Editor Comments

Dear Authors, please consider the following comments on your submission. First, there is something wrong with references in lines 339 and 341. Please correct. Second, and more importantly, figures are relatively small and thus it is difficult to see details. Please use the entire page width to show figures. This comment applies to most figures. Finally, I prefer to have figures embedded in the text in the vicinity when they discussed rather than grouped at the manuscript end. This makes reading the manuscript more easy. Please consider these comments while preparing the final version of the submission.

Thank you for these comments.

- We corrected the issue with the references in lines 339 and 341.
- We increased the sizes of the images to the full page width.
- And, we moved the figures to the approximate location where they are described in the text.

Reviewer 1

Review of: „From single Storm to Global Waves: A Global 2.5 km ICON Simulation of Weather and Climate“ The authors analyze a 4-year simulation of the ICOsahedral Non-hydrostatic model with a horizontal grid spacing of 2.5 km. The simulation has been done by porting ICON into GPUs in the ALPS machine as part of the EXCLAIM project. Through different analyses, the authors evaluate the simulation regarding the large-scale and mesoscale features of the simulations by comparing them with observations. The analyses point out that large-scale features are adequately represented (air temperature, precipitation, pattern of tropical cyclones, among others), but biases are observed in meso-scale or regional features. Overall, the manuscript is well written, and the amount of analysis is considerable. However, in the actual state of the manuscript, the little contextualization of the results within the current knowledge of km-scale Earth system community gives the impression that this is an isolated study. Moreover, it is not clear whether the authors aim to give some insights of the simulation or just present the simulations. I think that pointing out what the simulation of a horizontal grid spacing of 2.5 km gives compared to coarser or finer (if it exists) resolution will be a valuable contribution to the community working with km-scale Earth System models. Another criticism of this work is the word “climate simulation”, and I think that this has implications on the technical side. In the following lines, the major concerns about this manuscript are described.

Thank you very much for your constructive comments and feedback on our study. The comments largely helped to make our results more accessible and connected to existing literature. We carefully revised the manuscript according to your comments and replied to them in blue font in the pages below.

Major comments:

1. The manuscript gives the impression, starting with the title, that the authors will describe climate simulations. I struggled a bit with the word “climate” in this study for two reasons. One is the fact that the concept of climate involves different temporal variability, from synoptic to multidecadal. Second, a climate simulation should involve communication between components of the climate systems. While this is the case for land and atmosphere in this study, the ocean is not, i.e., sea-surface temperatures are prescribed. This means that the global atmosphere-only 4-year simulation presented in this study is not properly a climate simulation, i.e., too short for considering climate and only land-atmospheric interaction. These two points become relevant for the technical part. The throughput of this global atmosphere-only 4-year simulation is 0.25 simulated year per day (SYPD). This means that even in an atmosphere-only configuration, the throughput is much lower than what is expected when conducting climate simulations (ocean-land-atmosphere), which is 1SYPD. Also, this implies that with a dynamical ocean, the throughput will be lower. My suggestion is that the authors refer to the simulations as a 4-year atmosphere-only simulation rather than a climate one. Moreover, if the authors intend to provide

climate simulations in the future, it would be adequate to elaborate a bit more on how the authors will increase the throughput.

Thank you for raising this concern. We agree that 4 years is too short for sampling climatic processes. The throughput of the simulation could have been increased by using a larger number of cores (we added this to the revised document), but this should not be the deciding factor why a simulation should be called a climatic simulation, in our opinion. Similarly, running a coupled or AMIP-style setup has historically not been used to discriminate between climate or weather simulations but between coupled and uncoupled runs. Furthermore, running the model in a coupled setting will not necessarily make it slower, as we were not harnessing the full potential of the NVIDIA GH200 Grace Hopper superchips. The atmosphere and land were simulated exclusively on the Hopper GPUs. At the same time, the Grace CPUs were mostly idle, as it is difficult to efficiently use both GPUs and CPUs for concurrent or very tightly coupled processes in the atmosphere and on land. The ocean is much more loosely coupled and can thus be simulated using the CPUs (next to 4 GPUs, there are 4 CPUs with a total of 288 cores per node). Klocke et al. (2025) have done exactly this, and they mention that by doing so, they were able to run the ocean “essentially for free”.
Based on the short 4-year record, we decided to change the title of our manuscript and remove wording indicating that this is a climate simulation from the text. We, however, refer to the model setup as a climate setup to make it clear that this is a continuous run that produces weather largely uncorrelated with observed weather.

We agree with the reviewer, that there’s still considerable effort necessary to get close to 1 SYPD. However, our setup targeted computational efficiency and not throughput. Dipankar et al. (2026) show that the same model can achieve close to 0.6 SYPD on the same grid by running the setup on 640 nodes instead of the 240 nodes used here. Unfortunately, this lowers the computational efficiency. To make it clearer that this simulation was optimized for computational efficiency and to refer to Dipankar et al. (2026) for more detailed computational aspects and potentially further optimizations in throughput, we’ve added the following sentences:

L107-9: This setting was chosen for computational efficiency, and throughput can be increased by increasing the number of nodes. More information on the computational aspects of the specific model used for this simulation can be found in Dipankar et al. (2026).

2. I think that the manuscript will benefit tremendously if the authors describe more of the advances in the global storm-resolving and km-scale Earth System models. The first initiative of conducting global storm-resolving models was done by the Japanese community (Tomita et al., 2005; Miura et al., 2007; Satoh et al., 2008), developing the Non-hydrostatic ICosahedral Atmospheric Model (NICAM). Now, there are more centers developing global storm-resolving models (DYAMOND; Stevens et al., 2019; Satoh et al., 2020), and a few of them are coupling with the ocean, referred to as km-scale Earth System models (Hoheneggar et al., 2023; Rackow et al., 2025). So, there is a considerable spectrum of simulations using grid-spacing of 10 km or

less. In the ICON's universe, the model the authors are using, advances have been done to couple with the ocean (Hohenegger et al. 2023), conducting multidecadal climate simulations (Segura et al., 2025a), and recently, carbon cycle has been incorporated into the model, and a simulation was conducted using a horizontal grid-scaping of 1.25 km (Klocke et al., 2025). Having this in mind is important in order to contextualize the results of the simulation presented in this study. A simple question would be what is the added value of the 2.5 km simulation to the large universe of simulations being produced with coarser or finer (if they exist) horizontal grid spacing?

