Response letter for *Advancing the Capabilities for Efficient Hurricane-Centric Simulations with the Atmospheric Model ICON*

Response to reviewer #1:

Review for "Advancing the Capabilities for Efficient Hurricane-Centric Simulations with the Atmospheric Model ICON" by Senf and Cremer for GMD

Summary

This study addresses the issue of performing high-resolution, global-storm resolving simulations while balancing computing resource and storage requirements. The authors propose a method to mimic Lagrangian-like simulations by shifting grids along the path of hurricanes, the phenomenon of interest in this study. The focus is on the efficiency gain as well as possible spin-up effects.

Dear Fabian and Roxana, it was a pleasure to read the manuscript. It is well structured, nicely written, and informative. The figures are clear and use appropriate colormaps. I believe the study should be published after considering my comments below, which are all of minor nature or editorial. Best, Nadja

Dear Nadja, we sincerely thank you for your thorough and insightful review. Your detailed comments have been invaluable in strengthening our manuscript and clarifying important technical aspects of our work. We appreciate your constructive feedback and the care you took in reading and reviewing our study. Best, Fabian and Roxana

Minor comments

- Line 138: I do not understand what you mean by "ten revolutions". Could you elaborate on that? That is mostly my ignorance and not being familiar with tobac. Also why, do you need to coarse-grain the vorticity field? What does the convolution filter exactly do?

  We were somewhat unclear here and have revised the wording accordingly. We mean that, for the field of coarse-grained relative vorticity, a threshold of $\zeta_c = 10\,\mathrm{day}^{-1}$ was chosen to identify the hurricane track. This threshold was determined manually and represents a good compromise to detect the investigated hurricane early and to obtain a continuous track of good quality.

  The vorticity field shows vortex structures on many different scales. To identify the hurricane track, it is necessary to distinguish the hurricane vortex structure from vortices on smaller scales. This is achieved by applying a running-average filter, which smooths the small-scale structures and thereby facilitates identification of the hurricane.

  The convolution filter is a mathematical operator that is applied to the vorticity field to perform the smoothing. It computes a weighted sum of values in the neighborhood of each point in the field, where the weights are all equal for a running-average filter. This leads to a reduction of high-frequency components in the field and facilitates identification of the hurricane. We made this more explicit in the revised manuscript.

- Line 224: In what way are these spin-up effects clearly visible? Did you perform a similar analysis as you did later for your inner nests?

  We removed the corresponding sentence from the revised manuscript, as it could be misleading. Spin-up effects in the base run are not particularly interesting for the current study, and any associated issues can be easily avoided by excluding the first full day of base run data.

- Line 228: What is the impact of having the daily-updated SST?

  We have not conducted extensive and systematic analyses to quantify the impact of daily SST updates. Therefore, we are hesitant to make definitive statements about the effects on our simulation results.

  For the base simulations, we find that the SST setup can significantly influence the hurricane track. In Fig. 1 we show Paulette's best track (from HURDAT, white) compared to simulated tracks from the coarser simulation (right, R2B9) and the finer simulation (left, R2B10 base run). For the coarser simulation, better agreement with the observed track is achieved when SST is held temporally constant. However, initializing the simulation at a later time (+ 3 days) has a larger impact on track forecast improvement than the SST setup does. For the finer simulation, track forecast deviations are smaller, and differences between SST setups are less pronounced. Nevertheless, the qualitative picture remains the same. Further studies on this topic would be very interesting but are currently beyond the scope of what our study offers.

Figure 1: Paulette best track (from HURDAT, white) compared to simulated Paulette tracks from (right) coarser simulation (on R2B9) and (left) finer simulation (base run on R2B10). Tracks are differently colored by sst setup, and init time

- Line 254: Do I understand correctly, that I as a user can specify the targeted grid resolutions? Or is it required to follow the halving of the resolution, as typical in an online nest? Could I perform a coarser simulation at 1.2 km and then immediately jump to 100 m, for example?

  Yes, this functionality is in principle possible. However, the current implementation misses the distinction between the grid of the base run and the grid from which refinement is performed. This means that currently the refinement is always performed from the base run grid, by halving

the grid spacing. The addition of a second grid for refinement can easily be implemented and left for future extensions.

We have added a sentence in the outlook part of the conclusions section to highlight that further research on this topic is needed.

- Line 255: What is your estimated effort in making more than three nests possible? What is technical difficulty behind this? Also, is there a check, that the widths of the grids are not so large, that the inner grid would overlap with the nudging zone of LBCs of the outer grid? In your case, it is clear that this is not the case, assuming you are using the standard configuration of 14 grid cells as a nudging zone.

  The real technical difficulties will probably only become apparent during implementation and testing of the functionality. We assume that implementing more than three nesting levels would mainly require changes to the runscript template. There, the corresponding required files for IC, BC, and extpar must be selected. Additionally, namelist parameters may need to be specified as a list that extends sufficiently long. All these aspects must be thoroughly reviewed and tested.

  Indeed, we have not implemented additional checks for grid properties. We assume that users will verify and adjust the grid settings themselves making the method not failsafe in that respect.

  And yes, it is certainly sensible to leave sufficient space to the edge.

- Line 279: Is it correct, that the base runs were performed with ICON v2.6.6, while the grid segments were done with v2025-04-2. Do you expect any discrepencies between the two model versions that are important to your hurricanes?

  We cannot say for sure whether significant differences arise from using different versions. We performed the base runs with version 2.6.6 because this was the current version at the time we started to work on the base runs. Since we were satisfied with the results of the runs after a series of adjustments and tests, we decided to use these simulations as a reference and did not repeat them with other ICON versions. When we developed the method for higher-resolution segments, ICON version 2025-04-2 had just become available.

  The inconsistency in the ICON versions we used is thus a reflection of the temporal development of our study. We believe that such development in the scientific environment is not atypical and hope that the differences between the versions are not so large that they substantially affect the results of our study in a qualitative manner.

- Line 282: What was the impact of adding graupel and ice number mixing ratios to the ICs? Did you see a reduction in spin-up?

  It would certainly be fine to have ICON's own diagnostics compute graupel, hail, and particle concentrations for the first calculated segment if they are not provided. Microphysical spin-up would lead to consistent cloud variables being established after some time (perhaps hours). However, if one considers the transitions to the next segments, the absence of information about graupel, hail, and particle concentrations would lead to abrupt changes in these fields.

Our method, however, places emphasis on the structurally consistent evolution of all dynamic, thermodynamic, and cloud variables over time, at least in the Lagrangian analysis region. This also makes it possible to consistently track existing smaller-scale convective elements across segment transitions. The inclusion of graupel, hail, and particle concentrations in the ICs is so essential for us that we have not performed any further tests without these quantities in the ICs.

We added a sentence to the revised manuscript to clarify this point.

- Figure 7: What is the temporal resolution of the output, as it seem quite noisy to me, especially in the maximum 10 m wind speed. Would a running mean help here? Can you also plot the maximum 10 m wind speed and minimum surface pressure of the base run, which is the white track?

  We output all 2D variables (surface, radiation and vertically-integrated cloud variables) on the regular longitude-latitude grid every 5 minutes. Convective activity generates particularly high gustiness of the 10 m wind, which explains the fluctuations on the very short time scales.

  We have implemented all suggested updates to Figure 7. Using a running mean with a width of 2 hours, we smoothed the curves for minimum surface pressure and maximum wind speed. The original curves are retained as light gray shading in the background. Additionally, we have added the base run data as a black curve for comparison. We have updated the figure caption accordingly.

- Figure 9: Could you be more specific regarding the discontinuous jumps in the subplots? I am not sure I fully see them, maybe highlighting them with some circles could help, or you point to them more explicitly in the figure caption.

  We recognize that our original description was unclear, and we have revised it in the updated manuscript for greater clarity.

  To explain: The discontinuous jumps arise from the distinction between Eulerian and Lagrangian reference frames. In the Eulerian perspective, the individual segment domains are fixed in space and time, while the Lagrangian analysis domain moves with the hurricane across these fixed domains. Conversely, in the Lagrangian perspective, the analysis domain remains stationary while the segment domains move beneath it in the opposite direction. At reinitialization times, the segments jump to new positions, causing the segment domain boundaries to shift discontinuously relative to the Lagrangian analysis domain. These discontinuous jumps, visible as abrupt spatial displacements of the segment domains in the figure, are what Figure 9 further illustrates.

- Line 323: Was the subtimestepping of the dynamical core changed in your grid setups? If so, why? Because of stability reasons?

  The number of dynamical substeps is set to the default value of 5 for our simulations. We carefully check all logfiles again and did not find any indication for adaptive changes in the number of substeps.

  We changed the formulation in the revised manuscript.

- Line 336: Is your time step at 2.5 km grid resolution 6 seconds? If so, is this is because of stability reasons? I would have expected a higher time step.

  6 seconds is applied at the 1.2 km grid resolution. We reduced the time step from 10 seconds - the limit imposed by the horizontal-flow CFL criterion - to 6 seconds for stability reasons. For comparison, the 5 km and 2.5 km setups use time steps of 30 and 15 seconds, respectively. The high-resolution setup employs 150 vertical levels, representing a relatively fine vertical discretization. Combined with the large vertical wind speeds generated by intense deep convection, this fine vertical resolution appears to necessitate the smaller time step.

  We mentioned 1.2 km grid spacing in the text body, but we will clarify this point in the revised manuscript.

- Line 343: I am convinced that your method is saving computing resources, and believing right now your numbers, this would lead to a substantial decrease in computing resources. However, I am wondering if this is true and not a slight overstatement, given that in your 3-nest setup + narrow width, the hurricane is cut off across all resolutions, but strongest in the case of 300 m naturally. My thinking here is, if your statement with a factor of up to 175 is a bit too strong, because coming from a scientific perspective, I am not sure how much we would gain from the innermost nest with the narrow width? Hence, I would be curious to hear "more realistic/applicable" numbers.

  We revisited the speedup part in depth and redesigned the calculations accordingly. The most important change is that we dramatically reduced the size of the considered reference case: the classical limited-area setup. In the earlier submitted manuscript version we had built a reference domain that covered the entire Atlantic, making it usable for many different studies. We have now changed that. In the current manuscript we only consider the Atlantic region that actually contains the hurricane (in our case, Paulette). We defined this region with a bounding box and we believe that users of high-resolution simulations might choose a similar approach to keep the final domain as small as possible. As a result of these changes, the speed-up factors are reduced by a factor of four.

  Another important change concerns the introduction of the tube as a reference case. This was suggested by Reviewer #2 to better assess the performance impact of grid segmentation. In our presentation we have now adopted the following strategy: first, we discuss the speed-up obtained by moving from a bounding-box domain to a tube domain, which yields a factor of roughly 2–8 depending on the tube's width and, of course, on the compactness of the hurricane tracks. Next, we introduce segmentation of the tube, which typically adds another factor of about 2–6. This way of presenting the results leads to more moderate speedup figures being shown, encouraging potential users to consider the limitations more carefully. Among those limitations is the fact that the narrow domain does not contain important parts of the organized cloud system and therefore sometimes functions more like a dynamic downscaling approach.

- Figure 10: Any idea what is happening to segment 02 and 03? Why are the behaving quite different to the other segments in your analysis? Why do you see for some segments and

variables the zigzag pattern? What is your output frequency that your analysis is based on?

In response to comments from reviewers #2 and #3, we have revised the spin-up analysis in Figures 10 and 11. New simulations were performed in which an additional three hours were computed for each segment, creating a three-hour overlap with the subsequent segment. This analysis is now substantially more robust and can be better utilized for quantifying spin-up effects.

With the new method, the zigzag pattern has disappeared. We have also modified the figure to display only the later segments, which clearly illustrate the spin-up effects in a systematic manner. The output frequency for the analysis is 5 minutes, which is the same as the output frequency of all 2d variables.

- Figure 11: So, blue represents the 10th percentile, while red the 90th? Could you specify this in the figure caption? Was the analysis here and Figure 10 also done for the higher resolutions? Are you changing vertical levels? This would also impact your spin-up. Have you looked at the cloud quantities vertically resolved before integrating them? I would be curious if you see some time dependence of the spin-up across the vertical dimension.

  The color coding in Figure 11 represents different segments, with blue indicating earlier segments and red indicating later segments. We have moved the color coding description to the beginning of the figure caption to make this point clear from early on. Due to the new method for analyzing spin-up effects, data are no longer available for performing this analysis at higher resolutions. This is because we needed to rerun the simulations with a three-hour overlap for each segment, which was computationally prohibitive for the higher-resolution setups.

  The number of vertical levels increases from 70 to 150 when transitioning from the base run to the higher-resolution segments. During further horizontal refinement, the number of levels remains constant at 150. We have not performed spin-up analyses for the vertical structure and therefore cannot make statements about spin-up effects in the vertical dimension. To further investigate these questions, three-dimensional data with high temporal resolution would need to be stored.

Editorial comments

- Line 8: You usually refer to your tropical cyclones, as hurricanes, so to stay consistent I would also call them in this instance hurricanes instead of tropical cyclones.

  done.

- Line 45: add the reference to Zängl et al., 2014 after introducing ICON

  done.

- Line 55: I am being very picky here, but what do you define as “good”? Realistic? Accurate? Precise? In relation to what?

  changed to ”accurate”

- Line 66/202: You introduced ICON already so I would omit the Earth system part / weather and climate model part.

  mentioned parts omitted.

- Line 107: Regarding the merging of ICs: Could you point to your extensive discussion later? I stumbled across this while reading and marked it hoping you would provide more details later, which you did. So, a simple reference to the section would already help.

  done.

- Figure 3: R02B10 should be explained in the figure caption, because this is only later used in the text. Same goes for Figure 4, and therein L70.

  done.

- Line 207: replace “times” by x and “km” should be km^2.

  done.

- Figure 5: The tiny world map in the upper right corner shows the extent of the shown reflect solar radiation flux, right? I was wondering, if you can add as well the nests described in Lines 207ff, such that it is clear how large these base runs are. Either for both R02B09 and R02B10, or at least for the latter.

  We implemented a yellow outline for R2B10 in the inset map of Figure 5. However, we decided not to include R2B09, as it would have made the figure too cluttered and less clear. We have added text to the figure caption to clarify this point.

- Line 307: there is a word missing in “must make between . . . ”.

  slightly rephrased the sentence

- Line 372: Because you showcase a variety of resolutions, it would help to reiterate which resolution you are talking about in the case of the “outer nest”.

  link to Tab. 1 added and domain configuration specified in text body.

Advancing the Capabilities for Efficient Hurricane-Centric Simulations with the Atmospheric Model ICON

Fabian Senf[1] and Roxana Cremer[1]

[1]Leibniz Institute for Tropospheric Research, Leipzig, Germany.

Correspondence: Fabian Senf (senf@tropos.de)

Abstract. Global storm-resolving simulations with kilometer resolution are increasingly replacing traditional climate modeling approaches and show particular potential for resolving the dynamics and effects of deep moist convection. These modern modeling methods are moving toward sub-km scales, leading to extremely high energy and resource requirements. This makes iterations, parameter optimizations, and sensitivity studies no longer easily feasible. For the class of propagating, large-scale weather phenomena such as hurricanes, high-resolution limited-area approaches in combination with Lagrangian methods are therefore of interest, in which refined grids are shifted along the path of the phenomenon under consideration. To create this capacity for the ICON atmospheric model, this study develops a flexible workflow toolkit to enable efficient simulations of hurricanes on a sub-km scale. Our approach leverages the flexibility that ICON offers through the ability to create custom grids. The concept divides hurricane tracks into overlapping temporal windows of 12-24 hours and generates customized grid segments that follow the hurricane's movement. The technical implementation automates key components of the workflow, including hurricane tracking, flexible grid generation, and preparation and merging of initial conditions across successive segments. The application of the workflow is demonstrated using Hurricane Paulette (2020) as an example, for which high-resolution simulations with grid spacings down to 300 m were performed using different segment configurations. The results show that the hurricane track remains consistent with the base run and depends mainly on the across-track width of the chosen configuration, while intensity metrics such as minimum pressure and maximum wind speed show significant sensitivities to resolution in our multi-nested setups. The efficiency gains are significant compared to traditional approaches with fixed limited-area domains: replacing a fixed domain with a track-following tube reduces resource requirements by factors of 2–8, and segmenting the tube into short, frequently re-initialized intervals adds a further factor of 2–6. Analysis of re-initialization effects shows systematic but manageable impacts during segment transitions. Nevertheless, the efficiency gains achieved by our method are so substantial that they justify the acceptance of the re-initialization effects. Our segment-based approach in the hurricane-centric reference system now allows for more flexible regional hurricane simulations with the ICON model and more efficient investigation of new research questions regarding the sensitivity of hurricane cloud systems at very high resolutions.

Short Summary

Computer models for hurricane prediction are becoming increasingly detailed but require substantial computing resources. We developed a flexible approach that follows hurricanes as they move, applying high-resolution simulations only where needed. This method reduces computing costs by up to one order of magnitude compared to conventional fixed limited-

area domains, while achieving resolutions down to 300 meters. The approach enables more efficient hurricane research and improved understanding of tropical dynamics.

1 Introduction

Global storm-resolving simulations on a kilometer scale are increasingly replacing traditional climate modeling approaches (Satoh et al., 2018, 2019; Segura et al., 2025). These modern modeling methods show particular potential for resolving the dynamics of deep moist convective phenomena and their various forms of spatial organization, and the resulting interactions between small and large scales in a more physically realistic manner. Global storm-resolving simulations are now constantly evolving from kilometer to hectometer resolutions, enabling unprecedented detail in the representation of atmospheric processes (Hohenegger et al., 2023; Klocke et al., 2025). However, the development and calibration of these high-resolution approaches pose significant challenges (Shukla et al., 2010). They are extremely energy- and resource-intensive in terms of both computing resources and storage requirements. As a result, iterations, parameter optimizations, and sensitivity studies can no longer be easily performed (Mauritsen et al., 2022). For this reason, alternative and efficient approaches are needed to understand how well regional to locally limited phenomena are represented at high resolutions and standard parameter settings.

One typical solution to this problem is to use limited-area domains with high spatial resolution that refine the global model simulations in a predefined limited area and within a certain time period, thereby significantly reducing the computing resources required. These limited-area solutions can be designed to be consistent with global simulations by using initial and boundary conditions derived from them and working with the same set of physical parameterizations. Of course, the capability of limited-area approaches is inherently somewhat constrained, since interactions with boundary conditions can introduce errors and the upscale influence of small-scale processes onto the larger scales is not captured adequately. The atmospheric component of the Earth system model ICON (ICOsahedral Nonhydrostatic, see Zängl et al., 2014) already includes the possibility of simulating such regional domains with nested grid refinements (Stevens et al., 2020; Zängl et al., 2022; Hohenegger et al., 2023). These approaches are usually limited to spatially fixed domains that must be defined in advance. This concept is very well suited for spatially and temporally limited phenomena, such as the diurnal evolution of convection over land under rather calm wind conditions. However, the approach with fixed regional domains is less suitable when the weather phenomena under consideration propagate over large distances, such as hurricanes, and thus the covered area is as large as entire oceans. For this class of problems, high resolution Lagrangian approaches are interesting, in which the refined grids are shifted along the path of the phenomenon under consideration. Such approaches have been successfully applied for a long time for flexible high resolution hurricane forecasts in operational settings (Kurihara et al., 1998; Gopalakrishnan et al., 2002).

For predicting the path of hurricanes, km-scale model approaches have been shown to reach reduced track position errors (Dong et al., 2020). For accurate simulations of the structure and intensity of hurricanes, an explicit and well-designed representation of deep moist convective processes also appears to be key (Gao et al., 2023). It has been further shown for Hurricane Katrina, that its intensification and structure of its convective bands only improved sufficiently when a spatial resolution of around 1 km is achieved (Davis et al., 2008). Systematic investigations in Davis et al. (2011) have also shown that intensity

biases with resolutions around 1 km are reduced, especially for very high wind speeds, compared to coarser model resolutions. The study argues that it is essential to resolve cloud structures in hurricanes, but also that significant improvements between 100 m and 1 km horizontal resolution are questionable. These examples can be continued and provide evidence that a high spatial resolution is essential for realistic hurricane simulations and that entering the sub-kilometer scale may need further investigations.

Together, high spatial resolution and flexible adaptation of simulation domains can help enable novel, high-quality hurricane simulations while keeping computing costs within reasonable limits. Our efficient approach presented here contributes to this goal and advances the capabilities of ICON model. Redesigning the grid and data management in ICON and the associated communication patterns is so challenging, and the risks of negatively impacting the outstanding performance achieved in Klocke et al. (2025) are so high, that we have decided to implement a workflow-based approach to make regional hurricane simulations more flexible. This approach is an intermediate step on the way to a fully Lagrangian model concept and has the advantage that we can completely retain the proven and optimized communication patterns and data structures of ICON. Our approach allows us to take advantage of the flexibility that ICON offers through its ability to create user-defined grids. In this way, we can perform targeted and efficient hurricane-centric simulations down to the hectometer scale without having to re-optimize the efficiency of the underlying model. We are thus creating the possibility of using an approximate Lagrangian approach with the ICON versions available today to create innovative datasets for hurricane investigations in very high resolution. With this approach, we would like to encourage the scientific modeling community in general to give more consideration to co-design methods that promote the parallel development of sophisticated global simulation approaches and flexible and efficient regional refinements.

The paper is structured as follows: In Sect. 2, we present the methodology of our flexible refinement workflow in detail. We provide a conceptual overview together with detailed workflow diagrams and discuss the portability of our methods to different high-performance computing platforms. Sect. 3 demonstrates the application of the workflow using Hurricane Paulette (2020) as an example. Different refinement strategies are explored, and the results of high-resolution simulations are presented and discussed. There we also provide an analysis of the performance and resource requirements of our approach. Finally, Sect. 4 summarizes the key findings and outlines potential future directions for research in this area.

2 Methodology

2.1 General Overview

Figure 1 provides a general overview of the flexible refinement workflow used to create high-resolution hurricane simulations. The various steps are carried out sequentially. The first step is to create a base simulation of a hurricane. Both global and sufficiently large regional setups can be used for this purpose. In the example case presented in Sect. 3.1, a limited-area setup covering the entire tropical Atlantic is employed. Global storm-resolving simulations, such as the projections of the nextGEMS project (Segura et al., 2025), could also be used as base simulations. A wide range of data from the base simulation will be required in subsequent steps.

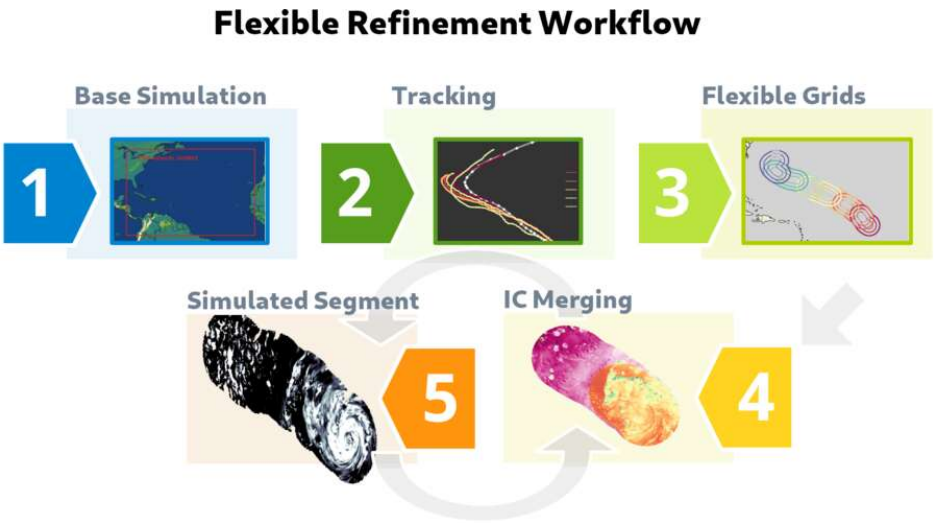

Figure 1. Overview of the flexible refinement workflow: Five steps are shown in different colored boxes and numbered accordingly. The steps of the workflow include (1) creating the base simulation, (2) identifying and tracking the hurricane, (3) creating the flexible grids and other necessary simulation data, (4) combining the initial conditions from the base and newly created simulation data, and (5) performing the actual high-resolution hurricane simulation. Gray arrows indicate the flow in the workflow diagram, with steps 4 and 5 being performed as iterative loop over the respective hurricane segments.

An important prerequisite is that the base simulation generates a hurricane and reproduces its characteristics sufficiently well. In the second step (see Fig. 1), the hurricane is identified in the base data and tracked over time. This tracking process creates a Lagrangian reference system, initiating the transition from a fixed domain to a flexible, hurricane-centric setup. It is important to note that we follow the hurricane as a large-scale structure which is not necessarily the same as shifting the domain according to the mean horizontal wind.

