Supporting Information:
Gaussian Process-Based Refinement of Dispersion Corrections

Jonny Proppe,[a,b,*,†] Stefan Gugler,[b,†] Markus Reiher[b,*]

September 10, 2019

[a]Departments of Chemistry and Computer Science, University of Toronto, Toronto, Ontario, Canada.
[b]Laboratory of Physical Chemistry, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland.
† These authors contributed equally.

The ROTA and CONF Sets

We sampled relative orientations of the ethyne–pentane dimer at 16 different centroid distances (3.5–10.0 Å, equidistant steps) without changing the internal coordinates of ethyne and pentane. To sample relative orientations of the ethyne–pentane dimer, we transformed the Cartesian nuclear coordinates of each monomer (x, y, z) into spherical coordinates,

$$r = \sqrt{x^2 + y^2 + z^2} \ , \tag{1}$$

$$\theta = \begin{cases} \tan^{-1}\left(\frac{y}{x}\right) & \text{if } x \neq 0 \\ \frac{\pi}{2} & \text{if } (x = 0 \wedge y \geq 0) \\ -\frac{\pi}{2} & \text{if } (x = 0 \wedge y \leq 0) \end{cases} , \tag{2}$$

and

$$\phi = \cos^{-1}\left(\frac{z}{r}\right) \text{ if } r > 0 \ . \tag{3}$$

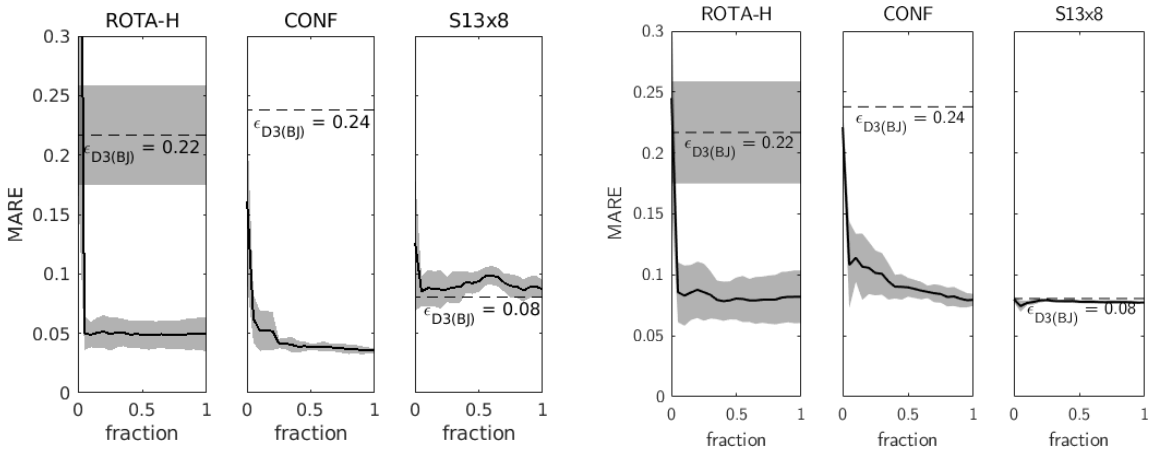
To ensure a uniform sampling of the rotation sphere, we drew the two random dummy variables A and B uniformly from the interval $(0, 1)$ and determined $\theta = 2\pi A$ as well as $\phi = \cos^{-1}(2B-1)$. Additionally, pentane was rotated around its internal main axis (defined by the carbon backbone) by angle ψ. For ethyne, this internal rotation was omitted as it belongs to the $D_{\infty h}$ point group. For each centroid distance, we sampled 120 rotamers (1,920 in total). All ethyne–pentane rotamers with a positive DLPNO-CCSD(T)/CBS[1–4] interaction energy (overall repulsive interaction) or with a minimum intermolecular interatomic distance of <1.5 Å were discarded, which left us with 1,100 dimers representing the ROTA set. For the CONF set, which comprises 44 entries, we sampled 50 ethyne–pentane rotamers with a fixed centroid distance of 3.5 Å as described above. Conformations of pentane were sampled randomly with RDKit.[5] Six dimers were discarded due to a positive DLPNO-CCSD(T)/CBS interaction energy.

The S13x8-T and S13x8-V Subsets

For the generation of the S13x8-T set, we removed all dimers with a relative intermolecular distance of $0.95r_{\text{eq}}$ and $1.10r_{\text{eq}}$ (here, r_{eq} refers to the intermolecular distance at equilibrium), respectively, from the S13x8 set. Furthermore, we removed all uracil–cyclopentane dimers and all ethene–pentane dimers from S13x8. The set difference of S13x8 and S13x8-T constitutes the S13x8-V set. S13x8-T and S13x8-V contain 66 and 38 dimers, respectively.