Thank you for emphasizing the history of global km-scale modeling. We expanded the text in our introduction about this topic with the suggested literature if it was not already included. Besides the comprehensive analysis of a large range of phenomena, the unique contribution of this study is to show how ICON with an NWP setup compares to ICON simulations that have been run previously. The major differences in, e.g., how ITCZ precipitation is simulated show that using different model physics has a large impact on the simulation. We included the following sentence to make this clearer:

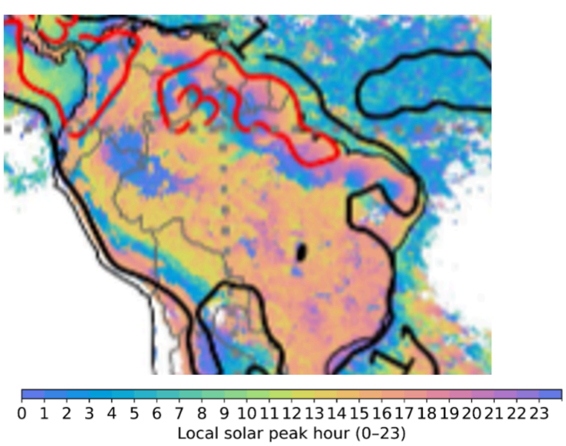
L81-3: We use a model setup that is closely aligned with ICON settings used in numerical weather prediction, which differs substantially in the used physics (i.g, land surface model, surface layer scheme, boundary layer scheme, microphysics, and radiation) compared to previously performed global km-scale ICON simulations e.g., Segura et al., 2025a).

3. I understand that this work is mostly to introduce the simulation, and future analysis will probably come in the future. However, I found a bit of disproportion in the explanation of the biases. Regarding precipitation and MCS's, the authors give a considerable explanation of how MCS's characteristics are biased compared to observations. Then, the authors go back to the large-scale, indicating that it's probable that most of the precipitation in the large-scale feature of precipitation is related to congestus and shallow clouds. Considering the other features analyzed in the manuscripts, the explanation most of the time is to cite other papers. For instance, the biases in temperature, according to the authors, are related to biases in soil moisture and are probably related to the lack of representation in the lateral flux of water. Nevertheless, there is no analysis of the soil moisture or the partition of energy between latent and sensible heat flux. Or are all the biases related to the lateral flux of water, or could it be related to local precipitation biases? In a similar direction are the tropical cyclones and maximum wind speed. The explanation of the biases lies in the difference between the time step of observations and the simulation. My suggestion is that the authors should reconsider which analyses they want to show. This could be based on what is actually shown in previous studies, selecting the one that provides new insights in the km-scale modeling community.

Relating the biases to previous studies is difficult, since this is the first time that ICON with NWP settings is run at a global scale. We relate biases to our historical experience with regional km-scale models (e.g., Barlage et al. (2023) show the importance of lateral groundwater fluxes for temperature and precipitation biases in the central U.S., similar to findings from Schlemmer et al. (2018) using the TERRA model). We carefully revised the paper to ensure that we can only hypothesize about the sources of some of the biases shown and base these hypotheses on existing literature. However, we believe that showing a large range of analysis and presenting where the model performs well and where major deficiencies are will help further development of ICON. Several studies are already underway that look into certain model biases in more detail, which are based on the results presented in this paper.

4. I wonder if the regionalization of the analysis concerning precipitation maximum and diurnal cycle provides insightful results, which does not seem to be the case. At the end, it seems that the good representation of such characteristics is region-dependent. In regional km-scale models, due to the spatial constraint of the boundary conditions, it is expected that precipitation characteristics in the region of simulation should be similar to observations inside the region. However, this is more complex for global atmosphere-only km-scale simulations. Small changes in the pattern of local winds, temperature, or soil moisture can change the precipitation characteristics locally. My suggestion here is to use a bigger domain to analyze the diurnal cycle of precipitation. For example, land or ocean, tropics and extratropics, similar to what was done in Segura et al. (2022) or Takasuka et al. (2024). This can also be done for the case of precipitation maximum.

Thank you for bringing up this point. Local/regional scale analysis is important because large-scale analysis that averages over areas with significant differences in the diurnal precipitation cycle cancels out regional patterns. E.g., the central U.S. has a well-studied JJA nocturnal precipitation peak that is related to nocturnal mesoscale convective systems that can be seen even if only one year of data is available. The small sampling uncertainties in Fig. 5 provide evidence that the diurnal cycle signal is above the noise level in most regions, and the results presented in Fig. 6 show very similar patterns compared to previously published studies that used longer records to analyze the diurnal cycle of precipitation over longer time periods. Fig. R1 shows a comparison of the peak precipitation diurnal cycle in our analysis to Pacini and Stevens et al. (2023) Fig. 4a. over the Amazon basin.



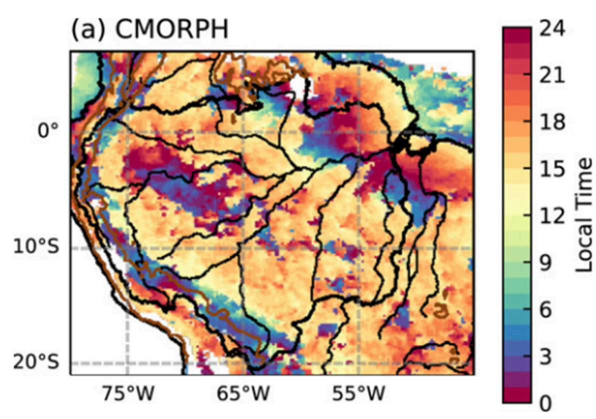


Fig. R1: Comparison of the peak diurnal cycle precipitation timing in our analysis (left based on IMERG; from our Fig. 6a) to Pacini and Stevens et al. (2023) Fig. 4a.

