

Response to Referee for: Fitting Elephants in the Density Functionals Zoo: Statistical Criteria for the Evaluation of DFT methods as a Suitable Replacement for Counting Parameters

Roberto Peverati¹

¹IJQC Special Issue

June 15, 2020

Dear Dr. Cavalleri,

I hereby resubmit my paper entitled “Fitting Elephants in the Density Functionals Zoo: Statistical Criteria for the Evaluation of DFT methods as a Suitable Replacement for Counting Parameters” with modifications that address all the concerns raised by the Referees. As such, I believe the manuscript is now ready for publication in the Authorea special issue of the International Journal of Quantum Chemistry.

I want to thank the Referees for the clear and very useful remarks. For your convenience, I report below a point-by-point reply to each of the concerns raised by the Referees, and how I’ve addressed them in the new manuscript:

Referee #1:

1) Multiple times (particularly in the abstract and introduction), the author alludes to an ongoing discussion in the DFT community that is biased towards seeing DFT methods that rely on a large number of fitted parameters as non-transferable to different systems compared to methods that rely on a smaller number of parameters. At times, the statements made about this seem rather anecdotal, as they lack any scientific citation. Some examples are:

“Counting parameters has become customary in the density functional theory community as a way to infer the transferability of popular approximations to the exchange–correlation functionals.”

“Among this latter school of thoughts, fitting xc functionals has achieved a somehow bad reputation, because parameters have been associated to overfitting and poor transferability”

“A corresponding frequently asked question in the DFT community is: “how many parameters does this functional have?”, implying that functionals with more than four or five fitted parameters are barely useful elephants in the DFT zoo.”

“Instead of playing the game of counting fitted parameters in “parametrized functionals” and compare them to hidden parameters in “zero-parameter” functionals, ...”

“The fight between counting parameters and analytical fits, ...”

The author needs to provide some references to support his statements. In addition, very strong wording is used in the article that in my opinion is not appropriate for

scientific publications and seems a bit exaggerated, such as “treacherous”, “war”, “fight”. I advise to adjust that type of language.

I agree with the Referee on this point, and I have modified the language accordingly. For example, “treacherous” has been replaced with “rugged”, the two instances of “war” have been replaced with “odds” and “debate”, “fight” has been replaced with “dispute”. Moreover, several of the stronger sentences highlighted by the Referee have been either removed or redacted, and I have provided more context and additional references to the issue of counting parameters.

2) A large portion of the article, namely Section 2, feels very disconnected to the main core of this work in Section 3, which answers the main question of this article and provides a statistical analysis based on calculated data.

Instead, the article fits a function with a single parameter to reproduce enhancement factors of popular DFT methods. It turns out that that parameter needs to rely on an immensely large number of decimals to achieve appropriate reproduction of the training function. Maybe I misunderstood this section, but while it looks like an interesting exercise, I do not see the relevance and the scientific insight. Expressing an enhancement factor of a complicated functional—which by itself may depend on a large number of parameters—through a fit function with one parameter does not really mean that that functional has now become dependent on just one parameter. Ultimately, the enhancement factor it was fitted against had been developed with multiple parameters. Regardless of whether the number of parameters in that original enhancement factor causes problems in applications or not, the new single-parameter fit still inherits the virtues and problems of the original method. In addition, the large number of decimals also makes me question how easy it is to use that parameter on a different machine and to reproduce exactly the same function. Is it not possible that some of those decimals already contain numerical noise? Finally, I called this section “disconnected” because the final important part of the manuscript never refers back to Section 2.

My recommendation is to shorten Section 2 and to mention it as an interesting “fun fact” with most information being moved to a Supporting Information.

While I understand and respect the position of the Referee on this issue, I partially disagree on this point, since I believe that the section on the 1-parameter fit (Section 2)—albeit light in spirit—is indeed an important exercise to prove the main point of this article, as well as to showcase the features of the new publishing platform, which I believe was in part the topic of this special issue. In order to improve the connection between Section 2 and Section 3—following also the suggestion of Referee #2—I have included a new figure (Fig. 4) and a new paragraph that address and discuss the lack of correlation between the degrees of freedom (number of parameters) and the average ranking (transferability) of a functional. As such, I believe the exercise in Section 2 belongs in the main text, and I would avoid moving it to the Supplementary Information.