In the third step, we split apart the hurricane track into individual elements, which we refer to as "*segments*". Based on these, flexible and user-defined grids are set up that overlap so that previous information can be used for subsequent simulations. All essential data like external parameters, lateral boundary and initial conditions (ICs) are created specifically for these grids. This completes the pre-processing workflow.

Steps four and five build the production workflow. Both steps are performed alternately for each simulation segment. Step four generates the respective ICs from which the atmospheric simulation is started. For the first segment, ICs are created solely from the base simulation. Step five conducts the actual high-resolution hurricane simulations in the respective grid segment and stores the necessary ICs for the next simulation. For the second and the subsequent segments, the loop goes back to step four and data from the previous segment is reused in the overlap area for the new simulation. Where there is no overlap between two subsequent grid segments, ICs from the base simulation are taken. Thus, step four produces a mixed initial state consisting of a merge of two information sources (see Sect. 2.3 for more details).

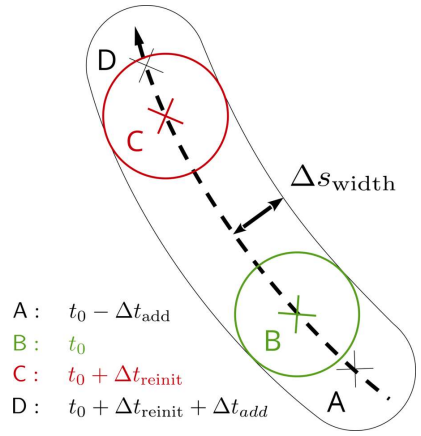

Figure 2. Schematic representation of a track segment. The reference track is represented by a thick, dashed black line. A grid segment with a width of $\Delta s_{\mathrm{width}}$ is constructed around it (thin solid black line). The spatial points A to D are determined by their temporal sequence along the track. The circles around points B and C indicate the areas of interest for analysis at the start and end time of the simulation within the segment. Points A and D show the extension of the grid segment by a predefined time Δt_{add}. For simplicity, only the outer nest of the track segment is shown. Further nests follow the same scheme, but lie inside the outer nest with smaller across-track widths.

2.2 Segment Concept

Our entire workflow is based on the idea of simulating a hurricane only within its vicinity along its temporal evolution. To accomplish this efficiently, we rely on a segment concept for which a reference track is divided into individual *track segments*. The length of the track segments is mainly determined by the envisioned re-initialization time interval, $\Delta t_{\mathrm{reinit}}$ (plus some additional buffer). Re-initialization data must be available from the base simulation for the chosen interval, e.g. 12 or 24 hours, to support the re-initialization of the higher-resolution simulations.

Figure 2 shows how a track segment is constructed. The points B and C correspond to the location of the hurricane center in the base simulations at times t_0 and $t_0 + \Delta t_{\mathrm{reinit}}$. The points A and D expand the track segment in the respective directions by a certain buffer distance to account for situations in which the hurricane moves slower or faster along its path in the higher-resolution simulations. For practical reasons, this buffer distance is also specified as a time interval, Δt_{add}. Combining re-initialization interval and buffer time, the selected track segment runs from point A to point D and thus covers the distance traveled by the hurricane e.g., from 21 UTC the previous day to 15 UTC ($\Delta t_{\mathrm{add}} = 3$ h; $\Delta t_{\mathrm{reinit}} = 12$ h).

A *grid segment* refers to a track segment that has been spatially expanded horizontally. It is constructed by expanding the selected track segment by a user-defined across-track width, $\Delta s_{\mathrm{width}}$, in all directions, forming a tube-like shape. Imagine a circular area with a radius of $\Delta s_{\mathrm{width}}$ and the center at the respective track point sweeping over the track segment. The circular section at a time t is defined later as Lagrangian analysis region for more detailed investigations (see Sect. 3.4).

Finally, the segment concept is applied to different start times t_i which link the k-th segment to an individual initialization time. For practical reasons, we count segments in relation to the initialization time of the reference t_{ini}. The start time of the

k-th segment is thus given by $t_k = t_{\mathrm{ini}} + k\ \Delta t_{\mathrm{reinit}}$ where $k = 0$ selects $t_0 = t_{\mathrm{ini}}$. If e.g. 24 hours of model spin-up are omitted from the reference data, the first meaningful segments can be found at $k = 1$ for $\Delta t_{\mathrm{reinit}} = 24\,\mathrm{h}$ and at $k = 2$ for $\Delta t_{\mathrm{reinit}} = 12\,\mathrm{h}$. However, it is up to the user to define which segments are the first meaningful segments in their simulation chain.

2.3 Workflow Schemes and Implementation Details

Section 2.1 provides an introductory description of the various stages incorporated within the workflow. Specifically, steps 1 through 3 are categorized under the pre-processing workflow, while steps 4 and 5 pertain to the production workflow. The automation of tasks using algorithms exhibits varying degrees of complexity. In the following, a detailed analysis is shown that details the implementation of the most complex algorithms used throughout the workflow.

Figure 3 provides a comprehensive flowchart of the grid generation algorithm. The starting point is the horizontal wind field at 10 m height from the base simulation that is converted to relative horizontal vorticity ζ and further coarse-grained with a running-average filter with a width of around 200 km. The coarse-grained vorticity ζ_c field allows to identify larger-scale cyclonic motions such as those characteristic for hurricanes. Detection and tracking of cyclonic features is done with the open-source python package *tobac* (Heikenfeld et al., 2019; Sokolowsky et al., 2024). A threshold criterion of $\zeta_c = 10\,\mathrm{day}^{-1}$ was used for identification, ensuring only the strongest motions were selected. Please note that a variety of other and more customized tools are available for identifying and tracking hurricanes, such as TempestExtremes (Ullrich et al., 2021) or the approaches by Kleppek et al. (2008) and Enz et al. (2023), which could have been applied here with equal justification. The preparation of the hurricane track data from base simulations is concluded by a matching algorithm. It utilizes location data of past hurricanes from HURDAT (Landsea and Franklin, 2013) to find the closest matching hurricane track in the base simulations (similar to Gutmann et al., 2018). This use of observational best-track data only provides additional insights for nudged simulations or simulations in a forecast mode, for free-running climate simulations a comparison to the observational track may not be physically meaningful.

The actual grid generation loop is entered with data from the best matching track, from a base grid (here, we take the grid from the finest reference nest) and from a configuration file that allows users to specify the along- and cross-track extend of the grid segments. The grid generation algorithm relies on DWD-ICON-Tools software (see Zängl et al., 2022, for grid-related aspects), facilitating the creation of flexible ICON grids through masks, crucial for our hurricane-centric workflow. For masking, a chosen base grid is read. Then, the track segment and its spatial extension by the across-track width, $\Delta s_{\mathrm{width}}$, are transferred to this grid to create a mask. The resulting mask is then used in the DWD-ICON-Tools grid generator to create the next refinement cutting out only the chosen grid segment. This builds the outermost nest of a flexible setup and serves as the basis for further refinements. If multiple nests are targeted, the track segment is transferred again, now to the just created outermost nest, and the lateral extension is performed again with a predefined, but smaller across-track width. This process is repeated several times until all nests for the respective refinement levels have been created. In Fig. 3, the shown loop determination criteria is $n = 3$ which means that a sequence of three nested grids with successively halved grid spacing are created. Finally, these nested grids resemble layers of an onion.

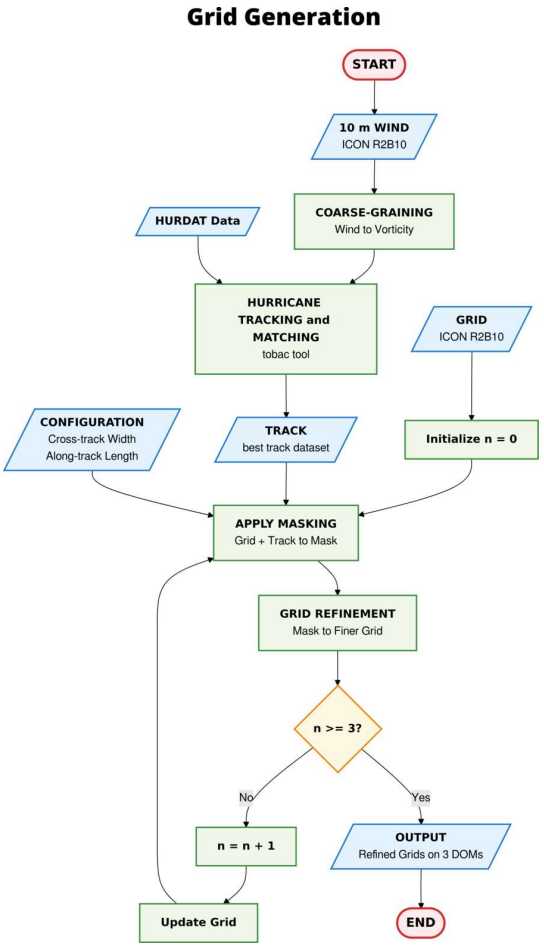

Figure 3. Detailed flowchart for the sequential grid generation process. Final start and end points are highlighted in red. Incoming and outgoing data are visualized as slanted blue boxes and methods that treat these data are shown as green rectangles. The index n counts the number of nested domains and thus determines the number of iterations in the grid generation loop. The term R2B10 indicates the triangular grid configuration of the ICON base run with a grid spacing of approximately 2.5 km.

Figures 4a and b split up the automated algorithm for IC generation into a pre-processing part and into a production part. For IC pre-processing, data from the base simulation, here noted by IC* as well as reference grid (source) and segment grid (target) are input. The horizontal regridding uses the tool cdo (Schulzweida, 2023) together with a conservative remapping to calculate regridded IC*. This dataset is input into a test ICON run that serves several purposes. With the test run (i) the integrity of the execution workflow is tested, (ii) performance-relevant aspects are investigated and resources for production runs are estimated and (iii) internal vertical interpolation is applied to consistently bring the regridded IC* from the reference levels to the possibly more refined target levels. After that point, the IC pre-processing workflow stops with the decision whether the considered segment is chosen to be the initial segment or not.

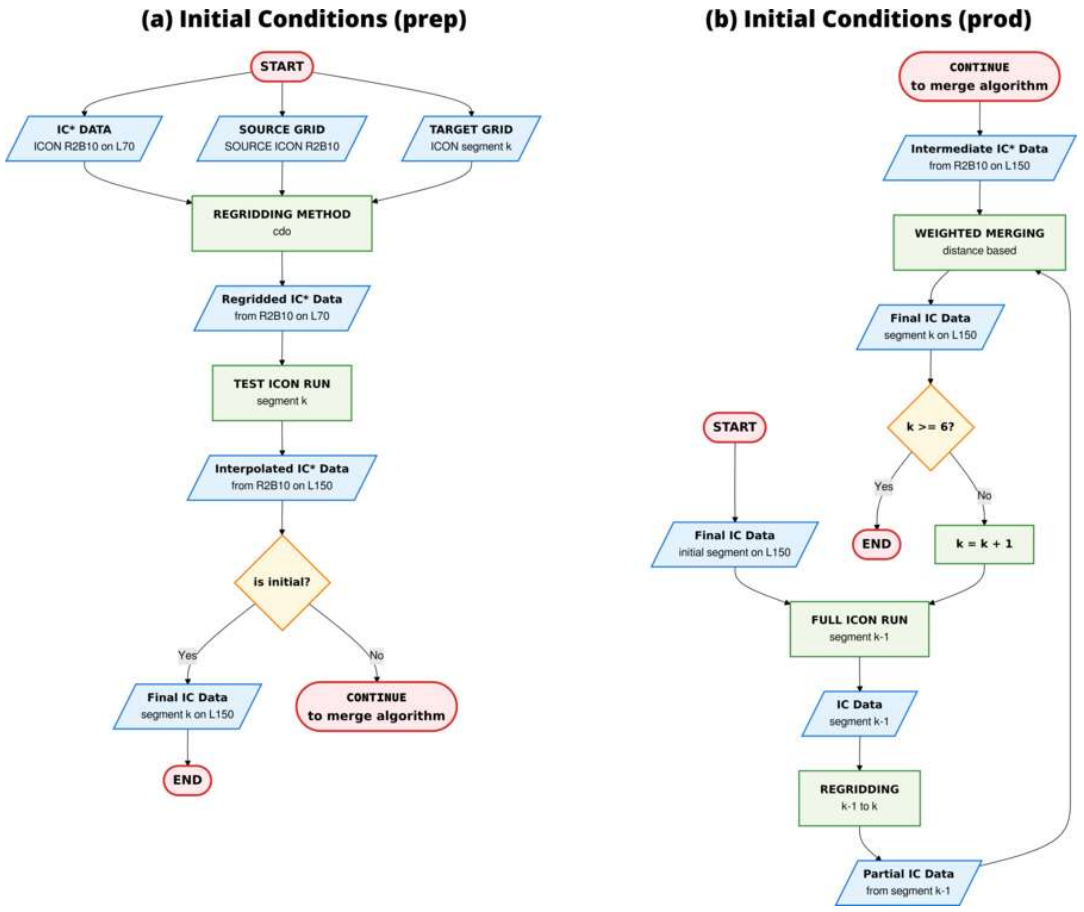

Figure 4. Detailed flowchart for (a) initial conditions pre-processing and (b) initial conditions production. Final and intermediate start and end points are highlighted in red. See Fig. 3 for meaning of the other elements. The index k counts the number of segments along the hurricane track. The loop terminates after a user-defined number of segments, e.g. here $k = 6$. The term L70 (L150) indicates the vertical grid configuration of the ICON base run with 70 (150) vertical levels.

In the IC production chain, regridded IC* are input into the full ICON run for the initial segment. At the end of the successful full ICON run, IC data for initializing the next segment are stored. Due to along-track shifts of the grid segments, these IC data only partially cover the region of the next segment. To compensate for this offset, a merging algorithm is applied that maps the partial IC data from segment $k-1$ to the grid for segment k in the overlap area. The remaining parts are filled with regridded IC* from the reference. The transition between the two IC data sources is smoothed by weights that decrease linearly with distance from the overlap boundary. Still, the mix of two IC datasets introduces a discontinuity in the initialization - a limitation that is discussed in depth in Sect. 3.5.

After grid and IC generation, the external parameters and boundary conditions are remapped onto the newly created grid segments. The former is directly interpolated from the original external data onto the respective grids yielding the highest

possible resolution of e.g. topography. The latter is exclusively obtained from the base run data and mapped using cdo on the respective lateral-boundary grids of the high-resolution simulations.

2.4 Portability and User Interfaces

The workflow toolkit is intended to run on super-computing platforms. It consists of a mix of python tools and bash scripts, the latter being submitted into a HPC scheduling system. The workflow was initially developed for the German Climate Computing Center (DKRZ) and its computing platform Levante and adjusted to its specific module environment and the Simple Linux Utility for Resource Management (SLURM) scheduler. The toolkit was then ported to the Jülich Supercomputing Center (JSC) Platform JUWELS, and its most important aspects were generalized. We therefore consider the software to be relatively platform-independent. In order to port the workflow toolkit, only the configurations of the corresponding module environment and the SLURM schedulers need to be properly set up for new platforms.

Users can readily configure the toolkit via a detailed configuration file, allowing them to set and modify parameters, as well as file and software dependencies, for different numerical experiments. We chose a TOML (Tom's Obvious, Minimal Language)-based configuration containing clearly structured elements and sections that can be easily edited with standard text editors. Additionally, we provide wrapper scripts to initiate the execution of entire workflows. It is possible, on the one hand, to start the pre-processing workflows of an entire series of consecutive grid segments. This results in the parallel execution of individual workflows that carefully handle their internal sequential dependencies. On the other hand, the production workflow wrappers can be executed. These trigger the execution of the ICON model in the selected grid segment with multiple restarts, if necessary, and merge the ICs before starting an ICON run for the next segment. All other user-relevant information can be found in the documentation of the latest software release (Senf, 2026c).

3 Application

3.1 Case Description

We now introduce a hurricane case chosen for demonstrating our flexible refinement workflow for hurricane-centric high resolution simulations. In 2020, Hurricane Paulette crossed the Atlantic and, with a total length of around 11,400 km, was one of the hurricanes with the longest track for that year. According to the National Hurricane Center's tropical cyclone report, Paulette formed as a tropical depression in the central Atlantic on September 7 and intensified into a tropical storm as it moved northwestward (National Hurricane Center, 2021). It reached category 1 hurricane status on September 13, intensified further into a category 2 hurricane in the early hours of September 14, and reached peak intensity of 90 kt (46 $\mathrm{m\,s^{-1}}$) and 965 hPa around 18 UTC on September 14. A snapshot of Paulette during its most intense phase is shown in Fig. 5a. Paulette finally transitioned into an extratropical cyclone at 12 UTC on September 16. During its first seven days of formation and main intensification period, Paulette's path was relatively straight, making it an ideal use case for our workflow methods.

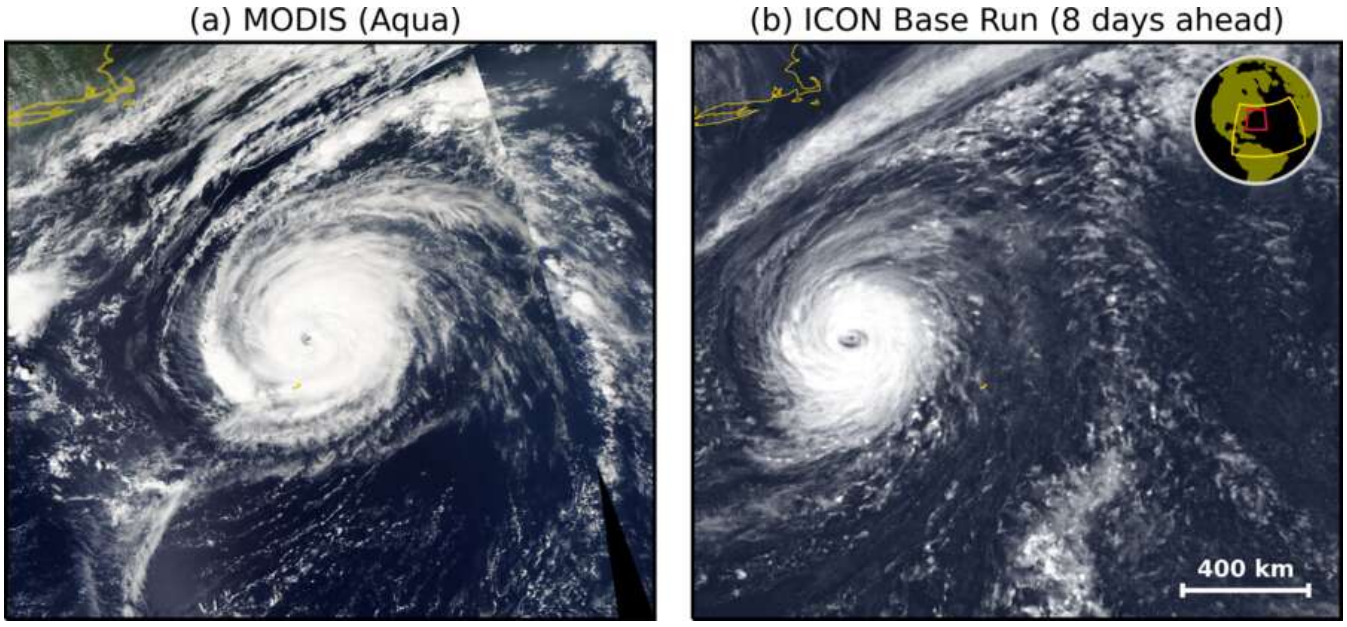

Figure 5. Top view of hurricane Paulette from (a) observations and (b) simulations. A regional cutout in the Atlantic ocean was defined reaching from -74.0 to -53.4 ° E and from 25.0 to 43.2 ° N. Subpanel (a) shows a true-colour image acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on the NOAA-AQUA satellite during an ascending orbit at overpass time on 14 September 2020, around 18:00 UTC. Subpanel (b) provides a simulation of the reflected solar radiation flux, normalized between 100 and 750 W m^{-2} from our ICON base run, which was initialized on 7 September 2020 at 0 UTC. The time, geographical region and projection have been matched between the observation and the simulation. The color scale and map in (b) have been adjusted to ensure visual consistency with (a). The length scale of 400 km is shown in the lower right corner of (b) and the inset in the upper right corner of (b) shows the location of the image cutout (red outline) and the inner base run domain (yellow outline) in the Atlantic ocean.

3.2 Base ICON Run

The ICON model solves hydrodynamic equations on a triangular grid (see Zängl et al., 2014; Dipankar et al., 2015), and can operate as a global or regional modeling system (Hohenegger et al., 2023; Müller et al., 2025). For our baseline run, we used the regional setup of the ICON model in the configuration with the numerical weather forecast (NWP) physics package (Zängl et al., 2014). The base simulations with an outer limited-area setup and grid spacings of 5 km forced by ERA5 data (Hersbach et al., 2020) were performed into which a second, inner domain with 2.5 km grid spacing was nested using one-way coupling.

For the outer nest, a spatial extent of 9900 $\times$ 6700 km^2 was chosen, which covers most of the tropical and subtropical Atlantic north of the equator. The outer nest was configured as R2B9 triangular ICON grid, corresponding to an equivalent grid spacing of 5 km. In this domain, convection processes and convective precipitation were parameterized following Bechtold et al. (2008) based on Tiedtke (1989). For the inner nest, the ICON R2B10 grid setup was chosen, corresponding to a grid spacing of approximately 2.5 km. The vertical resolution of the base run was set to 70 model levels in both nests, extending

from sea surface to a maximum height of 34 km. The inner nest is driven by the outer nest via a nudging zone close to the lateral boundary. However, due to one-way coupling, the inner nest does not provide feedback to the outer nest. In the inner nest, subgrid-scale convection was partially parameterized. The resolved model dynamics explicitly represent deep convective processes, while a parameterization approach still approximates shallow to mid-level convection. This setup has been shown to sufficiently reproduce realistic marine cloud structures and cloud radiative effects (Senf et al., 2020).

All other parameterizations from the NWP physics package were selected identically for both setups. Thus, all ICON simulations used the two-moment cloud microphysics scheme after Seifert and Beheng (2006). Subgrid-scale turbulent fluxes were calculated using a closure approach, as described by Baldauf et al. (2011), which includes the prognostic calculation of turbulent kinetic energy. The scheme ecRad was used for radiation calculations (Hogan and Bozzo, 2018). It performs radiative calculations in 14 solar and 16 terrestrial pre-defined spectral bands.

The simulations utilized in the following were initialized at 0 UTC on September 7 using the ERA5 data, a reanalysis dataset produced by European Centre for Medium-Range Weather Forecasts (ECWMF, Hersbach et al., 2020). This corresponds to a time when Paulette had already developed into a tropical depression. ERA5 data is available six-hourly as boundary data for the outer nest. For the base run, ICON version v2.6.6 is selected, which was the most recent ICON version at the time the base simulations were created. On top of this version, an update was developed that allows the sea surface temperature from ERA5 to be updated daily for all regional nests (Senf, 2026a). Therefore, SST fields are not kept constant at the initial SST values, but changed on a daily basis for our hurricane base simulations. Consequently, the actual effects of Paulette on the ocean surface are already included, which could indeed negatively impact the quality of simulating the hurricane development (Bender and Ginis, 2000). The output of the base run was configured so that data for initialization or re-initialization of higher-resolution runs is available every six hours. Furthermore, data to drive high-resolution runs via lateral boundary conditions is available at a resolution of one hour.

Figure 5b shows a visual impression of Hurricane Paulette simulated with our base run setup, at the time of the most intense hurricane period. The figure shows the reflected solar radiation at the top of the atmosphere for the inner R2B10 nest after an integration time of around eight days. For operational forecasts, Gao et al. (2023) reported a mean track error of approximately 300 km after 5 days of lead time. Here, a distance of approximately 400 km between our simulated hurricane center and the observation is achieved, which is a very positive result for a hindcast with around 8 days of lead time. Given the complexity and multiple interacting processes involved in hurricane formation, the reasonably good agreement was somewhat unexpected for the rather long forecast lead time. Although the variables being compared between the observation and simulation are not the same, making a quantitative comparison difficult, we still observe interesting differences in the morphology of the simulated hurricane compared to the observation. It appears that the spatial extent of the simulated anvil cloud associated with the hurricane is underestimated. Insufficient spatial resolution could be one reason for this discrepancy, among many others, and methods that lead to more spatially refined simulations, such as the workflow presented in this paper, would be an important step in further investigating this question of resolution dependence.

3.3 Exploration of Different Refinement Strategies

In the following, we will examine the refinement workflow presented in Sect. 2 in more detail for Hurricane Paulette as an example. As shown previously, the base run results reveal an intriguing hurricane development, motivating us to conduct further investigations with higher resolutions. This is exactly the starting point for the flexible refinement workflow presented in this paper. The base data is now entered into the flexible grid calculation alongside the calculated track of the simulated hurricane.