Specific comments:
-Line 24: The first ones in developing global storm-resolving models is the Japanese community, who developed NICAM (Tomita et al., 2005; Miura et al., 2007; Satoh et al., 2008).

We added this information to the introduction (L53), but did not change this paragraph since virtually all of the science and justification for km-scale climate modeling came from the regional modeling community.

- Lines 26-27: As I mentioned, global storm-resolving models were already conducted in Japan.

We decided not to mention these early efforts here, as the simulations were conducted over very short periods. We mention them in the third paragraph of the introduction. Similarly, we do not mention model developments in the late 1990s that paved the way for later non-hydrostatic modeling.

- Line 32: convection-permitting, storm-resolving, or km-scale atmosphere models? Aside from a philosophical/semantical question, try to use one nomenclature across the manuscript. I would prefer storm-resolving or km-scale atmosphere models. Convection-permitting could be arguable due to the type of convection being resolved: shallow, congestus-type, or deep.

Thank you for making us aware of this inconsistency. We remove the instances where we called these simulations “convection-permitting” with “km-scale”.

-Line 34: I would add Marsham et al. (2013)

We added Marsham et al. (2013).

-Line 35: What about Holloway et al. (2012) for precipitation extremes

Is added.

-Line 37: Also Marsham et al. (2013).

Is added.

-Line 44: I do not think that Hohenegger et al. (2009) showed that lateral groundwater fluxes help in representing soil moisture - precipitation characteristics. The authors evidenced that by explicitly resolving convection, there is a negative feedback between precipitation and soil moisture, i.e., dry soils can get precipitation. This is not observed when convection parameterizations are used. This has been studied using ICON (Lee and Hohenegger 2024), and was pointed out as a possible reason for a good representation of the terrestrial tropical rainbelt in ICON (Segura et al., 2022).

We were citing Hohenegger et al. (2009) to support the statement “Additionally, km-scale grids allow for a more faithful representation of land–atmosphere coupling…” but adding the citation at the end of the sentence was misleading. We moved the reference in the updated study. We also added (Lee and Hohenegger 2024) and (Segura et al., 2022) as references.

-Lines 49-50: Global km-scale atmosphere simulations were conducted already 20 years ago (Tomita et al., 2005; Miura et al., 2007; Satoh et al., 2008)

We added:
L53-4: 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).

- Line 62: Hohenegger et al. (2023) described ICON as a km-scale Earth System Model. Similar for IFS-FESOM/NEMO, which was described in Rackow et al. (2025). Segura et al. (2025a) showed that multidecadal simulations using km-scale Earth System Models (ICON and IFS-FESOM) are feasible.

We added these references.
L68-71: Ongoing developments extend these atmospheric configurations toward multidecadal coupled ocean-atmosphere-land models that include interactive ocean, sea ice, and biogeochemical components (Hohenegger et al., 2023; Segura et al., 2025b; Rackow et al., 2025) as well as the carbon cycle and aerosol emissions (Klocke et al., 2025).

-Line 70: A better term instead of “weather-to-climate” would be “meso- to large-scale systems” in this study. Climate has several connotations that are not totally addressed by the simulations in this study (interannual to decadal variability or a coupling system).

We changed the wording to “.. from single storms to global waves …” to align with the title of the paper.

-Line 95: Is this throughput enough to produce climate simulations?

Yes, if you are patient enough. We ran 80 years at 4 km grid spacing over North America at 0.17 SYPD. We chose the 0.25 SYPD primarily for efficiency rather than speed. We could have run faster using more nodes, but at lower computational efficiency. Dipankar et al. (2026) show that the same setup on 640 nodes can come close to 0.6 SYPD. Furthermore, there are ongoing efforts to further improve the performance of the model (Dipankar et al., 2026). We added the following to the updated manuscript.

L107-9: This setting was chosen for computational efficiency, and throughput can be increased by increasing the number of nodes. More information on the computational aspects of the specific model used for this simulation can be found in Dipankar et al. (2026).

-Line 227-229: So, the argument is that there is a lack of soil moisture due to a lack of lateral flux of water. My question is how much of this dry bias is related to a precipitation bias. In Figure 3g, there is a bias of 1.8 mm d-1 over North America, in the same region where the warm bias is. Moreover, I do not think that the lateral flux of water can explain the warm bias on a continental scale. Take, for example, the northeast Asia in the winter season (DJF) or the South America continent in JJA or SON. If the lateral flux of water is the culprit, could the authors elaborate a bit more on how it can explain biases on a continental scale?

This is based on our experience with continental-scale WRF simulations over North America (see Barlage et al. 2021). We saw similar dry and warm biases in the WRF simulations as we see here in ICON and performed a detailed investigation into the sources of these biases, finding that they are primarily caused by lateral groundwater flow. This lateral flow enhances the convergence of water in valleys, resulting in a net increase of evapotranspiration in large parts of the U.S. plains. This helps to mitigate the wet and dry bias to a large extent. This process is only relevant during summer in continental regions and has also been found over Eastern Europe in Schlemmmer et al. (2018). We updated this paragraph to improve to make this clearer.

-Line 238: Is 7.8 W m-2 referred to 6.8 W m-2 or 25 W m-2?

The 7.8 W m^{-2} reported by He et al. (2023) is intended to be compared with the overall median difference (6.8 W m^{-2}), whereas the 25 W m^{-2} value refers specifically to the noon peak in the diurnal cycle. We agree that this was unclear in the original text, and we have revised the sentence to remove this ambiguity.

