

Response to Reviewer-1

Review of: Global Monsoon in ICON: The scale-dependent response of Northern Hemisphere Monsoons

Authors: Praveen Kumar Pothapakula*, Andreas F. Prein, Anusha Sunkisala, and Anurag Dipankar

Journal: Weather and Climate Dynamics

Review: The authors are examining the fidelity of the ICOSahedral Non-hydrostatic (ICON) model in simulating the hydroclimate of the various regional monsoons in the tropics. They use a 10 year atmosphere-only simulation forced with observed SST to assess the model simulation of the seasonal cycle, interannual, intraseasonal variations and extreme rain events.

Overall the paper is well written and the figures are of high quality. I consider this paper to be important as it is familiarizing the readers with a relatively new model on its performance of the monsoon. The paper is easy to read and understand and I would support its publication after the authors address the following comments:

We thank the reviewer for the encouraging comments and for their constructive questions, which ranged from fundamental scientific arguments to helpful suggestions for improving the description of the datasets. We have now addressed all the comments with additional figures, analysis, and detailed reasoning. Please find our point-by-point responses below.

1. At the tested spatial grid spacings of 80, 40, and 10km don't believe that the nonhydrostatic feature is important at all. So, my question then is what is the purpose of running an expensive feature like the non-hydrostatic option for this study at all?

Thank you for raising this fundamental and very important question.

- We kindly refer to Liu et al. (2022), who investigated the non-hydrostatic (NH) vs. hydrostatic (H) dynamical core performance using the Simple Cloud-Resolving E3SM Atmosphere Model (SCREAM) at grid spacings from ~30 km to 10 km, with real-world experiments. E3SM, like ICON, uses an irregular grid and is a next-generation global modeling system. Their results revealed significant NH vs. H differences in the tropics, particularly over India, the Pacific Ocean, and the Indian monsoon core region (we refer to Fig.1 of Liu et al. (2022)). Even at 30 km resolution, the NH dycore produced 20% more rainfall over the Indian monsoon core compared to the H dycore, demonstrating that NH dynamics help reduce the dry bias over central India even at relatively coarse resolution. Liu et al. attributed this difference to changes in vertical velocity fields rather than surface moisture
- Motivated by these findings, we analyzed vertical velocity (w) across our three resolutions (80, 40, and 10 km) over the monsoon core regions. Our results (Fig. 1) show that updrafts are systematically stronger at finer grid spacing (10 km > 40 km > 80 km). Furthermore, these

stronger updrafts enhance the conversion of cloud water (q_c) to rain water (q_r), producing more precipitation (see response to comment #5). These results are consistent with Liu et al. (2022).

- Our experimental design (NH at all resolutions, grid spacing varying) does not allow us to separate the effects of grid spacing from those of NH dynamics. We therefore do not claim to have isolated the NH contribution. The detailed effects of NH vs. H are beyond the scope of this paper.

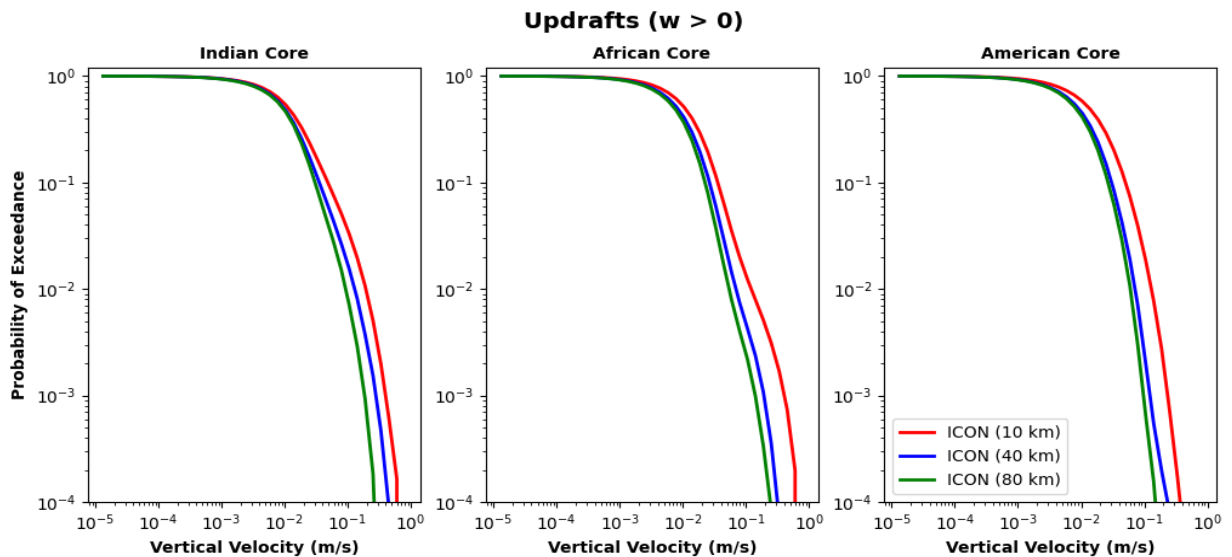


Figure 1: Vertical velocity (w) cumulative distribution function, or probability of exceedance on a logarithmic probability scale for various ICON simulations over the SAsiaM, NAfriM and NAmerM core monsoon regions at 850hPa level. The probability of exceedance plot highlights the intense updrafts with changing horizontal grid spacings. All the simulations are analysed at a common 30 km grid for fair comparison.

2. The tracks of the low pressure systems in Fig. 12 suggest that the results in the paper are representative of the 10-year integration period (2006-16) of ICON or sensitive to the algorithm used to diagnose these systems. For instance, Vishnu et al. (2020; <https://doi.org/10.1029/2020JD032977>) also find a fair number of tracks going northward towards Bangladesh from northern head Bay of Bengal as opposed to just northwestward tracks. So could you comment on this?

We thank the reviewer for this observation. To clarify whether the track characteristics in Fig. 12 of the manuscript are sensitive to (a) the 10-year integration period or (b) our in-house tracking algorithm, we conducted a series of sensitivity tests.

Specifically, we tested a range of sea-level pressure thresholds ($\Delta SLP = 2-4$ hPa) and maximum linking distances (150-250 km), encompassing and extending the values used in Vishnu et al. (2020). Here, ΔSLP is the pressure threshold relative to the local minimum used to define a low-pressure system, and linking distance is the maximum distance a system is allowed to move within 6 hours while still being considered the same storm. These values were chosen to assess whether reducing the linking distance would preferentially capture shorter, northward-propagating tracks from the head of the Bay of Bengal toward Bangladesh.

From these sensitivity tests, we found that while the total number of detected systems changes slightly, the overall spatial structure of the tracks, including the dominant northwestward propagation remains largely unchanged (Fig. 2). This indicates that the track orientation is not strongly sensitive to the specific parameter choices in our tracking algorithm.

The differences relative to Vishnu et al. (2020) are therefore more likely related to differences in the analysis period and dataset characteristics. Vishnu et al. (2020) derived their results from multi-decadal reanalysis datasets (1979–2019), which provide a climatological representation of LPS behavior. In contrast, our analysis is based on a shorter 10-year period (2006–2016), which may not fully capture the full range of interannual variability in track pathways. It is well known that LPS propagation is modulated by large-scale monsoon conditions and can vary substantially from year to year, a shorter sampling period may under-represent less frequent track types, such as northward-moving systems toward Bangladesh. Consistent with this interpretation, we observe from the IMD best track data that very few LPS propagate toward Bangladesh during our study period.

Overall, our analysis confirms that the algorithm parameter choices do not fundamentally alter the spatial structure of the detected tracks. We therefore conclude that the differences relative to Vishnu et al. (2020), particularly the absence of northward-propagating tracks toward Bangladesh are more likely attributable to sampling variability and interannual variability in LPS pathways, rather than to methodological inconsistencies in the tracking algorithm.