As a user, one can now choose from various important parameters. These parameters determine how many grid segments are created and their size. The first fundamental question concerns the spatial resolution. Starting with the resolution of the base run (R2B10 with 2.5 km), higher resolutions can be achieved through successive refinement. The following example uses setups with three nests, ranging from R2B11 with a grid spacing of 1.2 km to R2B13 with a grid spacing of approximately 300 m. The currently implemented workflow can handle fewer nests but would require additional software updates for using more than three nests. In principle, it is also possible to start initial refinement from grids finer than the base grid for which minor software updates would be needed.

Figure 6 shows various implementations of grid segmentations for Paulette. These implementations vary either the length of a segment along the hurricane track (white line) using the *"reinit"* parameter $\Delta t_{\mathrm{reinit}}$ or the across-track width $\Delta s_{\mathrm{width}}$. The different grid segments are visualized by different colors. When $\Delta t_{\mathrm{reinit}} = 12\,\mathrm{h}$ is selected (Fig. 6a and c), the grid segments appear shorter and more compact. However, more grid segments must be calculated sequentially to cover the same integration period. For the 12-hour re-initialization setup, it makes sense to start with segment 2 on September 8, 2020, at 0 UTC to avoid spin-up of the base run. If the integration ends after the 12th segment, as shown, the hurricane's development can be tracked for five days and 12 hours. The 24-hour re-initialization setup requires fewer segments. However, each segment covers a larger region and requires more computing power (see Figs. 6b and 6d). Each of the 24-hour setups shown starts with segment 1, and sequential integration up to segment 6 covers a period of six days.

Another important parameter determines the width of the segments across the track, $\Delta s_{\mathrm{width}}$. Ideally, the width would be chosen so that the hurricane development would not be significantly influenced by the lateral domain boundaries. In practice, computational constraints are significant, so responsible scientists must select the smallest feasible domain size. Given these two requirements, it is difficult to provide definite advice and determine in advance what the optimum width would be. Two possible configurations are provided in Fig. 6. First, the innermost high-resolution nest at R2B13 is selected so that the across-track width, measured as distance between the track center and the domain edge, is 100 km. Second, a wider setup was tested in which the inner nest was extended to 200 km. Table 1 shows the extent of the two outer high-resolution nests in detail.

High-resolution

ICON simulations were performed in the respective grid segments using the latest available version of ICON (v2025-04-2) (ICON partnership, 2025) at the time of the study. We chose 150 vertical levels (analogous to Heinze et al., 2017; Stevens et al., 2020), extending to an altitude of 34 km. Cloud microphysics and radiation were set up similarly to the base run. However, subgrid-scale turbulence was configured with a three-dimensional Smagorinsky-type mixing scheme and convection parame-

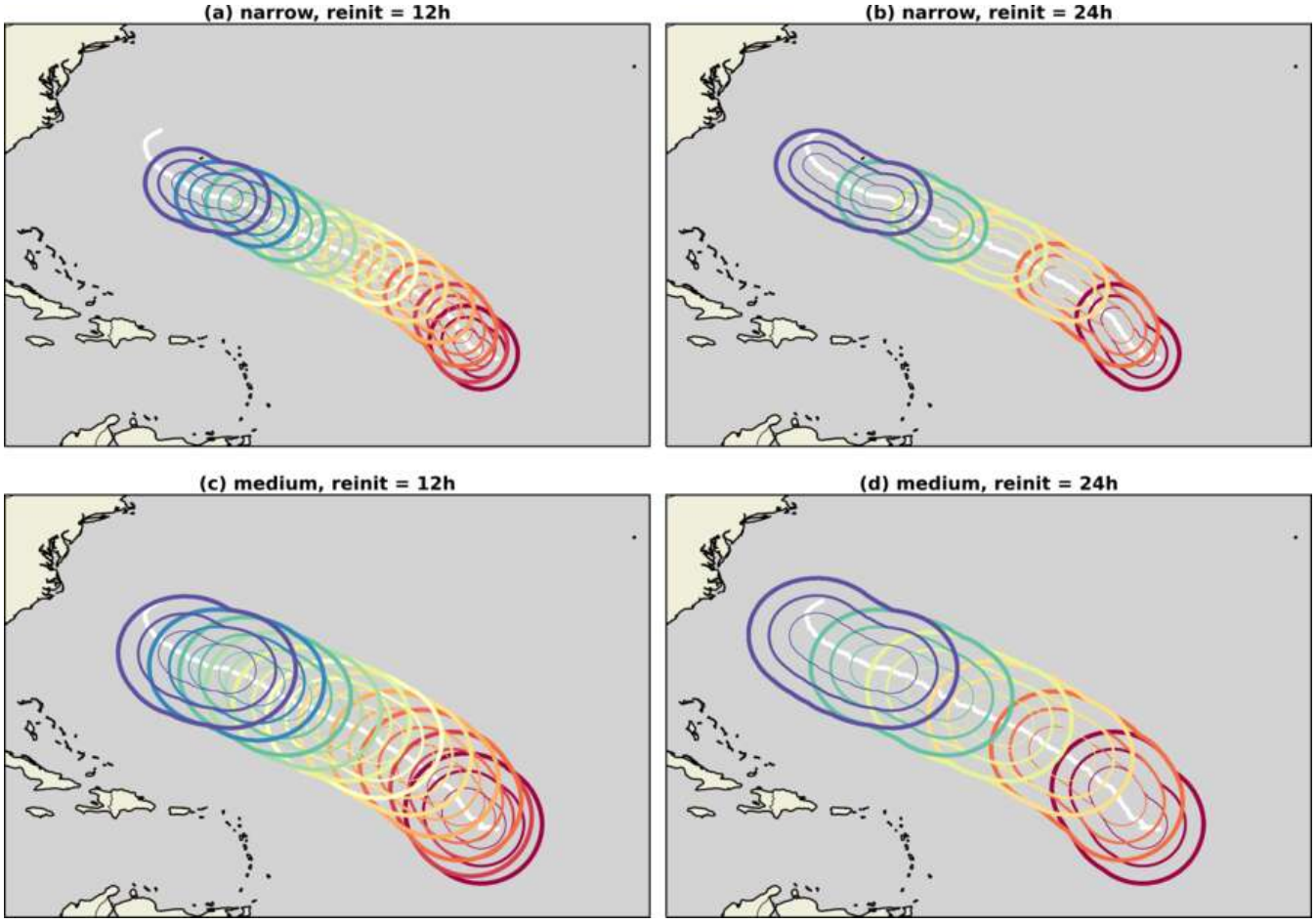

Figure 6. Grid segments prepared for hurricane Paulette (2020) with differently chosen re-initialization times and across-track widths. The narrow setup in (a) and (b) is compared to the medium setup in (c) and (d). Re-initialization at 12-hour intervals (a, c) leads to twice as many segments as re-initialization at 24-hour intervals (see Tab. 1). The track of the hurricane Paulette in the base simulation is plotted as a white solid line. Each segment consists of three nested domains (solid lines of decreasing thickness) colored by the initialization time, ranging from red on September 8 to blue on September 14.

terization was completely deactivated (following Stevens et al., 2020). To enable a consistent re-initialization of complete cloud microphysics, the ICON source code was modified to incorporate vertical interpolation of all microphysical moments (Senf, 2026d). This means that graupel mixing ratios and ice crystal number concentrations, for example, are available as fields when the high-resolution simulations begin and when they are re-initialized. Consequently, all cloud variables are available without discontinuities and abrupt changes throughout the entire simulation period. These initial fields are made available for all nests simultaneously, enabling a seamless continuation of the simulations when switching from one segment to the next.

Table 1. Comparison of grid and compute requirement parameters across re-initialization configurations and domains (DOM) listed in the different columns. The computational units (cu) refer to the product of the number of time steps and the number of horizontal grid cells and is a measure of the size of a limited area experiment.

property	reinit	narrow			medium			large	x-large
		DOM01	DOM02	DOM03	DOM01	DOM02	DOM03	DOM01	DOM01
grid	12h /24h	R2B11	R2B12	R2B13	R2B11	R2B12	R2B13	R2B11	R2B11
eq. resolution / m	12h /24h	1200	600	300	1200	600	300	1200	1200
width / km	12h /24h	300	200	100	500	350	200	750	1000
ave. # cells / 10^6	12h	0.35	0.76	1.2	0.79	1.8	3.1	-	-
	24h	0.46	1.1	1.8	0.96	2.3	4.2	1.8	3
cu / 10^9	12h	25	110	350	57	250	880	-	-
	24h	33	150	510	69	330	1200	130	210

Figure 7 shows examples of the results of two high-resolution ICON configurations with different grid segmentation. The narrow-size setup is compared with the medium-size setup, both of which have a 12-hour re-initialization period. The position of the simulated hurricane pressure minimum in the high-resolution run remains close to the reference track of the base run. There are no significant differences in minimum position for different simulation resolutions, but same grid configuration. However, as expected, the track of the wider setup deviates more strongly from the reference track. This is plausible because the position of the hurricane center in the wider setup is not as strongly constrained by the forcing at the lateral boundaries. The intensity of the simulated hurricane, as shown by the development of the pressure minimum in Fig. 7b and the maximum 10-meter wind speed in Fig. 7c, clearly depends on the resolution. This is evident from the greater spread of the curves (esp. in Fig. 7b), which is as large as that induced by different grid configurations. Better resolved convective activity generates particularly high gustiness of the 10 m wind, which explains the fluctuations on the very short time scales in Fig. 7c. However, a more detailed investigation of the resolution effects and their causes and consequences must be postponed to a follow-up study.

3.4 Prototyping Lagrangian Analysis

The grid remains static during the simulation of a grid segment, and the output is only available for statically defined grids. Output variables are written directly to the ICON grid. Additionally, variables can be internally interpolated in ICON to a regular longitude-latitude grid and then output on this grid. This allows variables to be stored on slightly coarser grids at a higher temporal frequency. Our workflow toolkit includes an option to automatically calculate suitable longitude-latitude grids for the corresponding flexible grid segments. This creates a suitable longitude-latitude grid for each grid segment on which standard analyses can easily be performed. Figure 8 shows an example of such an output, displaying 2-meter temperature,

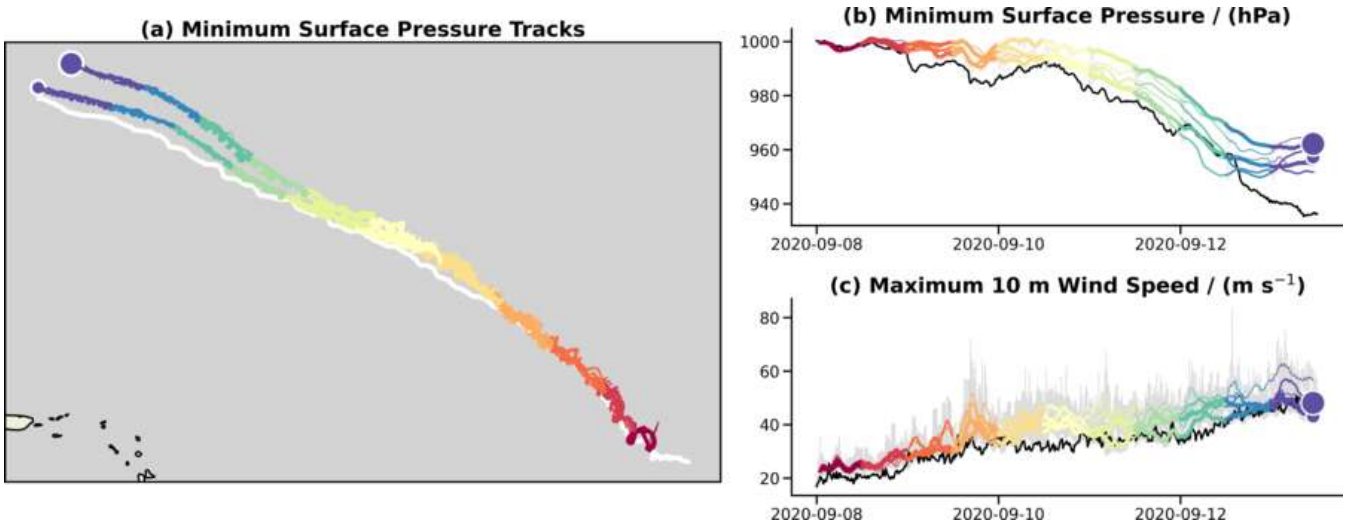

Figure 7. Simulated hurricane characteristics for grid configurations with differently chosen across-track widths. (a) The minimum pressure positions are plotted on a map, with colors representing the respective grid segments. The re-initialization time between successive segments is set to 12 hours. The narrow-size setup (small, solid circle at the end of the track) is compared to the medium-size setup (larger, solid circle at the end of the track). Different resolutions are plotted with different line thicknesses, ranging from thick for R2B11 to thin for R2B13, and they partially overlap. Minimum pressure positions from the base run are plotted in white. The minimum pressure and maximum 10-m wind speed are presented using similar color and line encoding as in (a). A running-average filter is applied to the curves in (b) and (c) to reduce noise. The original data is shown in light gray shading in the background. Data from the base run is shown in black.

10-meter wind speed, reflected solar radiation flux at the top of the atmosphere (TOA), and total column water vapor for a given point in time. The sequence shows three nested grids for each variable, with grid spacings of 1200, 600, and 300 m. The utilization of these representations facilitates the comparison of the effects of varying resolutions on the fine structure of the simulated hurricane. However, Fig. 8 also demonstrates that important parts of the hurricane may no longer be contained within the chosen domains, especially for the finest nest of 300 m grid spacing and rather long integration times of nearly five days. This illustrates the trade-offs responsible scientists must navigate between accuracy, resource requirements, and efficiency.

The visualizations on the static grid above have the typical characteristics of an Eulerian analysis, in which the reference points do not change over time. However, our hurricane-centric simulations make it possible to change the perspective by defining only the circle that moves over the reference track as the analysis area (see segment concept in Sect. 2.2). This transformation into a Lagrangian reference system results in a smooth and consistent composition of the hurricane development despite the inherent discontinuity because of the separate grid segments. For example, the evolution of ice water path (IWP) is shown in Fig. 9. Most of the ice produced is located northeast of the hurricane center. The Lagrangian analysis region is highlighted in color. The remaining parts of the segment, which are not used here, are shown in light colors. Interestingly, the Lagrangian approach results in a continuous and smooth path of the IWP. It gives the impression that the grid segments are

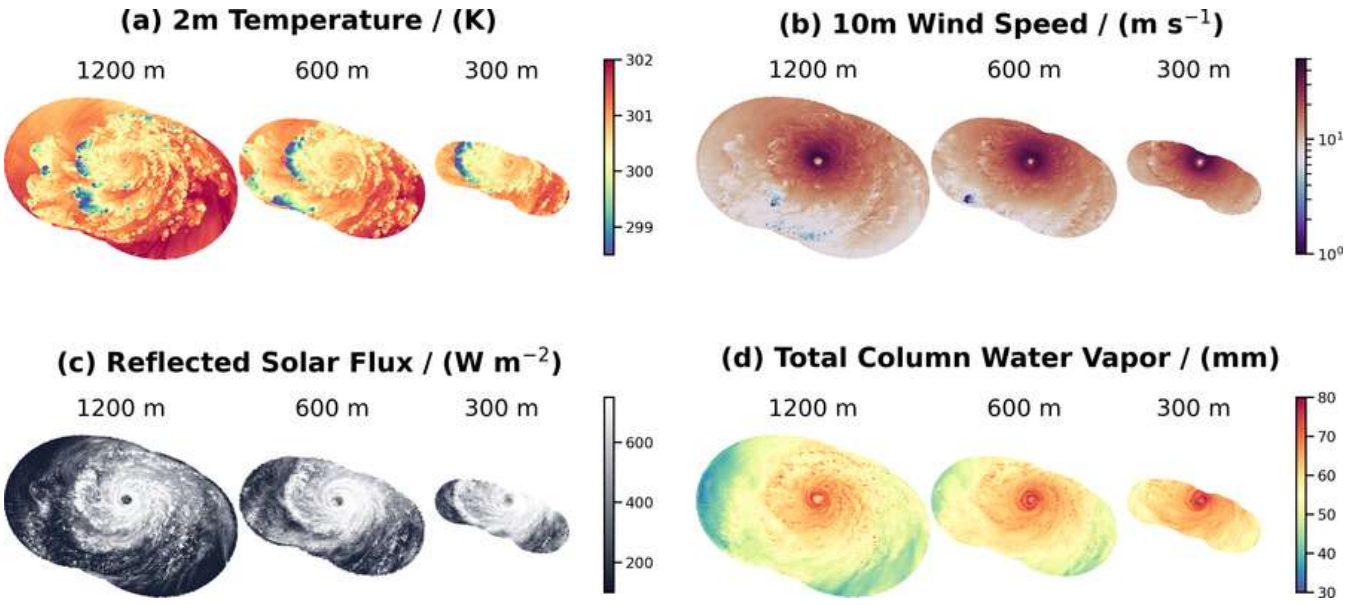

Figure 8. An overview of various variables, shown in the respective nests for a specific grid segment. The grid configuration was chosen so that the re-initialization time is 12 hours and the nests have an across-track width of 300, 200, and 100 km, respectively (narrow-size setup). The equivalent grid spacing is indicated above the respective nest. Variables shown are (a) 2 m temperature, (b) 10 m wind speed, (c) reflected solar radiation flux at TOA and (d) total column water vapor for segment 11 at 18 UTC on 12 September 2020.

sliding across the Lagrangian analysis area. Thus, data from different segments can be assembled to create a hurricane-centric analysis that consistently covers several days of hurricane development on a hectometer scale.

3.5 Performance and Shortcomings

The following discussion will cover the advantages and disadvantages of the method described so far. One significant advantage is that the target areas of the simulations are relatively small compared to the total region affected by the hurricane. This means that an acceptable number of calculation steps is being performed. Table 2 summarizes computational units (cu) as a measure of computational effort. Here, one cu refers to 1,000 time steps of fast physics (dynamical core is sub-stepped by a factor of 5), which are used to integrate one million horizontal grid cells forward. For the narrow-size case, cu increases from 480 to 700 when transitioning from the 12h-reinit setup to the 24h-reinit setup, which corresponds to an increase of approximately 45%. For the medium-size case, 1,190 and 1,610 cu are required respectively, representing a factor of between 2.3 and 2.4 more resources than the narrow-size case, showing an increase of 35% between 12h-reinit and 24h-reinit just because of the larger domain size of the 24h-reinit segments. Therefore, the 12h-reinit setup is significantly more favorable in terms of resource requirements, and doubling the width of the setup also leads to approximately a doubling of resource requirements. The costs for expanding the setup to large and extra-large sizes are also shown. However, when comparing the numbers, it should be

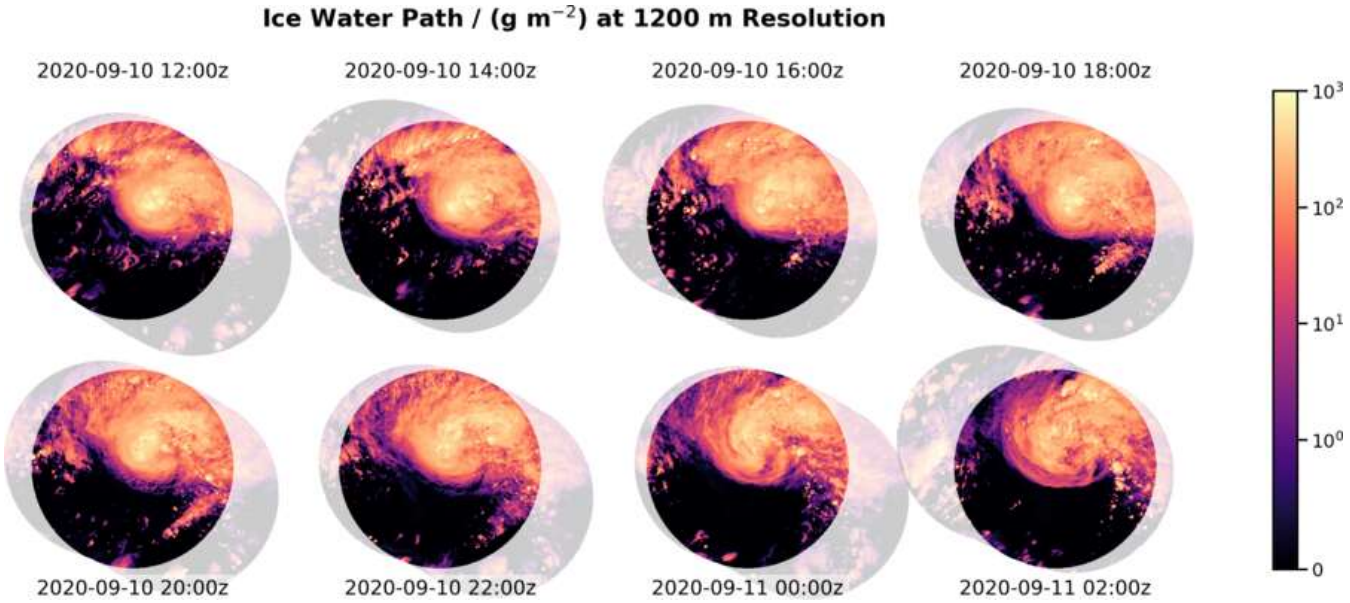

Figure 9. Ice water path is shown as a sequence in a Lagrangian framework. The boldly shaded areas highlight the circular Lagrangian analysis region that follows the track of the reference hurricane. The remaining parts of the respective grid segments are indicated by very light colors. Two-hour steps are made in the temporal sequence. Transitions between segments are visible between the first and second images, as well as between the second-to-last and last images, appearing as shifts in the domain boundaries. Within the Lagrangian analysis region, however, the ice water path evolves smoothly and continuously. Here, the medium-size grid configuration was chosen so that the shown nest (DOM01, 1.2 km grid spacing) has 500 km across-track width with re-initialization time of 12 hours.

noted that only the outermost nest with a grid spacing of 1.2 km was considered for the larger setups and that the complete setup consisting of three respectively refined nests requires 20 to 25 times more resources.

To assess how the savings of our method compare to classic regional, limited-area applications, two reference configurations are introduced in Tab. 2. The first is a tube-like setup that extends over the full hurricane track over the entire simulation period without splitting it into shorter segments. Such a tube shares the same cross-track geometry and nesting structure as our hurricane-centric segments, but is integrated as a single contiguous domain without re-initialization. The resource requirements for the tube, listed under the "6days" re-initialization interval, amount to approximately 2,640 and 5,440 cu for the narrow and medium three-nest configurations and 320 and 470 cu for the single-domain large and x-large setups, respectively. The second reference, labeled "classic" in Tab. 2, represents a conventional fixed limited-area domain and is constructed as the smallest regular longitude-latitude bounding box that fully encloses the tube geometry for the respective cross-track width. For the single-domain case at 1.2 km grid spacing, the classic limited-area approach amounts to approximately 800 cu; a three-nest configuration requires around 20,000 cu. Compared to this classic reference, the tube configurations already yield notable speedup factors of 7.6, 3.7, 2.5, and 1.7 for the narrow, medium, large, and x-large setups, respectively. Dividing the tube into short, frequently re-initialized segments adds a further factor of approximately 2 to 6, depending on the chosen across-

track width and re-initialization interval. Together, the segment configurations with frequent re-initialization achieve the largest speedups: factors of up to 42 and 17 for the narrow and medium three-nest setups with 12-hour re-initialization intervals, and up to 6.2 and 3.8 for the single-domain large and x-large setups with 24-hour re-initialization intervals (see Tab. 2). Considering the risk of model errors or suboptimal settings requiring reruns of high-resolution simulations, the use of our flexible workflow appears well justified.

Table 2. Comparison of compute requirement and compute cost parameters across re-initialization configurations and domain sizes. The reinit row "6days" indicates a nested tube configurations with effectively no re-initialisations. The term "classic" refers to a potential, fixed domain configuration such that a corresponding tube configuration could be accommodated. Single domain configurations are marked by $*$, all other domain configurations consist of three online coupled nests. The computational units (cu) refer to the product of physics timesteps and number of horizontal grid cells. The number of nodes and node hours refers to DKRZ Levante nodes with 128 cores for the 12h-reinit setup and to JSC JUWELS nodes with 48 cores for the 24h-reinit setup.

property	reinit	narrow	medium	large	x-large	classic
total cu / 10^9	12h	480	1 190	-	-	-
	24h	700	1 610	130*	210*	-
	6days	2 640	5 440	320*	470*	-
	-	-	-	-	-	800*, 20 000
potential speedup	-	**7.6 - 42**	**3.7 - 17**	**2.5 - 6.2***	**1.7 - 3.8***	1
total node hours (# nodes)	12h	2 900 (64)	6 300 (64)	-	-	-
	24h	13 600 (192)	32 800 (288)	3 100* (64)	4 400* (64)	-
total core hours / 10^6	12h	0.37	0.81	-	-	-
	24h	0.65	1.57	0.15*	0.21*	-

In addition to the hypothetical requirements, Tab. 2 also shows the actual costs of executing the respective setups. The 12h-reinit setup was calculated on the DKRZ Levante platform using an ICON build created with the Intel compiler without the thread-based parallelization via OpenMP. Production runs were performed on 64 nodes, generating costs of 2,900 and 6,300 node-hours for the narrow- and medium-size cases, respectively. This increase in costs by a factor of 2.2 roughly corresponds to the calculated increase in demand. The 24-hour reinit setup was calculated on the JSC JUWELS platform, which has fewer computing units than Levante, with 48 cores per node compared to 128. An ICON build was created and used on JUWELS with GCC, as well as without OpenMP. The narrow-size setup was calculated on JUWELS with 192 nodes, generating costs of 13,600 node-hours. The medium-size setup was calculated with 288 nodes and generated costs of 32,800 node-hours. Comparing the core hours values between Levante and JUWELS reveals that execution on JUWELS is 20–40 % less efficient. This difference is probably due to the different compilers and their respective optimization settings, rather than hardware differences.

