Bayesian inference analysis of transferable, united-atom, Mie λ-6 force fields for normal and branched alkanes. To be submitted to the Journal of Physical Chemistry, B.

Richard A. Messerly[1, a)]

Thermodynamics Research Center, National Institute of Standards and Technology, Boulder, Colorado, 80305

Keywords: Transferability, Force Field, Molecular Dynamics, Molecular Simulation

---

[a)]Electronic mail: richard.messerly@nist.gov

PURPOSE

The aim of this study is to demonstrate, using Bayesian inference, that a UA Mie force field cannot adequately predict VLE and compressed liquid pressures for normal and branched alkanes. For adequate prediction of VLE and compressed liquid pressures, we recommend using AUA models or perhaps Exp-6 or extended Lennard-Jones, i.e. 12-10-8-6.

OUTLINE

1. Introduction

2. Methods I

   (a) Simulation details

   (b) Force field

3. Case Study for alkanes

4. Methods II

   (a) Bayesian Analysis

   (b) Surrogate Model

   (c) Propagation of Uncertainty

5. Results

   (a) VLE and Compressed

      i. Parameter uncertainties

      ii. Propagation of uncertainties

6. Future Work

7. Conclusions

DETAILED OUTLINE

I. INTRODUCTION

1. Developing reliable fundamental equations of state (REFPROP) is an arduous task that relies on having high accuracy data over a wide range of state points

2. Reliable predictions for high pressure systems are important for many industrial applications

3. Recently, molecular simulation results have supplemented experimental data when developing fundamental equations of state

4. In addition, the 10th Industrial Fluid Properties Simulation Challenge is to predict the viscosities at high pressures of a highly branched alkane. Reliable PVT predictions are essential for this challenge.

5. The UA Mie potential has received significant attention for its ability to predict VLE without requiring an all-atom representation

6. However, the impact that modifying the repulsive exponent has on the higher pressure states has not been tested

7. The purpose of this study is to perform a rigorous Bayesian analysis as to the adequacy of a UA Mie potential for predicting both VLE and compressed liquid/supercritical pressures

II. METHODS I

A. Simulation Details

1. The compounds simulated in this study are ethane, propane, n-butane, n-octane, isobutane, isopropane, isohexane, isooctane, and neopentane

2. These compounds were selected as a sample set for 2,2,4-trimethylhexane, the compound studied in the 10th Industrial Fluid Properties Simulation Challenge

3. We perform NVT simulations in GROMACS along a supercritical isotherm and five isochores that correspond to saturated liquid densities

4. We use ITIC to convert the Udep and Z values obtained from the NVT simulations to rholsat and Pvsat

5. The specific state points are provided in supporting information

6. The GROMACS settings are provided in supporting information

B. Force field

1. For each compound studied, we use a united-atom representation as defined by the TraPPE (and Mie) force fields. For normal and branched alkanes this consists of CH3, CH2, CH, and C sites.

2. We use fixed bond-lengths, harmonic angles, cosine series for torsions, and exclude 1-4 nonbonded interactions

3. Bonded parameters are provided in supporting information

4. Nonbonded interactions for intermolecular and intramolecular sites separated by more than 3 bonds are represented using a Mie potential, which can be viewed as a generalized Lennard-Jones

5. (Maybe include Exp-6 and extended Lennard-Jones, i.e. 12-10-8-6)

III. CASE STUDY FOR ALKANES

1. Several force fields in the literature have been optimized to agree with VLE properties (TraPPE, Potoff, TraPPE-2, TAMie, Errington, AUA4)

2. TraPPE, Potoff, and AUA4 have parameters for each compound studied, while TraPPE-2 only has parameters for ethane, TAMie has parameters for all except isooctane and neopentane (containing a C group), and ErrExp-6 only has parameters for the n-alkanes

3. Figure: Z vs 1000/T for ethane, propane, n-butane, and n-octane

4. Figure: Z vs 1000/T for isobutane, isopropane, isohexane, isooctane and neopentane

5. PVT (Z) trends are inaccurate for both TraPPE and Potoff at high pressures, i.e. non-VLE conditions. Specifically, the TraPPE 12-6 under predicts while 16-6 over predicts

6. Although these results might suggest that a 14-6 potential would work best, recall that the TraPPE 12-6 does not accurately predict Pvsat.

7. Since TraPPE and Potoff use slightly different objective functions we want to perform an equivalent analysis for different values of lambda

8. Hypothesis that we want to test is that there does not exist a set of epsilon, sigma, and lambda that provides reasonable VLE, supercritical fluid, and compressed liquid (for ethane I already know this is not feasible. For n-alkanes I know that the 16-6 cannot accomplish this, but I am not sure about 14-6 or 15-6.)

9. AUA LJ 12-6 is much more accurate for ethane

10. AUA Mie 14-6 potential is not much better than UA Mie 16-6 for n-alkanes

11. AUA Exp-6 force field is not much better than UA Mie 16-6 for n-alkanes

12. No results for AUA4

IV. METHODS II

1. We use Bayesian inference with MCMC to quantify the uncertainty in the force field parameters

2. We perform direct molecular simulation for several reference force field parameter sets

3. We use MBAR-ITIC to predict rholsat and Pvsat with non-simulated force field parameters

4. We use MBAR to propagate the parameter uncertainties from VLE to Z of compressed liquids/supercritical

5. We perform this analysis for CH3 and CH2 sequentially and independently. In other words, the Markov Chain only samples a 2-dimensional space.