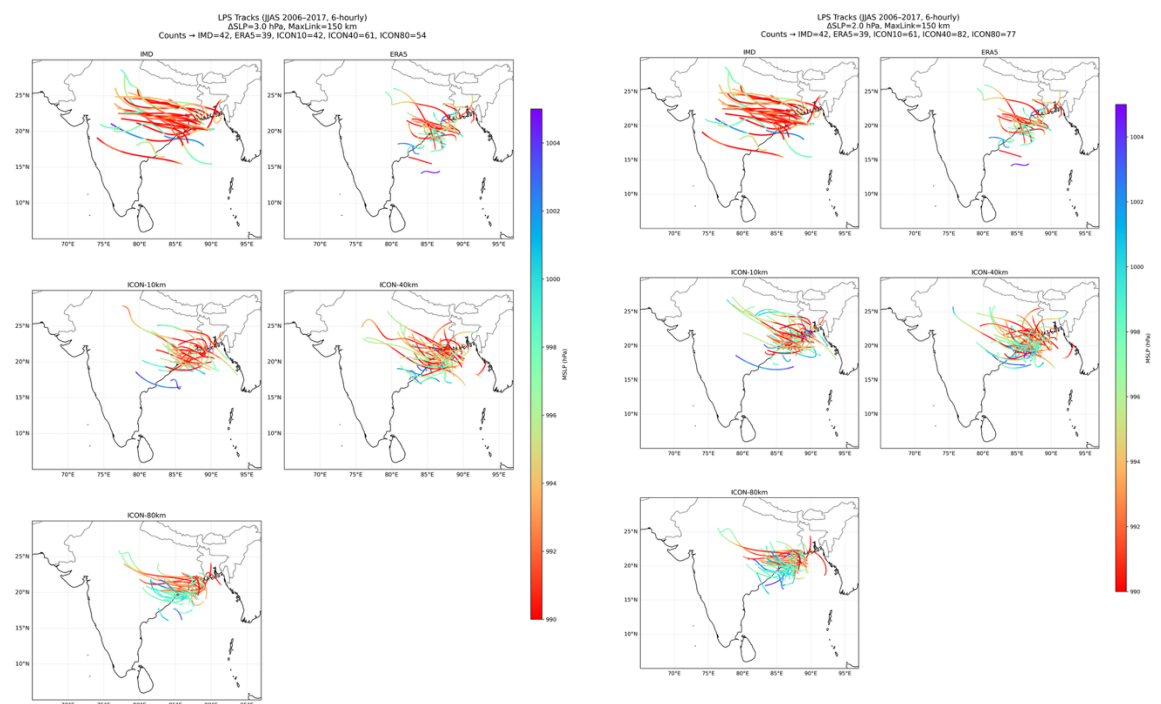


Figure 2 : Sensitivity tests with LPS tracking algorithm to test for tracks passing India and Bangladesh (left) by reducing the linking distance between two points on a track (right) changing the Δ SLP and fixing the linking distance to 150 km.

3. If the authors wish to compare the results with a variety of precipitation observations then they should also summarize their merits and disadvantages.

We appreciate the reviewer for pointing out this important information. We have now added relevant information discussing the merits and disadvantages of the observational/reanalysis products used in the study. (Appendix A, L715-L748)

- IMERG combines passive microwave estimates from multiple low-earth-orbital satellites with geostationary infrared (IR) observations and gauge data. Sophisticated techniques such as morphing and Kalman filter compensate for the limited low earth orbital sampling, enabling high-temporal (30-min) and spatial resolution. However, this dataset is limited due to its indirect estimation of the infrared channel for filling temporal gaps, and morphing techniques assume linear motion. Furthermore, gauge bias correction at monthly intervals is limited over the ocean, thus enhancing the probability for bias. Despite these limitations, IMERG represents the state-of-the-art in multi-satellite precipitation estimation over the tropics.
- MSWEP data uses a multitude of gauges (76,747 stations), satellites (CMORPH, GSMaP, TMPA), and reanalyses (ERA-Interim, JRA-55) into a global 0.1°, 3-hourly product. It has a full comprehensive land-ocean coverage with discharge-based bias correction (13,762 stations), and improved performance in mountains. Its 3-hourly temporal frequency misses sub-daily extremes, its reliance on gauge stations limits performance in regions of sparse network, discharge correction is indirect, complex methodology has assumptions.
- CMORPH, similar to GPM IMERG, retrieves precipitation from passive microwave sensors deployed on low-earth orbiting satellites, and uses cloud motion vectors from geo-stationary satellites to morph precipitation. The data is also a bias-corrected version against CPC gauges (land) and GPCC (ocean). CMORPH has high temporal resolution, global coverage, and has a good warm-season performance. However, like GPM, it uses morphing and thus assumes linear motion, has biases in complex terrain, underestimates heavy rain, and has variable regional performance.
- The GPCC provides gridded gauge-based precipitation analyses derived from quality-controlled in-situ station data over global land surfaces based on up to 84,800 stations. This product uses a large station network across the globe and has established quality control procedures (Becker et al., 2013, Schamm et al., 2014, Schneider et al., 2018). However, it is limited by the variable station density and bias correction for systematic gauge measuring errors. ((Becker et al., 2013 (DOI:10.5194/essd-5-71-2013); Schamm et al., 2014 (DOI: 10.5194/essd-6-49-2014); Schneider et al., 2018 (https://opendata.dwd.de/climate_environment/GPCC/PDF/GPCC_intro_products_v2_018.pdf))
- The CPC provides gauge based analysis of global daily precipitation over land- and surfaces. It uses optimal interpolation to combine data from over 30,000 rain gauges. It uses relatively large station network, rigorous quality control and provides estimations of gauge counts per grid cell, and accounts for orographic effects during interpolation. Like GPCC, this dataset is also limited by variable station density and has coarse spatial resolution.

- In addition to observational datasets, we used ERA5 which uses a vast amounts of historical observations (from satellites, radiosondes, aircraft, weather stations, and ships) with a numerical weather prediction model using data assimilation to create a physically consistent, globally complete dataset.
 - To validate our low pressure system tracks from ICON, we used IMD best track datasets to estimate the positions at 6 hourly intervals. It uses a multi-satellite approach for locating storm centers such as ndian geostationary satellites (Kalpana-1, INSAT-3A, and later INSAT-3D) along with polar-orbiting satellites from international partners. This dataset however relies completely on satellites, and hence the estimatons are biased especially during intense cloud days.
4. It would be nice to show the topography represented at all three grid resolutions.
- Thank you, we agree that topography is an important figure to show. We have now added this figure to the supplementary material (Fig. S1, zoomed figure in Fig. S2). Furthermore, we have added the discussion on resolution sensitivity in resolving the fine-scale topography in the manuscript. (updated in L141-L144)

