Review for Zhong et al., 2022: ***WRF-DL v1.0: A Bridge between WRF v4.3 and Deep Learning Parameterizations and its Application to Atmospheric Radiative Transfer***

This is a fairly well-written and pleasantly concise paper on a new module developed for WRF which, as I understand, can be used to communicate with any existing ML/DL libraries written in or accessible from Python, and easily implement ML-based parameterizations in WRF. While I lack the expertise to comment on computational and programming aspects of the coupler, I think it's clearly a good idea. The demonstration with a radiation emulator is relevant and useful. I have three more important points which should be addressed:

1. Before reading the review by RC1 I did not understand that the coupler can be used to communicate with any Python script, and not a specific DL library. For the non-technical reader like myself the basic principles, and how a user would go about to use this coupler in practice, should probably be clarified in Section 3.

2. If you are hoping that other people used this tool, should it not be added to a proper code repository like Github where it can be maintained?

3. I agree with RC1 about the evaluation in 5.2 being problematic, especially if you call it "accuracy". If you are measuring anything here, it is more likely "generalization" (it's not surprising that the complex models perform worse, they are overfitted), but it is altogether not particularly useful to compare a couple realizations of instantaneous fields. However, since the topic of the paper is the coupler and not the emulator, I think it's OK to keep this basic online evaluation if the limitations are mentioned.

Besides this, I only have minor comments, mostly concerning clarity and grammar, which are easily addressed. After the authors have done so, I recommend that the paper is published.

Minor comments

L69. ***"DL-based radiation emulators were studied most..."***
Technically the emulators in most of these papers were not DL, as they were based on shallow neural networks. You could change to "ML-based", or perhaps just leave it as it is and choose simplicity over correctness

L72: ***"...such as ECMWF IFS"***
RRTMG, the full radiation scheme you used in your study, is not used in the IFS. IFS uses ecRad (Hogan & Bozzo 2018), which is a modular radiation scheme that has the gas optics component from RRTMG, but is otherwise completely different. You could either not mention the IFS or you could add, "and its gas optics component is also used in the ECMWF IFS".

L99: ***"three-dimensional variables are converted to be indexed a (i,j,k)"***
This sentence is confusing, are you converting from 3D to 3D?

L120: ***"While GPUs are more powerful than CPU's"***
In my own tests, GPU's have not necessarily been faster for small networks, which might be used for parameterizations. You could instead write *"While GPUs are typically more powerful than CPUs for DL inference, both CPUs and GPUs are supported"*.

L135: ***"..for WRF compute processes" → "..as for WRF compute processes"***

L144. ***"This WRF-DL coupler is superior because.."***
Superior to what?

L151. The compiler version could be added.

L160-163. ***"Yayo et al. (2022) demonstrated that the bidirectional..."***
As far as I'm aware (the Yao study seems to be a preprint?), my own paper (Ukkonen, 2022) was the first to use bidirectional RNNs for radiative transfer and demonstrate they're more accurate than FNNs, so the paper should be mentioned here

Table 1: Does "layer number" refer to the number of bidirectional layers, meaning model E actually does 10 traversals through the vertical column (2*5)? I am asking out of interest, perhaps it need not be clarified in the paper.

Figure 6: You should mention the problem size (number of columns and their vertical resolution) in the caption.

L192. ***"emulators run" → "emulators are run" / "emulators were run"***

L193. End of sentence: ***"CPUs" → "GPUs"?***

Table 3: Problem size should be mentioned. I think all emulator studies should report the time per column in one way or another, and the hardware, so we can compare results between studies: it is a problem in many studies that only the speed-up is reported (but this is always relative, the original code can be slow)

L203. ***"When GPUs are used for inference, all DL-based emulators except model E are at least three times faster than the original RRTMG scheme"***
But what if you had used the same dedicated GPUs for RRTMG, would it still be faster? This comparison has some caveats which should be mentioned. It is unclear if it's the software (ML model) that is faster or the hardware.

L219: ***"model A. Model C". → "Model A and Model C"?***

L233-235: This statement could be clarified a little bit. For instance:
"..has the potential to be faster, or by training on observations or detailed high-resolution models, even more accurate than conventional parameterizations"

L238. ***"Some researchers..."*** This sentence is confusing. If you are referring to Fortran-Keras bridge, perhaps you mean:
"Some researchers used simple Fortran-based NN libraries, which could convert existing models trained in Python to models usable in the Fortran framework."

L241. Remove "art"

L242. *"..as the implementation of DL-based parameterization is achieved in Python"*
This sentence is a bit convoluted. *"...as it can communicate with Python-based frameworks"?*

References

Hogan, R. J. and Bozzo, A.: A flexible and efficient radiation scheme for the ECMWF model, Journal of Advances in Modeling Earth Systems, 10, 1990–2008, https://doi.org/https://doi.org/10.1029/2018MS001364, **2018**

Ukkonen, P.: Exploring pathways to more accurate machine learning emulation of atmospheric radiative transfer, Journal of Advances in Modeling Earth Systems, p. e2021MS002875, https://doi.org/10.1029/2021MS002875, **2022**