The immense increases in efficiency of our hurricane-centric setup however come at a price. Due to the limited size of the regional setups, the hurricane under consideration cannot develop completely freely and always remains influenced by the forcing at the lateral boundaries, which is stronger for a more narrow setup. In addition, the re-initialization procedure during the transition between segments introduces further challenges. The mix of ICs detailed in Section 2.3, which consist of different data sources inside and outside the overlap region of two subsequent segments, creates a discontinuity. This is smoothed out in the transition using a distance-weighted average. Despite this, the merged ICs can excite spurious sound and gravity waves, and re-initialization effects occur at the beginning of a simulation in a new segment. The following section evaluates the impact of these effects on simulation quality.

3.6 Analysis of Re-initialization Effects

To investigate the re-initialization behavior in a controlled manner, we conducted sensitivity experiments following the same grid configurations as listed in Table 1, but with the runtime of each segment extended by three hours. This creates a three-hour temporal overlap between consecutive segments, enabling a direct comparison of the atmospheric state from two successive segments at the same point in time and thus providing a pathway for directly quantifying re-initialization effects. To keep the analysis tractable, all sensitivity experiments were performed exclusively for the outermost nest (DOM01) at 1.2 km grid spacing. In addition, two parameter settings are compared throughout: the production setting with `init_latbc_from_fg=.FALSE.` (labeled "LBC flag OFF") and an updated setting with `init_latbc_from_fg=.TRUE.` (labeled "LBC flag ON"). The latter has been recommended by the ICON community since early 2026 to reduce model biases arising from mismatches between the initial lateral boundary conditions and the initial model state (see ICON Documentation, 2026).

In the following, we analyze the time series of thermodynamic and water-related atmospheric parameters for the first three hours in Fig. 10. There, the median values in the circular Lagrangian analysis regions (see Sect. 3.4) of selected surface and cloud variables are presented. For each variable, the instantaneous difference between segment k and the corresponding values from segment $k-1$ extended by 3 h is shown, allowing a direct quantification of re-initialization-induced deviations between consecutive segments. The figure was created for the medium-size, 12h-reinit DOM01 setup with 1.2 km grid spacing (see Tab. 1) . Light gray curves show the individual segment differences, and the thick solid (dashed) black curve gives their mean for the LBC flag OFF (ON) setting. The surface pressure in Fig. 10a shows a negative anomaly with peak values around -1.2 hPa after about 30 min of integration time for the LBC flag OFF setting. An opposing positive but slightly weaker anomaly of around 0.9 hPa is visible for the LBC flag ON setting, with the peak shifted to around 70 min. Both differences appear to have converged after approximately 2 hours, by which point they have decayed to near-zero values. The 2-m temperature curve in Fig. 10b also shows a systematic negative peak of around -0.06 K at 30 min for LBC flag OFF, and a positive peak of around 0.05 K at 75 min for LBC flag ON. Temperature differences likewise vanish after approximately 2 hours. The curves for 2-m specific humidity show positive differences for both LBC flag settings, with peak values of 0.07 $\mathrm{g\,kg^{-1}}$ at 45 min for LBC flag OFF and 0.05 $\mathrm{g\,kg^{-1}}$ at 120 min for LBC flag ON. The specific humidity difference for the LBC flag ON setting converges more slowly than for LBC flag OFF, with near-zero values not yet reached by 3 hours. The individual curves for total column water vapor (TCWV), liquid and ice water path (LWP, IWP) in Figs. 10d-f show considerable variability over

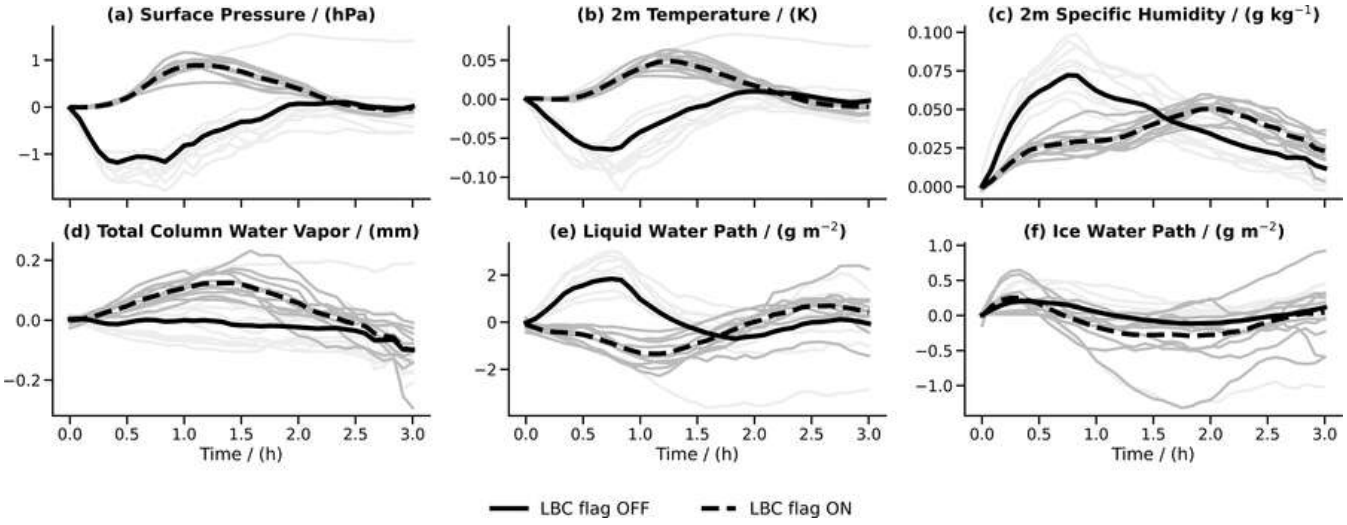

Figure 10. Temporal evolution of median values of surface and integrated cloud variables evaluated in the circular Lagrangian analysis region. The curves show the instantaneous difference between values from segment k and the corresponding values from segment $k-1$ extended by 3 hours, for (a) surface pressure, (b) 2 m temperature, (c) 2 m specific humidity, (d) total column water vapor, (e) liquid water path, and (f) ice water path. Results are shown for the medium setup with 12h-reinit. Gray curves show the evolution for individual segments, while the thick black curve shows the mean across all segments. Two parameter settings are compared: the experiment with the LBC flag deactivated (`init_latbc_from_fg=.FALSE.`), thin light gray and solid black lines) and the experiment with the LBC flag activated (`init_latbc_from_fg=.TRUE.`), thin gray and dashed black lines).

the three-hour window. The LBC flag ON setting tends to produce slightly larger mean peak values for TCWV and IWP, and smaller ones for LWP, and its peak consistently occurs later than for LBC flag OFF. Overall, the LBC flag ON setting tends to produce smaller peak differences for surface pressure and near-surface temperature, suggesting it may be the preferable choice for future applications, though its re-initialization effects appear to persist somewhat longer legitimating our choice of the LBC flag OFF setting for the production runs in this study.

Figure 11 provides a comprehensive statistical comparison of re-initialization effects across different configurations. As in Fig. 10, the analysis evaluates percentiles from segment k, here taken 30 minutes after initialization, and segment $k-1$ extended correspondingly. Nine percentiles (10–90% in 10% steps) of each field are computed across the Lagrangian analysis region separately, and their differences are taken. These per-segment differences are visualized as stacked histograms, with color coding by segment to enable direct comparison between setups.

The statistical results reveal distinct patterns for different atmospheric variables. Re-initialization effects are particularly visible when the values in the distributions significantly deviate from zero. For surface pressure, this pattern is clearly evident in Figs. 11a and 11e for the narrow setup. Surface pressure values scatter between -1 and -2 hPa. Comparing the effects of re-initialization time shows that longer intervals lead to slightly larger re-initialization effects in surface pressure (comparing Figs. 11a,e for narrow setup and Figs. 11i,m for medium setup). Domain size also influences these effects: larger domains

(Figs. 11q,u) show considerably less deviation from zero at 30 minutes, indicating that systematic re-initialization effects are less pronounced for larger domains. For temperature, similar effects are clearly visible. The value appear to be distributed to slightly lower magnitudes for the 12h-reinit setup than for the 24h-reinit setup. For large and x-large setups, temperature anomalies scatter around zero for all segments. Cloud variables such as LWP and IWP show significant spread in the narrow setup (Figs. 11c,d,g,h). As domain size increases, this variability decreases, likely due to more robust spatial statistics. The considerable variability in cloud variables masks potential re-initialization effects. Overall, while our hurricane-centric workflow introduces uncertainties through re-initialization effects, these remain sufficiently small and are outweighed by the substantial efficiency gains achieved.

4 Conclusions

This study presents a flexible workflow for the atmospheric model ICON that enables efficient high-resolution simulations in a hurricane-centric reference framework through a segment-based grid and pre-processing approach. The methodology demonstrates the general feasibility of conducting hectometer-resolution simulations of large, organized atmospheric phenomena like hurricanes in the tropics. Instead of investing in fixed and spatially extended regional domains throughout the entire simulation period, our flexible method follows the motion of a hurricane and generates tailored grid segments.

The technical implementation successfully automates key components of the workflow, including hurricane tracking, flexible grid generation, and initial condition preparation and merging across consecutive segments. The segment concept, which divides hurricane tracks into overlapping temporal windows of 12-24 hours, provides a practical framework for efficient utilization of computational resources. The workflow toolkit demonstrates portability across different HPC platforms, having been successfully deployed on both DKRZ Levante and JSC JUWELS systems with appropriate platform-specific configurations.

The application of the workflow is demonstrated for Hurricane Paulette (2020). High-resolution simulations with grid spacings down to 300 m were performed using different segment configurations. Variations in re-initialization intervals of 12h and 24h are explored that determine the respective along-track lengths of the segments. Additionally, different overall domain sizes are investigated by varying the across-track widths with tested configurations being rather narrow, medium-sized, large and extra large. The results indicate that the hurricane track remains consistent with the base run and mainly depend on the across-track size of the chosen configuration, while intensity metrics such as minimum pressure and maximum wind speed exhibit sensitivity to resolution in our multi-nested setups. A Lagrangian framework enables a hurricane-centric view, facilitating continuous analysis of variables like liquid and ice water path across the segmented simulations.

The increase in computational efficiency is substantial compared to traditional fixed-domain approaches. A track-following tube geometry alone reduces resource requirements by factors of 1.7–7.6 relative to an appropriately chosen longitude-latitude bounding box around the hurricane. Segmenting the tube into short, frequently re-initialized intervals adds a further factor of 2–6, yielding total speedups of up to 42 for the narrowest three-nest configuration. The setup with 12-hour re-initialization intervals proves to be more efficient than 24-hour intervals, requiring 35-45 % fewer computational resources while maintaining comparable simulation quality. Domain size scaling shows predictable resource scaling, with the twice as large medium-width

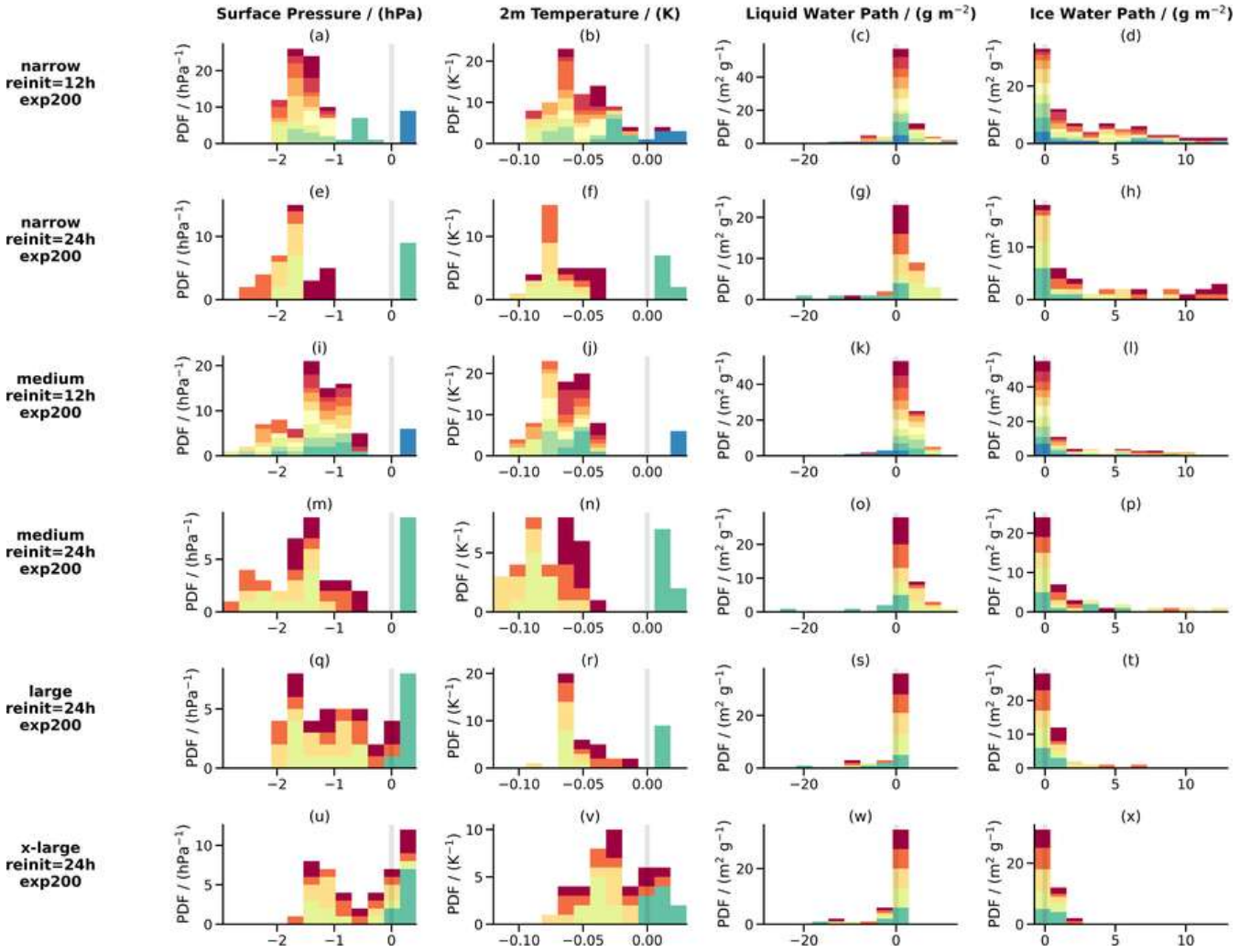

Figure 11. Statistical summary of potential re-initialization effects plotted as colored bars stacked on top of each other for each segment. The color coding is the same as in Fig. 6. At 30 minutes after initialization of segment k, nine percentiles (10–90% in 10% steps) of each field are computed across the Lagrangian analysis region separately for segment k and segment $k-1$ extended by 3 h, and their pairwise differences are taken. These per-segment differences are statistically summarized as probability density functions. The six rows show the respective setups, with the narrow and medium setups shown for re-initialization periods of 12 h and 24 h, respectively. The columns show surface pressure, 2 m temperature, liquid water path, and ice water path. The re-initialization analyses are only shown for the outer nest with a grid spacing of approximately 1 km.

configurations requiring approximately 2.3 times the resources of narrow configurations. However, the approach has notable limitations when the across-track width is set too narrowly, the domain may fail to encompass the entire phenomenon, a shortcoming that becomes especially evident for our narrow-size configurations. The analysis of re-initialization effects reveals

systematic but manageable impacts during segment transitions, with surface pressure showing negative anomalies on the order of 1 hPa that decay within 2 hours after re-initialization. The re-initialization effects are more pronounced in narrower domain configurations but remain sufficiently small to justify the substantial computational efficiency gains achieved by the hurricane-centric approach.

Several promising directions for methodological enhancement emerge from this work. On the technical side, extending the workflow to support more than three nesting levels would provide greater refinement flexibility and enable even finer resolution simulations. Similarly, offering greater flexibility in selecting the grid from which the refinement cycle is initiated (currently the grid of the base run) would allow for more diverse and independent nesting hierarchies. A further interesting area for development lies in improving the re-initialization procedure itself. Currently, the distance-weighted blending approach, while effective, could potentially be enhanced by integrating techniques from data assimilation, such as the Incremental Analysis Update (IAU) method available in the numerical weather prediction configuration of the ICON model. This approach might have the potential to substantially reduce spurious artifacts arising from the merging of initial conditions between consecutive segments. Additionally, future work could also investigate how the method performs over land, where re-initialization of subgrid-scale heterogeneities in land surface properties can be challenging and may introduce additional difficulties. Importantly, all such technical improvements must maintain spatial and structural consistency of simulated core variables across segment transitions such that Lagrangian analyses remain continuous and meaningful.

Beyond technical refinements, the computational efficiency gains achieved here enable new scientific applications. The method now makes high-resolution hectometer-scale simulations of additional hurricanes computationally feasible, opening opportunities for detailed dynamical analysis and process studies of a broader range of tropical storms. The reduced computational cost further enables simulation ensembles to be conducted in scientific studies, providing higher confidence in identifying systematic effects. Particularly noteworthy is the potential for perturbed parameter ensembles that would deepen our understanding of how model parameterizations influence the representation of convective organization, eyewall structure, and other small-scale phenomena in tropical cyclones. These applications demonstrate how advances in efficient simulation frameworks can open new avenues for investigating the cloud physics and dynamics of extreme atmospheric phenomena.

Code and data availability. The current version of the hurricane-centric workflow toolkit is available under https://doi.org/10.5281/zenodo.18271898 (Senf, 2026c). It depends on cdo (see Schulzweida, 2023), the DWD ICON tools available at https://gitlab.dkrz.de/dwd-sw/dwd_icon_tools (last accessed 27 January 2026, restricted access) and the extpar utility available at https://github.com/C2SM/extpar (last accessed 27 January 2026, toolkit tested with version v5.13). ICON open source release v2025.04 is published at https://doi.org/10.35089/wdcc/iconrelease2025.04 and a patch to enable consistent re-initialization of complete cloud microphysics is available under https://doi.org/10.5281/zenodo.18387047 (Senf, 2026d). ICON model v2.6.6 has restricted access and a patch to enable variable SSTs in a multiple nest setup is available under https://doi.org/10.5281/zenodo.18385022 (Senf, 2026a). Initial, boundary and grid data from the base run used to run the hurricane-centric workflow are available under https://www.wdc-climate.de/ui/entry?acronym=DKRZ_LTA_1376_dsg0001 (Senf, 2025). Plots and analysis were created mainly using Jupyter notebooks available under https://doi.org/10.5281/zenodo.19002039 (Senf, 2026b).

Author contributions. FS lead the development of the workflow and the drafting of the manuscript. RC contributed to the writing of the manuscript and the interpretation of the results.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The development was supported by the IFCES2 project funded by BMFTR (formerly known as BMBF - Federal Ministry of Education and Research) under grant number 16ME0689K as part of the SCALEXA funding initiative, which in turn was funded by the European Union - NextGenerationEU. Fabian Senf acknowledges funding from the CleanCloud project (grant: 101137639). We would like to thank Manoel Römmer for his comments and help with the workflow development. We are particularly grateful to Bernd Heinold, Ina Tegen, Jason Müller, Jan Kretzschmar, Nadja Omanovic, and two anonymous reviewers whose valuable comments contributed significantly to improving the manuscript. We also thank the ICON consortium for their continuous support and engagement in pushing ICON development to the next level. We gratefully acknowledge the computing time and storage resources granted on the supercomputer JUWELS at Jülich Supercomputing Centre (JSC) under project IFCES2-SCALEXA and on the supercomputer Levante at the German Climate Computing Centre (DKRZ) under project bb1376.

We would like to mention the use of DeepL to improve the grammatical and language quality of the manuscript and the use of github copilot for AI-assisted coding and debugging of the workflow toolkit.

References

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities, Mon. Wea. Rev., 139, 3887–3905, https://doi.org/10.1175/mwr-d-10-05013.1, 2011.

Bechtold, P., Köhler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M. J., Vitart, F., and Balsamo, G.: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales, Q. J. R. Meteorolog. Soc., 134, 1337–1351, https://doi.org/10.1002/qj.289, 2008.

Bender, M. A. and Ginis, I.: Real-Case Simulations of Hurricane–Ocean Interaction Using A High-Resolution Coupled Model: Effects on Hurricane Intensity, Mon. Weather Rev., 128, 917 – 946, https://doi.org/10.1175/1520-0493(2000)128<0917:RCSOHO>2.0.CO;2, 2000.

Davis, C., Wang, W., Chen, S. S., Chen, Y., Corbosiero, K., DeMaria, M., Dudhia, J., Holland, G., Klemp, J., Michalakes, J., Reeves, H., Rotunno, R., Snyder, C., and Xiao, Q.: Prediction of Landfalling Hurricanes with the Advanced Hurricane WRF Model, Mon. Weather Rev., 136, 1990 – 2005, https://doi.org/10.1175/2007MWR2085.1, 2008.

Davis, C., Wang, W., Cavallo, S., Done, J., Dudhia, J., Fredrick, S., Michalakes, J., Caldwell, G., Engel, T., and Torn, R.: High-Resolution Hurricane Forecasts, Comput. Sci. Eng., 13, 22–30, https://doi.org/10.1109/MCSE.2010.74, 2011.

Dipankar, A., Stevens, B., Heinze, R., Moseley, C., Zängl, G., Giorgetta, M., and Brdar, S.: Large eddy simulation using the general circulation model ICON, J. Adv. Model. Earth Syst., 7, 963–986, https://doi.org/10.1002/2015MS000431, 2015.

Dong, J., Liu, B., Zhang, Z., Wang, W., Mehra, A., Hazelton, A. T., Winterbottom, H. R., Zhu, L., Wu, K., Zhang, C., Tallapragada, V., Zhang, X., Gopalakrishnan, S., and Marks, F.: The Evaluation of Real-Time Hurricane Analysis and Forecast System (HAFS) Stand-Alone Regional (SAR) Model Performance for the 2019 Atlantic Hurricane Season, Atmosphere, 11, https://doi.org/10.3390/atmos11060617, 2020.

Enz, B. M., Engelmann, J. P., and Lohmann, U.: Use of threshold parameter variation for tropical cyclone tracking, Geosci. Model Dev., 16, 5093–5112, https://doi.org/10.5194/gmd-16-5093-2023, 2023.

Gao, K., Harris, L., Bender, M., Chen, J.-H., Zhou, L., and Knutson, T.: Regulating Fine-Scale Resolved Convection in High-Resolution Models for Better Hurricane Track Prediction, Geophys. Res. Lett., 50, e2023GL103 329, https://doi.org/https://doi.org/10.1029/2023GL103329, e2023GL103329 2023GL103329, 2023.

Gopalakrishnan, S. G., Bacon, D. P., Ahmad, N. N., Boybeyi, Z., Dunn, T. J., Hall, M. S., Jin, Y., Lee, P. C. S., Mays, D. E., Madala, R. V., Sarma, A., Turner, M. D., and Wait, T. R.: An Operational Multiscale Hurricane Forecasting System, Mon. Weather Rev., 130, 1830 – 1847, https://doi.org/10.1175/1520-0493(2002)130<1830:AOMHFS>2.0.CO;2, 2002.

Gutmann, E. D., Rasmussen, R. M., Liu, C., Ikeda, K., Bruyere, C. L., Done, J. M., Garrè, L., Friis-Hansen, P., and Veldore, V.: Changes in Hurricanes from a 13-Yr Convection-Permitting Pseudo–Global Warming Simulation, J. Climate, 31, 3643 – 3657, https://doi.org/10.1175/JCLI-D-17-0391.1, 2018.