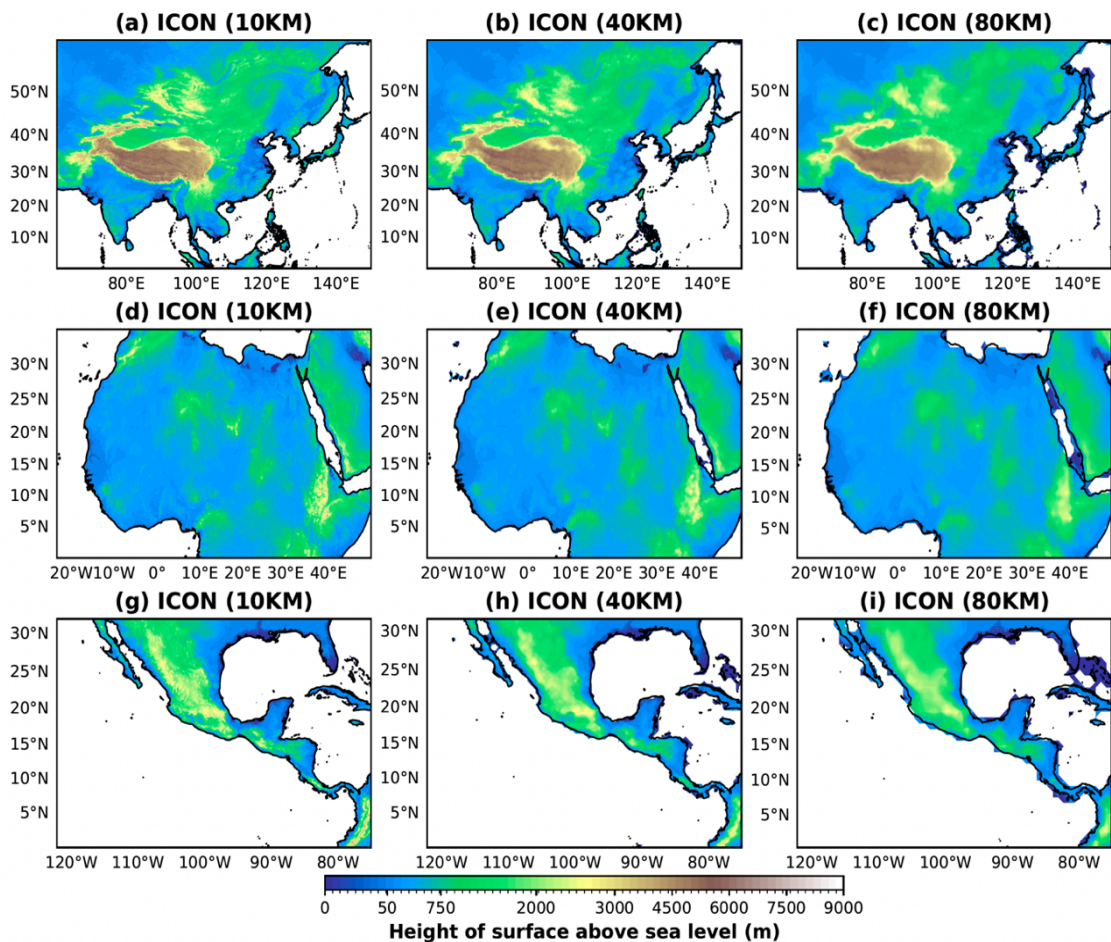


Figure 3 : Topography of Northern Hemisphere monsoon domains in various ICON simulations at 10 km, 40 km, and 80 km grid-spacing. (a-c) SAsiaM domain (d-f) WAFriM domain and (g-h) NAmErM domain

5. I would be curious to know if the partitioning of convective and stratiform precipitation changes across resolutions? Likewise does the fraction of high, medium, and low cloud cover change across resolutions for light, medium, and heavy rain events? Is there a difference in the precipitation efficiency and precipitation recycling ratio across resolutions?

Thank you for the comment. We have now added a detailed explanation, results, and description in the manuscript (L503-L521). Below is our detailed response to this question.

Partitioning of convective and grid-scale (resolved) precipitation: In Fig. 9 of our manuscript, we have shown the probability distribution of total precipitation, and in supplementary material Fig. S11, the distributions of convective precipitation across various simulations and monsoon domains. Our results show that the contribution of convective precipitation, especially for high-intense precipitation events (> 10 mm/day), decreases with decreasing grid spacings, indicating that the difference in the precipitation contributions across simulations is largely due to the resolved grid-scale precipitation.

To further enhance this important point, we have quantified the contribution of convective and grid-scale precipitation to the total precipitation in monsoon core regions across various simulations in the table below (Table 1).

Table 1: Quantification of convective (Vs) grid-scale precipitation

Resolution	South Asia % grid-scale (% conv)	North Africa (% grid-scale)/ (% conv)	North America (% grid-scale)/ (% conv)
10 km	46% (54%)	30% (70%)	10% (90%)
40 km	32% (68%)	16% (84%)	6% (94%)
80 km	28% (72%)	12% (86%)	4% (96%)

These results indicate that fine grid spacing induces more grid-scale precipitation compared to the coarse grid spacing simulation in all three monsoon cores. Notably, over the SAsiM (28% \rightarrow 46%) and NAFriM (12% \rightarrow 30%). Correspondingly, the convective fraction decreases. This indicates that higher resolution allows the model to generate more explicitly resolved precipitation, reducing the dominance of parameterized convection. However, it is important to note that the rate of increase in the grid-scale precipitation is not proportional to the decrease in convective precipitation. For example, over South Asia, the 10 km simulation's grid-scale's gain exceeds the convective loss by a factor of 1.7, and 1.5 over Africa. This indicates that fine grid spacing does not merely partition the existing precipitation but generates additional precipitation through explicitly resolved dynamics.

Sensitivity of Cloud Cover to Grid Spacing:

For the whole Monsoon Season (JJAS): Following your suggestion, we show the cloud cover from various ICON simulations on the model levels (120 levels) with corresponding estimated height above the sea level over the monsoon core region in Fig. 4. The results show that over SAsiM and NAFriM, the cloud cover is sensitive to grid spacing, and less so over NAmM. Particularly, we notice a systematic decrease in the low-level clouds and high-level clouds over these regions, with little impact on mid-level clouds.

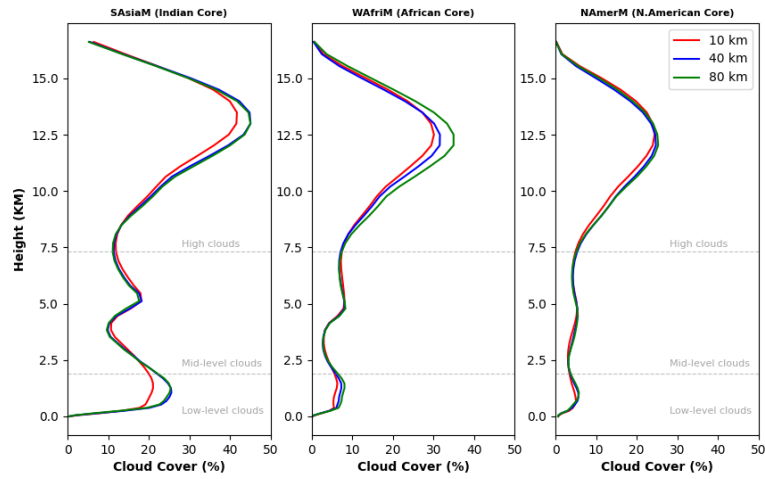


Figure 4: Sensitivity of total cloud cover (%) to grid spacing in ICON 10 km, 40 km, and 80 km simulations for the SAsiaM, WAfriM, and NAmerM domains. The cloud cover is averaged over the respective monsoon core regions and is plotted against model levels converted to height above sea level.

Extreme Events Cloud Cover: Investigating further into the cloud cover distribution over the core monsoon regions during intense precipitation events (median precipitation over core monsoon region > 5 mm/day) shows similar behavior of the average cloud distribution during the entire monsoon season, with the exception of NAmerM high clouds (Fig. 5).

