Uncertainty Quantification in Quantum Chemistry

Thomas Weymuth

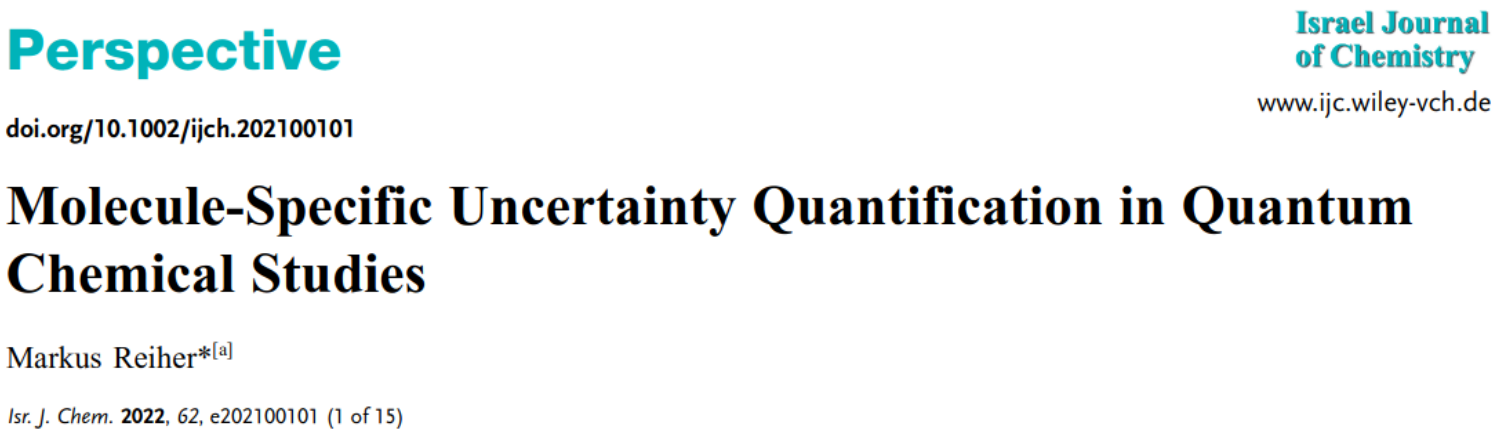
ETH Zurich

June 22, 2022

Introduction

- Quantum chemical methods rely on approximations
- Combined effect of these is difficult to assess

$\Rightarrow$ Every quantum chemical result is affected by some unknown error



Perspective

Israel Journal of Chemistry

www.ijc.wiley-vch.de

doi.org/10.1002/ijch.202100101

Molecule-Specific Uncertainty Quantification in Quantum Chemical Studies

Markus Reiher*[a]

Isr. J. Chem. **2022**, *62*, e202100101 (1 of 15)

Traditional Benchmarking?

- Knowledge of these errors is crucial for practical applications
- Traditional way of assessing the reliability of a quantum chemical method: (static) benchmarking
- This approach suffers from some shortcomings, which we will illustrate in this talk[1]

[1] T. Weymuth, M. Reiher, *arXiv:2204.06659* [physics.comp-ph], **2022** (*Phys. Chem. Chem. Phys.*, in press).

Current Situation

- Errors of quantum chemical calculations behave heteroscedastically

$\Rightarrow$ Small benchmark sets are not representative for most applications

- This was addressed by the creation of ever larger benchmark sets
- Example: set by Mardirossian and Head-Gordon[2] (4986 data points)
- All our conclusions are valid for any benchmark set

