Answer to Mr García Rodríguez, Referee #1

We thank Mr García-Rodríguez for reviewing the paper again. We answer point by point in the following with the reviewer's comments added in *red/italics*. Text added to the revised version of the manuscript is included here in *blue/italics*.

Thank you so much for the time taken to respond to my comments. However, I would like to address a few points in your response.
I understand that the model structure was inherited from the ICON model and that changing it was beyond the scope of this study. However, it is worth noting that GMD is a journal that emphasizes model development. Given that this article is categorized by you as a "Model Evaluation paper", it would be appropriate to include some discussion on it about computational aspects of the model being evaluated.

The guidelines clearly expose that:
"Model development papers in particular often include a large proportion of evaluation. Typically, this comprises a comparison of the performance of different model configurations or parameterizations."
"It is, however, common for pure evaluation papers to contain substantial conclusions about geoscience rather than about models, and such papers are not suitable for submission to GMD."

Considering these criteria, the current version of the manuscript has no discussion of the evaluation of the model. So, a more detailed discussion of these aspects would be more aligned with the journal's scope. Additionally, evaluate the ICON source code with FortranAnalyser and obtain a final score of "4.1" is a relatively high score compared to the existing climate models. Highlighting this result would help to contribute to the evaluation model and underline the importance of the quality code development. It would be beneficial to mention this explicitly in your manuscript.

The paper is not only a model evaluation paper in the strict sense given by the GMD guidelines cited by Mr García-Rodríguez. We have mainly focused on the tuning of the gravity-wave parameterizations, and have evaluated the result of this in a climatological sense, but also regarding stratospheric winter variability.
A model evaluation in a computational sense, as requested by Mr García-Rodríguez, to achieve robust conclusions about, e.g. scalability, would require additional simulations with different MPI configurations. This would introduce a completely different topic to the manuscript and would lead to a further enlargement.
We have added the following to Section 2.1 of the manuscript:

The model code used in this study is published on Zenodo (Kunze et al., 2024). It is based to a large extent on Fortran. We checked the code quality with FortranAnalyser (García-Rodríguez et al., 2024) and obtained a final score of 4.1, which is a relatively high score compared to the existing climate models (M. García-Rodríguez, personal communication, February 16, 2025).

Regarding the questions on software engineering practices, I acknowledge that the primary objective of your study is physical model evaluation rather than software development. However, I had expected a more detailed answer according to performance, maintainability, and scalability. Even if they were not the main focus of your study, they remain relevant to the long-term evolution of the model.
Software engineering aspects may not be the primary interest, but they are nonetheless significant in the broader context of model evaluation and development. While the study does not focus on software development, including some discussion on these aspects would enhance the manuscript's completeness and adherence to the standards expected by GMD.

There are ongoing efforts to increase the quality of the ICON model code, e.g. within the ICON consolidated concept (s. https://gitlab.dkrz.de/icon/icon-c). As these are ongoing works, in which none of the authors are involved, we do not want to discuss them in the current paper.

We have added the following to Section 2.1 of the manuscript:

The ICON model code is largely equipped with OpenACC directives and in special configurations ICON has been deployed on GPU computing architectures (Giorgetta et al., 2022). However, the upper atmosphere extension is not ported to GPU and we performed our model simulations on the CPU architecture of the DKRZ (German Climate Computing Center) HPC system. Using the message-passing interface with 20 cores and 2560 processors, one year of simulation requires a wall-clock time of 62 minutes.

References

García-Rodríguez, M., Añel, J. A., and Rodeiro-Iglesias, J.: Assessing and improving the quality of Fortran code in scientific software: FortranAnalyser, Software Impacts, 21, 100 692, https://doi.org/10.1016/j.simpa.2024.100692, 2024.

Giorgetta, M. A., Sawyer, W., Lapillonne, X., Adamidis, P., Alexeev, D., Clément, V., Dietlicher, R., Engels, J. F., Esch, M., Franke, H., Frauen, C., Hannah, W. M., Hillman, B. R., Kornblueh, L., Marti, P., Norman, M. R., Pincus, R., Rast, S., Reinert, D., Schnur, R., Schulzweida, U., and Stevens, B.: The ICON-A model for direct QBO simulations on GPUs (version icon-cscs:baf28a514), Geoscientific Model Development, 15, 6985–7016, https://doi.org/10.5194/gmd-15-6985-2022, 2022.

Kunze, M., Zülicke, C., Siddiqui, T. A., Stephan, C. C., Yamazaki, Y., Stolle, C., Borchert, S., and Schmidt, H.: Supplementary information on - UA-ICON with NWP physics package (version: ua-icon-2.1): mean state and variability of the middle atmosphere, https://doi.org/10.5281/zenodo.13927891, 2024.