Reviewer 2

In this manuscript, the authors have comprehensively shown the results of a multi-year ICON simulation with a NWP configuration and a 2.5-km horizontal mesh. The analyses focused on mean precipitation and radiative fields and global statistics of local-to-synoptic-scale weather phenomena such as MCSs, tropical cyclones, and equatorial waves, compared with several observational data sets. The evaluation of these aspects suggested that complied statistics of local winds and precipitation as well as the mean precipitation distribution are reproduced relatively well, whereas the simulation still has inevitable biases, as found in the morphology of MCSs, tropical waves (incl. the MJO), and radiative budgets.
I acknowledge that this is the first step of addressing weather and climate using this version of ICON, and that it is worth being reported as one of milestones. Meanwhile, some of the presented results are interpreted speculatively without concrete evidence. Also, I wonder what clear merits of this ICON simulation are, compared to preexisting km-scale regional climate simulations, because the results except for global mean fields weigh heavy on regional statistics of meso-to-synoptic-scale variability (rather than large-scale circulation), which could have been addressed by regional modeling. Furthermore, I feel that the Introduction does not precisely describe historical advances in global km-scale modeling, with too much emphasis on the recent trend. Based on these points, I think this paper must undergo major revision before it can be considered for publication.

We thank the reviewer for their insightful and constructive comments. We carefully revised our manuscript accordingly which helped to largegly improve the document. Please find our responses to your comments in blue font below.

[Major comments]

1. The authors have tried to interpret a source of the biases in several subsections: for example, reasons for high temperature biases over land, for lower differences at high wind speeds over several regions, and for underestimated equatorial waves. However, I wonder if their interpretation could be done by relatively narrow insights without sufficient evidence. While I do not intend to request very detailed analyses, it would be better to provide more hints about the emergence of biases for the future model improvement. The suggestions and/or issues are listed below.

Thank you for raising these points. One of our intentions in writing this paper was indeed to highlight areas of model deficiency to support future model development. An in-depth analysis of the identified deficiencies is not possible within this overview paper, but formulating testable hypotheses about the sources of the biases is highly valuable. Your points below helped to highlight potential sources of model errors that we have not listed in our initial manuscript.

* Impacts of cloud radiative forcing, in addition to the issues of land-atmosphere coupling, on the surface temperature bias (cf. Sections 3.1 and 3.2)
We added a misrepresentation of cloud radiative forcing as error source for near surface temperature and energy biases to Sections 3.1 and 3.2.

* Representation of extratropical cyclones around CNA, WCE, and eastern North America (cf. LL. 307-310)

We added the simulation of extra-tropical cyclones as a potential error source in the revised manuscript:

L338-9: The region with the most pronounced underestimation of UV10 extremes is Eastern North America (Fig. S4), which might be partly related to a low difference in tropical cyclone frequency, as we will see in section 3.5**, but regional differences in the simulation of extra-tropical cyclones could also contribute**.

* Equatorial Rossby waves are affected by coupling between moisture and dynamics (e.g., Yasunaga and Mapes, 2012. JAS; Yasunaga et al., 2019, JCLI; Nakamura and Takayabu, 2022, JAS), not just by dynamics.

Thank you for bringing this up! We added the three references and mention that Equatorial Rossby waves are affected by coupling between moisture and dynamics in the revised paper.

2. While the authors presented the geographical variability of mesoscale phenomena (e.g., MCSs, diurnal cycles) in Figures 5, 6, 7, 11, and 12, I wonder how different it is simulated when comparing the accumulated results from the regional km-scale modeling framework. It is true that the presented results have global aspects, but they should be somewhat described without global models. I would appreciate it if the authors could discuss added values of this study to address the above points.

We thank the reviewer for raising this question. We agree that many mesoscale features discussed here, such as MCS characteristics and precipitation diurnal cycles, have already been studied with regional km-scale models. We will clarify that the added value of this study lies in analyzing these processes within one consistent multi-year global km-scale simulation, which enables direct comparison across regions and links mesoscale phenomena to larger-scale circulation features and tropical–extratropical interactions that are difficult to assess in regional domains alone. We will also note that this study provides the foundation for follow-up work that is already in progress using the same modeling framework to investigate how mesoscale processes, such as convective outbreaks, feed back onto the larger-scale circulation.

We added the following text to the revised manuscript:

Introduction:
L50-2: While many of the mesoscale phenomena examined here have already been studied with regional convection-permitting models, the added value of the present study is that it evaluates them within one consistent multi-year global km-scale framework, enabling direct comparison across regions and assessment of their links to larger-scale circulation features.

