($0.8 \leq T^* \leq 1.2$) generated by NVT-Gibbs ensemble simulations (circles) in comparison with the results obtained by the pre-constructed two-phase PC surrogate model (lines). For this case, a polynomial of degree 6 and 20 quadrature nodes was used.

having uniform distribution. Specifically, the inputs m_i were parameterized by $\xi_i \sim \mathcal{U}(-1, 1)$, $i = 1, 2, \dots, n$ through

$$m_i(\xi) = \mu_i + \sigma_i \xi_i, \quad i = 1, 2, \dots, n, \quad (11)$$

where $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$, $\mu_i = (1/2)(a_i + b_i)$, and $\sigma_i = (1/2)(b_i - a_i)$, such that $m_i \sim \mathcal{U}(a_i, b_i)$. For a given vector of random inputs $\mathbf{m}(\xi)$ the output is denoted by $\mathbf{X}(\xi) = \mathbf{A}(\mathbf{m}(\xi))$. Note that $\mathbf{A}(\mathbf{m}(\xi))$ is the output of the MC simulation solved with the set of the input parameters $\mathbf{m}(\xi)$. $\mathbf{X}$ may correspond to any subset of the QoIs. In the single-phase case, $\mathbf{X}$ corresponds to the normalized equilibrium pressure (P^*), while in the two-phase scenario it corresponds to the normalized liquid- (ρ_{liq}^*) and vapor-phase saturation density (ρ_{vap}^*). The stochastic variables in the single-phase simulation are both the normalized Boltzmann's factor (β^*) and system number density (ρ^*). In the two-phase simulation, β^* is the only stochastic variable.

For this study, two independent surrogate models were generated to replace single-phase (canonical ensemble) and two-phase (NVT-Gibbs ensemble) MC simulations, simultaneously. MC simulations in canonical ensemble were run using 216 particles with 1×10^6 equilibrium MC steps, followed by 5×10^6 production MC steps. On the other hand, a total of 686 particles (343 particles uniformly distributed in each box at the initial configuration) were simulated in NVT-Gibbs ensemble with 10×10^6 and 20×10^6 equilibrium and production steps, respectively. In both ensembles, a cutoff radius $r_c = 0.49L$ was imposed while energy and pressure tail correction functions were introduced to account for the truncated interactions beyond r_c . Regarding the computational cost, each MC canonical ensemble run lasts for around 20 min using an Intel Xeon processor on a Dell workstation. On the same machine, a single NVT-Gibbs ensemble simulation requires around 150 min of computational time.

In the single-phase scenario, the Gaussian quadrature formula was employed to build the NISP matrix. A total of only 64 quadrature points (8 in each input parameters direction β^* and ρ^*) was used in order to construct the surrogate model that can reproduce the MC canonical ensemble results over a wide range of simulation conditions. According to this, the time needed to construct the surrogate model is 64 times the time needed by a single forward run of the canonical ensemble. However, once the surrogate model is constructed, any output can be evaluated in few seconds by running the surrogate model itself. Similarly, only 20 quadrature points were needed to generate the surrogate model that can produce the NVT-Gibbs ensemble results within seconds.

According to Fig. 1, the results obtained by the single-phase PC surrogate model are in excellent agreement with the P^* values directly calculated from the MC canonical ensemble simulations. In fact, this surrogate model can reproduce the MC canonical results within $\pm 3\%$ error along the whole range of $0.1 \leq \beta^* \leq 0.6$ and $0.1 \leq \rho^* \leq 0.8$. Similarly, the

FIG. 3. (a) Argon's experimental pressure values along supercritical isotherms ranging from 300 K to 700 K (symbols) fitted by the PC surrogate model of the MC canonical ensemble simulations (lines). (b) Argon's two-phase envelope (circles and diamonds represent experimental saturated vapor and liquid densities, respectively) fitted by the PC surrogate model of the NVT-Gibbs ensemble simulations (lines). (c) Deviation percentages from the true experimental pressure values along the fitted supercritical isotherms. (d) Deviation percentages from the true experimental saturated density in the two-phase region; dashed lines are added for eye guidance.

