Anonymous IJQC Reviewer[1]

[1]Affiliation not available

June 11, 2020

Abstract

This is a follow-on paper from the Hutchison group, expanding on some previous work looking at correlations of molecular energy from a variety of levels of theory with results from high-level ab initio calculations. A new addition in this paper is a small set of ML methods, a welcome addition to the forcefield and electronic structure methods usually used in comparisons of this kind.

The paper presents some interesting results, but is riddled with missing or misattributed data, typos, grammatical errors (particularly agreements for single and plural nouns) and errors in the references. The paper should be carefully corrected before resubmission.

The key omission in the paper is any attempt to provide confidence in the deductions made about the differences in accuracy between the methods compared. Confidence intervals on each of the estimators, estimates of success rates and their errors, and pairwise hypothesis tests, at a minimum, must be added before publication. With this data in hand the new version can make quantitative estimates of the differences between the methods.

1 Referee Report

When superiority of one set of results is asserted over another it is simply not acceptable to report raw performance numbers and state that the biggest/smallest is the best. All the results in this paper are sample results and therefore have an associated error, which must be reported. Given this error, when two methods are compared measures of the significance and the impact of the difference must be reported (hypothesis tests, effect sizes, confidence intervals etc.).

The R^2s shown throughout the manuscript have an error that can be calculated analytically - this must be provided along with the raw results. Confidence intervals for the estimators like mean and median must also be added.

Specific comments

Figure 2 was missing from the PDF.

What is the y-axis in Figure 3? Counts of molecules in that timing bin?

Figure 4: R^2 has a calculable error and should be included in the plot.

Figure 4: ANI family methods are not labelled/included in the plot, but are mentioned as being there in the text.

Figure 7: Add error for R^2.

Figure 8: Missing from the PDF

Test Set selection - “the training set was the first five conformers” - how were these conformers generated and ranked?

“any molecules with fewer than five conformers was omitted” - how many were omitted?

It is understood/guessable why some DLPNO calculations did not converge?

The statement “deriving accurate rankings..” is not supported by the previous sentence. Boltzmann weighting relies entirely on energy, so how is accurate ranking going to help improve “computational predictions”?

The text in “Comparison of single points vs. DLPNO-CCSD(T)” should be tabulated and is redundant with text elsewhere in the manuscript.

The writer appears to be getting to know Endnote’s referencing scheme:

Two citations in the Results page are in text form.

There are errors in the following citations: 7, 37, 59, 61 (content missing entirely here)

The SciPy citation, 59, is not the standard one - justification for using this citation should be provided.

The citation for pybel is a ‘?’

Comparison of timing - the details on the hardware should go into the Methods section.

Typos

xray - X-ray

Mllr-PLesset = Moller-Plesset

BATTY/n = BATTY

CPU Time in Table 1 cannot be 0.0. Perhaps each method should have time scaled by forcefield time set to 1?

2