Conclusion:
L531-3: This framework also provides a basis for follow-up studies investigating how mesoscale processes, such as convective outbreaks, feed back onto and modify the larger-scale circulation.

3. The descriptions about the advances in global km-scale modeling are heavily biased by the recent trend observed in Europe. This kind of activities started two decades ago in the Japanese community with the Nonhydrostatic ICosahedral Atmospheric Model (NICAM), and it has published many research articles describing its importance of both weather and climate modeling. To ensure the correct historical advances in science, please reorganize the Introduction with appropriate citations (please see also the specific comments).
[Specific and/or minor comments]

Thank you for raising this important issue. We added seminal papers from the NICAM developers to the revised paper and explicitly mention the two decades of leadership in this area. For instance, we added the following to the introduction:

L53-7: The field of global km-scale modeling was pioneered by the Japanese community (Tomita et al., 2005; Miura et al., 2007; Satoh et al., 2008) through the development of the Non-hydrostatic ICosahedral Atmospheric Model (NICAM). During the last two decades almost every major model developing center invested in the development of next-generation, non-hydrostatic global modeling capabilities. At the same time, advances in computer technology made it feasibly to run global km-scale simulations for more than just a few days.

Title: I feel that the title is exaggerated compared to the contents in the main text. What are "global waves", despite not fully mentioning planetary-scale waves? I would like the authors to reconsider the title to be consistent with the fact that the present study mainly addresses global statistics of meso-to-synoptic-scale features.

We thank the reviewer for this thoughtful comment. We agree that the original title may have overstated the scope of the manuscript, particularly through the phrase “global waves.” We therefore revised the title to **“From Single Storms to Large-Scale Waves: A Multi-Year**

Kilometer-Scale Global Simulation,” which we believe more accurately reflects the paper’s focus on mesoscale to large-scale features within a global kilometer-scale framework.

LL.13-14: I do not fully agree on this speculation. I wonder if the poor representation of convectively coupled equatorial waves is attributed to misrepresentation of thermodynamic-convection coupling (cf., Takasuka, Becker, and Bao, 2025).

Thank you for making us aware of the Takasuka et al. (2025) paper. We changed the sentence to:

L14-5: These biases might stem in part from a misrepresentation of thermodynamic-convection coupling.

And added the following discussion to the conclusions

L519-21: The biases presented here might also stem in part from a misrepresentation of thermodynamic–convection coupling, which can strongly influence convective organization and its coupling to larger-scale tropical variability (Takasuka et al., 2026).

LL.14-15: I wonder if the main text, especially the concluding section, does not provide the sufficient value of multi-year global km-scale simulations, even though it tells us some information about biases.

We agree. Running the DYAMOD III setup constrained the simulation length to four years. However, we provide uncertainty information and significance testing where possible and are confident that the main results that we highlight in this study are robust since model biases recur robustly within each of the four years (e.g., continental dry biases in summer, lack of tropical waves, underestimation of tropical MCS frequency). A longer simulation would have helped to understand the model's ability of capturing large-scale modes of variability (e.g., ENSO), which is not the focus of this study. We added the following sentence to the conclusion to emphasize this point.

L532-3: Future studies will also focus on longer integration periods since four years are not sufficient to sample climate variability.

LL.25-29: "historically restricted its use to..." has been true in the main stream, but NICAM already succeeded in this type of simulation. I would like the authors to reflect the historical review by Satoh et al. (2019) in the revised Introduction.

We included a more extensive summary of the groundbreaking work of the NICAM group in the introduction of the revised manuscript.

LL.36-37: Sato et al. (2009, JCLI) also showed the better representation of diurnal cycles in the global km-scale model.

We added Sato et al. (2009, JCLI) in the revised paper.

LL.37-39: Miura et al. (2007, Science) showed the first success of the realistic MJO simulation and associated tropical cyclogenesis, featured by the realistically simulated multi-scale convective organization.

We added Miura et al. (2007, Science) as a reference to this sentence.

LL.49-50: The same comment as that for LL.25-29.

We rewrote this paragraph accordingly.

L.60: These were already (i.e., before 2020s) shown by Miura et al. (2007, Science), Nasuno et al. (2008, JAS), Holloway et al. (2012, JAS)...

We added these three papers as references to this sentence.

LL.65-66: Miura et al. (2023, BAMS) have also pointed out this direction.

Thank you for making us aware of this paper. We added the reference to this sentence.