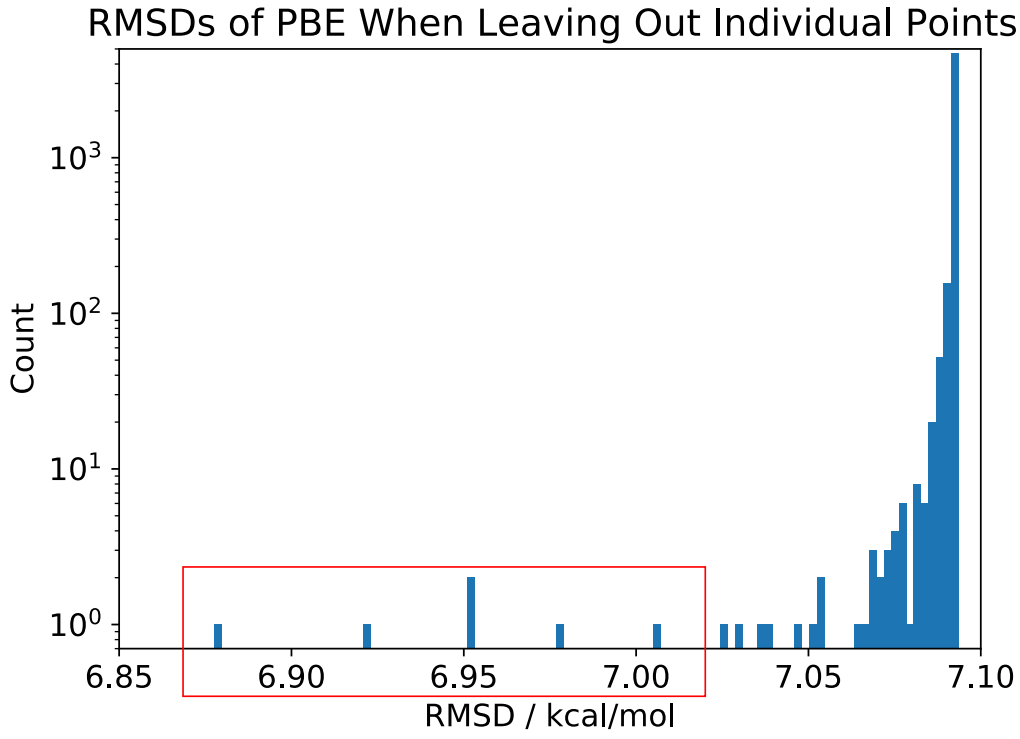
[2] N. Mardirossian, M. Head-Gordon, *Mol. Phys.*, **2017**, *115*, 2315.

Study Setup

- Key idea: subject benchmark set to statistical analysis to understand how conclusions w. r. t. accuracy and transferability depend on composition of set[3]
- Jackknifing: how strong is the effect of a single data point?

[3] T. Weymuth, M. Reiher, *arXiv:2204.06659* [physics.comp-ph], **2022** (*Phys. Chem. Chem. Phys.*, in press).