Heikenfeld, M., Marinescu, P. J., Christensen, M., Watson-Parris, D., Senf, F., van den Heever, S. C., and Stier, P.: tobac 1.2: towards a flexible framework for tracking and analysis of clouds in diverse datasets, Geosci. Model Dev., 12, 4551–4570, https://doi.org/10.5194/gmd-12-4551-2019, 2019.

Heinze, R., Dipankar, A., Henken, C. C., Moseley, C., Sourdeval, O., Trömel, S., Xie, X., Adamidis, P., Ament, F., Baars, H., Barthlott, C., Behrendt, A., Blahak, U., Bley, S., Brdar, S., Brueck, M., Crewell, S., Deneke, H., Girolamo, P. D., Evaristo, R., Fischer, J., Frank, C., Friederichs, P., Göcke, T., Gorges, K., Hande, L., Hanke, M., Hansen, A., Hege, H.-C., Hoose, C., Jahns, T., Kalthoff, N., Klocke,

D., Kneifel, S., Knippertz, P., Kuhn, A., van Laar, T., Macke, A., Maurer, V., Mayer, B., Meyer, C. I., Muppa, S. K., Neggers, R. A. J., Orlandi, E., Pantillon, F., Pospichal, B., Röber, N., Scheck, L., Seifert, A., Seifert, P., Senf, F., Siligam, P., Simmer, C., Steinke, S., Stevens, B., Wapler, K., Weniger, M., Wulfmeyer, V., Zängl, G., Zhang, D., and Quaas, J.: Large-eddy simulations over Germany using ICON: a comprehensive evaluation, Quart. J. Roy. Meteor. Soc., 143, 69–100, https://doi.org/10.1002/qj.2947, 2017.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. R. Meteorolog. Soc., 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Hogan, R. J. and Bozzo, A.: A Flexible and Efficient Radiation Scheme for the ECMWF Model, J. Adv. Model. Earth Syst., 10, 1990–2008, https://doi.org/https://doi.org/10.1029/2018MS001364, 2018.

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., Giorgetta, M., Gutjahr, O., Haak, H., Hanke, M., Ilyina, T., Jahns, T., Jungclaus, J., Kern, M., Klocke, D., Kluft, L., Kölling, T., Kornblueh, L., Kosukhin, S., Kroll, C., Lee, J., Mauritsen, T., Mehlmann, C., Mieslinger, T., Naumann, A. K., Paccini, L., Peinado, A., Praturi, D. S., Putrasahan, D., Rast, S., Riddick, T., Roeber, N., Schmidt, H., Schulzweida, U., Schütte, F., Segura, H., Shevchenko, R., Singh, V., Specht, M., Stephan, C. C., von Storch, J.-S., Vogel, R., Wengel, C., Winkler, M., Ziemen, F., Marotzke, J., and Stevens, B.: ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales, Geosci. Model Dev., 16, 779–811, https://doi.org/10.5194/gmd-16-779-2023, 2023.

ICON Documentation: Build and Run: Input Data — Lateral Boundary Conditions, https://docs.icon-model.org/documentation/buildrun/buildrun_input_data.html, last accessed: 1 July 2026, 2026.

ICON partnership: ICON release 2025.04, https://doi.org/10.35089/wdcc/iconrelease2025.04, dWD; MPI-M; DKRZ; KIT; C2SM, 2025.

Kleppek, S., Muccione, V., Raible, C. C., Bresch, D. N., Koellner-Heck, P., and Stocker, T. F.: Tropical cyclones in ERA-40: A detection and tracking method, Geophys. Res. Lett., 35, https://doi.org/https://doi.org/10.1029/2008GL033880, 2008.

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., Mazumder, P., Copik, M., Weber, B., Herten, A., Bockelmann, H., Hoefler, T., Hohenegger, C., and Stevens, B.: Computing the Full Earth System at 1km Resolution, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC '25, p. 125–136, Association for Computing Machinery, New York, NY, USA, ISBN 9798400714665, https://doi.org/10.1145/3712285.3771789, 2025.

Kurihara, Y., Tuleya, R. E., and Bender, M. A.: The GFDL Hurricane Prediction System and Its Performance in the 1995 Hurricane Season, Mon. Weather Rev., 126, 1306 – 1322, https://doi.org/10.1175/1520-0493(1998)126<1306:TGHPSA>2.0.CO;2, 1998.

Landsea, C. W. and Franklin, J. L.: Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format, Mon. Weather Rev., 141, 3576 – 3592, https://doi.org/10.1175/MWR-D-12-00254.1, 2013.

Mauritsen, T., Redler, R., Esch, M., Stevens, B., Hohenegger, C., Klocke, D., Brokopf, R., Haak, H., Linardakis, L., Röber, N., et al.: Early development and tuning of a global coupled cloud resolving model, and its fast response to increasing CO2, Tellus Series A: Dynamic Meteorology and Oceanography, 74, 346–363, https://doi.org/https://doi.org/10.16993/tellusa.54, 2022.

Müller, W. A., Früh, B., Korn, P., Potthast, R., Baehr, J., Bettems, J.-M., Bölöni, G., Brienen, S., Fröhlich, K., Helmert, J., Jungclaus, J., Köhler, M., Lorenz, S., Schneidereit, A., Schnur, R., Schulz, J.-P., Schlemmer, L., Sgoff, C., Pham, T. V., Pohlmann, H., Vogel, B., Vogel, H.,

Wirth, R., Zaehle, S., Zängl, G., Stevens, B., and Marotzke, J.: ICON: Toward Vertically Integrated Model Configurations for Numerical Weather Prediction, Climate Predictions, and Projections, Bull. Am. Meteorol. Soc., 106, E1017 – E1031, https://doi.org/10.1175/BAMS-D-24-0042.1, 2025.

National Hurricane Center: Tropical Cyclone Report: Hurricane Paulette (AL172020), Tech. rep., National Hurricane Center, Miami, FL, USA, https://www.nhc.noaa.gov/data/tcr/AL172020_Paulette.pdf, last access: 5 February 2026, 2021.

Satoh, M., Noda, A. T., Seiki, T., Chen, Y.-W., Kodama, C., Yamada, Y., Kuba, N., and Sato, Y.: Toward reduction of the uncertainties in climate sensitivity due to cloud processes using a global non-hydrostatic atmospheric model, Prog. Earth Planet. Sci., 5, 67, https://doi.org/10.1186/s40645-018-0226-1, 2018.

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., and Düben, P.: Global Cloud-Resolving Models, Curr. Clim. Change Rep., https://doi.org/10.1007/s40641-019-00131-0, 2019.

Schulzweida, U.: CDO User Guide, https://doi.org/10.5281/zenodo.10020800, 2023.

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., Brunner, L., Cheedela, S. K., Das, S., Denissen, J., Dragaud, I., Dziekan, P., Ekblom, M., Engels, J. F., Esch, M., Forbes, R., Frauen, C., Freischem, L., García-Maroto, D., Geier, P., Gierz, P., González-Cervera, A., Grayson, K., Griffith, M., Gutjahr, O., Haak, H., Hadade, I., Haslehner, K., ul Hasson, S., Hegewald, J., Kluft, L., Koldunov, A., Koldunov, N., Kölling, T., Koseki, S., Kosukhin, S., Kousal, J., Kuma, P., Kumar, A. U., Li, R., Maury, N., Meindl, M., Milinski, S., Mogensen, K., Niraula, B., Nowak, J., Praturi, D. S., Proske, U., Putrasahan, D., Redler, R., Santuy, D., Sármány, D., Schnur, R., Scholz, P., Sidorenko, D., Spät, D., Sützl, B., Takasuka, D., Tompkins, A., Uribe, A., Valentini, M., Veerman, M., Voigt, A., Warnau, S., Wachsmann, F., Wacławczyk, M., Wedi, N., Wieners, K.-H., Wille, J., Winkler, M., Wu, Y., Ziemen, F., Zimmermann, J., Bender, F. A.-M., Bojovic, D., Bony, S., Bordoni, S., Brehmer, P., Dengler, M., Dutra, E., Faye, S., Fischer, E., van Heerwaarden, C., Hohenegger, C., Järvinen, H., Jochum, M., Jung, T., Jungclaus, J. H., Keenlyside, N. S., Klocke, D., Konow, H., Klose, M., Malinowski, S., Martius, O., Mauritsen, T., Mellado, J. P., Mieslinger, T., Mohino, E., Pawłowska, H., Peters-von Gehlen, K., Sarré, A., Sobhani, P., Stier, P., Tuppi, L., Vidale, P. L., Sandu, I., and Stevens, B.: nextGEMS: entering the era of kilometer-scale Earth system modeling, Geosci. Model Dev., 18, 7735–7761, https://doi.org/10.5194/gmd-18-7735-2025, 2025.

Seifert, A. and Beheng, K. D.: A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description, Meteor. Atmos. Phys., 92, 45–66, https://doi.org/https://doi.org/10.1007/s00703-005-0112-4, 2006.

Senf, F.: Data for Nested Hurricane-Centric ICON Runs for Hurricane Paulette (2020), https://www.wdc-climate.de/ui/entry?acronym=DKRZ_LTA_1376_dsg0001, 2025.

Senf, F.: ICON v2.6.6 Patch: for flexible SSTs in nested domains, https://doi.org/10.5281/zenodo.18385022, 2026a.

Senf, F.: Jupyter Notebooks for Plotting Analysis of "Advancing Hurricane-Centric Simulations with ICON", Submission Release, https://doi.org/10.5281/zenodo.19002039, 2026b.

Senf, F.: Hurricane Centric Toolset Version v2026.01 - Common Initialization and Perturbation Experiment Capabilities, https://doi.org/10.5281/zenodo.18271898, 2026c.

Senf, F.: ICON v2025-04-2 Patch: for handling vertical interpolation in init of two-moment cloud microphysics, https://doi.org/10.5281/zenodo.18387047, 2026d.

Senf, F., Voigt, A., Clerbaux, N., Hünerbein, A., and Deneke, H.: Increasing Resolution and Resolving Convection Improve the Simulation of Cloud-Radiative Effects Over the North Atlantic, J. Geophys. Res. Atmos., 125, e2020JD032 667, https://doi.org/10.1029/2020JD032667, e2020JD032667 10.1029/2020JD032667, 2020.

Shukla, J., Palmer, T. N., Hagedorn, R., Hoskins, B., Kinter, J., Marotzke, J., Miller, M., and Slingo, J.: Toward a New Generation of World Climate Research and Computing Facilities, Bull. Am. Meteorol. Soc., 91, 1407 – 1412, https://doi.org/10.1175/2010BAMS2900.1, 2010.

Sokolowsky, G. A., Freeman, S. W., Jones, W. K., Kukulies, J., Senf, F., Marinescu, P. J., Heikenfeld, M., Brunner, K. N., Bruning, E. C., Collis, S. M., Jackson, R. C., Leung, G. R., Pfeifer, N., Raut, B. A., Saleeby, S. M., Stier, P., and van den Heever, S. C.: *tobac* v1.5: introducing fast 3D tracking, splits and mergers, and other enhancements for identifying and analysing meteorological phenomena, Geosci. Model Dev., 17, 5309–5330, https://doi.org/10.5194/gmd-17-5309-2024, 2024.

Stevens, B., Acquistapace, C., Hansen, A., , and Coauthors incl. Senf, F.: The Added Value of Large-eddy and Storm-resolving Models for Simulating Clouds and Precipitation, J. Meteor. Soc. Japan, 98, 395–435, https://doi.org/10.2151/jmsj.2020-021, 2020.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Mon. Wea. Rev., 117, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:acmfsf>2.0.co;2, 1989.

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., and Reed, K. A.: TempestExtremes v2.1: a community framework for feature detection, tracking, and analysis in large datasets, Geosci. Model Dev., 14, 5023–5048, https://doi.org/10.5194/gmd-14-5023-2021, 2021.

Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core, Quart. J. Roy. Meteor. Soc., 141, 563–579, https://doi.org/10.1002/qj.2378, 2014.

Zängl, G., Reinert, D., and Prill, F.: Grid refinement in ICON v2.6.4, Geosci. Model Dev., 15, 7153–7176, https://doi.org/10.5194/gmd-15-7153-2022, 2022.

Advancing the Capabilities for Efficient Hurricane-Centric Simulations with the Atmospheric Model ICON

Fabian Senf[1] and Roxana Cremer[1]

[1]Leibniz Institute for Tropospheric Research, Leipzig, Germany.

Correspondence: Fabian Senf (senf@tropos.de)

Abstract. Global storm-resolving simulations with kilometer resolution are increasingly replacing traditional climate modeling approaches and show particular potential for resolving the dynamics and effects of deep moist convection. These modern modeling methods are moving toward sub-km scales, leading to extremely high energy and resource requirements. This makes iterations, parameter optimizations, and sensitivity studies no longer easily feasible. For the class of propagating, large-scale weather phenomena such as hurricanes, high-resolution limited-area approaches in combination with Lagrangian methods are therefore of interest, in which refined grids are shifted along the path of the phenomenon under consideration. To create this capacity for the ICON atmospheric model, this study develops a flexible workflow toolkit to enable efficient simulations of ~~tropical cyclones~~ hurricanes on a sub-km scale. Our approach leverages the flexibility that ICON offers through the ability to create custom grids. The concept divides hurricane tracks into overlapping temporal windows of 12-24 hours and generates customized grid segments that follow the hurricane's movement. The technical implementation automates key components of the workflow, including hurricane tracking, flexible grid generation, and preparation and merging of initial conditions across successive segments. The application of the workflow is demonstrated using Hurricane Paulette (2020) as an example, for which high-resolution simulations with grid spacings down to 300 m were performed using different segment configurations. The results show that the hurricane track remains consistent with the base run and depends mainly on the across-track width of the chosen configuration, while intensity metrics such as minimum pressure and maximum wind speed show significant sensitivities to resolution in our multi-nested setups. The efficiency gains are significant compared to traditional approaches with fixed limited-area domains: ~~The hurricane-centric method~~ replacing a fixed domain with a track-following tube reduces resource requirements by factors of ~~13-175, depending on the area configuration~~2–8, and segmenting the tube into short, frequently re-initialized intervals adds a further factor of 2–6. Analysis of ~~spin-up~~ re-initialization effects shows systematic but manageable impacts during segment transitions. Nevertheless, the efficiency gains achieved by our method are so substantial that they justify the acceptance of the ~~spin-up~~ re-initialization effects. Our segment-based approach in the hurricane-centric reference system now allows for more flexible regional hurricane simulations with the ICON model and more efficient investigation of new research questions regarding the sensitivity of hurricane cloud systems at very high resolutions.

Short Summary

Computer models for hurricane prediction are becoming increasingly detailed but require substantial computing resources. We developed a flexible approach that follows hurricanes as they move, applying high-resolution simulations only where needed.

This method reduces computing costs by ~~factors of 13-175~~ up to one order of magnitude compared to conventional fixed limited-area domains, while achieving resolutions down to 300 meters. The approach enables more efficient hurricane research and improved understanding of tropical dynamics.

1 Introduction

Global storm-resolving simulations on a kilometer scale are increasingly replacing traditional climate modeling approaches (Satoh et al., 2018, 2019; Segura et al., 2025). These modern modeling methods show particular potential for resolving the dynamics of deep moist convective phenomena and their various forms of spatial organization, and the resulting interactions between small and large scales in a more physically realistic manner. Global storm-resolving simulations are now constantly evolving from kilometer to hectometer resolutions, enabling unprecedented detail in the representation of atmospheric processes (Hohenegger et al., 2023; Klocke et al., 2025). However, the development and calibration of these high-resolution approaches pose significant challenges (Shukla et al., 2010). They are extremely energy- and resource-intensive in terms of both computing resources and storage requirements. As a result, iterations, parameter optimizations, and sensitivity studies can no longer be easily performed (Mauritsen et al., 2022). For this reason, alternative and efficient approaches are needed to understand how well regional to locally limited phenomena are represented at high resolutions and standard parameter settings.

One typical solution to this problem is to use limited-area domains with high spatial resolution that refine the global model simulations in a predefined limited area and within a certain time period, thereby significantly reducing the computing resources required. These limited-area solutions can be designed to be consistent with global simulations by using initial and boundary conditions derived from them and working with the same set of physical parameterizations. Of course, the capability of limited-area approaches is inherently somewhat constrained, since interactions with boundary conditions can introduce errors and the ~~influence of larger-scale processes~~ upscale influence of small-scale processes onto the larger scales is not captured adequately. The atmospheric component of the Earth system model ICON ~~(ICOsahedral Nonhydrostatic)~~ (ICOsahedral Nonhydrostatic, see Zängl et al., 2014) already includes the possibility of simulating such regional domains with nested grid refinements (Stevens et al., 2020; Zängl et al., 2022; Hohenegger et al., 2023). These approaches are usually limited to spatially fixed domains that must be defined in advance. This concept is very well suited for spatially and temporally limited phenomena, such as the diurnal evolution of convection over land under rather calm wind conditions. However, the approach with fixed regional domains is less suitable when the weather phenomena under consideration propagate over large distances, such as hurricanes, and thus the covered area is as large as entire oceans. For this class of problems, high resolution Lagrangian approaches are interesting, in which the refined grids are shifted along the path of the phenomenon under consideration. Such approaches have been successfully applied for a long time for flexible high resolution hurricane forecasts in operational settings (Kurihara et al., 1998; Gopalakrishnan et al., 2002).

For predicting the path of hurricanes, km-scale model approaches have been shown to reach reduced track position errors (Dong et al., 2020). For ~~good~~ accurate simulations of the structure and intensity of hurricanes, an explicit and well-designed representation of deep moist convective processes also appears to be key (Gao et al., 2023). It has been further shown for Hur-

ricane Katrina, that its intensification and structure of its convective bands only improved sufficiently when a spatial resolution of around 1 km is achieved (Davis et al., 2008). Systematic investigations in Davis et al. (2011) have also shown that intensity biases with resolutions around 1 km are reduced, especially for very high wind speeds, compared to coarser model resolutions. The study argues that it is essential to resolve cloud structures in hurricanes, but also that significant improvements between 100 m and 1 km horizontal resolution are questionable. These examples can be continued and provide evidence that a high spatial resolution is essential for realistic hurricane simulations and that entering the sub-kilometer scale may need further investigations.

Together, high spatial resolution and flexible adaptation of simulation domains can help enable novel, high-quality hurricane simulations while keeping computing costs within reasonable limits. Our efficient approach presented here contributes to this goal and advances the capabilities ~~for the Earth system model ICON~~ of ICON model. Redesigning the grid and data management in ICON and the associated communication patterns is so challenging, and the risks of negatively impacting the outstanding performance achieved in Klocke et al. (2025) are so high, that we have decided to implement a workflow-based approach to make regional hurricane simulations more flexible. This approach is an intermediate step on the way to a fully Lagrangian model concept and has the advantage that we can completely retain the proven and optimized communication patterns and data structures of ICON. Our approach allows us to take advantage of the flexibility that ICON offers through its ability to create user-defined grids. In this way, we can perform targeted and efficient hurricane-centric simulations down to the hectometer scale without having to re-optimize the efficiency of the underlying model. We are thus creating the possibility of using an approximate Lagrangian approach with the ICON versions available today to create innovative datasets for hurricane investigations in very high resolution. With this approach, we would like to encourage the scientific modeling community in general to give more consideration to co-design methods that promote the parallel development of sophisticated global simulation approaches and flexible and efficient regional refinements.

The paper is structured as follows: In Sect. 2, we present the methodology of our flexible refinement workflow in detail. We provide a conceptual overview together with detailed workflow diagrams and discuss the portability of our methods to different high-performance computing platforms. Sect. 3 demonstrates the application of the workflow using Hurricane Paulette (2020) as an example. Different refinement strategies are explored, and the results of high-resolution simulations are presented and discussed. There we also provide an analysis of the performance and resource requirements of our approach. Finally, Sect. 4 summarizes the key findings and outlines potential future directions for research in this area.

2 Methodology

2.1 General Overview

Figure 1 provides a general overview of the flexible refinement workflow used to create high-resolution hurricane simulations. The various steps are carried out sequentially. The first step is to create a base simulation of a hurricane. Both global and sufficiently large regional setups can be used for this purpose. In the example case presented in Sect. 3.1, ~~two one-way nested domains with grid spacings of 5 and 2.5 km are employed to cover~~ a limited-area setup covering the entire tropical Atlantic

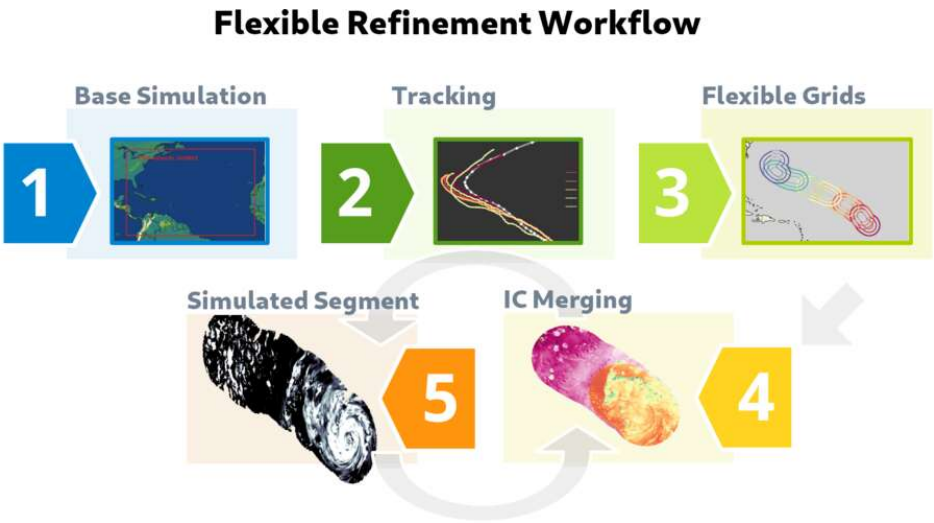

Figure 1. Overview of the flexible refinement workflow: Five steps are shown in different colored boxes and numbered accordingly. The steps of the workflow include (1) creating the base simulation, (2) identifying and tracking the hurricane, (3) creating the flexible grids and other necessary simulation data, (4) combining the initial conditions from the base and newly created simulation data, and (5) performing the actual high-resolution hurricane simulation. Gray arrows indicate the flow in the workflow diagram, with steps 4 and 5 being performed as iterative loop over the respective hurricane segments.

~~within a regional configuration~~is employed. Global storm-resolving simulations, such as the projections of the nextGEMS project (Segura et al., 2025), could also be used as base simulations. A wide range of data from the base simulation will be required in subsequent steps.

An important prerequisite is that the base simulation generates a hurricane and reproduces its characteristics sufficiently well. In the second step (see Fig. 1), the hurricane is identified in the base data and tracked over time. This tracking process creates a Lagrangian reference system, initiating the transition from a fixed domain to a flexible, hurricane-centric setup. It is important to note that we follow the hurricane as a large-scale structure which is not necessarily the same as shifting the domain according to the mean horizontal wind.

In the third step, we split apart the hurricane track into individual elements, which we refer to as "*segments*". Based on these, flexible and user-defined grids are set up that overlap so that previous information can be used for subsequent simulations. All essential data like external parameters, lateral boundary and initial conditions (ICs) are created specifically for these grids. This completes the pre-processing workflow.

Steps four and five build the production workflow. Both steps are performed alternately for each simulation segment. Step four generates the respective ICs from which the atmospheric simulation is started. For the first segment, ICs are created solely from the base simulation. Step five conducts the actual high-resolution hurricane simulations in the respective grid segment and stores the necessary ICs for the next simulation. For the second and the subsequent segments, the loop goes back to step four and data from the previous segment is reused in the overlap area for the new simulation. Where there is no overlap between

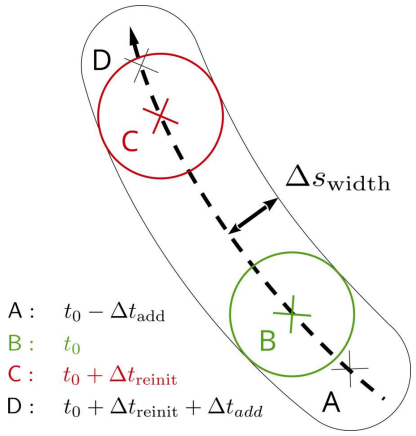

Figure 2. Schematic representation of a track segment. The reference track is represented by a thick, dashed black line. A grid segment with a width of $\Delta s_{\mathrm{width}}$ is constructed around it (thin solid black line). The spatial points A to D are determined by their temporal sequence along the track. The circles around points B and C indicate the areas of interest for analysis at the start and end time of the simulation within the segment. Points A and D show the extension of the grid segment by a predefined time Δt_{add}. For simplicity, only the outer nest of the track segment is shown. Further nests follow the same scheme, but lie inside the outer nest with smaller across-track widths.

two subsequent grid segments, ICs from the base simulation are taken. Thus, step four produces a mixed initial state consisting of a merge of two information sources ~~.~~ (see Sect. 2.3 for more details).

