Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2021-267-RC1, 2021

Reviewer comment on gmd-2021-267

Anonymous Referee #1

Referee comment on "Model calibration using ESEm v1.1.0 – an open, scalable Earth system emulator" by Duncan Watson-Parris et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-267-RC1, 2021

General comments:

Watson-Parris et al. describe a new tool called ESEm for model emulation and calibration, and demonstrate its use through a variety of examples. ESEm seems like a valuable addition to the climate modeling research community, and appears to be very flexible to allow the user to adapt it to a variety of purposes. It is also open-source and easily available via GitHub. In general the manuscript is well written and organized. Some of the details could be clarified or expanded on, and I have highlighted these in the specific comments below. Once these are addressed, I recommend the paper for publication in GMD.

Specific comments:

Figure 1: What is represented by the arrow going from observations to model data (via collocation)? Does that represent resampling/regridding? It almost makes it seem like the observations are used for training, which I believe is not the case.

Figure 1: Suggest adding “Calibration” somewhere to the flow chart, since it is a term mentioned in the caption and the text. Maybe in the same box as “Inference”?

Line 54: What determined the three options for emulation? Were other machine learning and/or statistical models considered?

Line 115: How were the parameter prior distributions determined in this case?

Line 118: Perhaps mixing terminology here: "five of the simulations are retained for testing" should that be validation instead of testing? Figure 1 and the accompanying description, as well as Line 138, use "validation". Though later you discuss all three sets (e.g., Line 161ff), so some clarification on terminology is needed.

Line 119: Which model fields are emulated? AAOD only or others as well?

Section 3: Is there anything built into ESEm to do hyper-parameter optimization/tuning? Or any suggested packages to automate that process? Suggest including that information somewhere in this section.

Line 150: Should that be Section 4?

Line 159: Is resampling required for using the ESEm package?

Line 167ff: An alternative to this could be to apply feature importance tests to the trained emulators; it would be interesting to see how the results compare to filtering parameters before training via BIC/AIC.

Figure 2 caption: This part is unclear: "that were not trained on these parameters"? Suggest re-wording the caption.

Line 219: Would Early Stopping also help with overfitting? As in, stopping training when the validation loss begins to increase beyond the training loss. How does that approach compare to dropout?

Line 255: I wonder if a simple ANN would perform better. You could also reduce the number of layers/nodes to reduce the number of trainable parameters.

Section 4 is nicely written to describe parameter calibration theory and how it is applied in practice for ESEm.

Line 386: Equation 5 is labeled twice.

Figure 4: The scatter outline is difficult to distinguish by eye, perhaps the border width could be increased? I would also suggest adding a line to the text to help the reader interpret the meaning of this second colorbar (red/blue values). Or perhaps changing the colorbar label to something more intuitive, as described in the figure caption.

Figure 5: Suggest changing “scatter points” to “crosses” or something similar, to be clear which is which.