-Lines 239-240: Is the overestimation in the frequency of shallow clouds the reason for a positive bias in downwelling shortwave radiation at the land-surface? Or is it for the ocean surface? If it is at the ocean surface, I can understand this argument for the subtropical region and eastern sides of the basins, but for the intertropical convergence zone, is that the case?

This sentence deals with solar incoming surface radiation over land regions compared to flux tower observations. Our simulation, and most km-scale simulations we are aware of (including the cited references), underestimate shallow clouds over mid-latitude land regions.

-Lines 245: What do you mean by surface evaporation or soil evaporation is not well represented? Do you mean the bulk formulas or the forcings as radiation, winds, or soil moisture? How is the ratio in the partition between latent and sensible heat flux? This would indicate that maybe there is a problem with the soil moisture.

We agree that the previous wording was too specific given the level of analysis presented here. From the current evaluation, we cannot determine whether the regional LE biases arise primarily from the bulk turbulent exchange formulation, from errors in the atmospheric forcing (e.g. radiation, wind), from soil moisture availability, or from the partitioning between latent and sensible heat fluxes. In particular, we did not separately diagnose the contributions of vegetation transpiration, soil evaporation, or interception evaporation, nor did we directly analyse the LE/H partitioning.

What our results show is that LE exhibits small median biases across stations, but with marked regional over- and underestimations. We therefore interpret these contrasted biases more cautiously as evidence of region-dependent limitations in the representation of land-surface controls on evaporation, potentially including surface heterogeneity, vegetation characteristics, and soil moisture availability, rather than attributing them to a single process.

We have revised the text accordingly:

L265-6: “This spatial compensation suggests region-dependent limitations in the representation of land-surface controls on evaporation, rather than a uniform model bias.”

-Lines 248-249: In Figures 2d,e, latent heat is slightly overestimated, but in Figures 2f,h, it seems that there is an underestimation. How is it possible? Is it because in the transition season, latent heat flux is overestimated?

Thank you for this comment. You are right that the original figure could suggest an inconsistency between panels 2d–e and panels 2f,h, and we re-examined this carefully.

First, these panels do not show the same level of aggregation. Panels 2d–e summarize the distribution of station biases when all stations and all seasons are pooled together, whereas panels 2f and 2h show the spatial distribution of noon biases for JJA and DJF only. Therefore, the latter do not represent the full distribution underlying panel 2e.

Second, during our re-analysis we identified a plotting issue in the original violin/boxplot panel. The x-axis limits were applied in a way that removed out-of-range values before the boxplot statistics were computed. For LE at noon, this excluded 35 strongly negative station biases below −100 W m^{-2}, which shifted the plotted median from the true value of +0.57 W m^{-2} to a more positive value of +4.51 W m^{-2}. We corrected this in the revised figure by keeping the full dataset for the statistical summaries and only restricting the visible plotting window afterward.

After correction, the all-season pooled LE bias is close to zero overall: the median station bias is +2.30 W m^{-2} for the mean difference and +0.57 W m^{-2} for the noon difference.

The apparent contrast with panels 2f and 2h is explained by seasonal compensation. When the noon LE bias is examined by season, the median station bias is:

DJF: −0.03 W m^{-2}
MAM: +9.01 W m^{-2}
JJA: −12.32 W m^{-2}
SON: +6.41 W m^{-2}

This shows that the pooled near-zero noon median results from compensation between a negative bias in JJA, an approximately neutral bias in DJF, and positive biases in MAM and SON. Therefore, the apparent discrepancy does not indicate a contradiction in the results, but rather reflects both seasonal aggregation effects and the corrected plotting issue in the original panel.

We have revised the figure accordingly.

-Lines 250-252: Could it be that the soil moisture biases are due to an underestimation of local precipitation?

Yes, but there is a positive feedback process that you kick off when you have too little precipitation. The main point of highlighting the lateral subsurface flow is that it increases local evaporation (moisture recycling) and therefore creates more favorable environments for forming precipitating storms, which in turn wet the soil. We were careful with our wording and consistently stated that parts of the bias are related to this process not being represented.

-Lines 252-253: What does it mean more site-specific?

We rephrased the sentence to:

L272-3: In contrast, there is more station-to-station variability for winter biases

-Lines 253-254: How do other variables make the result robust? Is it because warm biases is explained by surface heat fluxes?

We reworded this sentence to avoid misinterpretation:
L273-4: The spatial patterns and characteristics of surface flux differences are consistent with biases in temperature and precipitation.

-Lines 258-259: So, biases in shortwave radiation are not related to shallow clouds, as implied in Section 3.2, but rather related to deep convection.

The shortwave overestimation at the surface mentioned in Section 3.2 is a land feature (we made this clearer in the revised paper). As we show later in the paper, the overestimation of precipitation over equatorial ocean regions is partly due to a high fraction of rain falling from rather shallow clouds.

-Lines 259-260: Well-simulated land rainfall pattern in the tropics has been shown in Segura et al. (2022,2025b) with coarser horizontal grid spacing. So, does it mean that in ICON, rainfall patterns over tropical land are already well simulated, independent of the configuration?

This generalization is not possible. The model setup in Segura et al. (2022,2025b) is rather different from the setup used here, since we are using NWP physics that are different from most of the physics used in Segura et al. (2022,2025b).

-Lines 266-267: The double band of precipitation has been shown in coupled km-scale (Segura et al., 2022) and atmosphere-only km-scale simulations (Segura et al., 2025b).

We added these references.

-Line 274: Heavier precipitation intensities at hourly time-step than observations have also been observed in IFS-FESOM using horizontal grid spacing of 2.8 and 4.4 km without convective parameterization. Then, is the authors implying that using IMERG is not the best tool to characterize extreme rainfall events?

GPM-IMERGv7 and earlier versions are unreliable in representing grid-scale hourly precipitation rates and systematically underestimate intensities. The effective resolution of IMERG can be up to 100 km, meaning that heavy precipitation is smoothed spatially (Guilloteau et al. 2020). This can be shown when comparing IMERG precipitation with hourly station observations, as we did in this study or in Dominguez et al., (2024) and Yu et al., (2025).