3) Table 1 presents a ranking of functionals and supports the author’s main point, namely that the number of degrees of freedom does not tell us anything about a method’s applicability. The ranked functionals closely resemble previous recommendations based on traditional benchmarking, which complements those previous works nicely. The database used by the author relies heavily on existing databases and benchmark sets, with most of them also addressing noncovalent interactions. Indeed, one can see an improvement when some form of London dispersion correction is used. Therefore, I feel that some functionals are misrepresented, in particular methods such as revTPSS, M11, N12, M06 etc. Dispersion corrections for those methods have been put forward and they have been shown to work. This is also the

case for Truhlar’s methods, as shown by different groups, e.g. (Goerigk, 2015), or Head-Gordon et al in JCTC in 2016. I recommend to also show results for those dispersion-corrected versions. In fact, including those numbers will do the author a favour. The number of parameters will increase, but the results improve, further evidence supporting this paper’s main message.

As the reviewer suggested, I have added to the results (and to the new Table 2) data obtained with revTPSS-D3(BJ), M11-D3(BJ), M11-L-D3(0), N12-D3(0), M06-D3(0) and M06-2X-D3(0) (M06-L-D3(0) was already present), as well as N12-SX-D3(BJ). This brings the total number of considered functionals from 53 to 60. These new results do not drastically change the analysis of the results, but they somehow contribute to strengthen the main conclusion of the paper, as the Referee suggested.

4) Given that this is a full paper without page limitations, I feel it is not appropriate to refer to the supporting information of a previous paper for references of functionals or benchmark sets. That would be a disfavour to all the authors that developed and wrote those articles. They should be cited in this paper too.

5) I also find that referring to (Morgante & Peverati, 2019) for technical details is not appropriate. The readers do not want to look up a different paper for important information.

Both of these points have been addressed by adding a new column to Table 1 that gives the corresponding reference. I want to stress out that I am particularly sensitive to the issue of not giving proper citation to functionals when they are used in the literature, and therefore the citations corresponding to each of the functionals used in this study were already reported in the previous version of the manuscript, albeit in a less organized way (ref. 50 – 87 in the old manuscript). The necessity to present the reference in a fuzzy format was in part due to the inability to cite a reference within a table in the Authorea platform (according to Authorea support, the feature is planned, but not yet available). In the new version of the manuscript, I’ve manually added each citation to the last column of Table 1, hoping that they could then be linked in production, when the article gets converted to the galley proofs.

6) Where the author cites previous benchmark studies by Goerigk et al and Head-Gordon et al., he may also want to include Jan Martin’s latest contributions published in the recently published Journal of Physical Chemistry Festschrift in honour of Leo Radom(Santra et al., 2019) and in the Israel Journal of Chemistry(Martin & Santra, 2019).

I have included both references suggested by the Referee.

Referee #2:

1. In the first part of the manuscript the authors present their one-parameter fit of exchange-correlation functionals. Although I appreciate the point the authors are trying to make, the fact that Eq. 5 cannot be used in practice to evaluate the functional for arbitrary values of the density seems problematic to me. A functional cannot be reduced to a finite set of points. As a simple example, one can look at the existing approximations to the LDA correlation, as most of them are fits to the same set of points (the Monte Carlo data of Ceperley-Alder). These functionals are clearly not equivalent, as they give close, but different results. The authors do point out that a spline interpolation can be used, but in that case it’s hard to argue that such interpolation is a one parameter fit. I guess one could instead make the point that in the limit of infinite number of points (and infinite number of significant digits in the parameter), Eq. 5 is able to represent a given functional for arbitrary densities. Would this be correct? Maybe the authors could comment further on this?