Jackknifing Individual Data Points

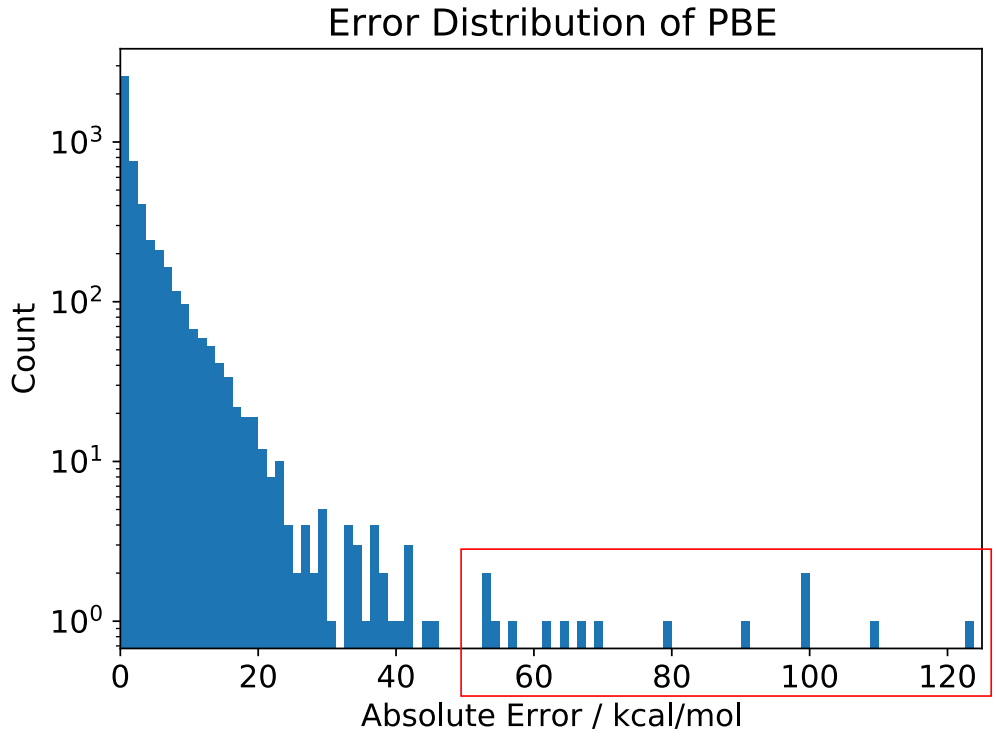


- Some points, when left away, lead to a markedly lower RMSD — these points have an unexpectedly large effect
- Leaving out the ten data points with largest error lowers the RMSD by 17 % (for PBE). What if these points were accidentally omitted?

Jackknifing Individual Data Points

A few additional data points can reduce or increase the currently assessed accuracy of a density functional up to 20 % even when a very large benchmark set is employed

Error Distribution



- Few points have very large errors — these have a large effect on the RMSD
- Large errors are almost always reported for atomization energies, *i.e.*, extensive quantities growing with system size

Error Distribution

Since errors are distributed very unevenly, even large benchmark sets cannot reliably estimate the error for a particular system; one needs to adopt a system-focused approach to uncertainty quantification

System-Focused Bayesian Uncertainty Quantification

- System-specific parametrization of density functionals including uncertainty estimation[4]
- Estimating prediction uncertainty of physico-chemical property models via bootstrapping[5]
- Quantifying the uncertainty of semiclassical D3 dispersion corrections[6]
- Error-controlled explorations of chemical reaction networks with Gaussian processes[7]

---

[4] G. N. Simm, M. Reiher, *J. Chem. Theory Comput.*, **2016**, *12*, 2762.

[5] J. Proppe, M. Reiher, *J. Chem. Theory Comput.*, **2017**, *13*, 3297.

[6] T. Weymuth, J. Proppe, M. Reiher, *J. Chem. Theory Comput.*, **2018**, *14*, 2480.

[7] G. N. Simm, M. Reiher, *J. Chem. Theory Comput.*, **2018**, *14*, 5238.

System-Focused Bayesian Uncertainty Quantification

- Kinetic modeling with propagation of free-energy uncertainties[8]
- Correct systematic errors in D3 dispersion corrections with Gaussian processes[9]
- Review on error assessment in computation chemistry[10]

---

[8] J. Proppe, M. Reiher, *J. Chem. Theory Comput.*, **2019**, *15*, 357.

[9] J. Proppe, S. Gugler, M. Reiher, *J. Chem. Theory Comput.*, **2019**, *15*, 6046.

[10] G. N. Simm, J. Proppe, M. Reiher, *Chimia*, **2017**, *71*, 202.

Conclusions and Outlook

- Fundamental challenge of static benchmarking: errors behave heteroscedastically
- For a benchmarking set representing large part of chemical space, aggregating all the errors into one overall error measure would make this error measure too large for some parts of chemical space, and too low for others.
- Simple but effective way to overcome this challenge: dynamic, *i.e.*, rolling and system-focused benchmarking
  - Create benchmark data for each particular application
  - Use fast ML model to predict error; retrain this model whenever necessary[11]

[11] M. Reiher, *Isr. J. Chem.*, **2022**, *62*, e202100101.

Acknowledgments

Thanks to:

- Prof. Markus Reiher
- All group members

Financial support:

- ETH Zurich
- Swiss National Science Foundation SNF