L.129: slow-evolving waves -> slow-evolving variability (because the MJO is not a dynamical wave...)

Is corrected.

LL.228-232: In addition to this problem, it seems that a large bias of radiation budget (at the TOA and SFC) has impacts on T2m bias. Also, how are the distributions of cloud radiative forcing?

Thank you for this comment. We agree that biases in cloud radiative forcing can contribute to the evolution of temperature biases and added the following text and Fig. R1 to the revised manuscript.

L247-9: Additionally, misrepresented cloud radiative forcings frequently (Fig. S1) contribute to surface temperature biases in models with convection parameterizations (e.g., Ahlgrimm and Forbes, 2012) and km-scale models (e.g., Sakradzija et al., 2020; Lucas-Picher et al., 2024).

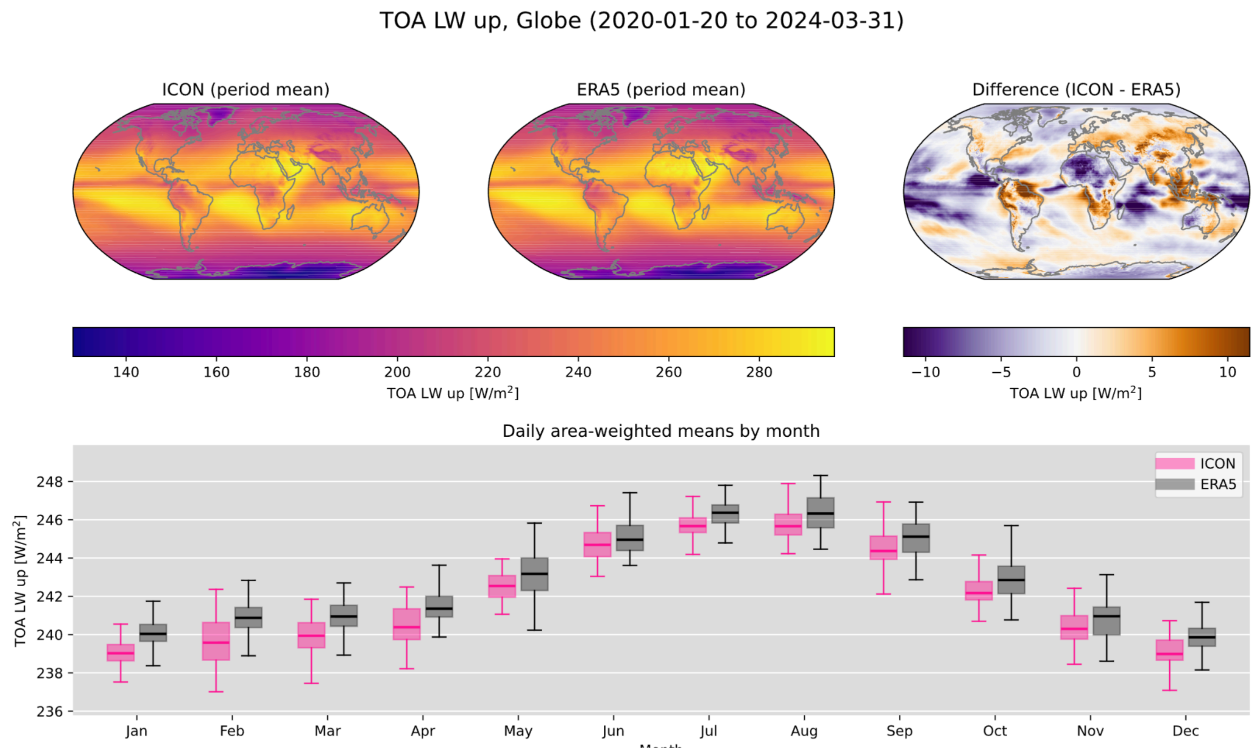

Fig. R1: Comparison of top of average atmosphere (TAO) outgoing lowngwave (LW) radiation in ICON (top left), ERA5 (top middle) and their difference (top right). Monthly average global mean statistics are shown in the lower panel.

LL.259-260: This does not necessarily hold true for all the regions; for example, over the North America and South America in the subtropics.

Thank you for raising this point. We rewrote this sentence to:

L279-82: Land precipitation is relatively well simulated overall, although dry differences emerge in several continental regions during boreal summer (Fig. 3g). In some of these regions, including parts of the central U.S. and Eurasia, the dry differences coincide with warm T2M differences (Fig. 1f), higher SWin, lower LE, and higher H, suggesting a contribution from land-atmosphere coupling and cloud radiation biases.