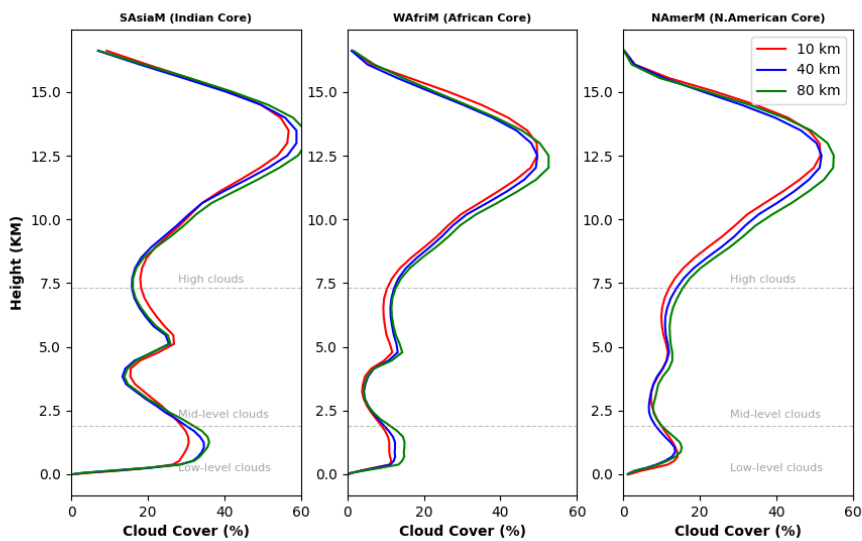


Figure 5: Same as Fig. 4, but for extreme precipitation event days over the Monsoon core regions.

Moderate Events Cloud Cover: For the moderate precipitation days (median precipitation over the core monsoon region > 2 mm/day and < 5 mm/day), the low-level clouds consistently decrease with fine grid spacing, while the mid-level clouds show a slight increase in fine grid spacing in SAsiaM and NAfriM core monsoon region (Fig. 6).

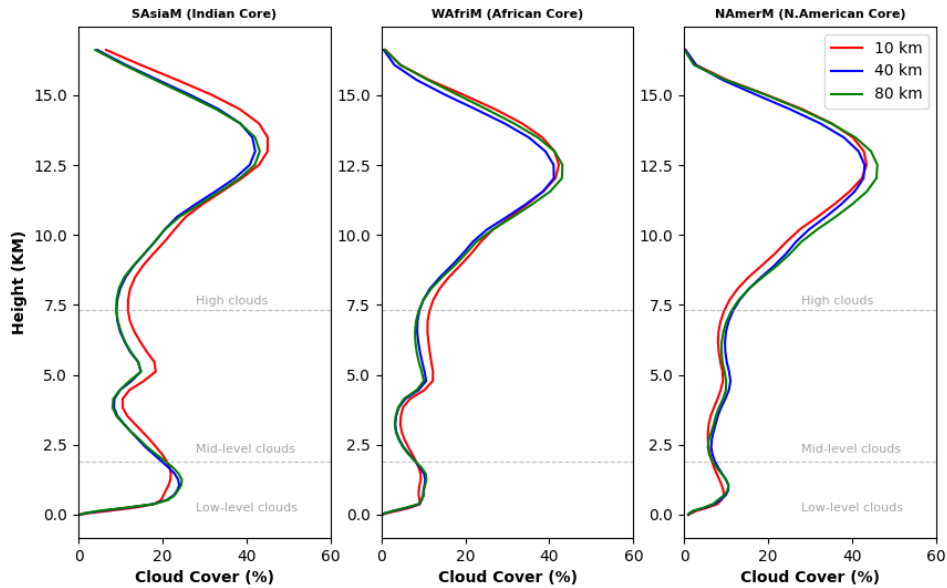


Figure 6: Same as Fig. 4, but for moderate precipitation event days over the Monsoon core regions.

Light Rain Event cloud Cover: Only over SAsiaM, the mid and high-level clouds are sensitive to the grid spacing during light rain days, however, the percentage contribution of these clouds is less than the medium and intense precipitation days (Fig. 7).

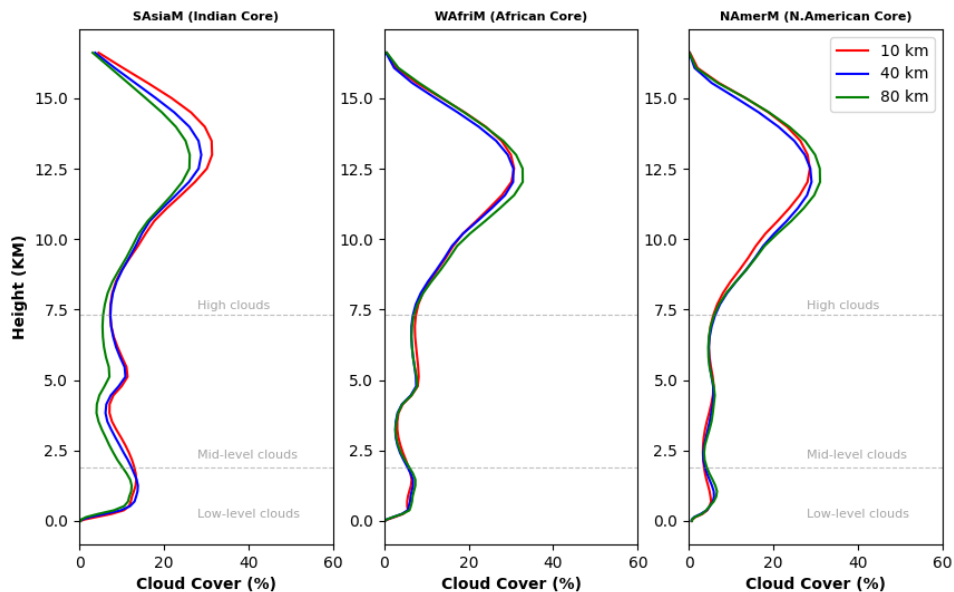


Figure 7: Same as Fig. 4, but for light precipitation event days over the Monsoon core regions.

Overall, the results indicate that the fine grid spacing results in decreased cloud cover, especially the low-level clouds and the high-level clouds during intense and moderate rainy days. The cloud cover over the SAsiaM and NAmM monsoon core regions is sensitive to grid spacing compared to WAfriM domain. The results are consistent with the previous studies, i.e., clouds become smaller and smaller with decreasing grid spacing (Langhans et al., 2012a, Prein et al., 2015).

Precipitation efficiency and precipitation recycling ratio: Here, we distinguish precipitation efficiency between cloud microphysics and the boundary layer, presenting results from two perspectives. Furthermore, we plan to investigate a detailed water budget analysis in a future

study; for this review, we use simple metrics to provide preliminary insights into these processes.

(a) Rain conversion in microphysics: We plot the prognostic variables, cloud water content (kg/kg) and rain water content (kg/kg) on model pressure levels (Fig. 8 and Fig. 9). Unfortunately, these variables are not saved on native model levels, so smooth vertical profiles are not expected due to interpolation onto limited pressure levels. The results indicate that cloud water content decreases in the mid-tropospheric levels with finer grid spacing over SAsiaM and NAfriM, while the results are less sensitive over NAmerM. We focus on mid-tropospheric properties as they are crucial for rain formation processes. Furthermore, across all simulations, rain water content increases with finer grid spacing, indicating that fine grid simulations more efficiently convert available cloud water into warm rain. A systematic increase in vertical velocity with finer grid spacing can enhance collision-coalescence and riming processes, thus leading to increased precipitation. Therefore, finer grid spacing enhances precipitation efficiency at the cloud microphysical scale.

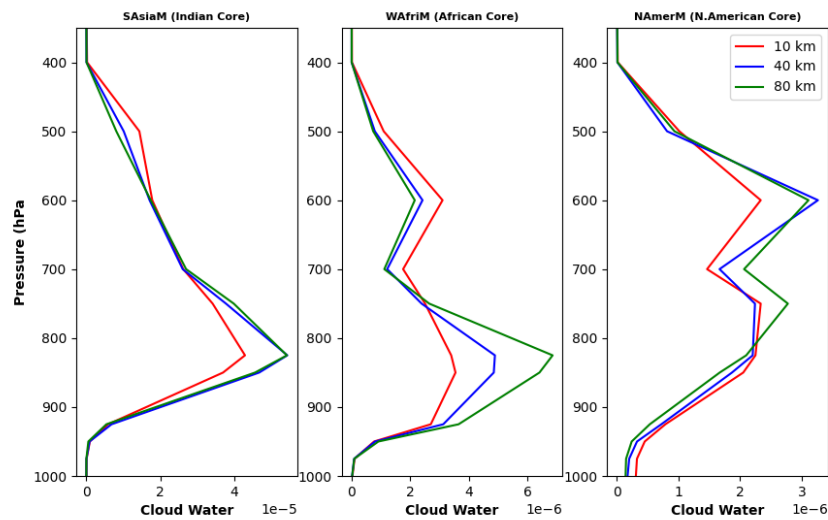


Figure 8: Sensitivity of cloud water (kg/kg) to grid spacing in ICON 10 km, 40 km, and 80 km simulations for the SAsiaM, WAFriM, and NAmerM domains. The cloud water is averaged over the respective monsoon core regions and is plotted against pressure levels level.