FIG. 4. Same as Fig. 3 but for krypton with supercritical isotherms ranging from 350 K to 700 K.

two-phase surrogate model can accurately estimate both ρ_{liq}^* and ρ_{vap}^* generated by NVT-Gibbs ensemble (Fig. 2) in the normalized temperature range of $0.8 \leq T^* \leq 1.1$, with a relative error of $\pm 2\%$.

B. Single-site LJ parameter optimization

As a direct application to the coupling presented in this article between PC and MC methods, a large-scale optimization for single-site LJ parameters was conducted for several small molecules. The recommended parameters (ϵ_{rec} , σ_{rec}) were obtained by minimizing the difference between various predictions from the surrogate against their corresponding experimental measurements. All the experimental data used in this paper were retrieved from NIST database.⁵⁴

For the starting point of the optimization, a physically valid guess is needed for both ϵ and σ . For this purpose, formulas obtained by fitting virial coefficient data were used.⁵⁵ These equations are dependent on the critical properties of the component itself,

$$\epsilon_0 = \frac{T_c k_B}{1.35} \text{ and } \sigma_0^3 = \frac{0.35M}{\rho_c N_A}, \quad (12)$$

where k_B is the Boltzmann's constant, N_A is the Avogadro's number, and M is the molecular weight, whereas T_c and ρ_c are the critical temperature and mass density, respectively.

The optimization process was performed using the MATLAB built-in function "fmincon," starting with the initial guesses (ϵ_0, σ_0) from Eq. (12). During the fitting process, the tuning parameters were subjected to certain lower and upper limits such that they can only attain values satisfying these

FIG. 5. Same as Fig. 3 but for xenon with supercritical isotherms ranging from 450 K to 750 K.

FIG. 6. Same as Fig. 3 but for methane with supercritical isotherms ranging from 300 K to 600 K.

constraints. Typical ranges used were ± 10 K for ε/k_B and ± 0.25 Å for σ . These constraints were forced for two main reasons. First, to make sure that the optimized parameters still agree with the theoretically expected quantities. Second, to stay within the limits of the constructed PC functions.

The optimization scheme aims to minimize the objective function defined as the difference between the surrogate model predictions and corresponding experimental data. The data used for matching were the single-phase vapor pressures along several supercritical isotherms ($T > T_c$ and $P > P_c$), in addition to the two-phase (liquid and vapor) saturation densities at different temperatures. Therefore, the objective function (J) is formulated as the discrete norm over the experimental data used,

$$J = \left\| \frac{P^{\text{sur}} - P^{\text{exp}}}{P^{\text{exp}}} \right\|_2 + \left\| \frac{\rho_{\text{vap}}^{\text{sur}} - \rho_{\text{vap}}^{\text{exp}}}{\rho_{\text{vap}}^{\text{exp}}} \right\|_2 + \left\| \frac{\rho_{\text{liq}}^{\text{sur}} - \rho_{\text{liq}}^{\text{exp}}}{\rho_{\text{liq}}^{\text{exp}}} \right\|_2. \quad (13)$$

Based on the described workflow above, LJ parameters for several molecules were optimized such that a good fit with experimental data was reached. At first, LJ parameters for three noble gases, namely Ar, Kr, and Xe, were obtained. In Figs. 3(a) and 3(b), experimental data used for tuning Ar parameters were plotted against the simulated data generated by the recommended values ($\varepsilon_{\text{rec}}/k_B$ and σ_{rec}). In addition, the relative misfits (in %) between each of these points in each phase were respectively reported (see Figs. 3(c) and 3(d)). According to these results, the MC simulation outputs for

FIG. 7. Same as Fig. 3 but for nitrogen with supercritical isotherms ranging from 400 K to 1100 K.

