

Anonymous referee #1

The authors provided some clarifications which make it easier to understand the proposed methodology and the underlying rationale. There are still paragraphs though that are quite hard to follow, for example the paragraph between l.76-83, which I only understood after two iterations between this paragraph and Table 1.

Answer: Thank you for noting the improvement in clarity. The paragraph mentioned is really just a initial outline. The approach is not intended to be understood at this stage. The approach is explained in much more detail throughout the paper. And we think there have to be some formulas in the description of the approach. These equations we did entirely move to the Table 1 to make them easier to find. Consequently, the interplay between written text and Table 1 is intentional. We think this separation helps also with reviewing and implementation of the whole method just by looking at Table 1 (after an initial read of the whole paper of course).

It is also helpful that the authors now state more explicitly that for the proposed methodology to work well, it is crucial that "all ParentModel_i should be closely yet meaningfully distributed around ParentModel₀". But this is where I still have the biggest concerns. While different examples of how this can be achieved are given, I still don't see how one can know in advance if the chosen regularization strategy does indeed entail closely yet meaningfully distributed fine-tuned models. There are some hyperparameters (e.g. 'n' or the number of epochs used for fine-tuning) related to the suggested regularization that must be chosen, and my feeling is that those can be quite important for the effectiveness of the proposed approach. How were they chosen here? My concern is that the authors tried a few different options and went with the ones that worked. But in a completely different setting where I don't know how good child models should look like, how can I make these choices and be sure in advance to end up in the case where the weight extrapolation is successful? The discussion of example 1 is quite useful to understand what can go wrong. Here, the poorly performing child models are still the minority, but how can I be sure that it's not a majority in my application?

Answer: Thank you for this question. The distribution of the ParentModel_i around ParentModel₀ is no hidden information as it is entirely generated without out of distribution (OOD) data. The distribution can be plotted and analyzed at for every weight or globally for the entire neural network by defining a suitable integral of the weight differences between the ParentModel_i and ParentModel₀. We now include this explicit suggestion into the manuscript:

“To choose a suitable approach for the ParentModel_i generation (Tab.~1, step 2a), the distribution of the ParentModel_i around ParentModel₀ can be easily plotted and analyzed, since this distribution does not depend on OOD data. The analysis could either focus on single weights or be based on a NN-wide integral view of weight differences between the ParentModel_i and ParentModel₀. ”

Thank you again for reviewing the manuscript and for providing your insights.

Anonymous referee #2

Minor comments:

-line 57: “these NN should”: you mean “these NN could”? Sentence 3 and sentence 5 in this paragraph are long and complicated – I don’t know if I really follow what you are trying to say. Try to simplify the sentences.

Answer: We specify “and these NN may be used on OOD data.” And we rephrased the respective sentences to:

“The reason is that many tasks do require NN, especially when these other approaches cannot be used. Examples of these tasks include reasoning and decision making, conditional generative AI, big data, and large classification problems such as image feature detection. These are applications in which extrapolating or averaging (cf. Hsieh, 2023) NN output will not give meaningful results. Our study focuses on problems where NN are used for various reasons (ease of use and implementation, existing workflows, compute limitations, invertibility, lack of alternatives, etc.), and these NN may be used on OOD data.”

-Table 1: 3. Weight-regression: 2nd point: Why mention R_a when it is not needed until 4. Weight-prediction? Mentioning R_a here is actually confusing.

Answer: We think it should be mentioned at this step. At this step you plan your predictors for the regression and the prediction. Here this is a reminder that the predictors for the weight-regression have to be known and prepared for the training region AND the target/OOD region as well. This you have to take into account when planing your set-up at this step.

-Line 112: “unified the inputs”: You mean you standardized the six inputs?

Answer: It is not about normalization or standardization. Here we state that the regression of the NN weights does not have more information than the training of the NN itself, i.e., the weight predictors are also part of the NN input. The whole sentence reads now: As we did not want to limit the parent NN by withholding information, we unified the inputs of parent NN training and weight-regression in experiment 2, i.e., the weight-predictors are deliberately included in the NN inputs.

The English used is sometimes colloquial. E.g., line 127 and 131 and elsewhere, the “, too” is not needed. Line 131 “so-called” is also not needed.

Answer: We redacted the manuscript accordingly.

-Line 136, the vertical slides are oriented east-west or north-to-south?

Answer: We specify “As input to the NN, we use 2d velocity fields representing vertical slices through the Atlantic Ocean along the latitude of 26°N”

-Line 136: “zonal velocity” here but “horizontal velocity” in the next sentence?

Answer: Its right as it is, meridional and zonal velocity are both horizontal velocities. We changed the sentence to “maps of horizontal velocity”. Before it was “horizontal velocity maps” - a bit confusing maybe.

-Fig.1: Might improve legibility (caused by overlapping lines) if the red and orange lines are replaced by symbols such as dots or crosses.

Answer: Symbols did not work as they lose a lot of detail – especially in the high frequency variability. We changed the plotting order of the lines and switched to dashed lines for the ground truth which is now plotted last. This did greatly improve the legibility. Thanks a lot for bringing this to our attention.

-Line 306: is it wind velocity or wind speed (line 327 and Fig.6 caption)?

also: -Line 319: what type of data? I have no idea what predictor variables are used to predict the wind speed.

Answer: It is zonal wind speed, i.e., a velocity. We specify “The training and evaluation periods span 2012–2017, using six-hourly data of the 10 meter zonal wind component, from two reanalysis products (ERA5 and CFSv2, cf. Sec. 3.3).”

-Line 327: I think "uncertainty" is redundant, since you already said “RMSE in wind speed”.

also: -Line 338: “of the uncertainty estimates” seems redundant.

also: -Fig.6 caption: “uncertainty” seems redundant.

Answer: In all these cases “uncertainty” is not redundant as the study we took the experiment from (Irrgang et al., 2020) the goal is to estimate (zonal) wind velocity uncertainty. In the respective chapter in our manuscript it is clearly stated that “this experiment assesses the proposed weight-prediction framework in the context of global wind velocity uncertainty estimation.” (Line 306)

Thank you for reviewing the manuscript again and providing the above suggestions.

Anonymous referee #3

I would like to thank the authors for their thorough and detailed responses to my queries. All of my questions and comments have been satisfactorily addressed, and I am now pleased to recommend this manuscript for publication.

Answer: Thank you for the positive feedback and for reading our paper again.

Uploaded files validation - Notification to the authors:

1. Kindly ensure that all materials belonging to the appendix are grouped within the respective 'Appendix' section and are not separated or interrupted by other parts of the manuscript.

done

2. Some tables contain coloured cells and/or coloured values. Please note that this will not be possible in the final revised version of the paper due to the HTML conversion. When preparing the final version, you may use footnotes or italic/bold font instead.

done