Gaussian Process-Based Refinement of Dispersion Corrections

Stefan Gugler [1], Jonny Proppe [1,2], and Markus Reiher [1]

[1] Laboratory of Physical Chemistry, ETH Zürich, Vladimir-Prelog-Weg 2, Zürich, Switzerland,
[2] Departments of Chemistry and Computer Science, University of Toronto, Toronto, Ontario, Canada.

Introduction

- We already quantified[1] statistical errors in Grimme's widely popular semiclassical dispersion correction model DFT-D3[2].
- In this work[3] we propose a self-improving, system-focused model based on Gaussian Process (GP) regression. It entails:
  - **systematic error corrections**
  - **uncertainty estimations**
- **Batch-wise variance-based sampling (BVS)** helps determining for which structures a reference calculation should be performed.

GP Regression

- GP regression is applied to learn a mapping $\mathcal{GP} : \mathbf{x} = \{x_1, \cdots, x_M\} \mapsto y$. Given N observations, mean and variance fully specify the GP:

$$\mathbb{E}\left[\hat{y}(\mathbf{x}_*)\right] \equiv \hat{\mu}_y(\mathbf{x}_*) = \mathbf{k}(\mathbf{x}_*, \mathbf{X})\Big(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \alpha_0 \mathbf{I}\Big)^{-1} \mathbf{y}$$

$$\mathbb{V}\left[\hat{y}(\mathbf{x}_*)\right] \equiv \hat{\sigma}_y^2(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}(\mathbf{x}_*, \mathbf{X})\Big(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \alpha_0 \mathbf{I}\Big)^{-1} \mathbf{k}(\mathbf{X}, \mathbf{x}_*)$$

  with $\mathbf{x}_*$ representing a new system, $\mathbf{X}$ the N feature vectors of the training set, $\mathbf{y}$ the target vector (observations), and the noise hyperparameter α_0.
- The kernel $k(\mathbf{x}_i, \mathbf{x}_j)$ is a measure of similarity between its arguments, analogously for $\mathbf{k}$ and $\mathbf{K}$.

Batch-wise Variance-based Sampling (BVS)

- Given a training set $\mathcal{T}_{N_\mathbf{x}}^{N_y}$ with $N_\mathbf{x}$ feature vectors, $\mathbf{x}$, and N_y observations, y, where $N_\mathbf{x} = N_y$, and a query set $\mathcal{Q}_{M_\mathbf{x}}$ with $M_\mathbf{x}$ feature vectors.
- In decreasing order of $\hat{\sigma}_y^2$, feature vectors from $\mathcal{Q}_{M_\mathbf{x}}$ are added to $\mathcal{T}_{N_\mathbf{x}}^{N_y}$.
- To reduce computational cost, the hyperparameter optimization to determine $\hat{\sigma}_y^2$ is only performed after each batch of size L added to the training set, generating $\mathcal{T}_{N_\mathbf{x}+L}^{N_y+L}$ and $\mathcal{Q}_{M_\mathbf{x}-L}$.
- This is repeated until for each $\mathbf{x}$ of the remaining $\mathcal{Q}_{M_\mathbf{x}}$, $\hat{\sigma}_y^2 < t$, where t is a predefined threshold.

Methodology

Data set Overview of the 1,248 molecular reference systems (dimers)

Set	#	Description
S13x8	104	dispersion-dominated subset (#34–46) of the S66x8 set
ROTA	1,100	ethyne–pentane dimers; varying relative orientations; centroid distances (3.5–10.0 Å)
CONF	44	ethyne–pentane dimers; varying relative orientations; pentane conformations; centroid distance of 5.2 Å

Electronic-Structure Calculations

- All calculations were carried out with ORCA 4.0.1. and are CP-corrected.
- PBE: ma-def-QZVPP basis set, def2-QZVP auxiliary basis set
- DLPNO-CCSD(T): aug-cc-pVT/QZ basis sets, aug-cc-pVQ/$\mathrm{5}$Z aux. basis sets
- Triple-ζ and quadruple-ζ DLPNO-CCSD(T) energies extrapolated to CBS.
- D3 corrections from DFTD3 with Becke–Johnson (BJ) damping scheme.
- asd

References

[1] Weymuth, T.; Proppe, J.; Reiher, M. Statistical Analysis of Semiclassical Dispersion Corrections, *Journal of Chemical Theory and Computation* **2018,** *14,* 2480-2494.

[2] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu, *The Journal of Chemical Physics* **2010,** *132,* 154104.

[3] Proppe, J.; Gugler, S.; Reiher, M. Gaussian Process-Based Refinement of Dispersion Corrections, **2019,** arXiv:1906.09342.