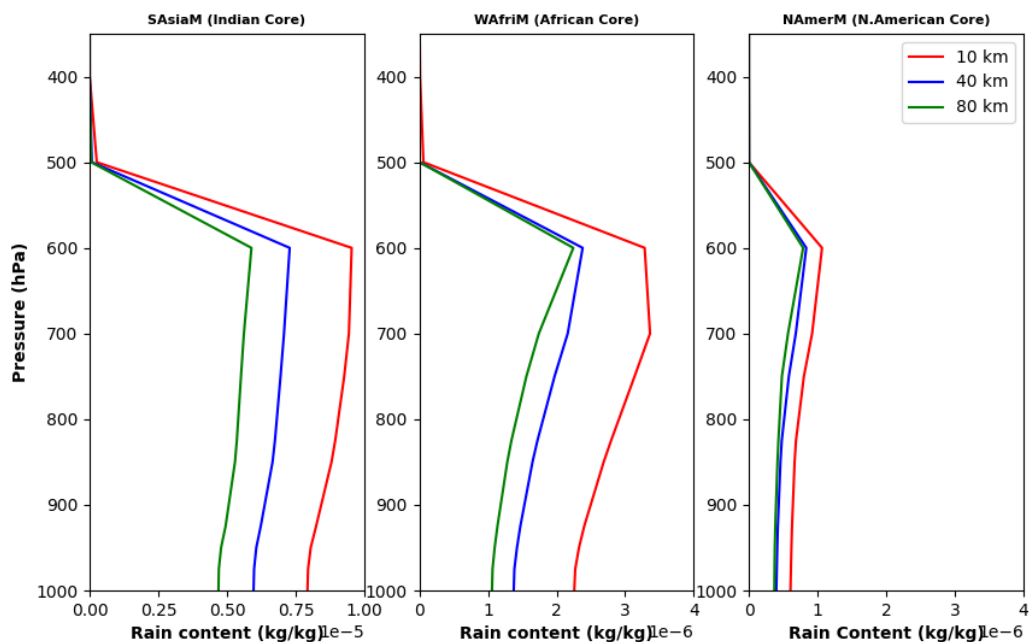


Figure 9: Same as Fig.8, but for rain content (kg/kg)

(b) Precipitation efficiency and recycling in lower boundary layer: We focus on the boundary layer efficiency and recycling of precipitation over the monsoon regions on monthly scales to account for the steady state. Precipitation Efficiency (PE) is calculated as $PE = P/(P+E)$, and the Recycling Ratio (RR) as E/P . The PE is derived from the surface moisture budget as an index of the fraction of total surface moisture flux (precipitation downward + evaporation upward) that is returned as precipitation. These simple, yet novel metrics provide preliminary insights into the processes.

Figure 10 shows that SAsiaM has the highest precipitation efficiency (~71%), followed by NAFriM (~60%) and NAmM (~50%). The African core shows a modest but systematic increase in PE at 10 km (+2%), suggesting improved efficiency at higher resolution, likely due to better-resolved land-atmosphere coupling and convective organization. In contrast, the SAsiaM and NAmM cores show little resolution dependence. For SAsiaM, this is likely because the large-scale monsoon circulation is so strong and moisture-rich that even coarse resolutions achieve high efficiency, leaving little room for further improvement. For North America, PE remains consistently lower (~50%) across all resolutions, possibly due to complex orography and a different convective regime that is less sensitive to resolution changes within the 80–10 km range. These regional differences indicate that the sensitivity of precipitation efficiency to resolution depends on the dominant physical processes in each monsoon core

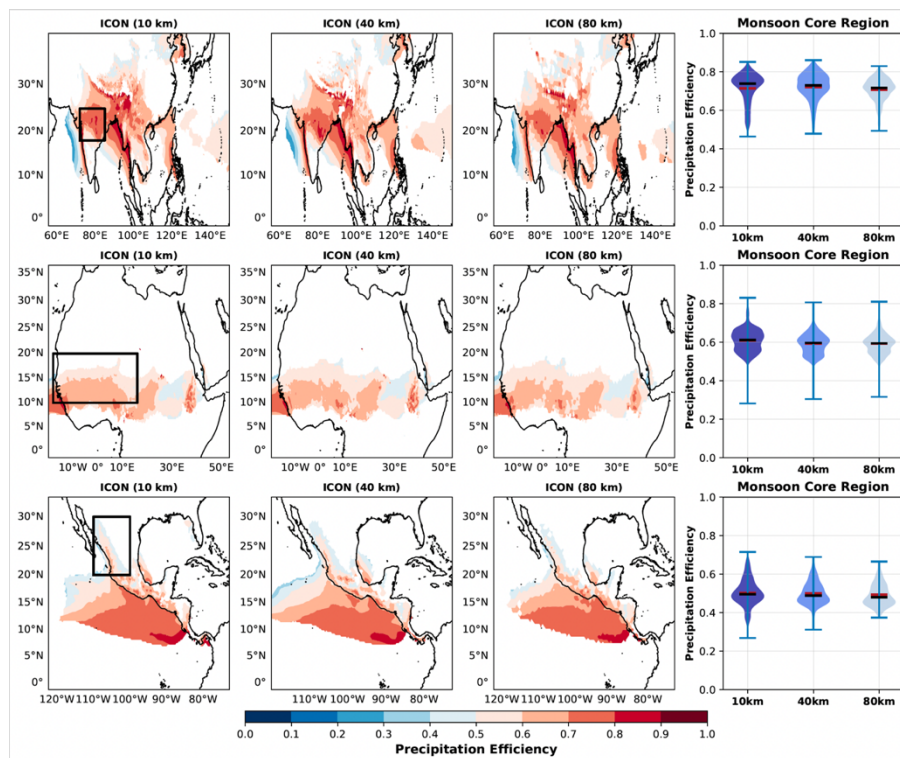


Figure 10: Sensitivity of PE to grid spacing in ICON 10 km, 40 km, and 80 km simulations for the SAsiaM, WAFriM, and NAmM domains. The violin plot on the right panel shows values over the respective core monsoon regions.