2.2 Segment Concept

Our entire workflow is based on the idea of simulating a hurricane only within its vicinity along its temporal evolution. To accomplish this efficiently, we rely on a segment concept for which a reference track is divided into individual *track segments*. The length of the track segments is mainly determined by the envisioned re-initialization time interval, $\Delta t_{\mathrm{reinit}}$ (plus some additional buffer). Re-initialization data must be available from the base simulation for the chosen interval, e.g. 12 or 24 hours, to support the re-initialization of the higher-resolution simulations.

Figure 2 shows how a track segment is constructed. The points B and C correspond to the location of the hurricane center in the base simulations at times t_0 and $t_0 + \Delta t_{\mathrm{reinit}}$. The points A and D expand the track segment in the respective directions by a certain buffer distance to account for situations in which the hurricane moves slower or faster along its path in the higher-resolution simulations. For practical reasons, this buffer distance is also specified as a time interval, Δt_{add}. Combining re-initialization interval and buffer time, the selected track segment runs from point A to point D and thus covers the distance traveled by the hurricane e.g., from ~~18~~ 21 UTC the previous day to 15 UTC ($\Delta t_{\mathrm{add}} = 3$ h; $\Delta t_{\mathrm{reinit}} = 12$ h).

A *grid segment* refers to a track segment that has been spatially expanded horizontally. It is constructed by expanding the selected track segment by a user-defined across-track width, $\Delta s_{\mathrm{width}}$, in all directions, forming a tube-like shape. Imagine a circular area with a radius of $\Delta s_{\mathrm{width}}$ and the center at the respective track point sweeping over the track segment. The circular section at a time t is defined later as Lagrangian analysis region for more detailed investigations (see Sect. 3.4).

Finally, the segment concept is applied to different start times t_i which link the ~~i~~k-th segment to an individual initialization time. For practical reasons, we count segments in relation to the initialization time of the reference t_{ini}. The start time of the ~~i~~k-th segment is thus given by ~~$t_i = t_{\mathrm{ini}} + i\,\Delta t_{\mathrm{reinit}}$ where $i = 0$~~ $t_k = t_{\mathrm{ini}} + k\,\Delta t_{\mathrm{reinit}}$ where $k = 0$ selects $t_0 = t_{\mathrm{ini}}$. If e.g. 24 hours of model spin-up are omitted from the reference data, the first meaningful segments can be found at ~~$i = 1$~~ $k = 1$ for $\Delta t_{\mathrm{reinit}} = 24\,\mathrm{h}$ and at ~~$i = 2$~~ $k = 2$ for $\Delta t_{\mathrm{reinit}} = 12\,\mathrm{h}$. However, it is up to the user to define which segments are the first meaningful segments in their simulation chain.

2.3 Workflow Schemes and Implementation Details

Section 2.1 provides an introductory description of the various stages incorporated within the workflow. Specifically, steps 1 through 3 are categorized under the pre-processing workflow, while steps 4 and 5 pertain to the production workflow. The automation of tasks using algorithms exhibits varying degrees of complexity. In the following, a detailed analysis is shown that details the implementation of the most complex algorithms used throughout the workflow.

Figure 3 provides a comprehensive flowchart of the grid generation algorithm. The starting point is the horizontal wind field at 10 m height from the base simulation that is converted to relative horizontal vorticity ζ and further coarse-grained with a ~~convolution filter~~ running-average filter with a width of around 200 km. The coarse-grained vorticity ζ_c field allows to identify larger-scale cyclonic motions such as those characteristic for hurricanes. Detection and tracking of cyclonic features is done with the open-source python package *tobac* (Heikenfeld et al., 2019; Sokolowsky et al., 2024). A ~~criterion of ten daily revolutions~~ threshold criterion of $\zeta_c = 10\,\mathrm{day}^{-1}$ was used for identification, ensuring only the strongest motions were selected~~based on this quantitative threshold~~. Please note that a variety of other and more customized tools are available for identifying and tracking hurricanes, such as TempestExtremes (Ullrich et al., 2021) or the approaches by Kleppek et al. (2008) and Enz et al. (2023), which could have been applied here with equal justification. The preparation of the hurricane track data from base simulations is concluded by a matching algorithm. It utilizes location data of past hurricanes from HURDAT (Landsea and Franklin, 2013) to find the closest matching hurricane track in the base simulations (similar to Gutmann et al., 2018). This use of observational best-track data only provides additional insights for nudged simulations or simulations in a forecast mode, for free-running climate simulations a comparison to the observational track may not be physically meaningful.

The actual grid generation loop is entered with data from the best matching track, from a base grid (here, we take the grid from the finest reference nest) and from a configuration file that allows users to specify the along- and cross-track extend of the grid segments. The grid generation algorithm relies on DWD-ICON-Tools software (see Zängl et al., 2022, for grid-related aspects), facilitating the creation of flexible ICON grids through masks, crucial for our hurricane-centric workflow. For masking, a chosen base grid is read. Then, the track segment and its spatial extension by the across-track width, $\Delta s_{\mathrm{width}}$, are transferred to this grid to create a mask. The resulting mask is then used in the DWD-ICON-Tools grid generator to create the next refinement cutting out only the chosen grid segment. This builds the outermost nest of a flexible setup and serves as the basis for further refinements. If multiple nests are targeted, the track segment is transferred again, now to the just created outermost nest, and the lateral extension is performed again with a predefined, but smaller across-track width. This process is repeated several times until all nests for the respective refinement levels have been created. In Fig. 3, the shown loop

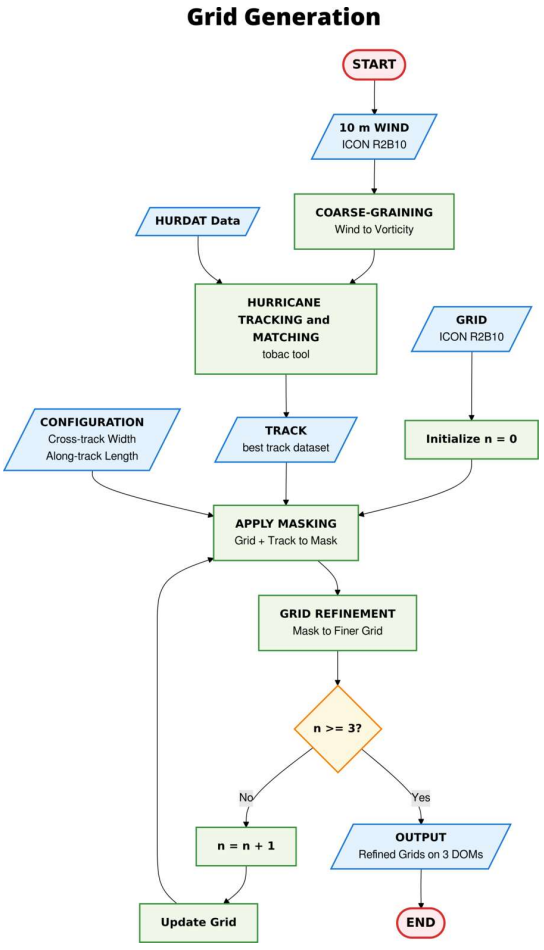

Figure 3. Detailed flowchart for the sequential grid generation process. Final start and end points are highlighted in red. Incoming and outgoing data are visualized as slanted blue boxes and methods that treat these data are shown as green rectangles. The index n counts the number of nested domains and thus determines the number of iterations in the grid generation loop. The term R2B10 indicates the triangular grid configuration of the ICON base run with a grid spacing of approximately 2.5 km.

determination criteria is $n = 3$ which means that a sequence of three nested grids with successively halved grid spacing are created. Finally, these nested grids resemble layers of an onion.

Figures 4a and b split up the automated algorithm for IC generation into a pre-processing part and into a production part. For IC pre-processing, data from the base simulation, here noted by IC* as well as reference grid (source) and segment grid (target) are input. The horizontal regridding uses the tool cdo (Schulzweida, 2023) together with a conservative remapping to calculate regridded IC*. This dataset is input into a test ICON run that serves several purposes. With the test run (i) the integrity of the execution workflow is tested, (ii) performance-relevant aspects are investigated and resources for production runs are estimated and (iii) internal vertical interpolation is applied to consistently bring the regridded IC* from the reference levels to

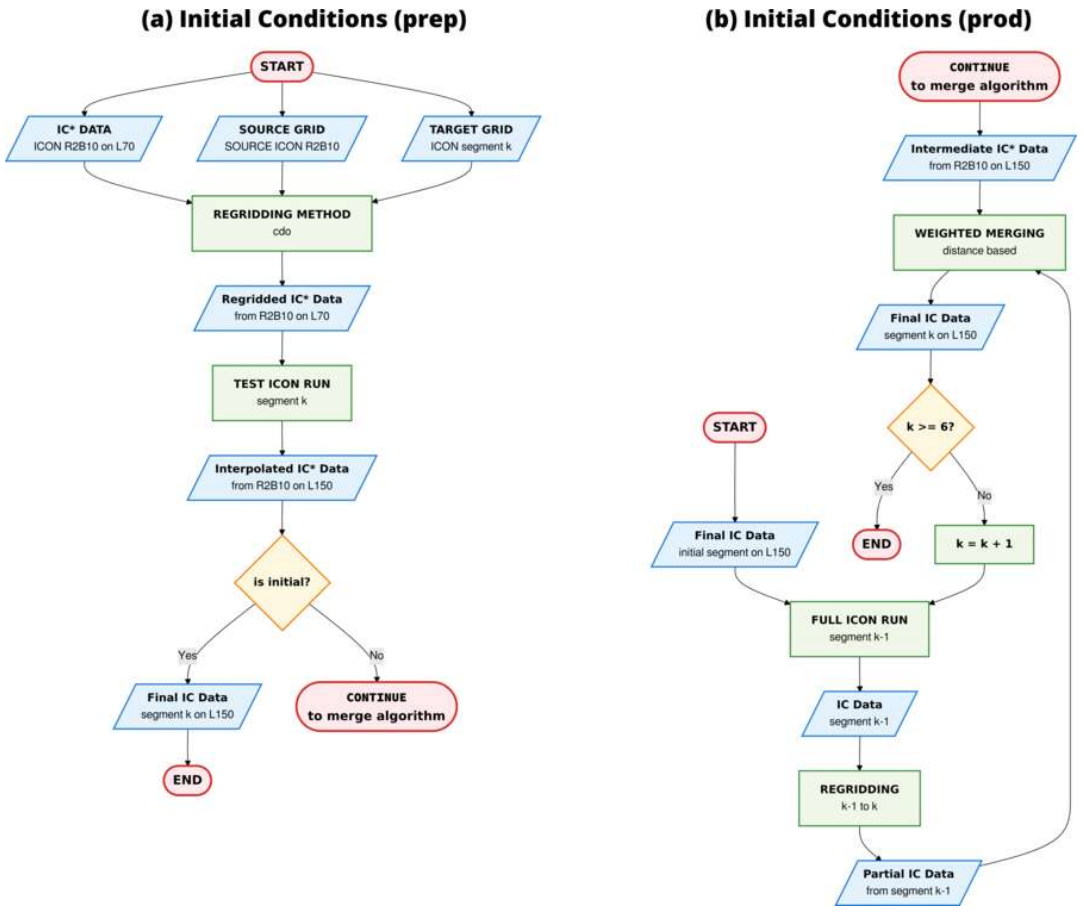

Figure 4. Detailed flowchart for (a) initial conditions pre-processing and (b) initial conditions production. Final and intermediate start and end points are highlighted in red. See Fig. 3 for meaning of the other elements. The index k counts the number of segments along the hurricane track. The loop terminates after a user-defined number of segments, e.g. here $k = 6$. The term L70 (L150) indicates the vertical grid configuration of the ICON base run with 70 (150) vertical levels.

the possibly more refined target levels. After that point, the IC pre-processing workflow stops with the decision whether the considered segment is chosen to be the initial segment or not.

In the IC production chain, regridded IC* are input into the full ICON run for the initial segment. At the end of the successful full ICON run, IC data for initializing the next segment are stored. Due to along-track shifts of the grid segments, these IC data only partially cover the region of the next segment. To compensate for this offset, a merging algorithm is applied that maps the partial IC data from segment $k-1$ to the grid for segment k in the overlap area. The remaining parts are filled with regridded IC* from the reference. ~~A distance-based weighting smooths the~~ The transition between the two IC data sources is smoothed by weights that decrease linearly with distance from the overlap boundary. Still, the mix of two IC ~~dataset~~ datasets introduces a discontinuity in the initialization - a limitation that is discussed in depth in Sect. 3.5.

After grid and IC generation, the external parameters and boundary conditions are remapped onto the newly created grid segments. The former is directly interpolated from the original external data onto the respective grids yielding the highest possible resolution of e.g. topography. The latter is exclusively obtained from the base run data and mapped using cdo on the respective lateral-boundary grids of the high-resolution simulations.

2.4 Portability and User Interfaces

The workflow toolkit is intended to run on super-computing platforms. It consists of a mix of python tools and bash scripts, the latter being submitted into a HPC scheduling system. The workflow was initially developed for the German Climate Computing Center (DKRZ) and its computing platform Levante and adjusted to its specific module environment and the Simple Linux Utility for Resource Management (SLURM) scheduler. The toolkit was then ported to the Jülich Supercomputing Center (JSC) Platform JUWELS, and its most important aspects were generalized. We therefore consider the software to be relatively platform-independent. In order to port the workflow toolkit, only the configurations of the corresponding module environment and the SLURM schedulers need to be properly set up for new platforms.

Users can readily configure the toolkit via a detailed configuration file, allowing them to set and modify parameters, as well as file and software dependencies, for different numerical experiments. We chose a TOML (Tom's Obvious, Minimal Language)-based configuration containing clearly structured elements and sections that can be easily edited with standard text editors. Additionally, we provide wrapper scripts to initiate the execution of entire workflows. It is possible, on the one hand, to start the pre-processing workflows of an entire series of consecutive grid segments. This results in the parallel execution of individual workflows that carefully handle their internal sequential dependencies. On the other hand, the production workflow wrappers can be executed. These trigger the execution of the ICON model in the selected grid segment with multiple restarts, if necessary, and merge the ICs before starting an ICON run for the next segment. All other user-relevant information can be found in the documentation of the latest software release (Senf, 2026c).

3 Application

3.1 Case Description

We now introduce a hurricane case chosen for demonstrating our flexible refinement workflow for hurricane-centric high resolution simulations. In 2020, Hurricane Paulette crossed the Atlantic and, with a total length of around 11,400 km, was one of the hurricanes with the longest track for that year. According to the National Hurricane Center's tropical cyclone report, Paulette formed as a tropical depression in the central Atlantic on September 7 and intensified into a tropical storm as it moved northwestward (National Hurricane Center, 2021). It reached category 1 hurricane status on September 13, intensified further into a category 2 hurricane in the early hours of September 14, and reached peak intensity of 90 kt (46 $\mathrm{m\,s^{-1}}$) and 965 hPa around 18 UTC on September 14. A snapshot of Paulette during its most intense phase is shown in Fig. 5a. Paulette finally

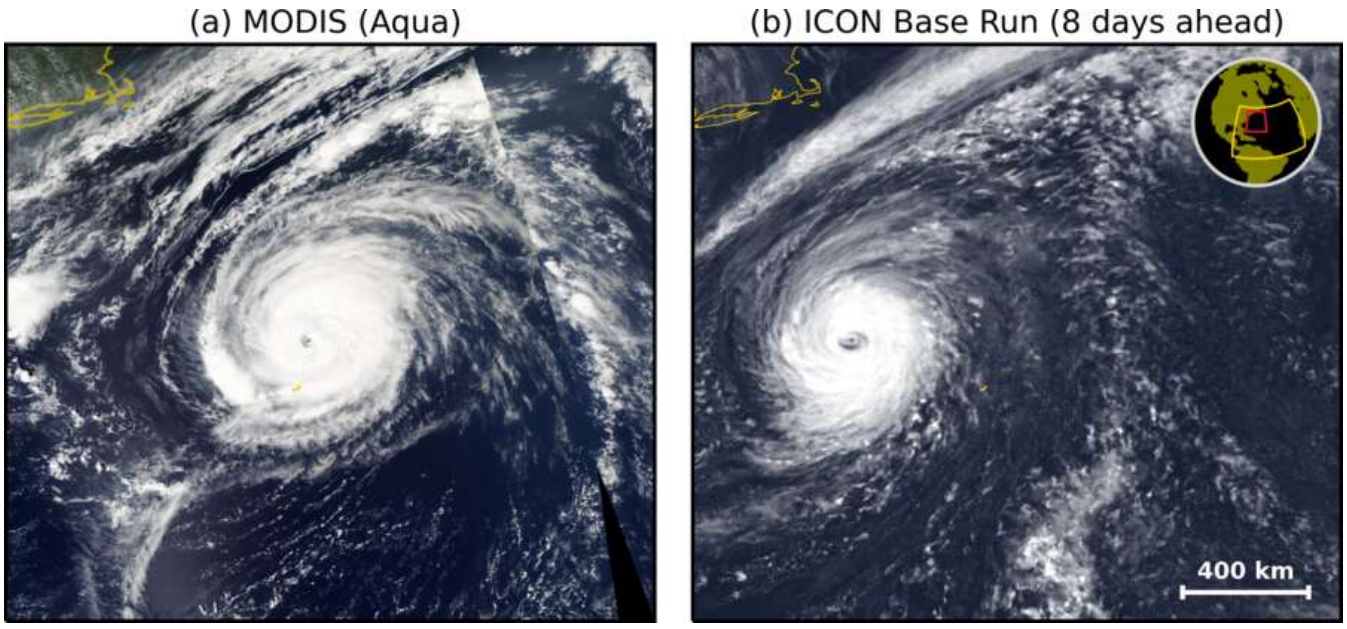

Figure 5. Top view of hurricane Paulette from (a) observations and (b) simulations. A regional cutout in the Atlantic ocean was defined reaching from -74.0 to -53.4 ° E and from 25.0 to 43.2 ° N. Subpanel (a) shows a true-colour image acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on the NOAA-AQUA satellite during an ascending orbit at overpass time on 14 September 2020, around 18:00 UTC. Subpanel (b) provides a simulation of the reflected solar radiation flux, normalized between 100 and 750 $\mathrm{W\,m^{-2}}$ from our ICON base run, which was initialized on 7 September 2020 at 0 UTC. The time, geographical region and projection have been matched between the observation and the simulation. The color scale and map in (b) have been adjusted to ensure visual consistency with (a). The length scale of 400 km is shown in the lower right corner of (b) and the inset in the upper right corner of (b) shows the location of the image cutout (red outline) and the inner base run domain (yellow outline) in the Atlantic ocean.

transitioned into an extratropical cyclone at 12 UTC on September 16. During its first seven days of formation and main intensification period, Paulette's path was relatively straight, making it an ideal use case for our workflow methods.

3.2 Base ICON Run

The ICON ~~weather and climate~~ model solves hydrodynamic equations on a triangular grid (see Zängl et al., 2014; Dipankar et al., 2015), and can operate as a global or regional modeling system (Hohenegger et al., 2023; Müller et al., 2025). For our baseline run, we used the regional setup of the ICON model in the configuration with the numerical weather forecast (NWP) physics package (Zängl et al., 2014). The base simulations ~~were performed in online~~ with an outer limited-area setup and grid spacings of 5 km forced by ERA5 data (Hersbach et al., 2020) were performed into which a second, inner domain with 2.5 km grid spacing was nested using one-way ~~nesting mode with two nested domains~~coupling.

For the outer nest, a spatial extent of 9900 ~~times~~ × 6700 $\mathrm{km^2}$ was chosen, which covers most of the tropical and subtropical Atlantic north of the equator. The outer nest was configured as R2B9 triangular ICON grid, corresponding to an equivalent grid

spacing of 5 km. In this domain, convection processes and convective precipitation were parameterized following Bechtold et al. (2008) based on Tiedtke (1989). For the inner nest, the ICON R2B10 grid setup was chosen, corresponding to a grid spacing of approximately 2.5 km. The vertical resolution of the base run was set to 70 model levels in both nests, extending from sea surface to a maximum height of 34 km. The inner nest is driven by the outer nest via a nudging zone close to the lateral boundary. However, due to one-way coupling, the inner nest does not provide feedback to the outer nest. In the inner nest, subgrid-scale convection was partially parameterized. The resolved model dynamics explicitly represent deep convective processes, while a parameterization approach still approximates shallow to mid-level convection. This setup has been shown to sufficiently reproduce realistic marine cloud structures and cloud radiative effects (Senf et al., 2020).

All other parameterizations from the NWP physics package were selected identically for both setups. Thus, all ICON simulations used the two-moment cloud microphysics scheme after Seifert and Beheng (2006). Subgrid-scale turbulent fluxes were calculated using a closure approach, as described by Baldauf et al. (2011), which includes the prognostic calculation of turbulent kinetic energy. The scheme ecRad was used for radiation calculations (Hogan and Bozzo, 2018). It performs radiative calculations in 14 solar and 16 terrestrial pre-defined spectral bands.

The simulations utilized in the following were initialized at 0 UTC on September 7 using the ERA5 data, a reanalysis dataset produced by European Centre for Medium-Range Weather Forecasts (ECWMF, Hersbach et al., 2020). This corresponds to a time when Paulette had already developed into a tropical depression. ~~Spin-up effects are clearly visible in the first hours of model simulations.~~ ERA5 data is available six-hourly as boundary data for the outer nest. For the base run, ICON version v2.6.6 is selected, which was the most recent ICON version at the time the base simulations were created. On top of this version, an update was developed that allows the sea surface temperature from ERA5 to be updated daily for all regional nests (Senf, 2026a). Therefore, SST fields are not kept constant at the initial SST values, but changed on a daily basis for our hurricane base simulations. Consequently, the actual effects of Paulette on the ocean surface are already included, which could indeed negatively impact the quality of simulating the hurricane development (Bender and Ginis, 2000). The output of the base run was configured so that data for initialization or re-initialization of higher-resolution runs is available every six hours. Furthermore, data to drive high-resolution runs via lateral boundary conditions is available at a resolution of one hour.

Figure 5b shows a visual impression of Hurricane Paulette simulated with our base run setup, at the time of the most intense hurricane period. The figure shows the reflected solar radiation at the top of the atmosphere for the inner R2B10 nest after an integration time of around eight days. For operational forecasts, Gao et al. (2023) reported a mean track error of approximately 300 km after 5 days of lead time. Here, a distance of approximately 400 km between our simulated hurricane center and the observation is achieved, which is a very positive result for a hindcast with around 8 days of lead time. Given the complexity and multiple interacting processes involved in hurricane formation, the reasonably good agreement was somewhat unexpected for the rather long forecast lead time. Although the variables being compared between the observation and simulation are not the same, making a quantitative comparison difficult, we still observe interesting differences in the morphology of the simulated hurricane compared to the observation. It appears that the spatial extent of the simulated anvil cloud associated with the hurricane is underestimated. Insufficient spatial resolution could be one reason for this discrepancy, among many others,

and methods that lead to more spatially refined simulations, such as the workflow presented in this paper, would be an important step in further investigating this question of resolution dependence.

3.3 Exploration of Different Refinement Strategies

In the following, we will examine the refinement workflow presented in Sect. 2 in more detail for Hurricane Paulette as an example. As shown previously, the base run results reveal an intriguing hurricane development, motivating us to conduct further investigations with higher resolutions. This is exactly the starting point for the flexible refinement workflow presented in this paper. The base data is now entered into the flexible grid calculation alongside the calculated track of the simulated hurricane.

As a user, one can now choose from various important parameters. These parameters determine how many grid segments are created and their size. The first fundamental question concerns the spatial resolution. Starting with the resolution of the base run (R2B10 with 2.5 km), higher resolutions can be achieved through successive refinement. The following example uses setups with three nests, ranging from R2B11 with a grid spacing of 1.2 km to R2B13 with a grid spacing of approximately 300 m. The currently implemented workflow can handle fewer nests but would require additional software updates for using more than three nests. In principle, it is also possible to start initial refinement from grids finer than the base grid for which minor software updates would be needed.

Figure 6 shows various implementations of grid segmentations for Paulette. These implementations vary either the length of a segment along the hurricane track (white line) using the *"reinit"* parameter $\Delta t_{\mathrm{reinit}}$ or the across-track width $\Delta s_{\mathrm{width}}$. The different grid segments are visualized by different colors. When $\Delta t_{\mathrm{reinit}} = 12\,\mathrm{h}$ is selected (Fig. 6a and c), the grid segments appear shorter and more compact. However, more grid segments must be calculated sequentially to cover the same integration period. For the 12-hour re-initialization setup, it makes sense to start with segment 2 on September 8, 2020, at 0 UTC to avoid spin-up of the base run. If the integration ends after the 12th segment, as shown, the hurricane's development can be tracked for five days and 12 hours. The 24-hour re-initialization setup requires fewer segments. However, each segment covers a larger region and requires more computing power (see Figs. 6b and 6d). Each of the 24-hour setups shown starts with segment 1, and sequential integration up to segment 6 covers a period of six days.