We added Guilloteau et al. (2020) in the revised manuscript.

Guilloteau, C. and Foufoula-Georgiou, E., 2020. Multiscale evaluation of satellite precipitation products: Effective resolution of IMERG. In Satellite Precipitation Measurement: Volume 2 (pp. 533-558). Cham: Springer International Publishing.

-Lines 273-300: It's difficult to get a clear message of the regionalization. Are the authors showing that the IMERG is not adequate for studying the diurnal cycle of precipitation? But according to the regionalization, this is not totally bad. There are, of course, errors in the satellite data, but overall, and in particular in the intensity (Figure 5), it seems good.

IMERGv7 can capture the diurnal cycle of precipitation amount well, but due to error cancellations. It has too high precipitation frequencies and too low intensities in line with what we show in Fig. 4. Importantly, IMERGv7 has regionally dependent biases and agrees better with observations in some regions (e.g. WCE). We added a closing paragraph to section 3.3 to make this clearer:

L326-8: In summary, we recommend using GPM-IMERGv7 hourly precipitation with caution in model evaluation studies, given its tendency to overestimate precipitation frequency and underestimate precipitation intensity. Where possible, in-situ hourly rain gauge observations should be used.

-Lines 308-310: Is the underestimation of maximum wind speed related to the frequency of tropical cyclones? Or is it that the tropical cycles in ICON tend to have lower wind speeds than observations?

As shown in Fig. 8b, the frequency of TCs is well captured by our simulation, except in the North Atlantic basin. This means that the lower extreme wind speeds in ICON are typically not related to TC frequencies.

Fig. 8c shows that TCs in our simulation have lower wind speeds than observed TCs.

-Line 310: “Most tropical”? According to Figure S3, there are not many stations in the tropics. And most of them are focused on the Maritime Continent.

We have 512 stations between +-20 deg latitude, and 264 are outside of the Maritime Continent (90 E to 150E).

-Lines 313-314: It’s very hard to see how the autocorrelation gives a proxy of the structure of the wind-producing storms. Is it possible to elaborate more?

Thank you for raising this point. This was not very well formulated. We changed the sentence to:

L341-3: Also, ICON reproduces the temporal autocorrelation of hourly winds well (Fig. S4), indicating that it captures the persistence and day-to-day temporal variability of near-surface wind fluctuations, including both the rapid decorrelation at short lags and the diurnal recurrence seen in the observations.

-Lines 315-319: Does it mean that km-scale models need to be evaluated differently?

The issues outlined in this part of the paper are generalizable for all weather and climate models, independent of their grid spacing. Km-scale models are in fact easier to compare with station based observations since they resolve processes at scales that are more comparable to station observations compared to coarser resolution models where scale differences are larger. We added the following to make this clear:

L348-50: Km-scale modes are generally more comparable to station observations because the scale differences between the simulated processes on the grid and point-based observations are smaller than in coarser-resolution models.

-Line 326: Did you expect this? Takasuka et al. (2024) also showed this behavior in NICAM.

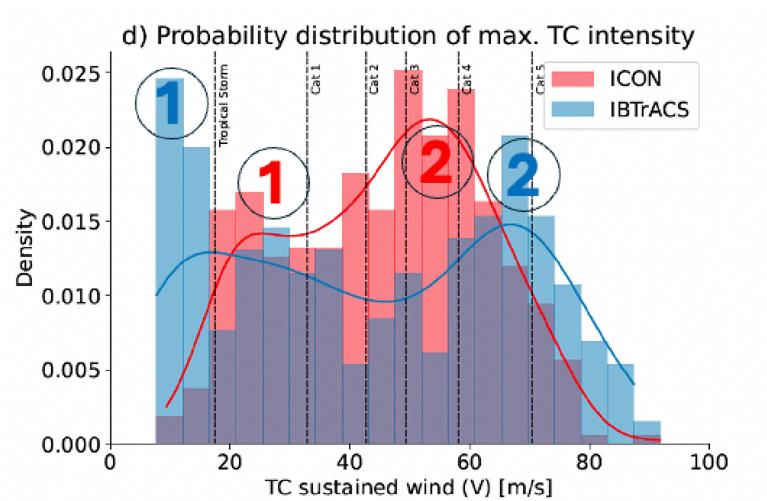
We were hoping for results that have about this quality. Thank you for making us aware of the Takasuka et al. (2024) study. We added it to the revised paper.

We also added the following text to emphasize differences between simulated wind and reported wind in IBTrACS.

L363-7: Similar to the comparison to station-based wind observations in the previous chapter, comparing modeled instantaneous wind fields to IBTrACS observations is challenging. The simulation provides gridded 10\,m near-surface winds, while IBTrACS reports storm-level maximum sustained winds from observational best-track analyses. They are therefore not directly equivalent, since the model resolves spatial wind structure and average wind speeds within grid cells, whereas IBTrACS summarizes cyclone intensity. This could explain parts of the low difference in wind speed seen in ICON.

-Lines 330-331: ICON simulates one peak, but observations have two. The difference in the distribution is due to the track method. This is the explanation for the first peak, but what about the second one?

Below is Fig. 8d with labeled peaks in IBTrACS and ICON. ICON has peak 2 shifted to lower wind speeds, and peak 1 is falling off too rapidly in ICON at 20 m/s. It is also important to note that densities are shown here, which means that peak 2 appears higher in ICON since tropical storms are not included in peak 1.



-Line 333: (Fig. 9), typo

Is corrected

-Line 339: typo

Is corrected

-Line 341: typo. Takasuka et al. (2024) also showed this.

Is corrected, and we included Takasuka et al. (2024).

-Lines 344: Takayabu (1994) was the first to use this method.

We added Takayabu (1994) in the revised manuscript.

-Line 351: Were Kiladis et al. (2009) the first ones to show this?