The recycling ratio in Fig. 11 shows that the NAFriM monsoon has the highest local moisture recycling (~70%), followed by SAsiaM (~40-45%). The NAmM core shows RR slightly above 1, which is consistent with the inclusion of oceanic grid points (Gulf of California) where evaporation exceeds precipitation. Resolution dependence is weak overall, though the NAFriM shows a slight decrease in recycling at 10 km. Taken together, these metrics reveal a consistent picture. Overall, precipitation efficiency and recycling ratios show only modest resolution

dependence (e.g., Africa PE: +0.02, RR: -0.05). This indicates that the substantial increase in total precipitation at 10 km, particularly in the heavy tail, is not primarily due to changes in local surface moisture partitioning. Instead, the increase is driven by stronger vertical velocities (w) that accelerate the conversion of cloud water to rain water, and by increased large-scale moisture advection into the monsoon core

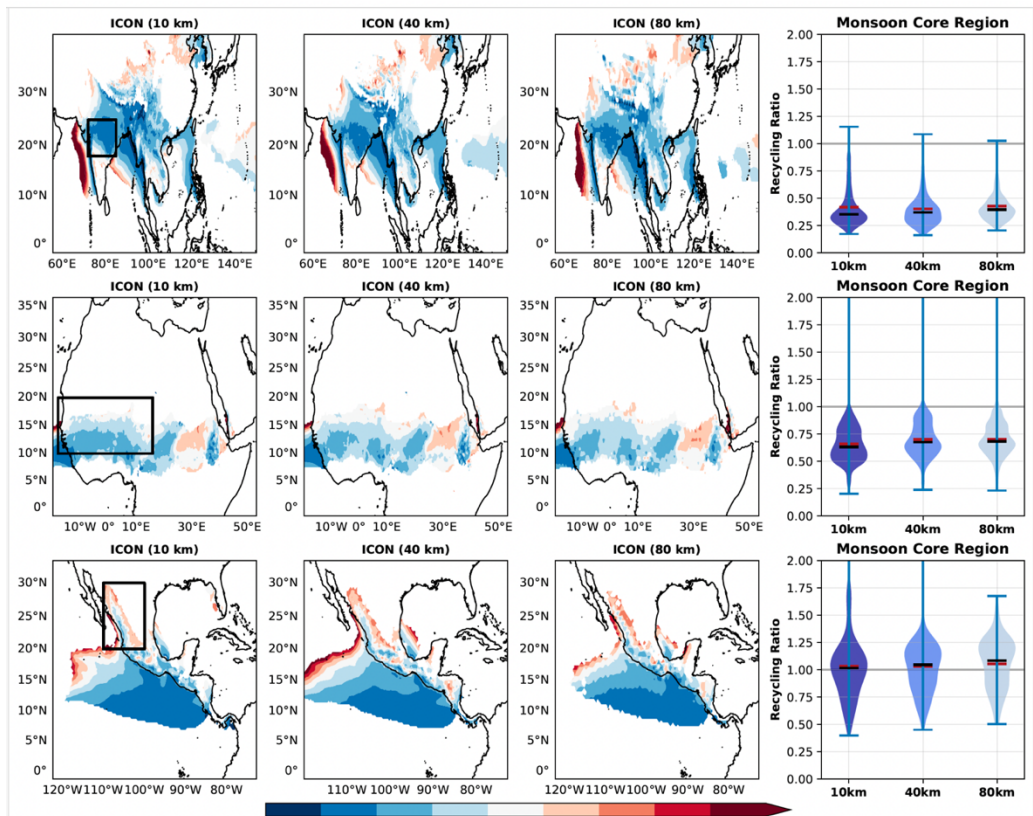


Figure 11: Same as Fig.10, but for RR.

6. The authors could provide more details on the models like what timesteps are used for the three models?
 - We have provided some of the important timesteps used in the model integration in the methodology section of the submitted manuscript (L135-L137). For clarity, we provide the full details here in a table. Furthermore, we have added additional timestep information for fast-physics processes in the revised manuscript.

Table 2: Various dynamic and physics time steps used in ICON simulations

Model Parameter	ICON 10 km	ICON 40 km	ICON 80 km
Dynamical timestep	60 sec	180 sec	360 sec
Radiation scheme	360 sec	360 sec	360 sec
Cloud cover scheme	60 sec	180 sec	360 sec
Convection scheme	60 sec	180 sec	360 sec

orographic gravity wave drag (sso)	60 sec	180 sec	360 sec
non-orographic gravity wave drag (gwd)	60 sec	180 sec	360 sec
Fast-physics processes	60 sec	180 sec	360 sec

- Fast-physics include: saturation adjustment, surface transfer scheme, land-surface scheme, boundary-layer/turbulent vertical diffusion scheme, and microphysics scheme.
- In all configurations, the timestep scales approximately linearly with grid spacing, with the 10 km simulation using a 60-second dynamical timestep, while the 40 km and 80 km simulations use 180 and 360 seconds, respectively.

Response to Reviewer-2

Review of: Global Monsoon in ICON: The scale-dependent response of Northern Hemisphere Monsoons

Authors: Praveen Kumar Pothapakula*, Andreas F. Prein, Anusha Sunkisala, and Anurag Dipankar

Journal: Weather and Climate Dynamics

Review: This contribution by Pothapakula et al. is a comprehensive evaluation study of ICON in a new version. It focuses on important aspects of the global monsoon system and concludes, that grid-spacing still influences the climate modeling performance. It is an interesting and important contribution, which I find well worth publication after revisions.

We thank the reviewer for the positive comments on our manuscript and for suggesting important points and clarifications. All the comments are addressed in detail, and furthermore, sensitivity tests were conducted to answer comments where necessary. Please find our point-by-point reply below

Still, I have a few comments and questions as detailed below.

1. The abstract is quite long and it is difficult to identify the main findings. Is it a goal to evaluate the new ICON core (line 7, but not mentioned below) or just the grid-spacing dependence of global/regional monsoon representation? What is “organized moderate variability”? What is an “excellent representation”? Line 23: increased → decreased. Line 25: What “underscores” the region-dependent sensitivity?

We agree with the reviewer, and now we have made the abstract shorter and focused, highlighting important findings of our manuscript.

Modified Abstract: The global monsoon system is a lifeline for two-thirds of the world’s population, as it is essential for tropical water security, food, and agriculture. However, its complex multiscale interactions challenge weather and climate models. This study investigates how horizontal grid spacing (80 km, 40 km and 10 km) in the ICOSahedral Non-hydrostatic (ICON) model affects both the mean and the variability of Northern Hemisphere monsoons across diurnal, intraseasonal, and interannual timescales.

All simulations show substantial skill in capturing the global monsoon system domain and its mean annual range of precipitation with a pattern correlation of > 0.7 and RMSE < 3 mm/day. For the key Northern Hemisphere regional monsoons, South Asia (SAsiaM), West Africa (WAfriM) and North America (NAmerM), ICON achieves an accuracy $> 80\%$ in capturing the observed monsoon domain. Crucially, the impact of grid spacing is strongly region-dependent and non-systematic. The finer grid spacing induces higher mean precipitation biases over continental SAsiaM, and WAfriM. Some of these biases are related to the intensity and location of moist monsoonal low-level jets, as well as their sensitivity to grid spacing. Furthermore, the

fine grid spacing overestimates monsoon precipitation variability at interannual and intraseasonal scales, including intense precipitation frequency (> 10 mm/day). This amplification stems primarily from enhanced grid-scale precipitation resulting from efficient microphysical processes, while convective precipitation exhibits limited sensitivity to grid spacing. Over NAmrM, biases are smaller and show minimal sensitivity to model grid spacing.

Increased intraseasonal variance (2-30 day band) in the 10 km simulation is linked to more intense low-pressure synoptic systems over SAsiaM and intense African easterly wave activity over WAFriM. All simulations agree on the diurnal precipitation peak timing, with the 10 km simulation marginally performing better over continents.

Our results demonstrate that fine grid spacing alone does not uniformly improve monsoon simulations. Some features, such as the precipitation diurnal cycle, are improved while existing biases in mean precipitation and variability are enhanced. This underscores the role of region-dependent sensitivity of grid spacing governing monsoon dynamics.

Questions:

(a) Is it a goal to evaluate the new ICON core (line 7, but not mentioned below) or just the grid-spacing dependence of global/regional monsoon representation?

This work aims to: (a) validate the performance of the ICON model for global and regional monsoon systems, and (b) identify the sensitivity of monsoon representation to horizontal grid spacing.

The refactoring of the dynamical core to GT4py is merely a code reimplementaion for computational efficiency and does not include any physical or dynamical changes from the original Fortran-based ICON version documented in Dipankar et al. (2026). Therefore, evaluating the GT4py core *per se* is not a goal of this study. Rather, we use the GT4py version as the computational platform to conduct our monsoon validation and grid-spacing sensitivity experiments.

