

Peer-Review for “Advancing the Capabilities for Efficient Hurricane-Centric Simulations with the Atmospheric Model ICON” by Senf and Cremer

Summary

This study presents a workflow enabling efficient high-resolution simulations of tropical cyclones with the ICON atmospheric model. Rather than using a single large fixed limited-area domain spanning an ocean basin, the method generates a sequence of tube-shaped unstructured grid segments that follow the hurricane track, each covering 12 to 24 hours of the storm's movement. The workflow automates hurricane tracking, grid generation and initial condition merging across successive segments via distance-weighted blending. All simulations are performed in a pure limited-area model chain: ERA5 drives a fixed outer regional domain (5 km), which drives a fixed inner domain (2.5 km), which in turn provides initial and boundary conditions for the series of high-resolution segments. The method is demonstrated on Hurricane Paulette (2020), with efficiency gains of 13 to 175 times compared to a conventional fixed-domain approach are reported. Spin-up effects at segment transitions are argued to be manageable as well.

The manuscript is well-written and the figures are well designed. Figs. 8 and 9 are excellent, and give a compelling illustration of what the method actually produces at hectometer scales. I recommend acceptance after the authors address the comments below, notably in how they define the spurious transient response to the segmentation.

Main Comments

1. The spin-up characterization in Sect. 3.5 (Figs. 10 and 11) is the paper's central methodological claim, but this diagnostic could be made more convincing by comparing every segment with its neighbour, rather than with its own initial state at $t = 0$. As I understand, the author's diagnostic of what constitutes “spin-up” mixes two contributions that are hard to separate: the physical evolution of the hurricane and any spurious adjustment from the inconsistent initial conditions. Fig. 11 has a related issue: the PDFs are colored by segment index, and the interpretation is that a shift of the red bars indicates systematic spin-up. However, later segments correspond to a more intense and more variable storm, so how could their larger anomalies be attributed to restart quality alone? Therefore, the more informative diagnostic would be to overlay the same physical quantity from two consecutive overlapping segments k and $k+1$ during their shared time window. Then, any discontinuity at the restart time could be attributed to the initialization. This could be diagnosed quickly, as it requires no additional runs: the overlapping data exists by construction.
2. A simpler alternative to the restart chain that is not discussed anywhere in the paper is a single tube-shaped domain covering the full hurricane track, run as a continuous simulation using standard ICON checkpoint restarts on the same fixed grid. This would eliminate all IC-merging artifacts and spin-up at segment boundaries while still remaining more efficient than any rectangular fixed domain. For Paulette's relatively straight track, the cell-count penalty compared to the 12h segment approach would be modest: the along-track extension adds cells, but the tube geometry retains the full efficiency advantage over a rectangle. I get the argument in favor of short segments (that is cells far ahead of and behind the storm are computed at every timestep, far from the cyclone), but whether this is a significant cost depends on track geometry, storm speed, and the chosen domain width. To me it is not so straightforward for a case like Paulette. Perhaps the authors could provide a brief quantitative cell-count+timestep estimate comparing both approaches for this case, and discuss under what conditions the restart chain is comparable, if not preferable due to the absence of restarting artefacts?

Minor Comments

1. Line 138: I am not familiar with this criterion of “ten daily revolutions” for identifying cyclonic features. What does this correspond to physically in terms of vorticity magnitude, and how sensitive is the tracking to this threshold?
2. Line 170: Is the distance-based weighting used to blend IC data from two sources at segment boundaries a linear function, or something more complex?
3. Line 193: “One of the longest tracks”, this makes this case particularly beneficial from the techniques, as opposed to a hurricane like Melissa that strengthened and grew while staying relatively static.
4. Line 214: Is the same partial parameterization carried through to the finest nests, down to 300m-resolution?
5. Line 225: I am not sure if this flag exists in the version of ICON the author used, but *lsstice* allows to prescribe hourly SST profiles. For a fast changing extreme event such as a cyclone, especially at 300m-resolution, this would be more accurate and could be mentioned for future implementation of this framework.
6. Line 280: As far as I understand, a warm start requires storing two timesteps for the time-marching scheme. Is this what is being used here, and blended in between segments using the distance-based weighting?
7. Figure 7a: it is somewhat difficult to differentiate the two cases, but I guess that is the point? If not, perhaps one could be made with some opacity or use another colormap?
8. Figure 7b/c: I also have a difficult time sorting out the various overlapping resolutions from each case. The segment appear very clearly though.
9. Figure 8: This is really well-made figure! Do I see correctly that the eye of the storm is weaker, more diffused in the highest-resolution case? This is a bit counter-intuitive, so I presume this is caused by the proximity with the domain’s boundaries?
10. Figure 9: Another beautiful figure, but I struggle to understand the author’s point here: I cannot see any discontinuity with the background (opaque vs transparent contours), which looks in fact quite smooth.
11. Line: 423: The amount of computational gain is real, especially for 12h-restarts. But eventually, as the cost diminishes due to further restricting the spatio-temporal bounds of the simulations, is the solution not too-strictly determined by the lateral/initial conditions? Can the 300m-resolution run sufficiently long to allow for the development of fine-scale physics, or has it become a sophisticated downscaling technique? I realise that there is no simple answer to this question, and that it is all a matter of trade-off, but it is worth discussing in the conclusion.