No. Kiladis et al. (2009) is a review paper, not the first study to show that convection and equatorial waves can reinforce one another. The review itself points back to earlier observational and theoretical work, including studies such as Takayabu (1994). We added Takayabu (1994) to the citation.

-Line 357: southern hemisphere or Southern Hemisphere?

We corrected this consistently throughout the paper.

-Line 364: underestimation of MCS's sizes seems similar in the ocean and on land. It's difficult to see any big differences in the Violin plot.

We agree that this difference is subtle and have removed this sentence.

-Line 371-372: In ICON, MCS's velocities are independent of the size, which is not the case in observations. Is there a hypothesis about this?

We assume that you refer to results that observed short-lived storms moving more slowly than long-lived storms, whereas ICON has both storm types move similarly fast. The reason for this difference is unclear, and we did not want to hypothesize about its causes.

-Lines 379-380: I will also point out that the morphology of precipitation between observation and ICON is totally different. In observations, there is a tendency to sustain precipitation in the 20% of the lifetime; the long-lived ones over land are more efficient in doing this, till 40%. On the other hand, simulated MCSs do not do it. They have a rapid decay in terms of precipitation. It seems that long-lived ones are doing it, but during the peak of their growth.

This is a good point. We added the following text to the paper.

L415-7: ICON simulates mean precipitation within ±20 % of observed values during the growth phase, with the exception of land-based long-lived systems, which show a high difference of up to 40 %. **Long-lived MCSs maintain high mean precipitation rates during the growth phase of the storm (especially over tropical land), while simulated MCSs show a decline in mean precipitation even early on during their lifetime. This results in s**imulated mean MCS precipitation becoming increasingly low-biased over the life cycle of storms ending at low differences of up to -60 % during the decaying phase.

-Lines 388-389: What is the source of bias?

We replaced "source" with "causes".

-Lines 394-395: Most of the biases in non-MCS clouds are over land and a little over ocean.

We added that the non-MCS cold clouds have a low frequency bias mostly over land.

-Line 394: Fig S10 or Fig S8

Thank you for finding this typo. We meant Fig. S8.

-Lines 396-401: The authors could refer to Segura and Hohenegger (2024) regarding cumulonimbus, congestus, and shallow clouds and their relationship with precipitation. In their work, shallow clouds only represent 8% of the total amount of precipitation, while congestus accounts for mostly 46%. Is this result going in the same direction? Based on your analysis of the surface energy, shallow clouds are probably small over land, due to more shortwave radiation into the surface. Another question is how much of the precipitation satellite data can observe the shallow or congestus precipitation.

Thank you for this helpful suggestion. We agree that Segura and Hohenegger (2024) is highly relevant here and cited it in the revised manuscript. Their results are broadly consistent with our interpretation in that shallow clouds contribute only a small fraction of total precipitation, while congestus clouds contribute much more substantially. Additionally, GPM-IMERGv7 is useful for large-scale precipitation statistics, but it is not well suited to robustly quantify precipitation specifically from shallow or congestus clouds. Shallow warm-rain precipitation is especially uncertain. We added the following text to the manuscript to reflect this:

L435-8: These results are broadly consistend with Segura and Hohenegger (2024) who showed that a large fraction of tropical precipitation originates from shallow and particularly congestus clouds. GPM-IMERGv7 is not well suited to robustly quantify precipitation specifically from shallow or congestus clouds due to deficiencies in detecting precipitation from warm rain clouds (North et al., 2022; Huffman et al., 2023).

-Lines 432-434: It is difficult to consider a 4-year atmosphere-only km-scale simulation as a climate simulation, through the throughput and the simulated time-scale variability. However, my impression is that this simulation tends to fill the gap between simulating meso-scale and large-scale features of the atmosphere.

Here we refer to the divide between weather prediction and climate model development as outlined in Randal and Emanuel (2024) by building collaborations between the two communities which we have done in the EXCLAIM project between ETH and Meteo Swiss and other partners. We change the following wording to make sure that we do not give the impression that the current development is achieving this goal completely:

"....represents a major milestone **in** bridging the long-standing divide between numerical weather prediction (NWP) and climate modeling…"

"...represents a major milestone **toward** bridging the long-standing divide between numerical weather prediction (NWP) and climate modeling…"

-Line 440: Weren't the surface heat fluxes showing the strong biases?

We agree and added the following to this sentence.

L485-6: … although biases such as an overestimation of sensible heat fluxes over land during the warm season persist

-Lines 447-448: Is this something new?

While similar results have been found before, we believe it is important to emphasize the robust simulation of hourly precipitation in the simulation. There are many studies that attribute the heavier hourly precipitation in km-scale models compared to remote sensed precipitation products (e.g., GPM-IMERG) as model biases, while they are largely deficiencies in these observational products. We added the following sentence to make this point clearer.

L494-7: Importantly, model analyses that focus on the intensity or frequency of hourly precipitation should use in-situ observation records where possible since these precipitation characteristics are not well captured in GPM-IMERG-v7 and reanalysis products as we show here and in previous work (e.g., Dominguez et al., 2024).

-Lines 461-462: Could the authors give an estimation of how much comes from shallow clouds? I think most of the precipitation comes from congestus clouds.

We cannot differentiate between shallow and congestus clouds with the current analysis. However, we added that the shown differences are also likely related to deficiencies in GPM-IMERG-v7 precipitation estimates.

-Lines 466-467: Convective Rossby waves are also controlled by column water vapor (Nakamura and Takayabu 2022).

We added this reference to the revised paper.

-Lines 469-470: But atmosphere-only simulations can simulate convective coupled equatorial waves (Falko and Rios-Berrios 2021, Takasuka et al., 2024, Ortega et al., 2026)

We added Takasuka et al., (2024) and Ortega et al., (2026) and modified the text accordingly.

-Line 477: variability? I did not see any time-scale variability analysis.

We removed climate variability from the sentence.