To the authors' knowledge, this is the first extensive validation of the ICON model (in any configuration) for global and regional monsoon studies. We have now removed the "ICON GT4py" statement from the abstract to avoid confusion and to focus clearly on the validation and grid-spacing analysis. A description of the GT4py ICON code is provided in the methodology section for completeness, but the scientific conclusions do not depend on the specific code implementation.

(b) **What is organized moderate variability:** Our study shows that the variability associated with intraseasonal scale (30-90 day and 2-30 day) bands increases with fine grid spacing. To determine if intense precipitation in fine grid spacing simulations disproportionately influences the variability estimates, we applied a $\log(1+P)$ transformation to all precipitation (P) datasets prior to band-pass filtering. This transformation reduces the statistical leverage of intense rainfall, while preserving the timing of precipitation events.

After transformation, the systematic increase of variance persists across both the 30--90 and 2--30 day bands over SAsiaM and WAFriM domains. Crucially, the transformed variance ratios are below unity for 2--30 day, revealing that all simulations underestimate organized

precipitation variability when intense precipitation is statistically controlled, while the ratio of 30--90 day in fine grid spacing remains near unity.

We define "**organized moderate variability**" as the precipitation variability associated with statistically controlling (i.e., down-weighting) intense precipitation events, which largely originate from convective systems. This distinction allows us to separate variability contributed by extreme events from more moderate, organized precipitation processes.

For the abstract to remain focused and concise, we have removed the statement about statistically controlled precipitation transformation. A detailed discussion is provided in Section 3.4 of the manuscript.

(c) Excellent representation of diurnal cycle?: Now the statement in the abstract reads as "All simulations agree on the diurnal precipitation peak timing, with the 10 km simulation marginally performing better over continents. (L24-L25)

(d) We have corrected the increasing grid spacing to fine grid spacing (L25).

2. Has the simple linear interpolation (line 120) of monthly SST data to daily data some impact on monsoon onset/offset? In the pre-monsoon phase this should slightly underestimate SST warming and this slightly increases the land-sea contrast. Perhaps this can explain a few days of monsoon onset/offset bias?

Thank you for this very important suggestion. We completely agree with the reviewer, and accordingly conducted a sensitivity experiment to address this point.

In the manuscript, we used sea surface temperatures from the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for the AMIP simulations. For a sensitivity experiment, we used the ESA Climate Change Initiative (ESA-CCI) daily SST product, whereas the AMIP SST is available at a monthly resolution. Figure 1 shows the absolute SST magnitude over the extended SasiaM (Fig. 1a-c), WafriM (Fig. 1d-f), and NAmM (Fig. 1g-i) regions, along with the difference between AMIP and ESA datasets during the pre-monsoon season (MAM). Over the Indian Ocean, AMIP SST is approximately 0.5–1°C colder compared to ESA, with similar cooling observed over the tropical Atlantic off the coast of NAmM. Over the Pacific, the differences are mixed, while AMIP is colder in the western Atlantic.

As the reviewer pointed out, the underestimation of AMIP SST compared to ESA during the pre-monsoon period may be further enhanced by the interpolation from the monthly to daily scale. This would likely increase the land-sea contrast during the monsoon initiation phase and could potentially lead to an early onset. This is particularly relevant over SasiaM and NAmM due to their strong dependence on surrounding oceans, specifically the land-ocean temperature contrast. Figure 2 confirms this expected behavior. Over the SasiaM domain, especially over the eastern Indian subcontinent, the daily-updated (ESA SST) ICON simulation shows a delay in monsoon onset of approximately 6–10 days (Fig. 2d) compared to the monthly-updated (AMIP SST) simulation. This is also true for the WafriM domain, where the daily-updated SST

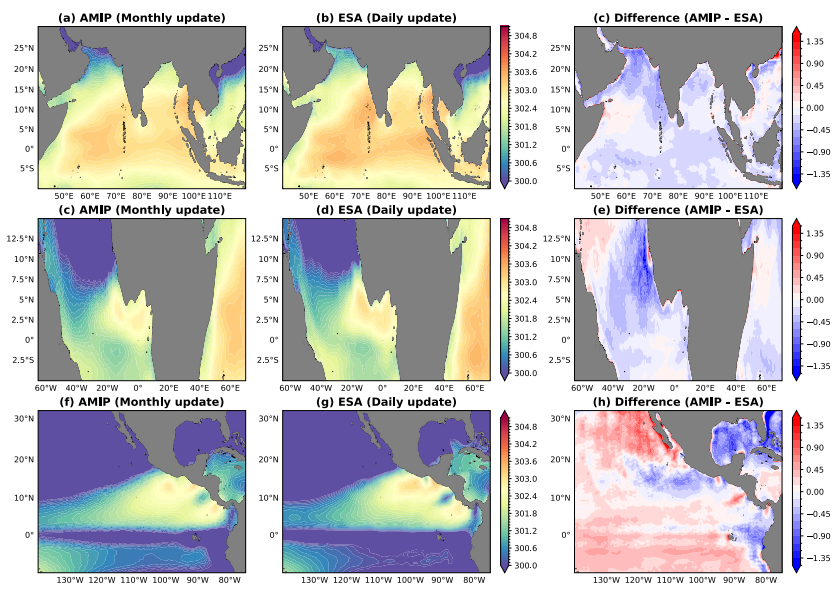


Figure 1: Sea surface temperature (SST) boundary conditions during the pre-monsoon season used in the ICON reference simulation and a sensitivity experiment. (a) Monthly AMIP SST (reference simulation) over South Asia. (b) Daily ESA SST (sensitivity experiment) over South Asia. (c) Spatial difference in SST (AMIP minus ESA) over South Asia. (d-f) Same as (a-c), but for the Africa region. (g-i) Same as (a-c), but for the North America region.

simulation shows a delayed south-to-north propagation, particularly over the western Sahel region. We would like to bring to the attention of the reviewer that, due to computational constraints, our sensitivity experiment only includes 4 simulated years.

These results confirm that our ICON monsoon onset simulations in the manuscript are sensitive to the monthly-updated SST boundary conditions. However, since our comparison of grid spacing sensitivity (10 km, 40 km, and 80 km) uses the same SST boundary dataset, our results remain valid for quantifying the role of grid spacing. We have noted this sensitivity in the manuscript in L364-L366.

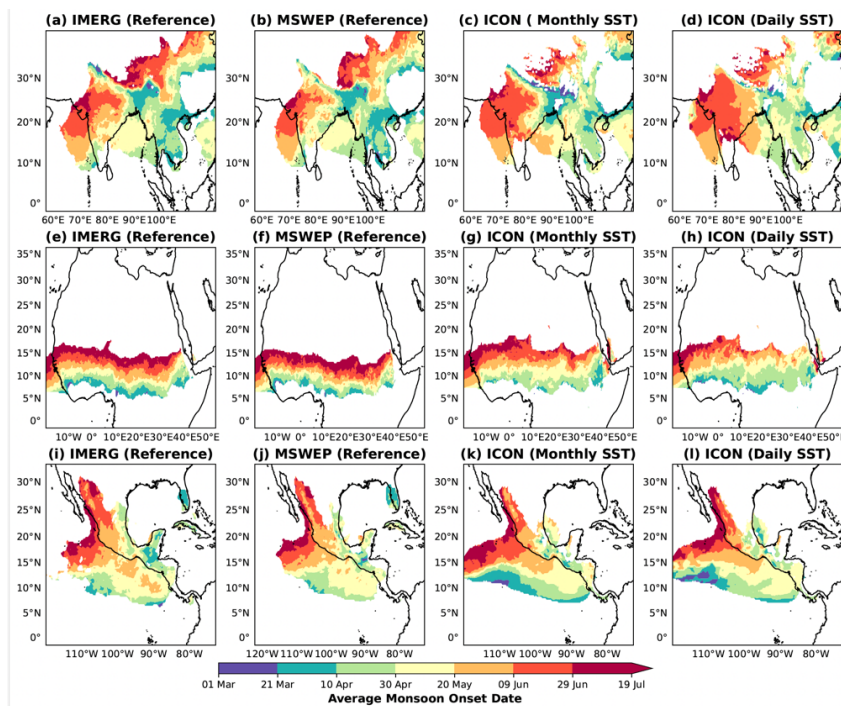


Figure 2: Spatial patterns of monsoon onset dates over the (a-d) SAsiaM, (e-h) WAFriM and the (i-l) NAmM domain in IMERG, MSWEP and two ICON simulations. The Monthly SST update is the reference ICON simulation used in the manuscript, while the daily SST update is a sensitive experiment with ICON using the ESA high resolution dataset (10 km).

