Benchmarking Simulations against the ThermoML Database: Neat Liquid Densities and Static Dielectrics

Kyle A. Beauchamp*,[1] Julie M. Behr*,[1] Patrick B. Grinaway,[1] Bas,[1]
Kenneth Kronlein,[2] Michael R. Shirts,[3,*] and John D. Chodera[1,†]

[1]*Memorial Sloan-Kettering Cancer Center, New York, NY, USA*
[2]*NIST*
[3]*Department of Chemical Engineering, University of Virginia, Charlottesville, VA 22094-0471*

(Dated: November 26, 2014)

I. ABSTRACT

Useful atomistic simulations require accurate depictions of solvent. Simple experimental observables, such as density and static dielectric constants, offer straightforward targets for evaluating forcefield quality. Here we examine the possibilty of benchmarking atomistic models against the NIST ThermoML database of physicochemical measurements, which curates thousands of density, dielectric, and other measurements. We present a detailed benchmark of the GAFF forcefield against measurements extracted from ThermoML and discuss the extent of available data for neat liquids. We show that empirical polarizability models correct systematic biases inherent in predicting dielectric constants with fixed-charged forcefields. Combining our dataset with the Virtual Chemistry benchmark set provides an extensive benchmark suite for liquid properties.

II. INTRODUCTION

Intro

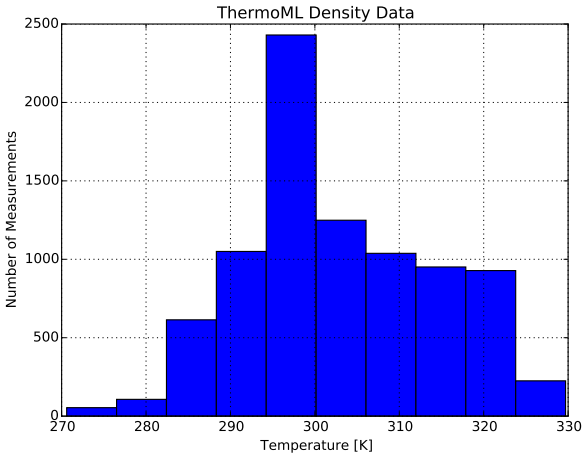
III. RESULTS

A. Neat Liquid Measurements in ThermoML

To assess the feasibility of benchmarking organic molecule forcefields against ThermoML, we performed a number of queries to summarize the data content of ThermoML. Our aim is to explore neat liquid data with functional groups relevant to drug-like molecules. We therefore applied the following series of filters: Druglike elements (H, N, C, O, S, P, F, Cl, Br), Heavy Atom Count (≤ 10), Temperature [K] ($270 \leq T \leq 330$), and Pressure [kPA] ($100 \leq P \leq 102$). After applying these filters, we also assume that all pressures within this range are one atmosphere. We also assume that temperatures can be rounded to one decimal place. These approximations are motived by common data entry errors; for example, an experiment performed at water's freezing point at ambient pressure might be entered as either 101.325 kPA or 100 kPA, with a temperature of either 273 K or 273.15 K.

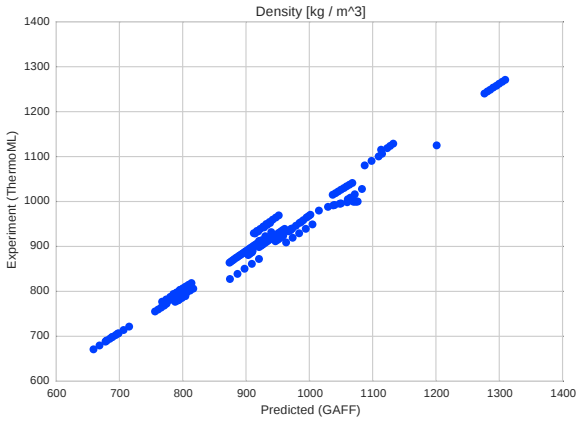
Filter	Mass Density	Static Dielectric
0. Full	308248	4170
1. Druglike Elements	48909	847
2. Heavy Atoms	35465	773
3. Temperature	18374	609
4. Pressure	8247	399
5. Aggregate T, P	3396	382

TABLE I. ThermoML Statistics

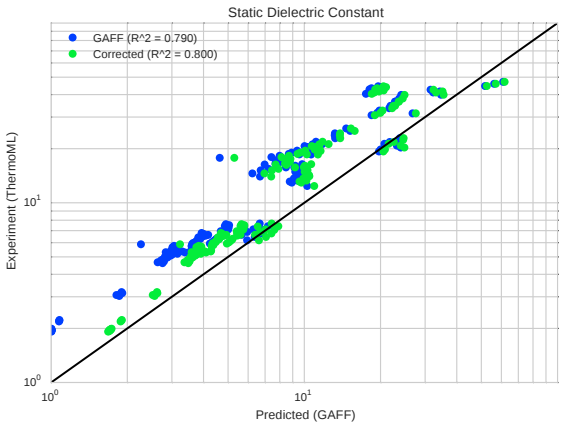


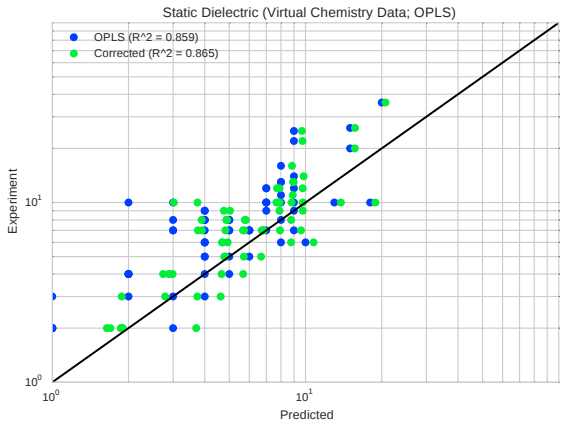
* michael.shirts@virginia.edu
† jchodera@mskcc.org

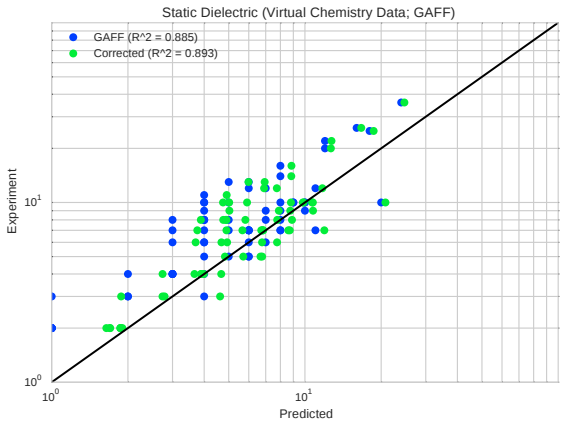
B. Benchmarking GAFF against ThermoML: Mass Density



C. Benchmarking GAFF against ThermoML: Static Dielectric







D. Benchmarking GAFF against ThermoML: Static Dielectric

IV. DISCUSSION

A. Fitting Forcefields to Dielectric Constants

Cite work by Chris Fennell on CCl4. How to get a internally consistent model. Static dielectric constant includes electronic polarization, need to subtract it out. Cite TIP4PEW paper. Could use frequency-dependent dielectric, but empirical polarizability models are highly accurate. Cite LPW CCL4 paper. Use idea of a molecular series CCL0 -¿ CCL1 ... CCL4 as example of internally inconcistent model when directly fitting dielectric constant.

B. ThermoML as a Data Source

Pro: Automated, Curated, Growing, Free, Paper trail
Cons: Requires Parsing, Data Entry Errors