-Lines 481-482: It's difficult to see this given the number of projects using km-scale Earth System Models. nextGEMS produced an 8-month simulation (coupled) using IFS-FESOM with a horizontal grid spacing of 2.8 km (Rackow et al., 2025). Using ICON, a two-month simulation was produced with a horizontal grid spacing of 2.5 km (Hohenegger et al., 2023). Recently, ICON's simulations with a horizontal grid spacing of 1.25 km were conducted, in which the carbon cycle was included (Klocke et al., 2025). The simulation in this study does not show adequate throughput to conduct climate simulations, even if an atmosphere-only configuration is used. Rather, I would say that it's going to fill the gap between the meso- and the large-scale circulation in the atmosphere.

We rephrased the last sentence to avoid a wrong impression. We meant to highlight the joint community activity in developing next generation weather and earth system modeling capabilities.

L534-6: While notable challenges remain in representing tropical variability, mesoscale convective organization, and land-atmosphere coupling, the demonstrated capability of our simulation and other km-scale global modeling efforts establish a new foundation for next-generation Earth system modeling and the transition toward truly unified weather–climate simulations.

References

-

Tomita, H., Miura, H., Iga, S., Nasuno, T., & Satoh, M. (2005). A global cloud-resolving simulation: Preliminary results from an aqua planet experiment. Geophysical Research Letters, 32(8), 1–4. https://doi.org/10.1029/2005GL022459

-
Miura, H., Satoh, M., Tomita, H., Noda, A. T., Nasuno, T., & Iga, S. I. (2007). A short-duration global cloud-resolving simulation with a realistic land and sea distribution. Geophysical Research Letters, 34(2), 2–6. https://doi.org/10.1029/2006GL027448
-
Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T., & Iga, S. (2008). Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations. Journal of Computational Physics, 227(7), 3486–3514. https://doi.org/10.1016/j.jcp.2007.02.006
-
Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., Düben, P., Judt, F., Khairoutdinov, M., Klocke, D., Kodama, C., Kornblueh, L., Lin, S. J., Neumann, P., Putman, W. M., Röber, N., Shibuya, R., Vanniere, B., Vidale, P. L., … Zhou, L. (2019). DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. Progress in Earth and Planetary Science, 6(1). https://doi.org/10.1186/s40645-019-0304-z
-
Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., & Düben, P. (2019). Global Cloud-Resolving Models. Current Climate Change Reports, 5(3), 172–184. https://doi.org/10.1007/s40641-019-00131-0
-
Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., Bao, J., Bastin, S., Behravesh, M., Bergemann, M., Biercamp, J., Bockelmann, H., Brokopf, R., Brüggemann, N., Casaroli, L., Chegini, F., Datseris, G., Esch, M., George, G., … Stevens, B. (2023). ICON-Sapphire : simulating the components of the Earth System and their interactions at kilometer and subkilometer scales. Geoscientific Model Development, 779–811. https://doi.org/10.5194/gmd-2022-171
-
Rackow, T., Pedruzo-Bagazgoitia, X., Becker, T., Milinski, S., Sandu, I., Aguridan, R., Bechtold, P., Beyer, S., Bidlot, J., Boussetta, S., Deconinck, W., Diamantakis, M., Dueben, P., Dutra, E., Forbes, R., Ghosh, R., Goessling, H. F., Hadade, I., Hegewald, J., … Ziemen, F. (2025). Multi-year simulations at kilometre scale with the Integrated Forecasting System coupled to FESOM2.5 and NEMOv3.4. Geoscientific Model Development, 18(1), 33–69. https://doi.org/10.5194/gmd-18-33-2025
-
Segura, H., Pedruzo-Bagazgoitia, X., Weiss, P., Müller, S. K., Rackow, T., Lee, J., Dolores-Tesillos, E., Benedict, I., Aengenheyster, M., Aguridan, R., Arduini, G., Baker, A. J., Bao, J., Bastin, S., Baulenas, E., Becker, T., Beyer, S., Bockelmann, H., Brüggemann, N., … Stevens, B. (2025a). nextGEMS: entering the era of kilometer-scale Earth system modeling. Geoscientific Model Development, 18(20), 7735–7761. https://doi.org/10.5194/gmd-18-7735-2025
-
Klocke, D., Frauen, C., Engels, J. F., Alexeev, D., Redler, R., Schnur, R., Haak, H., Kornblueh, L., Brüggemann, N., Chegini, F., Römmer, M., Hoffmann, L., Griessbach, S., Bode, M., Coles, J., Gila, M., Sawyer, W., Calotoiu, A., Budanaz, Y., … Stevens, B. (2025). Computing the Full Earth System at 1km Resolution. Proceedings of the

International Conference for High Performance Computing, Networking, Storage, and Analysis, SC 2025, 125–136. https://doi.org/10.1145/3712285.3771789

-

Segura, H., Hohenegger, C., Wengel, C., & Stevens, B. (2022). Learning by Doing: Seasonal and Diurnal Features of Tropical Precipitation in a Global-Coupled Storm-Resolving Model. Geophysical Research Letters, 49(24), 1–10. https://doi.org/10.1029/2022GL101796

-

Takasuka, D., Kodama, C., Suematsu, T., Ohno, T., Yamada, Y., Seiki, T., Yashiro, H., Nakano, M., Miura, H., Noda, A. T., Nasuno, T., Miyakawa, T., & Masunaga, R. (2024). How Can We Improve the Seamless Representation of Climatological Statistics and Weather Toward Reliable Global K-Scale Climate Simulations? Journal of Advances in Modeling Earth Systems, 16(2). https://doi.org/10.1029/2023MS003701

-

Marsham, J. H., Dixon, N. S., Garcia-Carreras, L., Lister, G. M. S., Parker, D. J., Knippertz, P., & Birch, C. E. (2013). The role of moist convection in the West African monsoon system: Insights from continental-scale convection-permitting simulations. Geophysical Research Letters, 40(9), 1843–1849. https://doi.org/10.1002/grl.50347