This is definitely an important point, and the Referee is correct in the assumption that the fits are exact in

the limit of infinite number of points. I have modified the main text to reflect this important clarification. The new portion of the text now reads (new and relevant parts highlighted here in *italic*):

This numerical representation becomes exact in the limit of infinite number of points, $N \rightarrow \infty$. As previously demonstrated (Peverati 2011), a grid of just simply $N=20$ points is practically sufficient to describe the enhancement factors of most exchange GGA functionals (e.g. PBE (Perdew 1996) and B88 (Becke 1988)) with sub-milliHartrees precision, when used in conjunction with a well-behaved interpolation between the points—such as a cubic or univariate spline. For a handful of more complicated functionals (e.g. SOGGA11 (Peverati 2011)) a slightly finer grid of $N=100$ points will suffice to achieve accuracies of $\sim 10^{-6}$ Hartrees. [...]

[...] It is important to notice that eq. 5 only reproduces the position of the points, while the spline interpolation is still required to obtain a continuous function over the considered interval (*an exact fit would require $N \rightarrow \infty$, and therefore an infinitely long encoding parameter*).

2. Another question that is not entirely clear to me regarding the one-parameter fit is if or how such procedure could be carried over to hybrid functionals and in particular to range-separated hybrids. I understand that this goes beyond the scope of the paper, but considering the popularity and importance of such functionals, I would appreciate if the authors could share any ideas they might have on this subject.

The extension of the 1-parameter fit to hybrid and range-separated hybrid functionals is non-trivial. While on one hand this could be done by encoding up to four extra parameters into the fit, on the other hand this would defy the infinite number of points limit discussed above, and it would represent just an *ad hoc* construction, as opposed to a true fit of the functional for any arbitrary density. I have modified the discussion at the end of Section 2 to provide more details on this issue. The new portion of the text now reads (new and relevant parts highlighted here in *italic*):

Extension to meta-GGA exchange–correlation functionals, as well as to functionals with more complex forms sitting on higher rungs of Jacob’s ladder, *could be achieved with various degrees of difficulty. For example, meta-GGA functionals depend on at least three variables that cannot be decoupled (e.g. the density, its gradient, and the kinetic energy density), and therefore they require higher dimensional interpolations. The interpolation using multi-dimensional grids and appropriate functions is not problematic, especially using available python libraries. A slightly more complicated case is the case of hybrid functionals (e.g. functionals that include a fraction of Hartree–Fock exchange), for which the parameter that represents the fraction of HF exchange could be encoded in the procedure, for example at the beginning of the sequence. For range-separated hybrid functionals, more complicated ad hoc procedure must be designed.* However, since representing functionals with one parameter has no inherent benefit for DFT as a method, going beyond the simple proof-of-principle described above has very little scientific merit and is not explored further in this context. A more rewarding endeavor is the search for a procedure that does not rely on counting the number of parameters to evaluate the transferability of functionals, as presented in the next section.

3. Looking at Table 3, there’s a somewhat obvious connection that could be made between Sections 2 and 3. At first glance, there seems to be no correlation between the ranking of the functionals and the number of degrees of freedom. This would be further proof of the inadequacy of counting parameters as a proxy for functional transferability.

The Referee is indeed correct on this lack of correlation. I have used this fact to present and comment a new Figure (Fig. 4) at the end of Section 3. This new material is useful to tie the results with the data in Section 2, as also part of the effort to address the concern raised by Referee #1. I want to particularly

thank the Referee for this particularly useful suggestion.

4. I couldn't find the ASCDB UE C.csv file that is used in the Jupyter notebook to generate Table 1 in the associated data. Also, it seems that the weights defined in Eq. 11 are hard-coded in the notebook. The manuscript explains how they should be calculated, but it is not possible to check the values without further data.

I have included the csv file in the github repository and the Binder webpage that is now associated with Table 1 (my intention was to do this in the first place as well, but I guess something went wrong with the Authorea platform and it wasn't where it was supposed to be, I apologize for this issue).

5. Finally, a very minor point. Although it has become customary to cite Ref. 53 of the manuscript when referring to the PW91 GGA, the functional was first described in (Perdew., 1991). I mention this because I've seen a few times people mixing up the PW91 GGA with the PW92 LDA as they assume (correctly!) that the year appearing in the functional abbreviation corresponds to the year in which the corresponding paper was first published.

I want to thank the Referee for this clarification, which was indeed a mistake in my previous manuscript (although a common one, as the Referee suggests). I have now cited the correct paper, alongside the 1992 paper that is customary cited for the PW91 GGA functional, as the Referee suggests.