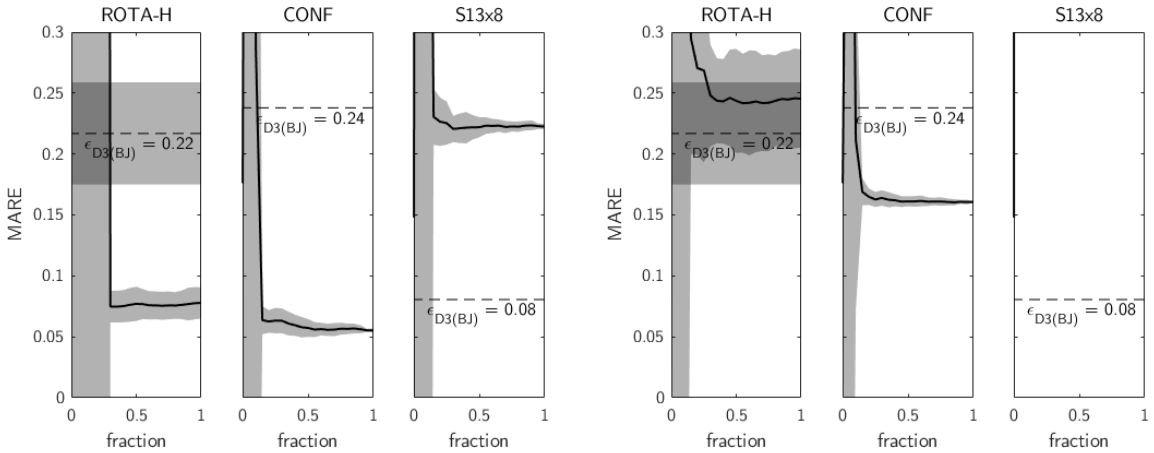
*corresponding authors: jonny.proppe@utoronto.ca, markus.reiher@phys.chem.ethz.ch

Additional Learning Curves



a) Matérn-½ **b)** Matérn-3/2

Figure 1: Learning curves (solid lines) of the D3(BJ)-GP model for PBE with respect to three disjoint test sets (ROTA-H, CONF, and S13x8). The learning curves (as measured by the MARE) refer to eigD3(BJ) features and two kernels, **a)** Matérn-½ and **b)** Matérn-3/2. The D3(BJ)-GP model was trained on the ROTA-T set, a random sample of 1,000 instances drawn from the ROTA set (the remaining 100 instances are contained in the holdout set ROTA-H). The training set fractions were taken in 20 equidistant steps between 1 % and 100 %. Furthermore, the entire training set (1,000 data points) was drawn 10 times, resulting in standard deviations indicated by the gray bands. The errors of the corresponding D3(BJ)-corrected PBE interaction energies, denoted $\epsilon_{D3(BJ)}$, are shown as horizontal dashed lines.



a) histD3(BJ) **b)** eigD3(BJ)

Figure 2: Learning curves (solid lines) of a linear correction model for PBE-D3(BJ) interaction energies with respect to three disjoint test sets (ROTA-H, CONF, and S13x8). The learning curves (as measured by the MARE) refer to histD3(BJ) (left) and eigD3(BJ) (right) features. The linear model was trained on the ROTA-T set, a random sample of 1,000 instances drawn from the ROTA set (the remaining 100 instances are contained in the holdout set ROTA-H). The training set fractions were taken in 20 equidistant steps between 1 % and 100 %. Furthermore, the entire training set (1,000 data points) was drawn 10 times, resulting in standard deviations indicated by the gray bands. The errors of the corresponding D3(BJ)-corrected PBE interaction energies, denoted $\epsilon_{D3(BJ)}$, are shown as horizontal dashed lines.

Successively adding quadratic, cubic, and quartic terms to the linear model (Figure S2) dramatically worsens the predictions on ROTA-H for an ensemble of training sets containing 100 ROTA data each. For 1,000 training data, the cubic model performs best, but still worse than our GP models (with respect to both mean and standard deviation of the learning curves). Furthermore, a constant model with only one variable parameter yields worse predictions compared to the linear model (two variable parameters) for both 100 and 1,000 training data. By contrast, the constant model performs best for predictions on the S13x8 among the polynomial models considered (constant, linear, quadratic, cubic, and quartic), but still worse than our GP models (with respect to both mean and standard deviation of the learning curves). This result is independent of the size of the training set.