-

Hohenegger, C., Brockhaus, P., Bretherton, C. S., & Schär, C. (2009). The soil moisture-precipitation feedback in simulations with explicit and parameterized convection. Journal of Climate, 22(19), 5003–5020. https://doi.org/10.1175/2009JCLI2604.1

-

Lee, J., Hohenegger Edited by Paul A Dirmeyer, C. I., & Dickinson, R. E. (2024). Weaker land-atmosphere coupling in global storm-resolving simulation. https://doi.org/10.1073/pnas

-

Holloway, C. E., Woolnough, S. J., & Lister, G. M. S. (2012). Precipitation distributions for explicit versus parametrized convection in a large-domain high-resolution tropical case study. Quarterly Journal of the Royal Meteorological Society, 138(668), 1692–1708. https://doi.org/10.1002/qj.1903

-

Segura, H., Bayley, C., Fievét, R., Glöckner, H., Günther, M., Kluft, L., Naumann, A. K., Ortega, S., Praturi, D. S., Rixen, M., Schmidt, H., Winkler, M., Hohenegger, C., & Stevens, B. (2025b). A Single Tropical Rainbelt in Global Storm-Resolving Models: The Role of Surface Heat Fluxes Over the Warm Pool. Journal of Advances in Modeling Earth Systems, 17(7). https://doi.org/10.1029/2024MS004897

-

N. Takayabu, Y. (1994). Large-Scale Cloud Disturbances Associated with Equatorial Waves. Journal of the Meteorological Society of Japan. Ser. II, 72(3), 433–449. https://doi.org/10.2151/jmsj1965.72.3_433

-

Nakamura, Y., & Takayabu, Y. N. (2022). Convective Couplings with Equatorial Rossby Waves and Equatorial Kelvin Waves. Part I: Coupled Wave Structures. Journal of the Atmospheric Sciences, 79(1), 247–262. https://doi.org/10.1175/JAS-D-21-0080.1

-

Segura, H., & Hohenegger, C. (2024). How Do the Tropics Precipitate? Daily Variations in Precipitation and Cloud Distribution. Journal of the Meteorological Society of Japan, 102(5), 525–537. https://doi.org/10.2151/jmsj.2024-028

- Sebastián Ortega, Hans Segura, Victor C. Mayta, et al. Convectively Coupled Equatorial Waves in a Global Storm-Resolving Model. ESS Open Archive . February 07, 2026.
-

Judt, F., & Rios-Berrios, R. (2021). Resolved Convection Improves the Representation of Equatorial Waves and Tropical Rainfall Variability in a Global Nonhydrostatic Model. Geophysical Research Letters, 48(14), 1–10. https://doi.org/10.1029/2021gl093265

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).

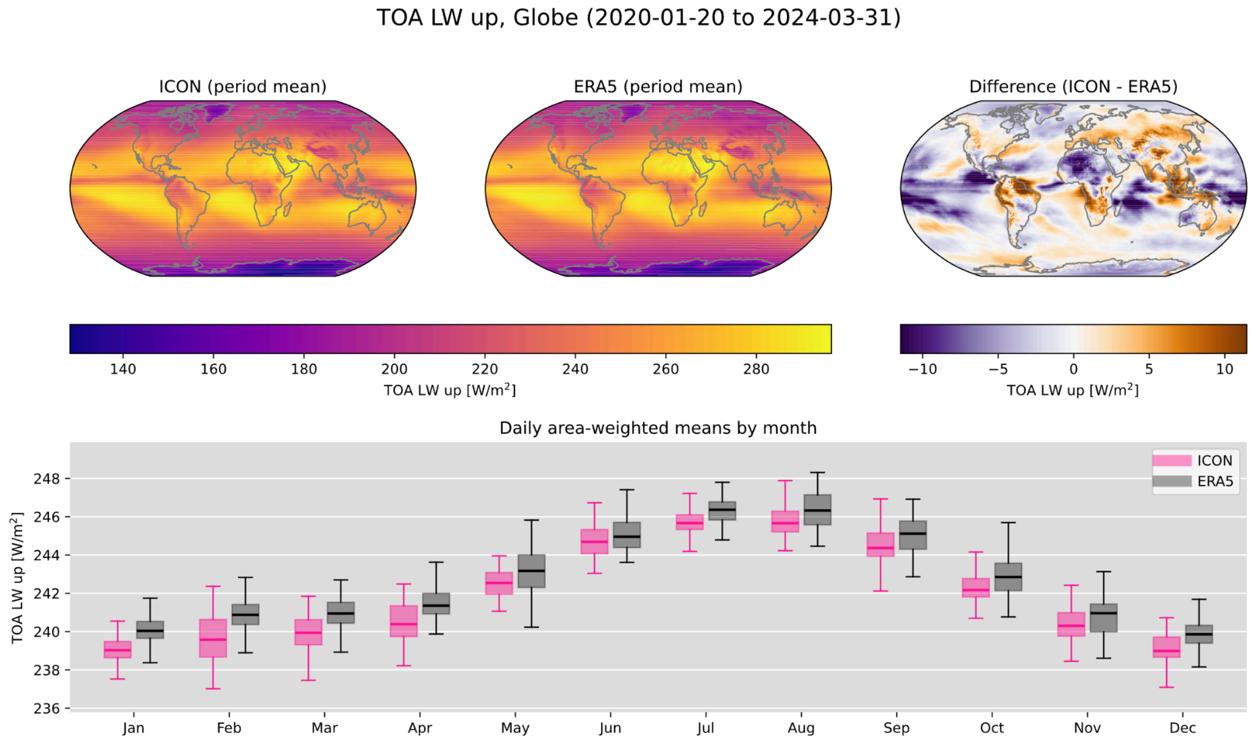


Fig. R1: Comparison of top of average atmosphere (TAO) outgoing lowngwave (LW) radiation in ICON (top left), ERA5 (top middle) and their difference (top right). Monthly average global mean statistics are shown in the lower panel.

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.**