Machine Learning-Enabled Optimization of Force Fields for Hydrofluorocarbons

Bridgette J. Befort[a], Ryan S. DeFever[a], Edward J. Maginn[a], Alexander W. Dowling[a*]

[a]*Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States*
adowling@nd.edu

Abstract

In this work surrogate assisted optimization is utilized to calibrate predictive molecular models, called force fields, used in molecular simulations to reproduce the liquid density of a hydrofluorocarbon refrigerant molecule. A previous calibration workflow which relied on Gaussian process regression models and large Latin hypercube samples to screen force field parameter space is updated to include Bayesian optimization methods to efficiently guide the search for force field parameters. In comparison to the previous work, the Bayesian-based calibration workflow is able to find a parameter set which results in a lower objective function value than the original workflow after evaluating approximately 50% fewer parameter sets. It is envisioned that this updated workflow will facilitate rapid force field optimization enabling screening of vast molecular design space.

Keywords: Bayesian optimization, Gaussian process regression, Molecular simulation

1. Introduction

Molecular simulation is a powerful tool for studying the thermodynamic and dynamic properties of materials. For example, molecular simulation shows great promise for screening vast molecular design spaces which could be expensive or infeasible to probe experimentally. However, to utilize molecular simulation in this capacity requires accurate predictive molecular models, called force fields. Force fields use a functional form and parameters to describe the potential energy of a system and are utilized in classical molecular simulations to model intra- and intermolecular interactions. Developing generalized, or transferable, force fields to describe large swaths of chemical space has historically been a laborious endeavour, often taking months to years to complete. Though these off-the-shelf force fields offer accurate predictions for some systems, they inevitably lack accuracy across the extraordinary range of molecules found in the natural and synthetic world. Further optimization of force field parameters is often necessary to ensure the model has the required accuracy for the molecules and properties of interest [1]. Thus, force field optimization represents a bottleneck to applying molecular simulation to new systems.

Emerging computational frameworks promise to greatly accelerate the calibration of highly accurate, physics-based force fields from experimental data. Efforts to calibrate force fields encompass gradient-based, stochastic search, analytical, and *ad hoc* optimization approaches [2]. Often, a barrier to efficiently calibrating force fields is the expense of calculating the objective function, which quantifies the difference between the simulation prediction and experimental value of a property of interest. The time

requirement of simulations, ranging from minutes-to-hours (*e.g.,* liquid density calculations) to days-to-weeks (*e.g.,* vapor-liquid equilibrium), often makes the objective function calculation cost prohibitive. This cost only increases as more objectives, state points (*e.g.*, temperatures, pressures), and parameters are incorporated in the optimization procedure. Recently, machine learning (ML) methods have been harnessed to address this challenge by mapping microscopic coordinates to a microscopic potential. This functionality enables ML force fields [3], but their black box nature often prevents physical insights that can be gained from a physics-based functional form and parameters of a traditional force field.

ML techniques also facilitate surrogate assisted optimization. Recently, we developed Gaussian process regression (GPR) and support vector machine models to emulate molecular simulations to calibrate force fields [2], hereafter referred to as the JCIM workflow. Our surrogate assisted optimization JCIM workflow successfully screened millions of potential parameter sets, generated through iterative batches of space filling Latin hypercube sampling (LHS). Compared to a force field calibrated via hand-tuning, our workflow enabled the screening of $\mathcal{O}(10^3)$ times more parameter sets while requiring five times fewer simulations. However, this semi-automated workflow required generating large parameter set samples each iteration as well as user input to select which parameter sets to consider in the next iteration (*i.e.*, batch of molecular simulations).

In this work, we explore automating our ML-enabled force field calibration framework by leveraging Bayesian optimization (BO) [4] to intelligently propose new parameter sets. Instead of relying on large $\mathcal{O}(10^5 - 10^6)$ LHS batches of parameter sets and user-driven decisions to screen parameter space, BO automatically balances the search for optimal parameter sets between regions which improve the molecular simulation agreement with experimental data (exploitation) and regions which, if sampled, will reduce the uncertainty of the ML model (exploration). As a demonstration case, we optimize force field parameters for a hydrofluorocarbon (HFC) refrigerant molecule, difluoromethane (HFC-32). HFCs are a motivating application because they are subject to recent mandates which require the phaseout of high global warming potential (GWP) materials [5]. Molecular simulation can aid in the sustainable implementation of this phaseout, but this is contingent upon highly accurate force field models for HFCs. Here, we consider optimization of a force field for HFC-32 which is widely used in many refrigerants and will play a significant role in the transition to next-generation refrigerants due to its low GWP relative to other commonly used HFCs.