Descriptors

- **eigD3(BJ)**: Decreasing eigenvalues of $E_{IJ}^{\mathrm{D3(BJ)}}$ with 0-padding:

$$E_{IJ}^{\mathrm{D3(BJ)}} = \sum_{n=6,8} s_n \frac{C_n^{IJ}}{R_{IJ}^n + \left(a_1 \sqrt{C_8^{IJ}/C_6^{IJ}} + a_2\right)^n} \quad \forall I \neq J,\ 0 \text{ else}$$

- **histD3(BJ)**: 16-dimensional energy histogram from DFTD3:

$$e_m^{\mathrm{histD3(BJ)}} = \sum_{\substack{I>J \\ r_m^{\min} < R_{IJ} \leq r_m^{\max}}} \sum_{n=6,8} s_n \frac{C_n^{IJ}}{R_{IJ}^n + \left(a_1 \sqrt{C_8^{IJ}/C_6^{IJ}} + a_2\right)^n}$$

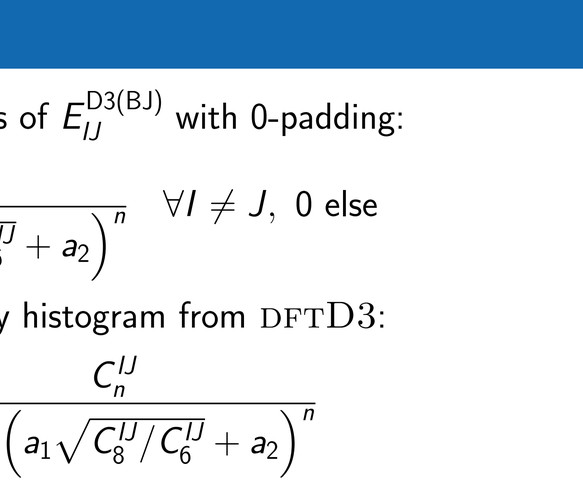
  with R_{IJ}, the distance of atoms I and J, $C_{\{6,8\}}^{IJ}$, the dispersion coefficients, $s_6 \triangleq 1$, and a_1, a_2, and s_8 the PBE-dependent, empirical parameters.
- We used the isotropic Matérn-$1/2$ kernel (exponential kernel):

$$k(\mathbf{x}_i, \mathbf{x}_j) = \alpha_1 \exp\left\{-\alpha_2 \|\mathbf{x}_i - \mathbf{x}_j\|\right\}$$

Results and Discussion

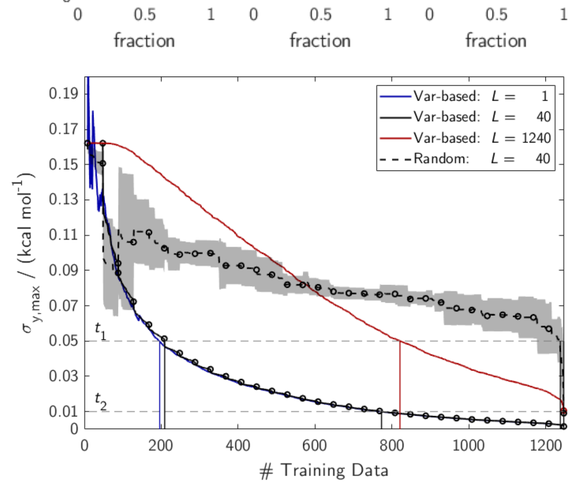
Learning Curves: Intra- vs. extrapolation

- D3(BJ)-GP for PBE in histD3(BJ)
- MARE: mean abs. rel. error
- Train: ROTA-T (1,000 of ROTA)
  Test: ROTA-H (rest), CONF, 13x8
- 20 steps of adding data, hyperparam. opt. each time
- linear model systematically worse



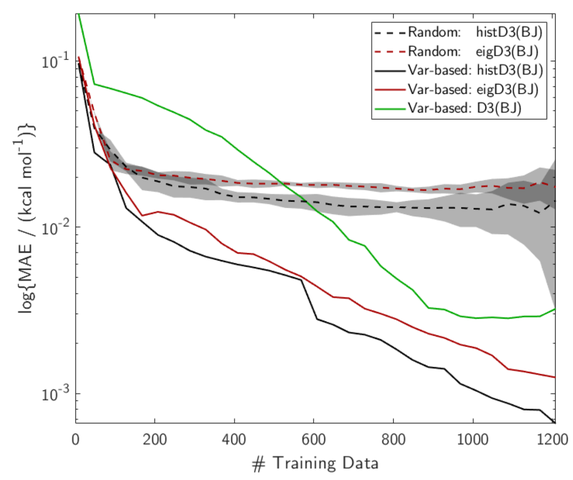
BVS on all data

- D3(BJ)-GP for PBE in histD3(BJ)
- Max. prediction uncertainty, $\sigma_{\mathrm{y,max}}$
- Eight random initial draws
- BVS with different L and random
- Hyperparam. opt. each L
- $L = 1248$ overestimates $\sigma_{\mathrm{y,max}}$
- $L = 40$ almost as good as $L = 1$



Prediction error on all data

- D3(BJ)-GP performance in eig/histD3(BJ) for random sampling and BVS ($L = 40$)
- Error measured in log MAE
- Eight random initial draws
- histD3(D3) outperforms eigD3(D3), and both outperform D3(BJ)
- Random sampling worse than BVS



Conclusion & Outlook

- BVS accelerates the learning process and keeps the training informative
- We obtain a **system-focused, self-improving model** for dispersion interactions equipped with confidence intervals that is almost as efficient as their D3(BJ) complements.
- The D3-GP workflow is also applicable to other functionals, damping schemes, or dispersion corrections, e.g. DFT-D4.

Acknowledgments

J.P. acknowledges funding through an *Early Postdoc.Mobility* fellowship by the Swiss National Science Foundation (project no. 178463). S.G. and M.R. are grateful for financial support by the Swiss National Science Foundation (project no. 200021_182400).