3. Significance: (line 330) "indicates that 40 km performs best". This statement is based on a relatively small number of simulation years and the distributions shown in the violin plots (Fig. 4) are well overlapping. I assume here there is no conclusion about best grid-spacing possible.

Thank you for bringing this to our attention. The reviewer is correct that this statement is based on a relatively short simulation period (10 years) for quantifying monsoon onset bias. However, while acknowledging the sample size, we conducted significance tests, and the results shown in the violin plot are statistically significant.

Nevertheless, we agree with the reviewer that concluding the ICON 40 km run performs "best" is too strong a statement, given that the onset median (bias) for the 10 km, 40 km, and 80 km runs are 7 (7), 3 (5), and 12 (12) days, respectively, differences that are relatively small. We have therefore rephrased this sentence as follows:

"This indicates that the 10 km and 40 km run marginally improve the onset bias compared to 80 km. The difference in the onset bias is very small between 10 km and 40 km grid spacings, highlighting a non-monotonic relationship with grid spacing". (L339-L341)

Furthermore, in the manuscript, we have highlighted that the grid spacing sensitivity is non-monotonic, and thus we did not intend to identify a clear winner among the ICON simulations.

4. Why do you exclude the southern hemispheric monsoons from the evaluation, but show them in the Figs?

We have acknowledged in the manuscript (L280) that we focus only on the tropical monsoon systems in the Northern Hemisphere (NH) for this study. We agree with the reviewer that Fig. 1 and Fig. 2 show the performance of Southern Hemisphere (SH) monsoons in addition to NH monsoons.

However, we would like to emphasize that our inclusion of both NH and SH domains is solely to demonstrate the representation of monsoon domains in terms of spatial extent and to quantify spatial correlation, RMSE, and skill scores. Including a detailed discussion of SH monsoon characteristics, such as onset, variability, and process-oriented understanding, would significantly complicate the manuscript and make it excessively lengthy. As the reviewer can see, the manuscript already provides a very detailed analysis of the NH monsoon with extensive supporting details.

While we sincerely wanted to include SH discussions, the manuscript already struggles with maintaining a focused and concise narrative. In our next detailed study, we will provide results and descriptions for the SH monsoon. For the benefit of readers, we have emphasized in L273-L275 that:

"For the purposes of this article, we focus on tropical monsoon systems in the Northern Hemisphere, but we show a preliminary investigation of monsoon domain spatial extent from various ICON simulations for Southern Hemisphere monsoons. A detailed investigation of the characteristics of SH monsoons is, however, beyond the scope of this article." (L281-L283)

5. l632: I guess, here decreased not increased grid-spacing is meant? Does this worsening performance of precipitation simulation over ocean imply that there is a tuning issues in the model? Later (l647) you state that the precipitation contribution by the convection parameterization independent of the chosen grid-spacing. The parameterization should contribute less at 10 km grid-spacing.

Yes, the reviewer is correct to point out that in we wrote "increased grid spacing" instead of "decreased grid spacing." We have now corrected this sentence to state that fine grid spacing leads to precipitation biases in oceanic regions (L661).

Furthermore, we agree with the reviewer regarding model tuning issues. The current setup is built on the previous expertise of the authors' model setup, mainly for regional simulations (Pothapakula et al., 2020; Soerlund et al., 2020). The authors have not performed any tuning experiments for the simulations used in the current manuscript. Systematic tuning requires a large consortium effort, and unfortunately, we have limited resources. Given the decent performance of ICON, the authors have documented its skill in capturing global monsoon systems.

Nevertheless, we plan to adopt objective tuning using a Linear Meta-Model optimisation (LiMMo) to derive an optimized model configuration in future studies. Specifically, we aim to tune the most important parameters for the surface energy budget, which affect turbulent heat fluxes. These parameters include the resistances to the fluxes in the laminar boundary layer: `rlam_heat`, `cr_bsmin`, `rat_lam`, `rat_sea`, and `rsmin_fac` in ICON. (We kindly refer the reviewer to lines 193–195 in the manuscript by the ICON-CLM community: <https://doi.org/10.5194/egusphere-2025-4726>)

Regarding the contribution from the deep and shallow convection parameterization scheme, we agree with the reviewer that the 10 km grid spacing should ideally contribute less to the total precipitation. In response to Reviewer 1, we performed a detailed analysis of precipitation partitioning between convective and grid-scale (resolved) precipitation (we kindly refer the reviewer to Response to Reviewer 1, Point 5). Over the monsoon core regions, we notice that the 10 km convective scheme contributes relatively less compared to the coarse grid simulation; however, it is important to note that the rate of increase in grid-scale precipitation is not proportional to the rate of decrease in convective precipitation. This is also very true over oceanic regions, where the contribution from stratiform precipitation dominates while the rate of decrease from the convection scheme is slow. Hence, we see such strong biases over the oceans.

Furthermore, rain conversion, especially in the lower troposphere, is very efficient, indicating efficient microphysical processes in the fine grid spacing simulation (10 km) compared to the coarser simulation (80 km) (detailed analysis in response to reviewer 1, point 5). We believe these factors contribute to excessive precipitation in fine grid spacing simulations, both in core regions and over oceans. This indicates a need for tuning efforts in turbulence, microphysics, and boundary layer process.

Minor issues:

l133: Tergen → Tegen

We apologize for this typo. Now this issue is fixed. (L137)

l165: "V" is horizontal (add) wind vector (and thus should be bold). Also here and in the following "kg/m-s" is uncommon. Correct would be kg/(ms) but we are used to kg/ms too. At line 441,2 you use inconsistently but correctly kg m⁻¹ s⁻¹.

Thank you for assisting us in getting the units correct. We have now kg/ms throughout the manuscript, and corrected the units. Also, the V is turned to bold to represent a vector.

l289: "skill scores" – usually a reference prediction/simulation (like a random, persistent ... simulation) is used in the definition of skill scores. In this sense you do not consider skill scores, just scores.

We thank the reviewer for this comment. While we acknowledge the strict predictive definition of "skill score," our study is a climatological evaluation (10-year simulation vs. observed climatology), not an operational prediction. In the climatological literature, metrics such as POD, FAR, and Bias Score are routinely referred to as "skill scores" when evaluating model performance against observed climatology. To avoid any potential confusion, we have clarified in L289 that we use *verification metrics* (POD, FAR, Bias Score, Accuracy) to evaluate model skill in capturing monsoon domain extent. (updated in L162)

l325: "common bias" – references?

Thank you, we have now added this essential reference.

Senan, Retish, et al. "Impact of springtime Himalayan–Tibetan Plateau snowpack on the onset of the Indian summer monsoon in coupled seasonal forecasts." *Climate Dynamics* 47.9 (2016):

l389: dipole → dipole bias?

Thank you, corrected.

l656: "... variability show_s"

The sentence is corrected.

l837: incomplete reference

Now the reference is complete.