Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-79-RC2, 2019

Geoscientific Model Development Discussions

Interactive comment on "Fast domain-aware neural network emulation of a planetary boundary layer parameterization in a numerical weather forecast model" *by* Jiali Wang et al.

Anonymous Referee #2

Received and published: 6 June 2019

This is a very interesting work. To the best of my knowledge, the authors are the first researchers who applied machine learning techniques to PBL parameterization. Their results are thought-provoking. Of course, the final evaluation of the developed NN must be performed in parallel runs of WRF with the original PBL parameterization and with the developed NN. However, it is an issue for a separate paper. I believe that this paper should be accepted after some revision and clarifications.

General comments:

1. It is not clear from the text if the authors developed a NN emulation of the PBL parameterization. NN emulation has the same inputs (sometimes augmented by additional metadata) and the same outputs as the original (in this case YSU) parameterization. Is this the case for the presented study?

2. Domain-aware NN is a confusing term. Which domain are the authors talking about (geographic domain, domain covered by inputs in the input space, etc.)? From the sentence in the paper: "a key drawback of the naïve FFN is that it does not consider the underlying PBL domain structure, such as the patterns that are locality specific and the vertical dependence between different vertical levels of each profile", it can be concluded that it is about vertical correlations between different vertical levels. Probably "domain-aware" name is misplaced (see also comment 5).

3. I cannot completely agree with the aforementioned (in comment 2) sentence from the paper. First, FFN does accounts for vertical correlations in profiles because all level components of profile are built from the same neurons of the previous hidden layer. Second, in this particular study, as is explained in the text, all outputs (including all vertical components of the same profile) are normalized independently, which significantly reduces the sensitivity of FFN (and any NN) to vertical correlations between levels. To preserve vertical dependencies, a profile should be normalized as a whole but not each component independently.

4. The time length of data set to be used for training is not a valuable and universal characteristic. It depends on representativeness of data set, i.e. on how well the variety of atmospheric states is represented in the training set, or how well the domain of input space is sampled. For example, including in the training set more grid points would enrich it with new/different atmospheric states and made more representative. It may shorten the time length of data required for training.

5. Most/all problems with applications to neighboring grid points can be alleviated or completely removed if all grid points (entire grid) are included in training set together with some metadata, i.e. if lat, lon, and the terrain conditions (elevation etc.) are included as additional inputs at each grid point. The NN trained in such a way I'd call

“domain-aware NN”.

Specific comments:

1. Is the sizes of the input and output layers are 16 + 85 = 101? (= near-surface variables) and 85 (= 17 vertical levels $\times$ 5 output variables).

2. It is not clear from the text how 17-level profiles produced by NN are integrated with WRF profiles in total 38 level profiles?

3. How many hidden layers have different NNs that the authors use?

4. How many trained parameters (NN weights) has each of these NNs?

5. How many records has the training set that is used?

6. Did not the authors try to train a shallow NN with the same number of weights as the best DNN has?

Without all this information it is difficult to understand why NNs with different architectures perform so differently.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-79, 2019.