LL.266-271: Figure 2a shows the weak precipitation band in the Southern Hemisphere very near the equator. I think this can be a glimpse of a double ITCZ... Also, how about mentioning the reproducibility of precipitation bands in the mid-latitude?

We decided to keep the text about the improved ITCZ as is since we use the simulations presented in Segura et al. (2025) as a benchmark and clearly improve their default simulation without minimum wind speed adjustment (see Fig. 1). We also added a note that the location of mid-latitude precipitation bands that are associated with storm tracks is reasonably well simulated.

L.271: Note that Takasuka et al. (2024, JAMES) showed that resolving the double ITCZ problem was achieved by the reconsideration of microphysics.

We added the findings of Takasuka et al. (2024, JAMES) to the revised paper.

LL.291-292: I'm a bit surprised at this, because the phase lag of diurnal cycles of precipitation has been found in IMERG, which uses passive sensors for the estimation of precipitation, compared to radar-based precipitation products.

Thank you for raising this point. We added a reference to hayden and Liu (2021) that showed such a lag and mention it in the main article.

Hayden, L. and Liu, C., 2021. Differences in the diurnal variation of precipitation estimated by spaceborne radar, passive microwave radiometer, and IMERG. Journal of Geophysical Research: Atmospheres, 126(9), p.e2020JD033020.

LL.321-322: What is a possible reason for the low bias over the North Atlantic Basin? I wonder if this could be related to the underestimation of easterly waves (similar to equatorial waves, as shown in Fig. 9)

We started investigating this bias since it also appears in other km-scale global models and currency investigate how African Easterly Waves are aligned with MCSs in West Africa. It is to early to make a statement about this research in the current paper and we did not want to hypothesise about the reasoning of this bias.

LL.327-328: Is this attributed to the poor representation of rapid intensification? Also, Baker et al. (2024, GRL) should be cited somewhere in this paragraph, because they already showed benefits of km-scale models in representing tropical cyclones.

We have RI events present in our simulations but did not have room in this publication to investigate how these compare to observations. Based on the presented results the differences shown might partly be related to a mismatch in sustained maximum wind speed reported in IBTrACKS and the instantaneous 10 m wind speed used in the model. We added this explanation to the revised manuscript.

We also added Baker et al. (2024, GRL) to the introduction where we discuss TC simulations in km-scale models.

L.322: Equatorial waves -> Equatorial waves and the Madden-Julian oscillation (because MJO is not a equatorial wave.)

We adapted the section heading accordingly.

L.339, L.341: Please cite appropriate references.

Thank you for making us aware of this. We corrected the references.

Section 3.7: It would be better to provide a brief description about the criteria for the detection of MCSs. Readers may not read other paper carefully that introduces the methodology.

We added the definition of MCSs in the introduction:

L216-20: MCSs are defined as in Prein et al. (2024) following four criteria: 1) Continuous Tb ≤ 241 K area ≥ 40,000 km2 for ≥ 4 hours; 2) maximum hourly precipitation beneath the Tb ≤ 241 K area > 10 mm h−1 for ≥ 4 hours; 3) hourly precipitation volume > 20,000 km2 mm h−1 at least once during the MCS lifetime; 4) minimum Tb < 225 K at least once during the MCS lifetime.

LL.359-360: The same comment as that for LL.13-14.

We mention thermodynamic-convection coupling and refer to Takasuka, Becker, and Bao, (2025) in this paragraph in the revised manuscript.

LL.400-401: Is this also related to underestimated shortwave incoming (especially over ocean)? I wonder if water clouds have higher albedo and thus prevent radiation from reaching the surface.

We thank the reviewer for this helpful point. We agree that underestimated surface shortwave radiation over the ocean may reflect biases in cloud radiative effects, for example, overly reflective or overly abundant water clouds that reduce the amount of radiation reaching the surface. Because sea surface temperatures are prescribed in our simulation, this radiative bias does not feed back onto SST, but it may still indicate deficiencies in the simulated cloud field and associated surface energy budget. We clarified this point in the revised manuscript.

We added the following to the revised manuscript:

L444-6: Excess precipitation from clouds with warm cloud tops indicates overly active warm-cloud processes or too slow glaciation of tropical clouds in the ICON microphysics scheme. **This is consistent with the underestimated surface shortwave radiation over tropical oceans (Fig. 2c) that could be caused by overly reflective or overly abundant water clouds.**