FIG. 8. Same as Fig. 3 but for carbon monoxide with supercritical isotherms ranging from 200 K to 500 K.

FIG. 9. (a) Supercritical isotherms of carbon dioxide data (symbols) ranging from $T = 600$ K to $T = 950$ K fitted by the single-phase PC surrogate model (lines). (b) Large deviation in two-phase region between experimental data (symbols) in comparison with the PC model fit (lines). Contrary to the previous molecules, representing CO₂ molecule by merely an LJ sphere is not sound.

Ar can be reproduced in few seconds using the constructed PC surrogates within $\pm 5\%$ deviation for the majority of the investigated data points. It is also evident that at supercritical conditions, larger misfits were found at lower pressures; in other words, at temperatures closer to the critical temperature. In a similar fashion, higher deviations were observed while estimating the gas phase densities in the two-phase envelope. Such behavior can be attributed to the misfits inherited by the PC surrogates from the high statistical uncertainties of the MC simulations in the vapor phase.

Similar trends, but generally with higher misfits, were reached for both Kr and Xe (Figs. 4 and 5, respectively) when compared to Ar case. Furthermore, experimental data

of bigger but rather simple molecules were matched too. For instance, results from PC surrogates showed excellent agreement with CH₄ data (Fig. 6). The resulting deviation was less than $\pm 5\%$ for all data points studied, except for few vapor phase densities.

In addition to CH₄, both CO and N₂ were modeled as single-site LJ particles. Interestingly, the data for both molecules were reproduced with the recommended parameters in rather good agreement (Figs. 7 and 8), despite the simplicity of the representing molecular model. Clearly, the match between experimental data and simulation results for CO and N₂, especially in the two-phase region, was not as good as the match obtained for any of the noble gases considered or CH₄.

FIG. 10. (a) Supercritical isotherms of ethane data (symbols) ranging from 450 to 625 K fitted by the single-phase PC surrogate model (lines). (b) Large deviations in two-phase region similar to CO₂ case.

TABLE I. Theoretical and recommended single-site LJ model parameters.

Molecules	ϵ_0/k_B (K)	$\epsilon_{\text{rec}}/k_B$ (K)	σ_0 (Å)	σ_{rec} (Å)
Ar	111.6200	116.4030	3.5128	3.3911
Kr	155.1704	162.2714	3.7696	3.6264
Xe	214.6170	224.6165	4.1089	3.9484
CH ₄	141.1585	147.7146	3.8555	3.7309
CO	98.4148	91.6006	3.7696	3.5876
N ₂	93.4756	98.3203	3.7317	3.6272

A further, more challenging, step was taken trying to fit experimental data for CO₂ and C₂H₆. However, and as expected, the spherical LJ model failed to capture the complex nature of such molecules (Figs. 9 and 10). This misfit was dramatically manifested in the two-phase region. The reason for such failure is due to the limitation of the single-site model itself, as both CO₂ and C₂H₆ are linear molecules instead of spherical. Moreover, CO₂ partial charges, which play a vital role in determining its thermodynamic behavior, were not considered by the LJ model used.

To sum up, the large-scale optimization process for the LJ parameters was successfully demonstrated for several noble gases, in addition to CH₄, CO, and N₂ molecules. On the other side, attempting to use the spherical non-polar LJ model to represent more complex molecules has failed. Therefore, the recommended LJ parameters for only the well-fitted molecules were reported (Table I).

IV. CONCLUSION

A novel approach combining MC molecular simulation with the PC expansions method was presented. This combination is particularly successful because MC simulations require very long computational time while having few input variables, which suits well the PC expansion formulation. The resulting surrogate models built via PC expansion functions allowed extremely fast and sufficiently accurate reproduction for both single- and two-phase data of LJ fluids. In addition, an application example for the combination of these two methods was demonstrated through conducting a large-scale optimization of LJ model parameters for few molecules. A summary of these recommended parameters is available in Table I. Future work may consider sophisticated potential models that can fairly represent structural and polar molecules. Moreover, MC simulation results through the surrogate models are to be used to feed thermodynamic information needed by flow simulators in porous media.