2. Methods

2.1. Force Field Model

This work calibrates a classical molecular mechanics force field for HFC-32 with the functional form:

$$U(\boldsymbol{r}) = \sum_{\text{bonds}} k_r(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} v_n[1 + \cos(\eta\phi - \gamma)] + \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_i \sum_{j>i} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

Here, U is the potential energy and $\boldsymbol{r}$ is the vector of position coordinates within the configuration space. The first three terms in this equation represent intramolecular interactions and the fourth term represents Coulombic intermolecular interactions. The parameters in these terms are not calibrated and are reported in [2]. The final term of this force field functional form contains σ_{ij} and ε_{ij} which parameterize the Lennard-Jones potential describing the van der Waals repulsion-dispersion intermolecular interactions between atoms i and j. Here, we focus on rapidly generating an accurate force field, *i.e.*, improving $U(\boldsymbol{r})$, for HFC-32, by calibrating the like-interaction Lennard-Jones parameters, σ_{ii} and ε_{ii}, to reproduce experimental HFC-32 liquid density. These parameters are calibrated for the three atom types (C, F, and H) of HFC-32, resulting in six total fitting parameters. A description of the model, parameters, and general system setup can be found in [2]. System changes in this work include: first, performing simulation using the LAMMPS molecular dynamics package and, second, expanding the parameter bounds as follows (σ in Å, ε in kcal/mol): $3.0 \leq \sigma_C \leq 4.0$, $2.5 \leq \sigma_F \leq 3.5$, $1.7 \leq \sigma_H \leq 2.7$, $0.06 \leq \varepsilon_C \leq 0.16$, $0.04 \leq \varepsilon_F \leq 0.14$, $0.0 \leq \varepsilon_H \leq 0.05$. While this paper only focuses on optimizing the intermolecular Lennard-Jones parameters, which are the least accurate when force fields are traditionally parameterized using quantum calculations, we emphasize the proposed BO calibration workflow is applicable to any parameters in the force field.

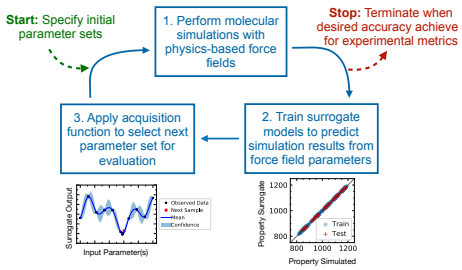

Figure 1: Overview of the proposed BO-enabled force field calibration workflow.

2.2. Bayesian Optimization Workflow

The goal of this work is to refine $U(\boldsymbol{r})$ by optimizing force field parameters, ζ, such that the objective function, $f(\zeta)$, is minimized. For this case study, $\zeta = (\sigma_C, \sigma_F, \sigma_H, \varepsilon_C, \varepsilon_F, \varepsilon_H)$. Figure 1 shows the BO-enabled force field calibration workflow. First, ten initial parameter sets are generated via LHS. In step one, molecular simulations compute the liquid density, $\mathbf{y}^{sim}$, of HFC-32 from $U(\boldsymbol{r})$ at multiple state points of interest. Depending upon the quality of $U(\boldsymbol{r})$, $\mathbf{y}^{\mathrm{sim}}$ may or may not be close to the experimental values, $\mathbf{y}^{exp}$, and this discrepancy is quantified as the mean squared error objective function, $f(\zeta) = \sum_{i=1}^{n}\left(y_i^{sim}(\zeta) - y_i^{exp}\right)^2$, where n is the number of state points considered. After the initial molecular simulations, for each parameter set, $\mathbf{y}^{sim}$ and the subsequent objective function $f(\zeta)$ are computed from the simulation output. At some state points the parameters used in the simulation are so poor that the vapor, not liquid, density of HFC-32 is the simulation result, leading to a discontinuity in objective function value that may be difficult for a surrogate model to capture. To prevent this discontinuity, if a simulation outputs a density lower than the critical density of HFC-32, the simulation density is reported as the critical density. This formulation results in a sufficiently poor objective function value, indicating a poor parameter set, while preventing a discontinuity. In step two, a GPR model is trained to predict $f(\zeta)$ as a function of calibrated parameters ζ such that:

$$\widehat{f(\zeta)} = GP(m(\zeta), \mathrm{cov}(\zeta, \zeta))$$

where $\widehat{f(\zeta)}$ is the surrogate model prediction of $f(\zeta)$, the objective function, GP is the GPR model, $m(\cdot)$ is the GPR mean function, and $\mathrm{cov}(\cdot,\cdot)$ is the covariance (kernel) function. In step three a BO acquisition function is optimized to determine the next parameter set to evaluate. Here, the expected improvement (EI) acquisition function is used to select the next optimal parameter set for simulation. EI is defined as the expected value of $\max(g(\bar{\boldsymbol{\zeta}}) - g(\boldsymbol{\zeta}^+), 0)$ where $g(\cdot)$ is a surrogate model, $\bar{\boldsymbol{\zeta}}$ are the next parameters for evaluation, and $\boldsymbol{\zeta}^+$ represents the best-known parameter set. When $g(\cdot)$ is a GPR model, the analytic form of EI is:

