Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2021-114-RC2, 2021

Comment on gmd-2021-114

Anonymous Referee #2

Referee comment on "Robustness of neural network emulations of radiative transfer parameterizations in a state-of-the-art general circulation model" by Alexei Belochitski and Vladimir Krasnopolsky, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-114-RC2, 2021

Unfortunately, the paper does not convince me in the current form. The paper is separated into two parts. The first part shows that a neural networks emulator that was developed for one model can be used within a different model and obtain reasonable results. This is interesting in principle. However, the new model is changed significantly to allow for the application of the old emulator (e.g. in terms of vertical resolution) and the quality of the solution with the neural network is not evaluated sufficiently. The second part is a long discussion on what should be done to obtain emulators that are working in different environments. While the discussion is interesting in principle, it is not backed up by the results of the first part (e.g. as only a single (shallow) neural network has been tested). The arguments of the discussion are intuitive, but most of them are not backed up by powerful references or by numerical experiments. I therefore think that the paper is not suitable for GMD in the current form. I hope the authors can improve the paper along the following suggestions.

Major comments:

The figures 1-3 do not leave me convinced that the neural network configuration is good enough. Can you show results that would convince me that the GFS version with 64 vertical levels is producing something useful (e.g., by comparing it against the default model with 127 vertical levels)? Can you show more diagnostics that compare GFS and HGFS for quantities that are no mean fields or at least provide plots of variability? Can you show vertical profiles that compare the neural network emulator and the default radiation scheme to visualise how different the solutions are?

Each sub-section of section 4 should show a clear connection to the numerical simulations. If this is not possible, the content should be either removed or backed up with a theoretical study. Section 4.1 should become part of the introduction. Section 4.2 should

be backed up by experiments because it otherwise remains speculative. The argument that SNNs are better than DNNs as they are more generalisable is not backed up with simulations and we also do not learn how non-linear the ML solution needs to be from the paper. Section 4.3 is also not backed up by any experiments. Section 4.4 and section 4.5 read like chapters of a textbook and are not very relevant for a paper that is not training new emulators.

Minor comments:

The description of radiation schemes in general is rather long with machine learning only starting at l.55

The neural network solutions that are used in the paper are not described in sufficient detail in Section 2.

l.65: I do not think that this very general statement holds. I am sure that I could break forecast skill with an unbiased, random, uncorrelated error if I wanted to.

l.77: I do not understand "blackbody cloud"

l.118: Why do you need to store the entire training data set for regression trees?

Figure 1: I assume that this does not cover the entire year 2018 but rather 1st Feb 2018 – 1st Feb 2019, correct?

Figure 2 and 3 are not "zonal means" as indicated in the captions. The vertical coordinate is not the "model level number".

l.208: " hybrid deterministic-statistical GFS" -- Why do you call it "hybrid"?

l.248: This should be moved to the introduction.

l.185: Yes, OK, the number of possible model states is increased. But how do you know that this is leading to difficulties in the emulation? I do not think that this can be shown mathematically.

l.281: It is hard for me to believe that there is no universal theorem for DNNs