

Learning structure-energy relationships for the prediction of molecular crystal structures

Graeme M. Day

School of Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom

g.m.day@soton.ac.uk

The discovery of new functional materials can be guided by computational screening, particularly if the structure of a material can be reliably predicted from its chemical composition. For this application, we have been developing the use energy-structure-function maps [1] of the crystal structures available to a molecule. These maps help understand the properties of predicted crystal structures and their energetic stabilities. However, the use of these methods is still limited by the computational cost of crystal structure prediction (CSP), most of which is associated with the calculation of the relative energies of predicted crystal structures using energy models that are sufficiently accurate to provide reliable energetic rankings. To accelerate these methods, we have been developing machine learning approaches to predict high quality energies (e.g. from solid state density functional theory) from structures that have been generated with computationally efficient energy models. These approaches rely on statistical models, in our case Gaussian Process Regression, to relate lattice energies to geometric descriptors of crystal structures. The talk will discuss two approaches that we have developed: learning of total energies calculated using solid state density functional theory [2,3], and a fragment-based approach [4] where we learn high level dimer energies, which are used to build up the total lattice energies of predicted structures.

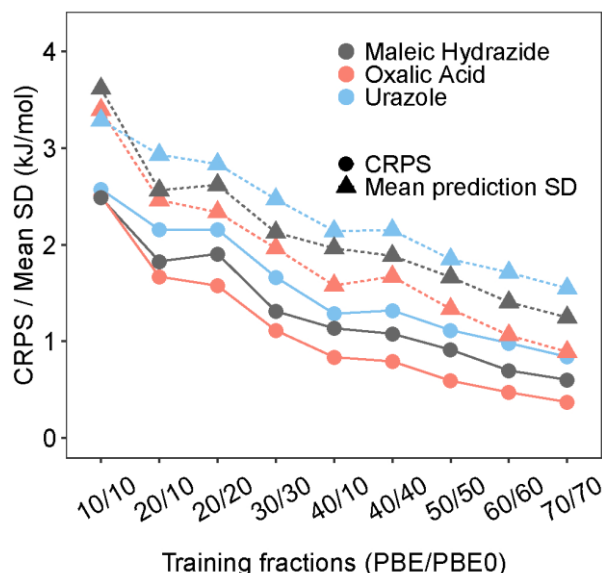


Figure 1. Continuous Ranked Probability Score (CRPS) and model uncertainty for a two-level multi-fidelity prediction of the energies of the predicted crystal structures of three small organic molecules. Training fractions are given as X/Y, where X is the percentage of structures used to train the lower level of the prediction (here, calculated lattice energies at the PBE level) and Y is the percentage of those structures used in the first level that are used to train the higher level of model (calculated lattice energies using the PBE0 functional) [3].

[1] Day, G. M. and Cooper, A. I. (2018) *Adv. Mater.*, 30, 1704944

[2] Musil, F, De, S., Yang, J., Campbell, J. E., Day, G. M. and Ceriotti, M. (2018) *Chem. Sci.*, 9, 1289-1300.

[3] Egorova, E., Hafizi, R., Woods, D. C. and Day, G. M. (2020) *J. Phys. Chem. A*, 124, 8065–8078.

[4] McDonagh, D., Skylaris, C.-K. and Day, G. M. (2019) *J. Chem. Theory Comput.*, 15, 2743–2758

Keywords: crystal structure prediction, machine learning, Gaussian process regression