Additional Results of BVS-Guided Active Learning

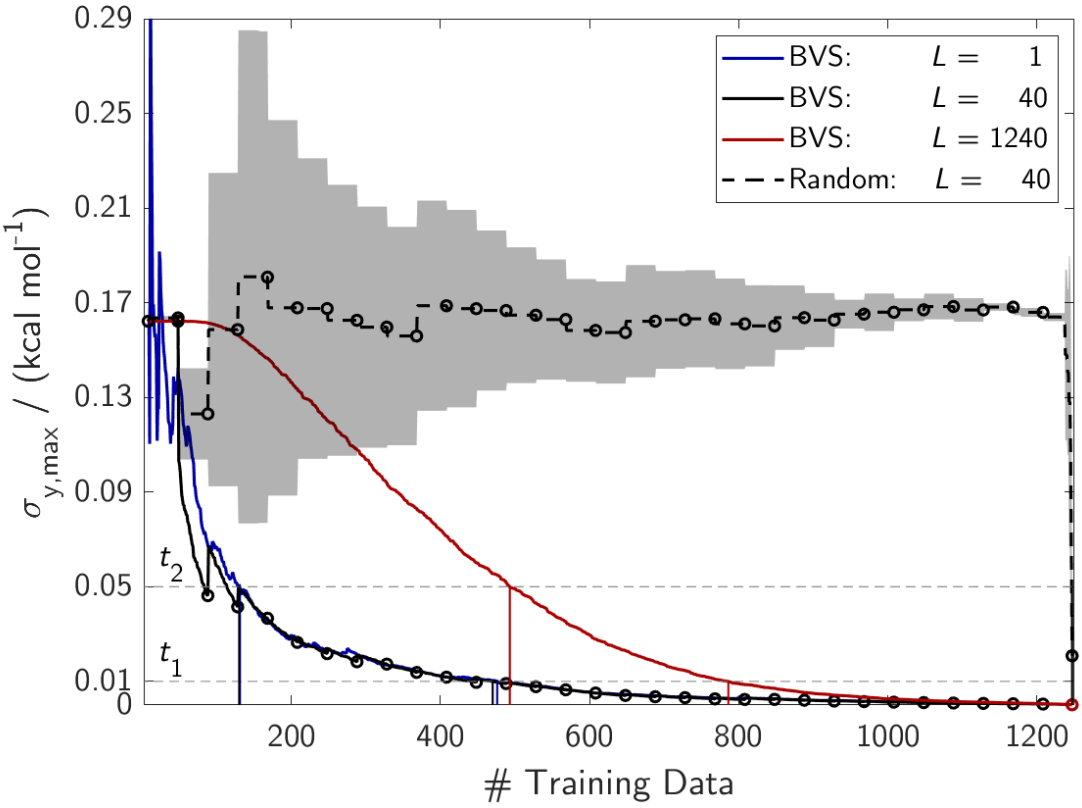


Figure 3: The maximum prediction uncertainty as a function of the number of training data. It is measured as the maximum standard deviation of the GP posterior distribution, $\hat{\sigma}_{y,\max}$. The initial training set consisted of 8 dimers, which were drawn randomly from the overall data set (S13x8 + ROTA + CONF). The remaining 1,240 dimers served as pool of potential training data (query set). The maximum prediction uncertainty relates to the current query set. Two sampling strategies were considered, BVS with three different batch sizes ($L = 1$, blue line; $L = 40$, solid black line; and $L = 1{,}240$, red line) and random sampling with a batch size of $L = 40$ (black dashed line and gray band obtained from 5 repeated draws). Intersections with two possible accuracy thresholds, t_1 and t_2 (gray dashed lines), are highlighted by vertical lines. We employed the Matérn-3/2 kernel and the histD3(BJ) featurization to produce this figure.

Distance Intervals of the histD3(BJ) Feature Vector

Table 1: Minimum and maximum interatomic distances, $r_m^{\min}$ and $r_m^{\max}$, as applied to the 16 elements of the histD3(BJ) vector. These distance intervals have been adopted without alteration from Grimme's DFTD3 code.[6,7] All distances are reported in Å.

m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$r_m^{\min}$	0.0	1.5	2.0	2.3	2.7	3.0	3.3	3.7	4.0	4.5	5.0	5.5	6.0	7.0	8.0	9.0
$r_m^{\max}$	1.5	2.0	2.3	2.7	3.0	3.3	3.7	4.0	4.5	5.0	5.5	6.0	7.0	8.0	9.0	10.0

References

[1] Riplinger, C.; Neese, F. An Efficient and near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method, *J. Chem. Phys.* **2013,** *138,* 034106.

[2] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural Triple Excitations in Local Coupled Cluster Calculations with Pair Natural Orbitals, *J. Chem. Phys.* **2013,** *139,* 134101.

[3] Neese, F. The ORCA Program System, *WIREs Comput. Mol. Sci.* **2012,** *2,* 73-78.

[4] Neese, F. Software Update: The ORCA Program System, Version 4.0, *WIREs Comput. Mol. Sci.* **2018,** *8,* e1327.

[5] Landrum, G. "RDKit: Open-Source Cheminformatics", 2006.

[6] 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, *J. Chem. Phys.* **2010,** *132,* 154104.

[7] Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory, *J. Comput. Chem.* **2011,** *32,* 1456-1465.