$$EI(\bar{\boldsymbol{\zeta}}) = \begin{cases} (\mu_*(\bar{\boldsymbol{\zeta}}) - g(\bar{\boldsymbol{\zeta}}^+) - \xi)\Phi(z) + \sigma_*(\bar{\boldsymbol{\zeta}})\phi(z), & \sigma_*(\bar{\boldsymbol{\zeta}}) > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$z(\bar{\boldsymbol{\zeta}}) = \begin{cases} \dfrac{\mu_*(\bar{\boldsymbol{\zeta}}) - g(\bar{\boldsymbol{\zeta}}^+) - \xi}{\sigma_*(\bar{\boldsymbol{\zeta}})}, & \sigma_*(\bar{\boldsymbol{\zeta}}) > 0 \\ 0, & \text{otherwise} \end{cases}$$

where μ_* is the prediction mean, σ_* is the variance, $\Phi(z)$ and $\phi(z)$ are the cumulative and probability density functions for the standard normal distribution, respectively, $g(\bar{\boldsymbol{\zeta}}^+)$ is the largest measured valued $((\bar{\boldsymbol{\zeta}}^+) = \mathrm{argmax}(g(\bar{\boldsymbol{\zeta}}_1), \ldots, g(\bar{\boldsymbol{\zeta}}_n))$, and ξ is an adjustable parameter describing the exploration and exploitation mode [4]. Upon applying the EI acquisition function, a new parameter set is generated which is used in new simulations, and the workflow continues iteratively until the desired simulation accuracy is reached or no improvement can be achieved in the objective function.

3. Results

We begin by comparing our automated BO workflow to our prior semi-automated JCIM workflow. Figure 2 plots the best (lowest) objective function value found after simulating 110 trial parameter sets (*i.e.,* initial ten LHS parameter sets plus one hundred EI-generated samples) in the BO workflow compared to the best objective function values found after evaluating 200, 400, 600, and 800 total parameter sets in the JCIM workflow. Figure 2 shows the improvement in objective function value for both workflows as more parameter sets are evaluated and the surrogate models are trained on more data. After 101 parameter sets are evaluated in the BO-based workflow, the

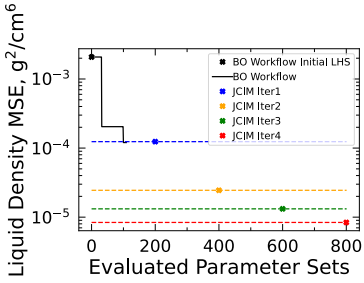

Figure 2: Comparison of the best objective function value after evaluating a certain number of parameter sets using the proposed BO workflow (black **x**, black line) versus our prior JCIM workflow (colored **x**, dashed lines).

objective function is 1.20E-4 g^2/cm^6 while the lowest objective function for the JCIM workflow's initial 200 parameter sets, which were generated via LHS, was 1.24E-4 g^2/cm^6. This indicates that the BO workflow can achieve a lower objective function value after evaluating approximately 50% fewer parameter sets, and therefore performing less simulations, than a space filling sample of parameter space. We hypothesize this is the result of the adaptive nature of BO acquisition function, which can effectively explore and exploit parameter space to more efficiently find optimal parameter sets. We expect that as more simulations are performed, the GPR models in the BO workflow will improve such that this workflow will additionally require fewer simulations to surpass the objective function values found after evaluating 400, 600, and 800 samples in the JCIM workflow.

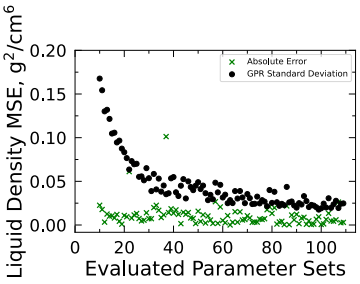

Figure 3: Absolute error between GPR model and simulation (green x, left y-axis) and GPR prediction uncertainty (black •, right y-axis) change as more parameter sets are evaluated.

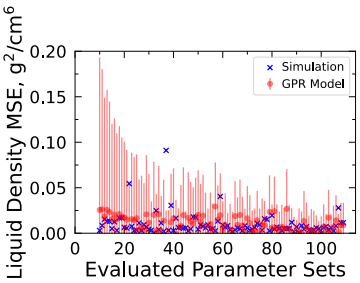

Figure 4: Comparison of GPR model prediction mean (red •) and standard deviation (red | , error bar) versus molecular simulation results (blue x) as more parameter sets are evaluated.