ACKNOWLEDGMENTS

This research is funded by King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia.

¹A. Firoozabadi, *Thermodynamics of Hydrocarbon Reservoirs* (McGraw Hill Professional, USA, 1999).

²A. Firoozabadi, *Thermodynamics and Applications of Hydrocarbons Energy Production* (McGraw Hill Professional, USA, 2015).

³M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, USA, 1989).

⁴R. L. Rowley, *Statistical Mechanics for Thermophysical Property Calculations* (Prentice Hall, USA, 1994).

⁵D. A. McQuarrie, *Statistical Mechanics* (University Science Books, California, USA, 2000).

⁶D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, San Diego, CA, 2001).

⁷P. Ungerer, V. Lachet, and B. Tavittian, "Applications of molecular simulation in oil and gas production and processing," *Oil Gas Sci. Technol.* **61**, 387–403 (2006).

⁸P. Ungerer, C. Nieto-Draghi, V. Lachet, A. Wender, A. Lelladi, A. Boutin, B. Rousseau, and A. H. Fuchs, "Molecular simulation applied to fluid properties in the oil and gas industry," *Mol. Simul.* **33**, 287–304 (2007).

⁹G. M. Torrie and J. P. Valleau, "Monte Carlo study of a phase-separating liquid mixture by umbrella sampling," *J. Chem. Phys.* **66**, 1402–1408 (1977).

¹⁰G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *J. Comput. Phys.* **23**, 187–199 (1977).

¹¹J. P. Valleau, "The Coulombic phase transition: Density-scaling Monte Carlo," *J. Chem. Phys.* **95**, 584–589 (1991).

¹²J. P. Valleau, "Density-scaling: A new Monte Carlo technique in statistical mechanics," *J. Comput. Phys.* **96**, 193–216 (1991).

¹³J. P. Valleau, "Density-scaling Monte Carlo study of subcritical Lennard-Jonesium," *J. Chem. Phys.* **99**, 4718–4728 (1993).

¹⁴K. Kiyohara, "Thermodynamic scaling Gibbs ensemble Monte Carlo: A new method for determination of phase coexistence properties of fluids," *Mol. Phys.* **89**, 965–974 (1996).

¹⁵J. P. Valleau, "Temperature-and-density-scaling Monte Carlo: Methodology and the canonical thermodynamics of Lennard-Jonesium," *Mol. Simul.* **31**, 223–253 (2005).

¹⁶J. P. Valleau, "Temperature-and-density-scaling Monte Carlo: Isothermal-isobaric thermodynamics of Lennard-Jonesium," *Mol. Simul.* **31**, 255–275 (2005).

¹⁷J. S. Wang and R. Swendsen, "Non-universal critical dynamics in Monte Carlo simulations," *Phys. Rev. Lett.* **58**, 86–88 (1987).

¹⁸A. M. Ferrenberg and R. H. Swendsen, "New Monte Carlo technique for studying phase transitions," *Phys. Rev. Lett.* **61**, 2635–2638 (1988).

¹⁹A. Ferrenberg and R. Swendsen, "New Monte Carlo technique for studying phase transitions," *Phys. Rev. Lett.* **63**, 1658 (1989).

²⁰A. Lyubartsev, A. Martsinovski, S. Shevkunov, and P. Vorontsov-Velyaminov, "New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles," *J. Chem. Phys.* **96**, 1776–1785 (1992).

²¹F. A. Escobedo and J. J. Pablo, "Expanded grand canonical and Gibbs ensemble Monte Carlo simulation of polymers," *J. Chem. Phys.* **105**, 4391–4394 (1996).

²²E. Marinari and G. Parisi, "Simulated tempering: A new Monte Carlo scheme," *Europhys. Lett.* **19**, 451–458 (1992).