Another important parameter determines the width of the segments across the track, $\Delta s_{\mathrm{width}}$. Ideally, the width would be chosen so that the hurricane development would not be significantly influenced by the lateral domain boundaries. In practice, computational constraints are significant, so responsible scientists must select the smallest feasible domain size. Given these two requirements, it is difficult to provide definite advice and determine in advance what the optimum width would be. Two possible configurations are provided in Fig. 6. First, the innermost high-resolution nest at R2B13 is selected so that the across-track width, measured as distance between the track center and the domain edge, is 100 km. Second, a wider setup was tested in which the inner nest was extended to 200 km. Table 1 shows the extent of the two outer high-resolution nests in detail.

~~After grid generation, as described in Sect. 2.2, the external parameters, ICs, and boundary conditions are remapped onto the newly created grid segments. The latter two are obtained from the base run data and prepared for the high-resolution simulations.~~ High-resolution ~~simulations also require vertical refinement. For our examples, we chose 150 vertical levels~~

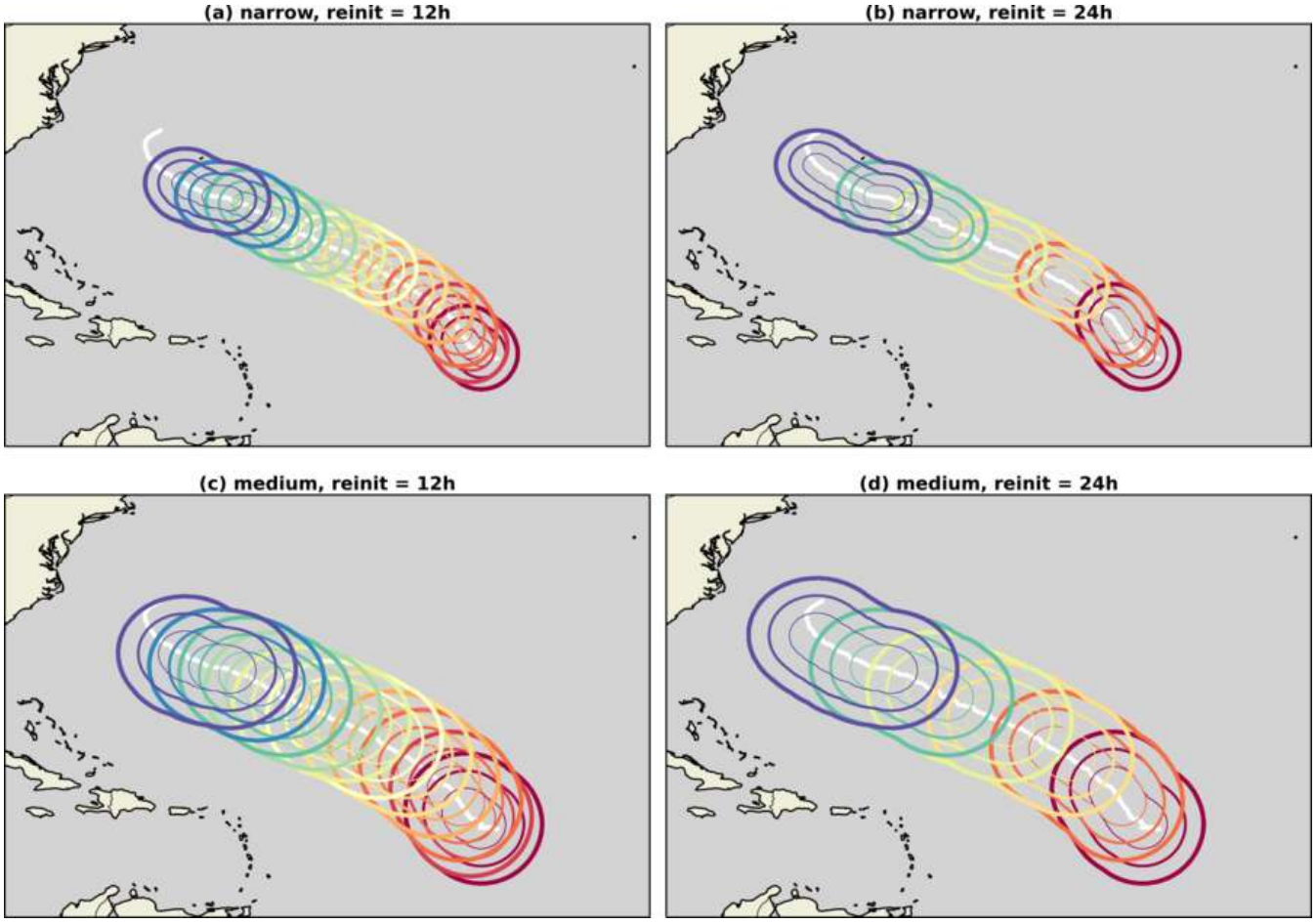

Figure 6. Grid segments prepared for hurricane Paulette (2020) with differently chosen re-initialization times and across-track widths. The narrow setup in (a) and (b) is compared to the medium setup in (c) and (d). Re-initialization at 12-hour intervals (a, c) leads to twice as many segments as re-initialization at 24-hour intervals (see Tab. 1). The track of the hurricane Paulette in the base simulation is plotted as a white solid line. Each segment consists of three nested domains (solid lines of decreasing thickness) colored by the initialization time, ranging from red on September 8 to blue on September 14.

~~(analogous to Heinze et al., 2017; Stevens et al., 2020), extending again to an altitude of 34 km. We apply the ICON internal interpolation routines for the required vertical interpolation of the initial data to ensure consistency with the vertically discretized hydrostatic balance in the model. The interpolation is carried out as part of a model test run.~~

~~High-resolution~~ ICON simulations were performed in the respective grid segments using the latest available version of ICON (v2025-04-2) (ICON partnership, 2025) at the time of the study. We chose 150 vertical levels (analogous to Heinze et al., 2017; Stevens et al., extending to an altitude of 34 km. Cloud microphysics and radiation were set up similarly to the base run. However, subgrid-scale turbulence was configured with a three-dimensional Smagorinsky-type mixing scheme and convection parameterization

Table 1. Comparison of grid and compute requirement parameters across re-initialization configurations and domains (DOM) listed in the different columns. The computational units (cu) refer to the product of the number of time steps and the number of horizontal grid cells and is a measure of the size of a limited area experiment.

property	reinit	narrow			medium			large	x-large
		DOM01	DOM02	DOM03	DOM01	DOM02	DOM03	DOM01	DOM01
grid	12h /24h	R2B11	R2B12	R2B13	R2B11	R2B12	R2B13	R2B11	R2B11
eq. resolution / m	12h /24h	1200	600	300	1200	600	300	1200	1200
width / km	12h /24h	300	200	100	500	350	200	750	1000
ave. # cells / 10^6	12h	0.35	0.76	1.2	0.79	1.8	3.1	-	-
	24h	0.46	1.1	1.8	0.96	2.3	4.2	1.8	3
cu / 10^9	12h	25	110	350	57	250	880	-	-
	24h	33	150	510	69	330	1200	130	210

was completely deactivated (following Stevens et al., 2020). To enable a ~~warm start~~ consistent re-initialization of complete cloud microphysics, the ICON source code was modified to incorporate vertical interpolation of all microphysical moments (Senf, 2026d). This means that graupel mixing ratios and ice crystal number concentrations, for example, are available as fields when the high-resolution simulations begin and when they are re-initialized. Consequently, all cloud variables are available without discontinuities and abrupt changes throughout the entire simulation period. These initial fields are made available for all nests simultaneously, enabling a seamless continuation of the simulations when switching from one segment to the next.

Figure 7 shows examples of the results of two high-resolution ICON configurations with different grid segmentation. The narrow-size setup is compared with the medium-size setup, both of which have a 12-hour re-initialization period. The position of the simulated hurricane pressure minimum in the high-resolution run remains close to the reference track of the base run. There are no significant differences in minimum position for different simulation resolutions, but same grid configuration. However, as expected, the track of the wider setup deviates more strongly from the reference track. This is plausible because the position of the hurricane center in the wider setup is not as strongly constrained by the forcing at the lateral boundaries. The intensity of the simulated hurricane, as shown by the development of the pressure minimum in Fig. 7b and the maximum 10-meter wind speed in Fig. 7c, clearly depends on the resolution. This is evident from the greater spread of the curves (esp. in Fig. 7b), which is as large as that induced by different grid configurations. Better resolved convective activity generates particularly high gustiness of the 10 m wind, which explains the fluctuations on the very short time scales in Fig. 7c. However, a more detailed investigation of the resolution effects and their causes and consequences must be postponed to a follow-up study.

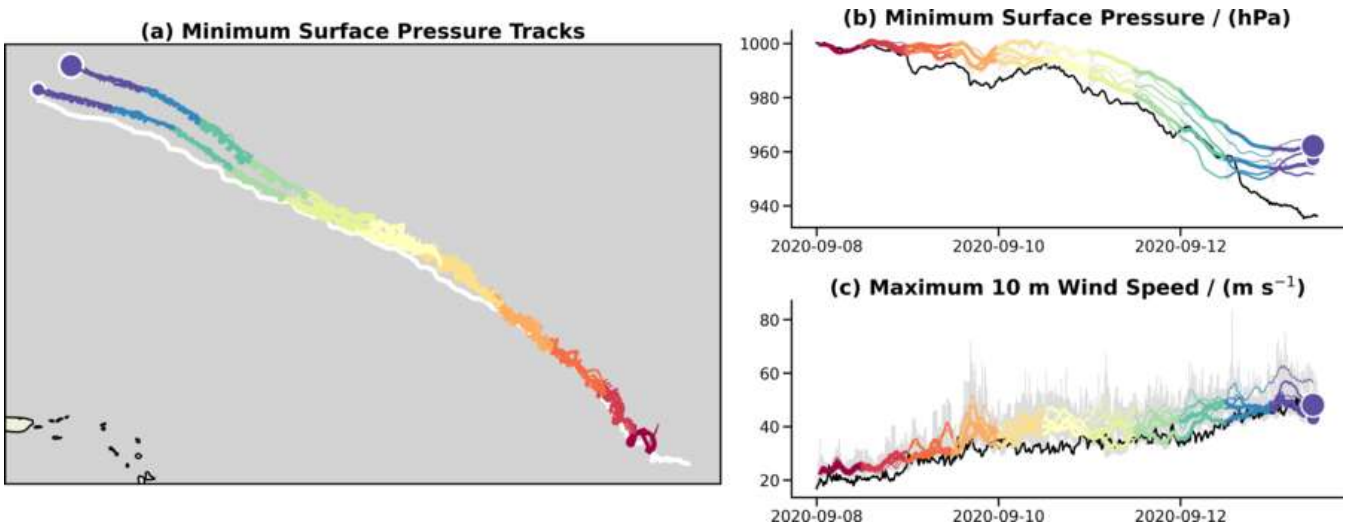

Figure 7. Simulated hurricane characteristics for grid configurations with differently chosen across-track widths. (a) The minimum pressure positions are plotted on a map, with colors representing the respective grid segments. The re-initialization time between successive segments is set to 12 hours. The narrow-size setup (small, solid circle at the end of the track) is compared to the medium-size setup (larger, solid circle at the end of the track). Different resolutions are plotted with different line thicknesses, ranging from thick for R2B11 to thin for R2B13, and they partially overlap. Minimum pressure positions from the base run are plotted in white. The minimum pressure and maximum 10-m wind speed are presented using similar color and line encoding as in (a). A running-average filter is applied to the curves in (b) and (c) to reduce noise. The original data is shown in light gray shading in the background. Data from the base run is shown in black.

3.4 Prototyping Lagrangian Analysis

The grid remains static during the simulation of a grid segment, and the output is only available for statically defined grids. Output variables are written directly to the ICON grid. Additionally, variables can be internally interpolated in ICON to a regular longitude-latitude grid and then output on this grid. This allows variables to be stored on slightly coarser grids at a higher temporal frequency. Our workflow toolkit includes an option to automatically calculate suitable longitude-latitude grids for the corresponding flexible grid segments. This creates a suitable longitude-latitude grid for each grid segment on which standard analyses can easily be performed. Figure 8 shows an example of such an output, displaying 2-meter temperature, 10-meter wind speed, reflected solar radiation flux at the top of the atmosphere (TOA), and total column water vapor for a given point in time. The sequence shows three nested grids for each variable, with grid spacings of 1200, 600, and 300 m. The utilization of these representations facilitates the comparison of the effects of varying resolutions on the fine structure of the simulated hurricane. However, Fig. ~~7~~ 8 also demonstrates that important parts of the hurricane may no longer be contained within the chosen domains, especially for the finest nest of 300 m grid spacing and rather long integration times of nearly five days. This illustrates the ~~compromises~~ trade-offs responsible scientists must ~~make~~ navigate between accuracy, resource requirements, and efficiency.

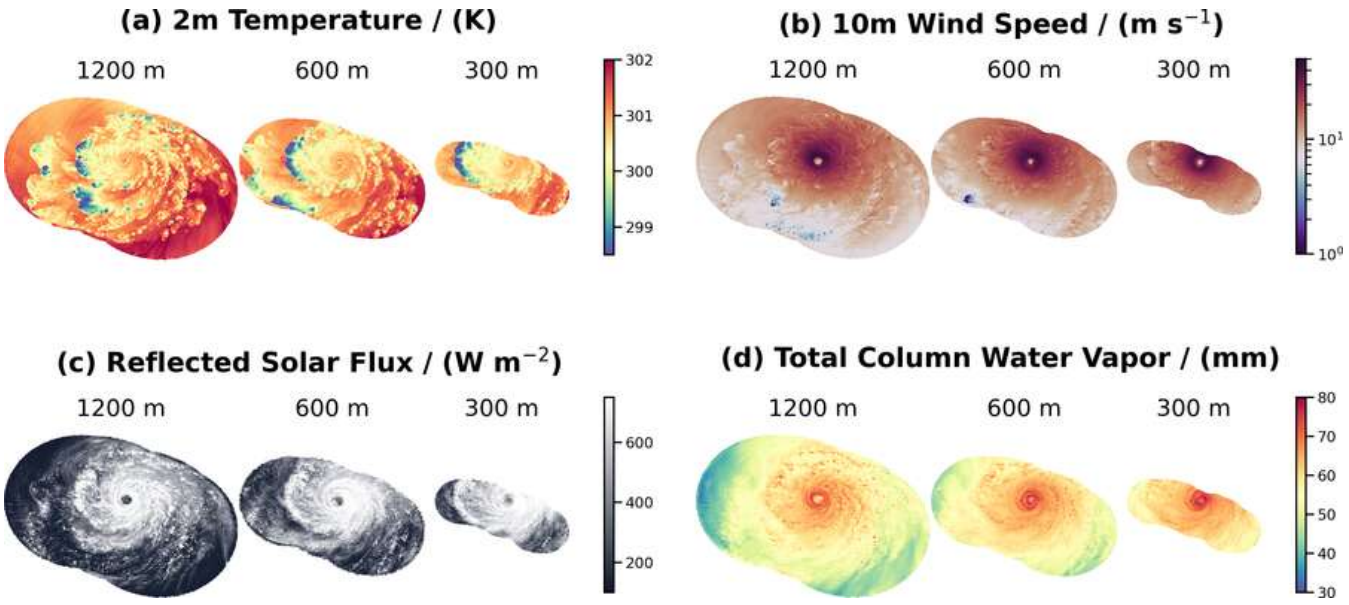

Figure 8. An overview of various variables, shown in the respective nests for a specific grid segment. The grid configuration was chosen so that the re-initialization time is 12 hours and the nests have an across-track width of 300, 200, and 100 km, respectively (narrow-size setup). The equivalent grid spacing is indicated above the respective nest. Variables shown are (a) 2 m temperature, (b) 10 m wind speed, (c) reflected solar radiation flux at TOA and (d) total column water vapor for segment 11 at 18 UTC on 12 September 2020.

The visualizations on the static grid above have the typical characteristics of an Eulerian analysis, in which the reference points do not change over time. However, our hurricane-centric simulations make it possible to change the perspective by defining only the circle that moves over the reference track as the analysis area (see segment concept in Sect. 2.2). This transformation into a Lagrangian reference system results in a smooth and consistent composition of the hurricane development despite the inherent discontinuity because of the separate grid segments. For example, the evolution of ice water path (IWP) is shown in Fig. 9. Most of the ice produced is located ~~northwest~~ northeast of the hurricane center. The Lagrangian analysis region is highlighted in color. The remaining parts of the segment, which are not used here, are shown in light colors. Interestingly, the Lagrangian approach results in a continuous and smooth path of the IWP. It gives the impression that the grid segments are sliding across the Lagrangian analysis area. Thus, data from different segments can be assembled to create a hurricane-centric analysis that consistently covers several days of hurricane development on a hectometer scale.

3.5 Performance and Shortcomings

The following discussion will cover the advantages and disadvantages of the method described so far. One significant advantage is that the target areas of the simulations are relatively small compared to the total region affected by the hurricane. This means that an acceptable number of calculation steps is being performed. Table 2 summarizes computational units (cu) as a measure of computational effort. Here, one cu refers to 1,000 time steps of fast physics (dynamical core is ~~adaptively~~ sub-stepped by a

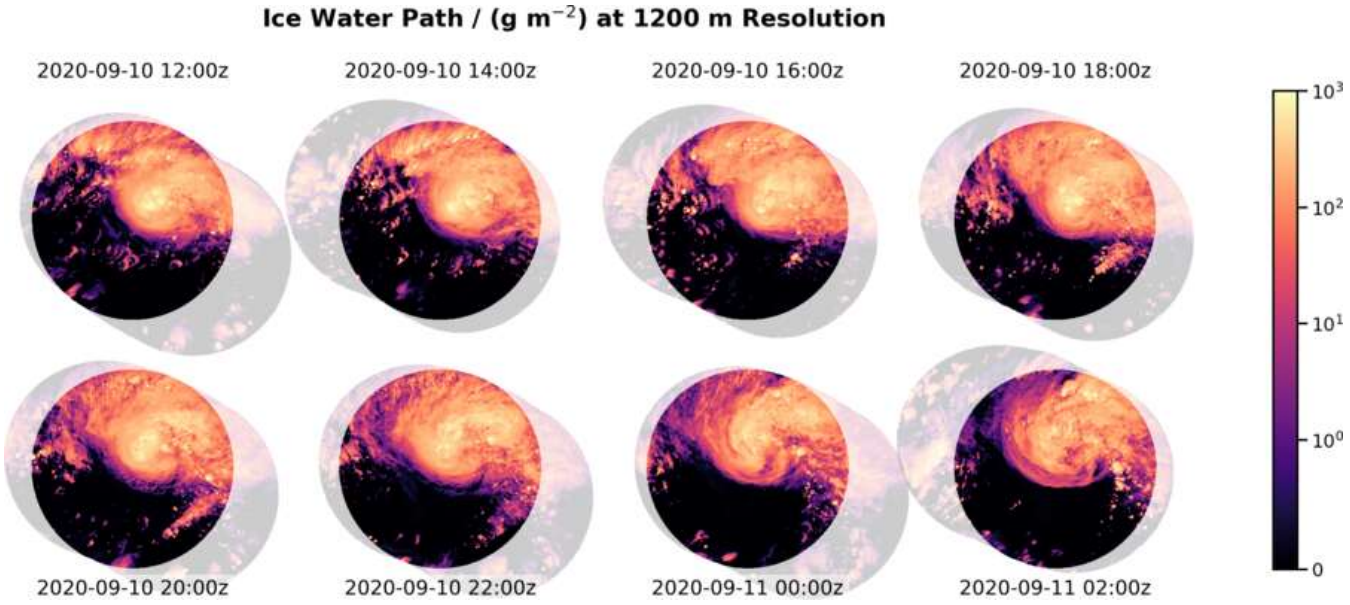

Figure 9. Ice water path is shown as a sequence in a Lagrangian framework. The boldly shaded ~~regions~~ areas highlight the circular Lagrangian analysis ~~area~~ region that follows the track of the reference hurricane. The remaining parts of the respective grid ~~segment~~ segments are indicated by very light colors. Two-hour steps are made in the temporal sequence. ~~Discontinuous jumps~~ Transitions between segments are visible between the first and second images, as well as between the second-to-last and last images~~. However, there is no discontinuity~~ appearing as shifts in the domain boundaries. Within the Lagrangian analysis region, however, the ice water path evolves smoothly and continuously. Here, the medium-size grid configuration was chosen so that the shown nest (DOM01, 1.2 km grid spacing) has 500 km across-track width with re-initialization time of 12 hours.

factor of 5~~to 7~~), which are used to integrate one million horizontal grid cells forward. For the narrow-size case, cu increases from 480 to 700 when transitioning from the 12h-reinit setup to the 24h-reinit setup, which corresponds to an increase of approximately 45%. For the medium-size case, 1,190 and 1,610 cu are required respectively, representing a factor of between 2.3 and 2.4 more resources than the narrow-size case, showing an increase of 35% between 12h-reinit and 24h-reinit just because of the larger domain size of the 24h-reinit segments. Therefore, the 12h-reinit setup is significantly more favorable in terms of resource requirements, and doubling the width of the setup also leads to approximately a doubling of resource requirements. The costs for expanding the setup to large and extra-large sizes are also shown. However, when comparing the numbers, it should be noted that only the outermost nest with a grid spacing of 1.2 km was considered for the larger setups and that the complete setup consisting of three respectively refined nests requires 20 to 25 times more resources.

~~The question, however, is whether do the savings through~~ To assess how the savings of our method compare to ~~those of~~ classic regional, limited-area applications, ~~in which the area must be large enough to encompass the entire duration of the hurricane. Taking the baseline run of our workflow as a reference, we find that it requires 6.3 million grid cells at a grid spacing of 2.5 km. A first refinement step (factor 4) with a simultaneous reduction of the target domain by 20 % would result~~

in approximately 20 million grid cells. Integrating this setup forward with a time step of six seconds for six days would result in 86, 400 time steps and a resource requirement of approximately 1,700 cu . Compared to large and extra-large setups, this would increase requirements by factors of 13 and eight, respectively(see two reference configurations are introduced in Tab. 2. The first is a tube-like setup that extends over the full hurricane track over the entire simulation period without splitting it into shorter segments. Such a tube shares the same cross-track geometry and nesting structure as our hurricane-centric segments, but is integrated as a single contiguous domain without re-initialization. The resource requirements for the tube, listed under the "6days" re-initialization interval, amount to approximately 2,640 and 5,440 cu for the narrow and medium three-nest configurations and 320 and 470 cu for the single-domain large and x-large setups, respectively. The second reference, labeled "classic" in Tab. 2). While using our methods here leads to a significant gain in efficiency, it must nevertheless be weighed against the disadvantages of the method. For further refinement steps, however,computational efficiency dramatically increases. Assuming further refinements to 600 and 300 m involving halving the grid spacing, doubling the number of time steps and reducing the domain by 20, represents a conventional fixed limited-area domain and is constructed as the smallest regular longitude-latitude bounding box that fully encloses the tube geometry for the respective cross-track width. For the single-domain case at 1.2 % (combined factor of 6.4), the costs would be approximately 11, 000 and 71, km grid spacing, the classic limited-area approach amounts to approximately 800 cu; a three-nest configuration requires around 20,000 cu, respectively. A nested setup of this magnitude would therefore have a total of around 84, 000 cu, which, cu. Compared to this classic reference, the tube configurations already yield notable speedup factors of 7.6, 3.7, 2.5, and 1.7 for the narrow, medium, large, and x-large setups, respectively. Dividing the tube into short, frequently re-initialized segments adds a further factor of approximately 2 to 6, depending on the selected setup, corresponds to an additional requirement of a factor of approximately 50 to 175 chosen across-track width and re-initialization interval. Together, the segment configurations with frequent re-initialization achieve the largest speedups: factors of up to 42 and 17 for the narrow and medium three-nest setups with 12-hour re-initialization intervals, and up to 6.2 and 3.8 for the single-domain large and x-large setups with 24-hour re-initialization intervals (see Tab. 2). Considering the risk of model errors or suboptimal settings requiring reruns of high-resolution simulations, the use of our flexible workflow appears well justified.

In addition to the hypothetical requirements, Tab. 2 also shows the actual costs of executing the respective setups. The 12h-reinit setup was calculated on the DKRZ Levante platform using an ICON build created with the Intel compiler without the thread-based parallelization via OpenMP. Production runs were performed on 64 nodes, generating costs of 2,900 and 6,300 node-hours for the narrow- and medium-size cases, respectively. This increase in costs by a factor of 2.2 roughly corresponds to the calculated increase in demand. The 24-hour reinit setup was calculated on the JSC JUWELS platform, which has fewer computing units than Levante, with 48 cores per node compared to 128. An ICON build was created and used on JUWELS with GCC, as well as without OpenMP. The narrow-size setup was calculated on JUWELS with 192 nodes, generating costs of 13,600 node-hours. The medium-size setup was calculated with 288 nodes and generated costs of 32,800 node-hours. Comparing the core hours values between Levante and JUWELS reveals that execution on JUWELS is 20–40 % less efficient. This difference is probably due to the different compilers and their respective optimization settings, rather than hardware differences.