GPR model improvement for the BO workflow is seen in Figure 3, which shows the absolute error between the GPR model prediction of the objective and the actual simulation result for each trial parameter set. The standard deviation in the GPR model prediction is plotted and shows a decreasing trend. The discrepancy between the GPR model prediction and actual simulation result also shows a decreasing, although less obvious, trend. This indicates GPR model improvement as more training data and regions of parameter space are sampled via the guidance of BO. Figure 4 compares the GPR model predictions and the actual simulation results of the objective for each evaluated parameter set, with the GPR model uncertainty plotted as error bars. This figure provides an example of how the BO-based workflow balances exploration and exploitation; BO suggests parameter sets which either reduce the value of the objective function or provide more information about the objective function in specific regions of parameter space, reducing the GPR variance in those regions. Figure 4 shows how GPR model predictions for certain trial parameter sets result in objectives very close to simulation results (exploitation) while other predictions are significantly different than simulations and have high uncertainty, indicating exploration. Thus, instead of relying on GPR models to generate batches of new trial parameter sets in the original workflow, the BO-based workflow systematically samples parameter sets to gain information about the parameter regions which result in the lowest objective function values. We believe this approach both reduces the burden of the user by automatically selecting new parameter sets to sample and improves the efficiency of the workflow by decreasing the number of parameter set samples and subsequent simulations required to calibrate force fields.

Further analysis is required to benchmark the BO-based calibration workflow. The BO results shown in Figure 2 were generated using the gradient based L-BFGS (scipy) optimizer to calibrate the GPR model hyperparameters for the first 80 iterations. Then, the hyperparameter optimization failed due to a Cholesky factorization error (the GP kernel became negative semi-definite) and was switched to an ADAM (BOTorch) optimizer. Reproducing this result using only the BOTorch optimizer is a work in progress; with only the BOTorch optimizer, the GPR model error remains $\mathcal{O}(10^{-2})$ whereas the JCIM workflow was $\mathcal{O}(10^{-4})$ after 200 samples. Ongoing work is investigating the differences in trained GP hyperparameters and overall BO performance using these two optimizers. Additional opportunities for further improving this framework include determining the minimum amount of data and initial parameter set samples necessary for efficient GPR model improvement, exploring various formulations for the objective function, kernel function, and the vapor-liquid density discontinuity, and evaluating the capabilities of various GPR model optimizers. We expect these analyses to improve the overall efficiency of the BO calibration workflow.

4. Conclusions

In this paper, we prototype a fully automated BO framework for force field calibration. Results show that after evaluating 101 parameter set samples with the BO workflow, the lowest mean squared error between simulation and experimental values for the liquid density of HFC-32 is 1.20E-4 g^2/cm^6. This objective was 0.04E-4 g^2/cm^6 smaller than the best objective found in the initial 200 parameter set LHS used in our prior JCIM workflow. This result suggests the BO techniques enhance the efficiency of force field calibration. Additionally, BO has enabled automated sampling of parameter space removing the need for user decisions for generating trial parameter sets. We expect that as more parameter sets are sampled, improvement within the GPR models will continue to show that fewer simulations will yield equally accurate force fields as the original workflow. Future work includes systematically benchmarking the BO workflow performance to determine the best practices for implementing BO techniques within the calibration workflow. Ultimately, this framework can be used to develop accurate force fields for multiple HFCs and other classes of molecules for which accurate molecular models are lacking.

5. Acknowledgements

We acknowledge support from the National Science Foundation grant CBET-1917474 and the University of Notre Dame. BB acknowledges support from the Richard and Peggy Notebaert Premier Fellowship.

6. References

[1] Wang, J., & Kollman, P. A. (2001). Automatic parameterization of force field by systematic search and genetic algorithms. *J. Comp. Chem.*, *22*(12), 1219-1228.
[2] Befort, B. J., DeFever, R. S., Tow, G. M., Dowling, A. W., & Maginn, E. J. (2021). Machine Learning Directed Optimization of Classical Molecular Modeling Force Fields, *J. Chem. Inf. Model*, 61(9), 440-4414.
[3] Unke, O. T., Chmiela, S., Sauceda, H. E., Gastegger, M., Poltavsky, I., Schütt, K. T., Tkatchenko, A., & Müller, K. R. (2021). Machine learning force fields. *Chemical Reviews*.
[4] Wang, K., & Dowling, A. W. (2022). Bayesian optimization for chemical products and functional materials. *Current Opinion in Chemical Engineering*, *36*, 100728.
[5] Environmental Protection Agency. The Ozone-Depleting Substances Phaseout: 2020-2030.