²³C. J. Geyer and E. A. Thompson, "Annealing Markov chain Monte Carlo with applications to ancestral inference," *J. Am. Stat. Assoc.* **90**, 909–920 (1995).

²⁴D. Frenkel, "Speed-up of Monte Carlo simulations by sampling of rejected states," *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17571–17575 (2004).

²⁵A. Kadoura, A. Salama, and S. Sun, "A conservative and a hybrid early rejection schemes for accelerating Monte Carlo molecular simulation," *Mol. Phys.* **112**, 2575–2586 (2014).

²⁶A. Kadoura, A. Salama, and S. Sun, "Speeding up Monte Carlo molecular simulation by a non-conservative early rejection scheme," *Mol. Simul.* **42**, 241 (2015).

²⁷S. Sun, A. Kadoura, and A. Salama, "An efficient method of reweighting and reconstructing Monte Carlo molecular simulation data for extrapolation to different temperature and density conditions," in *13th International Conference on Computational Science, Procedia Computer Science, Barcelona, Spain, 5–7 June 2013* (Elsevier, 2013), pp. 2147–2156.

²⁸A. Kadoura, S. Sun, and A. Salama, "Accelerating Monte Carlo molecular simulations by reweighting and reconstructing Markov chains: Extrapolation of canonical ensemble averages and second derivatives to different temperature and density conditions," *J. Comput. Phys.* **270**, 70–85 (2014).

²⁹A. Kadoura, A. Salama, and S. Sun, "Switching between the NVT and NpT ensembles using the reweighting and reconstruction scheme," in *15th International Conference on Computational Science, Procedia Computer Science, Reykjavik, Iceland, 1–3 June 2015* (Elsevier, 2015), pp. 1259–1268.