A. Bayesian Analysis

1. By quantifying the uncertainty in epsilon and sigma for a given lambda, we can determine if the Mie potential is adequate for reproducing VLE and compressed liquid/supercritical pressures

2. Posterior includes saturated liquid density and vapor pressure

3. Markov Chain Monte Carlo is used to sample

4. Details for MCMC are provided in supporting information (i.e. number of steps for burn-in and production, frequency that step sizes are updated, resulting acceptance percentages, etc.)

5. The parameter uncertainty is propagated when predicting high pressures

B. Surrogate Model

1. A Bayesian analysis is computationally too expensive if direct molecular simulations are performed for every MCMC step

2. As demonstrated in a previous publication, MBAR can reweight the configurations that are sampled from different force fields without direct simulation

3. In other words, a set of reference force fields are simulated for each molecule and MBAR is used instead of direct simulation for each MCMC step

4. As demonstrated in a previous publication, MBAR is used to predict Udep and Z while ITIC is used to convert Udep and Z to rholsat and Pvsat

5. ITIC state points are fit to rectilinear and Antoine equation to interpolate rholsat and Pvsat, this allows for comparison with experimental data at hundreds of temperatures

6. The likelihood includes the experimental uncertainties but, more importantly, the numerical uncertainties. In other words, the numerical uncertainties account for the uncertainties that arise from the simulations themselves, the MBAR reweighting, the ITIC algorithm, and fitting to rectilinear and Antoine.

7. To leave no room for doubt in our conclusions, we use very conservative (and empirical) estimates of numerical uncertainty for rholsat and Pvsat (see Supporting Information)

C. Propagation of Uncertainty

1. From the MCMC parameter sets, we randomly sample a subset of 100-1000 parameter sets

2. We use MBAR to predict Z for each of those parameter sets

3. We plot these results as a histogram to determine the 95% credible interval for Z at each state point

V. RESULTS

A. VLE and Compressed

1. Parameter uncertainties

1. Figure: The uncertainty regions for CH3, CH2, CH, and C. I can include 14-6, 15-6, and 16-6. Perhaps I will only do this rigorous analysis for CH3 or for CH3 and

CH2. Probably not for all. I could include the results from the alternative posterior (excluding Pvsat and including high pressures) but then it might be out of place in this section.

2. Bayes factors demonstrate that, for VLE, a 15-6 or 16-6 potential are favored significantly more than a 14-6 (could include 17-6 or 18-6 as well)

3. CH2 credible regions overlap considerably between propane, n-butane, and n-octane

4. By comparing the Bayes factor of a transferable CH2 site and three independent CH2 sites we observe that the CH2 sites are indistinguishable

5. Statement about CH credible regions for isobutane, isopentane, and isohexane

6. Statement about C credible region for isooctane and neopentane

2. *Propagation of uncertainties*

1. Figure: Uncertainties in rholsat and Pvsat for n-alkanes. Include 14-6, 15-6, 16-6.

2. Figure: Uncertainties in rholsat and Pvsat for branched alkanes. Include only 16-6.

3. Clearly the uncertainties are fairly conservative due to the relatively large numerical uncertainties we assigned in the posterior

4. Figure: Z vs 1000/T for n-alkanes where the error bars represent the Bayesian uncertainties from VLE. Include 14-6, 15-6, and 16-6 results.

5. The 16-6 potential is not able to predict both VLE and compressed liquid/supercritical pressures

6. VLE is much worse for 14-6, about the same for 15-6

7. Condensed liquid pressures are slightly better for 14-6 and 15-6 but still over predict

8. Figure: Z vs 1000/T for branched alkanes. Results are only included for the 16-6 potential.

9. Same results as for normal alkanes.

1. We modify the posterior by excluding the Pvsat data and including the REFPROP correlations at high pressures

2. Figure: I can either include the parameter uncertainties here or back in the Parameter Uncertainties section. I could even move this to supporting information

3. Bayes ratios show the evidence for different values of λ

4. We recommend that lower values of λ be favored

VI. FUTURE WORK

1. As observed in the case study, the AUA approach typically provides more reliable estimates at high pressures.

2. At higher pressures you need the hydrogens, the higher the shift in the bond-length the better.

3. For example, notice that the AUA LJ (TraPPE-2) model is better than the AUA Mie (TAMie) and AUA Exp-6 (ErrExp-6), despite having fewer fitting parameters. This is because they use a much larger bond displacement.

4. An alternative method to AUA is to use an extended Lennard-Jones potential, 12-10-8-6, that has the flexibility of a Mie potential but without the steep barrier

VII. CONCLUSIONS

1. Although the UA Mie potential provides great improvement over the UA LJ 12-6 potential for VLE, it drastically over predicts pressures for supercritical fluids and compressed liquids (actually for supercritical fluids it over predicts at high densities and under predicts at low densities. Not sure I want to open up that can of worms.)

9

2. By performing a rigorous statistical analysis, we verify that no set of ϵ and σ can adequately predict VLE and high pressures (I need a better way to refer to this) for a 16-6, 15-6, or 14-6 potential.