General comments

This paper deals with evaluation of ML emulators of radiative parameterizations in General Circulation Models. In particular, authors aim to verify if the results of model calculations with original parameterization and with its ML emulator are identical. Comparisons have been performed between results obtained with GFS at C96 horizontal resolution and an hybrid deterministic-statistical GFS (HGFS), in which the original parameterization schemes have been replaced with shallow-NN-based emulators of LW and SW RRTMG developed for CFS.

The problem has been explained clearly, the topic discussed in this work is very interesting, so the manuscript is potentially suitable for publication. It furnishes a good contribution in the context of existing literature. However, before I can recommend the publication, several issues must be addressed.

My main concern is about the organization of the paper, since in the present form it is not so easy to be read. In particular:

- The Introduction is too long and does not allow the reader to have a clear frame of the problem under study. I suggest to split it into two parts: one about the main concepts, a synthesis of the state of art, and the aims of the present work; another one about all the other technical details of the emulators developed at ECMWF, NCEP et al...
- The title of Section 2 does not reveal the content. It is true that authors explain the differences between the two models, but they also explain how they have modified the setup of GFS in order to perform the tests described.
- The discussion section is interesting, but the concepts described are quite generical and
not specifically linked to the results obtained in the present work. In particular, concepts expressed in subsection 4.1 would fit better into the Introduction. Subsection 4.2 is mainly theoretical: I suggest to shorten it and to move it in another section.

- Analysis of results is mainly qualitative and performed only through analysis of Figures. I suggest to add some Tables with numerical values of specific indicators (bias, rmse etc...) in order to have a detailed quantification of the differences. In other words, Results section must be strengthened, since it is quite weak in the present form.
- Similarly, the Conclusion section is too generical and quite weak.

Specific comments

Line 60: What do you mean with “results are physically identical”?

Lines 70-71: I suggest to introduce a sentence to make a better link between the concept expressed in the paragraph 65-70 with the one expressed afterwards.

Lines 151-154: In my opinion, this paragraph is not properly linked with the previous text, but rather seems a “stand alone” one.


Line 211: “e-folding” is scientific jargon. I suggest to avoid it.

Lines 203-208: “In addition ... or HGFS”. This paragraph does not fit well into the “Results” section. I would move it into the previous section.

Line 207: “with radiative transfer emulators developed in 2011”. This sentence is too generical, please use more precise terms.

Line 216-218: “Note, that... experiments”. I think that this information is not so relevant for the reader, being more related to management policies of the supercomputing center.

Figures 1-2-3: Put a proper label on horizontal and vertical axes. Put the unit of measure on the colorbar.
Line 230: Explain the meaning of TOA (Top of Atmosphere).

Figures 2-3 (caption): “Vertical coordinate shows model level number.” It seems to me that latitude values are shown on the vertical axis, please clarify.

Line 279. Acronym SNN is introduced here for the first time. Explain the meaning.