- ³⁰R. Ghanem and J. Red-Horse, "Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach," *Physica D: Nonlinear Phenomena*, **133**, 137–144 (1999).
- ³¹O. Knio and O. Maitre, "Uncertainty propagation in CFD using polynomial chaos decomposition," *Fluid Dyn. Res.*, **38**, 616–640 (2006).
- ³²I. Sraj, M. Iskandarani, A. Srinivasan, W. C. Thacker, J. Winokur, A. Alexanderian, C. Y. Lee, S. S. Chen, and O. M. Knio, "Bayesian inference of drag parameters using AXBT data from Typhoon Fanapi," *Mon. Weather Rev.*, **141**, 2347–2367 (2013).
- ³³I. Sraj, K. T. Mandli, O. M. Knio, C. N. Dawson, and I. Hoteit, "Uncertainty quantification and inference of Mannings friction coefficients using DART buoy data during the Tōhoku tsunami," *Ocean Modell.*, **83**, 82–97 (2014).
- ³⁴A. Alexanderian, J. Winokur, I. Sraj, A. Srinivasan, M. Iskandarani, W. C. Thacker, and O. M. Knio, "Global sensitivity analysis in an ocean general circulation model: A sparse spectral projection approach," *Comput. Geosci.*, **16**, 757–778 (2012).
- ³⁵R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach* (Springer, New York, 1991).
- ³⁶O. L. Maitre and O. Knio, *Spectral Methods for Uncertainty Quantification* (Springer, New York, 2010).
- ³⁷O. L. Maitre, O. Knio, H. Najm, and R. Ghanem, "A stochastic projection method for fluid flow. I. Basic formulation," *J. Comput. Phys.*, **173**, 481–511 (2001).
- ³⁸O. L. Maitre, H. N. Najm, P. Pebay, R. Ghanem, and O. M. Knio, "Multi-resolution-analysis scheme for uncertainty quantification in chemical systems," *SIAM J. Sci. Comput.*, **29**, 864–889 (2007).
- ³⁹H. N. Najm, B. Debusschere, Y. Marzouk, S. Widmer, and O. L. Maitre, "Multi-resolution-analysis scheme for uncertainty quantification in chemical systems," *Int. J. Numer. Methods Eng.*, **80**, 789–814 (2009).
- ⁴⁰B. D. Phenix, J. L. Dinaro, M. A. Tatang, J. W. Tester, J. B. Howard, and G. J. McRae, "Incorporation of parametric uncertainty into complex kinetic mechanisms: Application to hydrogen oxidation in supercritical water," *Combust. Flame*, **112**, 132–146 (1998).
- ⁴¹M. T. Reagan, H. N. Najm, P. P. Pbay, O. M. Knio, and R. Ghanem, "Quantifying uncertainty in chemical systems modeling," *Int. J. Chem. Kinet.*, **37**, 368–382 (2005).
- ⁴²F. Rizzi, H. N. Najm, B. J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O. M. Knio, "Uncertainty quantification in MD simulations. Part I: Forward propagation," *Multiscale Model. Simul.*, **10**, 1428–1459 (2012).
- ⁴³F. Rizzi, H. N. Najm, B. J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O. M. Knio, "Uncertainty quantification in MD simulations. Part II: Bayesian inference of force-field parameters," *Multiscale Model. Simul.*, **10**, 1460–1492 (2012).
- ⁴⁴F. Rizzi, R. Jones, B. J. Debusschere, and O. M. Knio, "Uncertainty quantification in MD simulations of concentration driven ionic flow through a silica nanopore. Part I: Sensitivity to physical parameters of the pore," *J. Chem. Phys.*, **138**, 194104 (2013).
- ⁴⁵F. Rizzi, R. Jones, B. J. Debusschere, and O. M. Knio, "Uncertainty Quantification in MD simulations of concentration driven ionic flow through a silica nanopore. Part II: Uncertain potential parameters," *J. Chem. Phys.*, **138**, 194105 (2013).
- ⁴⁶P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos, "Data driven, predictive molecular dynamics for nanoscale flow simulations under uncertainty," *J. Phys. Chem. B*, **117**, 14808–14816 (2013).
- ⁴⁷M. Salloum, K. Sargsyan, R. Jones, H. N. Najm, and B. Debusschere, "Quantifying sampling noise and parametric uncertainty in atomistic-to-continuum simulations using surrogate models," *Multiscale Model. Simul.*, **13**, 953–976 (2015).
- ⁴⁸K. Sargsyan, B. Debusschere, H. N. Najm, and Y. Marzouk, "Bayesian inference of spectral expansions for predictability assessment in stochastic reaction networks," *J. Comput. Theor. Nanosci.*, **6**, 2283–2297 (2009).
- ⁴⁹A. Z. Panagiotopoulos, "Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble," *Mol. Phys.*, **61**, 813–826 (1987).
- ⁵⁰A. Panagiotopoulos, N. Quirke, M. Stapleton, and D. Tildesley, "Phase equilibria by simulation in the Gibbs ensemble: Alternative derivation, generalization and application to mixture and membrane equilibria," *Mol. Phys.*, **63**, 527–545 (1988).
- ⁵¹B. Smit, P. Smedt, and D. Frenkel, "Computer simulations in the Gibbs ensemble," *Mol. Phys.*, **68**, 931–950 (1989).
- ⁵²A. Z. Panagiotopoulos, "Direct determination of fluid phase equilibria by simulation in the Gibbs ensemble: A review," *Mol. Simul.*, **9**, 1–23 (1992).
- ⁵³N. Wiener, "The homogeneous chaos," *Am. J. Math.*, **60**, 897–936 (1938).
- ⁵⁴E. W. Lemmon, M. O. McLinden, and D. G. Friend, "Thermophysical Properties of Fluid Systems" in *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, edited by P. J. Linstrom and W. G. Mallard (National Institute of Standards and Technology, Gaithersburg MD, 2015), <http://webbook.nist.gov>.
- ⁵⁵S. I. Sandler, *An Introduction to Applied Statistical Thermodynamics* (Wiley, USA, 2010).