Table 2. Comparison of compute requirement and compute cost parameters across re-initialization configurations and domain sizes. The reinit row "6days" indicates a nested tube configurations with effectively no re-initialisations. The term "classic" refers to a potential, fixed domain configuration ~~extending across the whole Atlantic ocean~~such that a corresponding tube configuration could be accommodated. Single domain configurations are marked by $*$, all other domain configurations consist of three online coupled nests. The computational units (cu) refer to the product of physics timesteps and number of horizontal grid cells. The number of nodes and node hours refers to DKRZ Levante nodes with 128 cores for the 12h-reinit setup and to JSC JUWELS nodes with 48 cores for the 24h-reinit setup.

property	reinit	narrow	medium	large	x-large	classic
total cu / 10^9	12h	480	1 190	-	-	-
	24h	700	1 610	130*	210*	-
	6days	2 640	5 440	320*	470*	-
	-	-	-	-	~~1 700~~	800*, ~~84~~20 000
potential speedup	-	~~120~~ 7.6 - ~~175~~42	~~50~~ 3.7 - ~~70~~17	~~13~~2.5 - 6.2*	~~8~~1.7 - 3.8*	1
total node hours (# nodes)	12h	2 900 (64)	6 300 (64)	-	-	-
	24h	13 600 (192)	32 800 (288)	3 100* (64)	4 400* (64)	-
total core hours / 10^6	12h	0.37	0.81	-	-	-
	24h	0.65	1.57	0.15*	0.21*	-

The immense increases in efficiency of our hurricane-centric setup however come at a price. Due to the limited size of the regional setups, the hurricane under consideration cannot develop completely freely and always remains influenced by the forcing at the lateral boundaries, which is stronger for a more narrow setup. In addition, the re-initialization procedure during the transition between segments introduces further challenges. The mix of ICs detailed in Section 2.3, which consist of different data sources inside and outside the overlap region of two subsequent segments, creates a discontinuity. This is smoothed out in the transition using a distance-weighted average. Despite this, ~~spin-up~~ the merged ICs can excite spurious sound and gravity waves, and re-initialization effects occur at the beginning of a simulation in a new segment. ~~To evaluate~~ The following section evaluates the impact of these effects on simulation quality~~, we will now analyze the temporal evolution and statistical distributions of key atmospheric variables during the first three hours of each simulation segment . This analysis will allow us to quantify the magnitude of the spin-up effects, understand how they vary systematically across different domain configurations, and determine if they remain small enough to justify the computational efficiency gains of our approach.~~

~~To better understand the spin-up process through our~~

3.6 Analysis of Re-initialization Effects

To investigate the re-initialization ~~strategy~~behavior in a controlled manner, we conducted sensitivity experiments following the same grid configurations as listed in Table 1, but with the runtime of each segment extended by three hours. This creates a

three-hour temporal overlap between consecutive segments, enabling a direct comparison of the atmospheric state from two successive segments at the same point in time and thus providing a pathway for directly quantifying re-initialization effects. To keep the analysis tractable, all sensitivity experiments were performed exclusively for the outermost nest (DOM01) at 1.2 km grid spacing. In addition, two parameter settings are compared throughout: the production setting with `init_latbc_from_fg=.FALSE` (labeled "LBC flag OFF") and an updated setting with `init_latbc_from_fg=.TRUE.` (labeled "LBC flag ON"). The latter has been recommended by the ICON community since early 2026 to reduce model biases arising from mismatches between the initial lateral boundary conditions and the initial model state (see ICON Documentation, 2026).

In the following, we analyze the time series of thermodynamic and water-related atmospheric parameters for the first three hours in Fig. 10. There, the median values in the circular Lagrangian analysis regions (see Sect. 3.4) of selected surface and cloud variables are presented. For each variable, ~~its initial value was subtracted, such that all curves start at zero and any deviation can be analyzed easily. The time series allow to relate the initial development of different segmentsto each other. Systematically similar behavior can be attributed in part to spin-up effects.~~ the instantaneous difference between segment k and the corresponding values from segment $k-1$ extended by 3 h is shown, allowing a direct quantification of re-initialization-induced deviations between consecutive segments. The figure was created for the ~~outer nest of the~~ medium-size, 12h-reinit ~~setup and shows the relative progression of the different grid segments with different line colors shifting from blue to red with an initialization time at September 8, 2020 at 0 UTC and subsequent re-initializations each 12 hours~~DOM01 setup with 1.2 km grid spacing (see Tab. 1) . Light gray curves show the individual segment differences, and the thick solid (dashed) black curve gives their mean for the LBC flag OFF (ON) setting. The surface pressure in Fig. 10a shows a negative anomaly ~~for grid segment 4 and larger~~ with peak values around ~~-1.3~~ -1.2 hPa after about 30 min of integration time ~~. This anomaly appears to have decayed~~ for the LBC flag OFF setting. An opposing positive but slightly weaker anomaly of around 0.9 hPa is visible for the LBC flag ON setting, with the peak shifted to around 70 min. Both differences appear to have converged after approximately 2 hours~~. The spread of all surface pressure curves after 3 hours is approximately 2 hPa, which is about twice as large as the spin-up effects at approximately 30 min~~, by which point they have decayed to near-zero values. The 2-m temperature curve in Fig. 10b also ~~appears to show~~ shows a systematic negative peak of around ~~-0.08~~ -0.06 K at 30 min ~~, and the~~ for LBC flag OFF, and a positive peak of around 0.05 K at 75 min for LBC flag ON. Temperature differences likewise vanish after approximately 2 hours. The curves for 2-m specific humidity ~~for the later segments peaks~~ show positive differences for both LBC flag settings, with peak values of 0.07 g kg^{-1} at 45 min ~~with a value of approximately 0.1 g~~for LBC flag OFF and 0.05 g kg^{-1} ~~. The~~ at 120 min for LBC flag ON. The specific humidity difference for the LBC flag ON setting converges more slowly than for LBC flag OFF, with near-zero values not yet reached by 3 hours. The individual curves for total column water vapor (TCWV), liquid and ice water path (LWP, IWP) in Figs. 10d-f show considerable variability ~~at 3 hours, and systematic spin-up effects are hardly noticeable in comparison to this variability. Qualitatively, slightly enhanced values are visible for latter variables. It can therefore be concluded that the spin-up described here is of a more dynamic nature~~over the three-hour window. The LBC flag ON setting tends to produce slightly larger mean peak values for TCWV and IWP, and smaller ones for LWP, and its peak consistently occurs later than for LBC flag OFF. Overall, the LBC flag ON setting tends to produce smaller peak differences for surface pressure and near-surface temperature, suggesting it may be the preferable choice

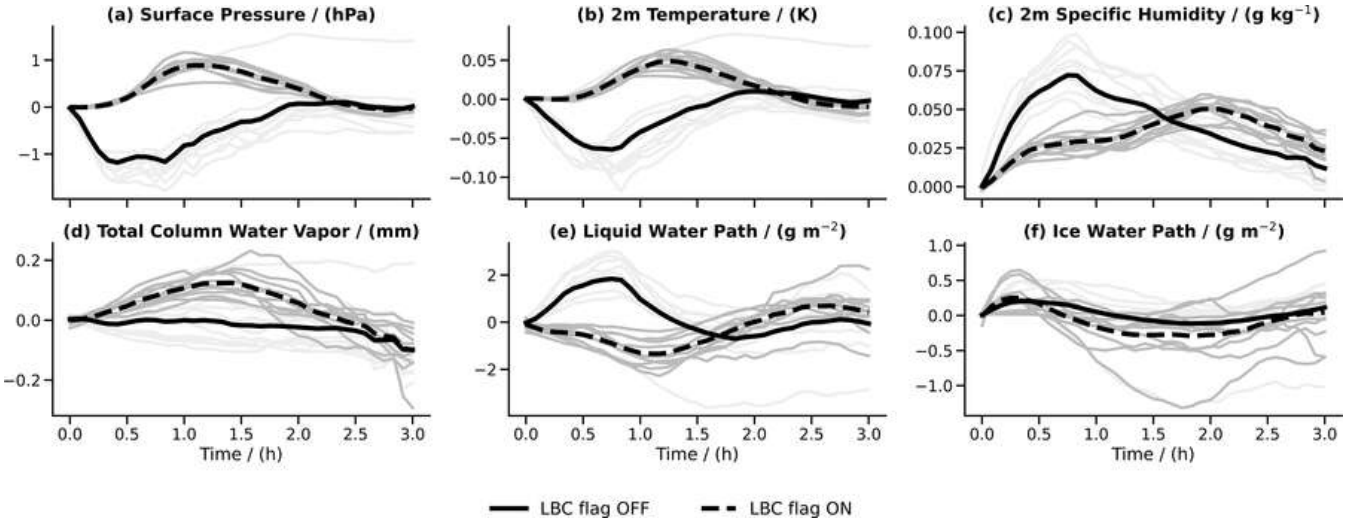

Figure 10. Temporal evolution of ~~the respective~~ median values of surface and integrated cloud variables evaluated in the circular Lagrangian analysis region. The ~~respective colored~~ curves show the ~~development of~~ instantaneous difference between values from segment k and the corresponding values from segment $k-1$ extended by 3 hours, for (a) surface pressure, (b) 2 m temperature, (c) 2 m specific humidity, (d) total column water vapor, (e) liquid water path, and (f) ice water path. Results are shown for the ~~grid segments of the~~ medium setup with 12h-reinit. ~~For all~~ Gray curves show the evolution for individual segments, while the ~~anomaly relative to~~ thick black curve shows the ~~respective values at the segment init is shown~~mean across all segments. ~~The gray vertical lines mark zero and two hours~~ Two parameter settings are compared: the experiment with the LBC flag deactivated (`init_latbc_from_fg=.FALSE.`), thin ~~dashed~~light gray and solid black lines) and ~~0.5 hours~~ the experiment with the LBC flag activated (~~thick solid~~`init_latbc_from_fg=.TRUE.`)~~after init~~, ~~which are used for subsequent analyses~~thin gray and dashed black lines).

for future applications, though its re-initialization effects appear to persist somewhat longer legitimating our choice of the LBC flag OFF setting for the production runs in this study.

Figure 11 provides a comprehensive statistical comparison of ~~spin-up~~ re-initialization effects across different configurations. ~~The analysis follows a three-step process: First, nine percentiles (10-90~~As in Fig. 10, the analysis evaluates percentiles from segment k, here taken 30 minutes after initialization, and segment $k-1$ extended correspondingly. Nine percentiles (10–90% in 10% steps) ~~were computed for selected variables, as their relative behavior showed similar patterns. Second, anomalies at 30 minutes were calculated relative to a reference value (the mean of values at 0 and 2 hours). Finally, histograms were created separatelyfor each grid segment and then stacked~~ of each field are computed across the Lagrangian analysis region separately, and their differences are taken. These per-segment differences are visualized as stacked histograms, with color coding by segment to enable direct comparison between setups.

The statistical results reveal distinct patterns for different atmospheric variables. ~~Spin-up effects due to re-initialization~~ Re-initialization effects are particularly visible when the ~~distribution of values for the initial segments (blue colors) deviates from the distribution of the later grid segments (red colors)~~values in the distributions significantly deviate from zero. For

surface pressure, this pattern is clearly evident in Figs. 11a and 11e for the narrow setup. Surface pressure values ~~for the initial grid segment scatter around zero, while for later segments, a distribution shifted into the negative range emerges~~scatter between -1 and -2 hPa. Comparing the effects of re-initialization time shows that longer intervals lead to slightly larger ~~spin-up~~ re-initialization effects in surface pressure (comparing Figs. 11a,e for narrow setup and Figs. 11i,m for medium setup). Domain size also influences these effects: larger domains (Figs. 11~~m,~~q,u) show considerably less deviation from zero at 30 minutes, indicating that systematic ~~spin-up~~ re-initialization effects are less pronounced for larger domains. For temperature, ~~domain size effects are also~~ similar effects are clearly visible. The ~~mode of the distribution shifts from -0.1 K in the narrow setup (Fig. 11b) to -0.07 K in the medium setup(Fig. 11j).~~ value appear to be distributed to slightly lower magnitudes for the 12h-reinit setup than for the 24h-reinit setup. For large and x-large setups, temperature anomalies ~~for later segments~~ scatter around zero ~~. The initial spin-up (blue bars) shows positive anomalies between 0.05 and 0.1 K.~~ for all segments. Cloud variables such as LWP and IWP show significant spread in the narrow setup (Figs. 11c,d,g,h). As domain size increases, this variability decreases, likely due to more robust spatial statistics. The considerable variability in cloud variables masks potential ~~spin-up~~ re-initialization effects. Overall, while our hurricane-centric workflow introduces uncertainties through ~~spin-up~~ re-initialization effects, these remain sufficiently small and are outweighed by the substantial efficiency gains achieved.

4 Conclusions

This study presents a flexible workflow for the atmospheric model ICON that enables efficient high-resolution simulations in a hurricane-centric reference framework through a segment-based grid and pre-processing approach. The methodology demonstrates the general feasibility of conducting hectometer-resolution simulations of large, organized atmospheric phenomena like hurricanes in the tropics. Instead of investing in fixed and spatially extended regional domains throughout the entire simulation period, our flexible method follows the motion of a hurricane and generates tailored grid segments.

The technical implementation successfully automates key components of the workflow, including hurricane tracking, flexible grid generation, and initial condition preparation and merging across consecutive segments. The segment concept, which divides hurricane tracks into overlapping temporal windows of 12-24 hours, provides a practical framework for efficient utilization of computational resources. The workflow toolkit demonstrates portability across different HPC platforms, having been successfully deployed on both DKRZ Levante and JSC JUWELS systems with appropriate platform-specific configurations.

The application of the workflow is demonstrated for Hurricane Paulette (2020). High-resolution simulations with grid spacings down to 300 m were performed using different segment configurations. Variations in re-initialization intervals of 12h and 24h are explored that determine the respective along-track lengths of the segments. Additionally, different overall domain sizes are investigated by varying the across-track widths with tested configurations being rather narrow, medium-sized, large and extra large. The results indicate that the hurricane track remains consistent with the base run and mainly depend on the across-track size of the chosen configuration, while intensity metrics such as minimum pressure and maximum wind speed exhibit sensitivity to resolution in our multi-nested setups. A Lagrangian framework enables a hurricane-centric view, facilitating continuous analysis of variables like liquid and ice water path across the segmented simulations.

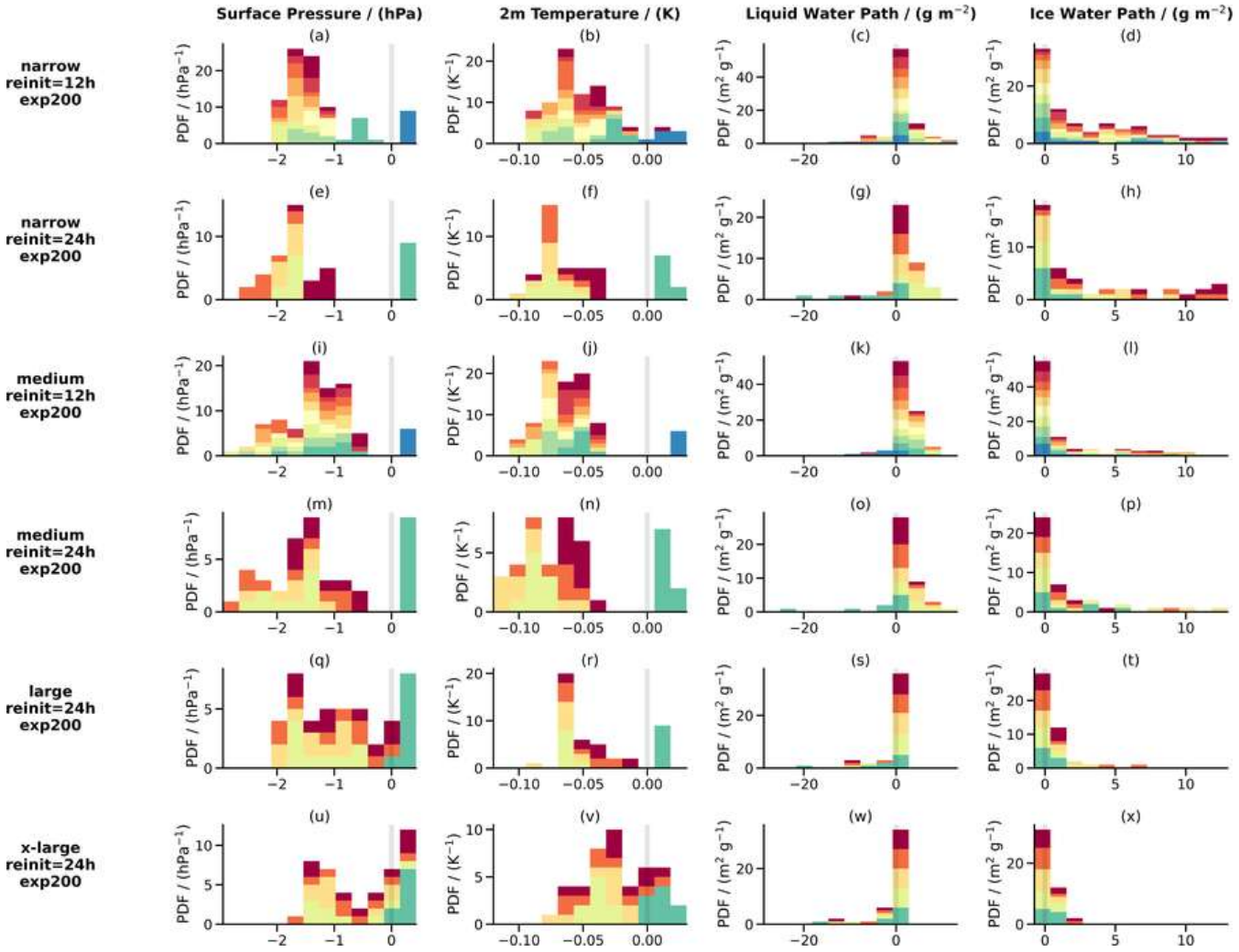

Figure 11. Statistical summary of potential ~~spin-up~~ re-initialization effects plotted as colored bars stacked on top of each other for each segment. ~~For~~ The color coding is the ~~time of 0.5 hours~~ same as in Fig. 6. At 30 minutes after ~~the respective init~~initialization of segment k, nine ~~different~~ percentiles ~~in the range between 10 and 90%~~ (10–90% in ~~steps of~~ 10% steps) ~~were calculated in~~ of each field are computed across the Lagrangian analysis region separately for segment k and ~~converted into anomalies relative to a reference value. The reference value was chosen as the average between the values at times 0~~ segment $k-1$ extended by 3 h, and ~~2 hours~~their pairwise differences are taken. ~~All values obtained in this way were~~ These per-segment differences are statistically summarized as probability density functions~~and plotted as colored bars stacked on top of each other for each segment~~. The six rows show the respective setups, with the narrow and medium setups shown for re-initialization periods of 12 h and 24 h, respectively. The columns show surface pressure, 2 m temperature, liquid water path, and ice water path. The ~~spin-up~~ re-initialization analyses are only shown for the outer nest with a grid spacing of approximately 1 km.

The increase in computational efficiency is substantial compared to traditional fixed-domain approaches. ~~The hurricane-centric method~~ A track-following tube geometry alone reduces resource requirements by factors of ~~13–175 depending on domain~~

~~configuration, with the narrow setup giving the largest gain in computational efficiency~~1.7–7.6 relative to an appropriately chosen longitude-latitude bounding box around the hurricane. Segmenting the tube into short, frequently re-initialized intervals adds a further factor of 2–6, yielding total speedups of up to 42 for the narrowest three-nest configuration. The setup with 12-hour re-initialization intervals proves to be more efficient than 24-hour intervals, requiring 35-45 % fewer computational resources while maintaining comparable simulation quality. Domain size scaling shows predictable resource scaling, with the twice as large medium-width configurations requiring approximately 2.3 times the resources of narrow configurations. However, the approach has notable limitations when the across-track width is set too narrowly, the domain may fail to encompass the entire phenomenon, a shortcoming that becomes especially evident for our narrow-size configurations. The analysis of ~~spin-up~~ re-initialization effects reveals systematic but manageable impacts during segment transitions, with surface pressure showing negative anomalies on the order of 1 hPa that decay within 2 hours after re-initialization. The ~~spin-up~~ re-initialization effects are more pronounced in narrower domain configurations but remain sufficiently small to justify the substantial computational efficiency gains achieved by the hurricane-centric approach.

Several promising directions for methodological enhancement emerge from this work. On the technical side, extending the workflow to support more than three nesting levels would provide greater refinement flexibility and enable even finer resolution simulations. Similarly, offering greater flexibility in selecting the grid from which the refinement cycle is initiated (currently the grid of the base run) would allow for more diverse and independent nesting hierarchies. A further interesting area for development lies in improving the re-initialization procedure itself. Currently, the distance-weighted blending approach, while effective, could potentially be enhanced by integrating techniques from data assimilation, such as the Incremental Analysis Update (IAU) method available in the numerical weather prediction configuration of the ICON model. This approach might have the potential to substantially reduce spurious artifacts arising from the merging of initial conditions between consecutive segments. Additionally, future work could also investigate how the method performs over land, where re-initialization of subgrid-scale heterogeneities in land surface properties can be challenging and may introduce additional difficulties. Importantly, all such technical improvements must maintain spatial and structural consistency of simulated core variables across segment transitions such that Lagrangian analyses remain continuous and meaningful.

Beyond technical refinements, the computational efficiency gains achieved here enable new scientific applications. The method now makes high-resolution hectometer-scale simulations of additional hurricanes computationally feasible, opening opportunities for detailed dynamical analysis and process studies of a broader range of tropical storms. The reduced computational cost further enables simulation ensembles to be conducted in scientific studies, providing higher confidence in identifying systematic effects. Particularly noteworthy is the potential for perturbed parameter ensembles that would deepen our understanding of how model parameterizations influence the representation of convective organization, eyewall structure, and other small-scale phenomena in tropical cyclones. These applications demonstrate how advances in efficient simulation frameworks can open new avenues for investigating the cloud physics and dynamics of extreme atmospheric phenomena.

Code and data availability. The current version of the hurricane-centric workflow toolkit is available under https://doi.org/10.5281/zenodo.18271898 (Senf, 2026c). It depends on cdo (see Schulzweida, 2023), the DWD ICON tools available at https://gitlab.dkrz.de/dwd-sw/dwd_icon_tools (last accessed 27 January 2026, restricted access) and the extpar utility available at https://github.com/C2SM/extpar (last accessed 27 January 2026, toolkit tested with version v5.13). ICON open source release v2025.04 is published at https://doi.org/10.35089/wdcc/iconrelease2025.04 and a patch to enable consistent re-initialization of complete cloud microphysics is available under https://doi.org/10.5281/zenodo.18387047 (Senf, 2026d). ICON model v2.6.6 has restricted access and a patch to enable variable SSTs in a multiple nest setup is available under https://doi.org/10.5281/zenodo.18385022 (Senf, 2026a). Initial, boundary and grid data from the base run used to run the hurricane-centric workflow are available under https://www.wdc-climate.de/ui/entry?acronym=DKRZ_LTA_1376_dsg0001 (Senf, 2025). Plots and analysis were created mainly using Jupyter notebooks available under https://doi.org/10.5281/zenodo.19002039 (Senf, 2026b).

Author contributions. FS lead the development of the workflow and the drafting of the manuscript. RC contributed to the writing of the manuscript and the interpretation of the results.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The development was supported by the IFCES2 project funded by BMFTR (formerly known as BMBF - Federal Ministry of Education and Research) under grant number 16ME0689K as part of the SCALEXA funding initiative, which in turn was funded by the European Union - NextGenerationEU. Fabian Senf acknowledges funding from the CleanCloud project (grant: 101137639). We would like to thank Manoel Römmer for his comments and help with the workflow development. We are particularly grateful to Bernd Heinold, Ina Tegen, Jason Müller, ~~and~~ Jan Kretzschmar, Nadja Omanovic, and two anonymous reviewers whose valuable comments contributed significantly to improving the manuscript. We also thank the ICON consortium for their continuous support and engagement in pushing ICON development to the next level. We gratefully acknowledge the computing time and storage resources granted on the supercomputer JUWELS at Jülich Supercomputing Centre (JSC) under project IFCES2-SCALEXA and on the supercomputer Levante at the German Climate Computing Centre (DKRZ) under project bb1376.

We would like to mention the use of DeepL to improve the grammatical and language quality of the manuscript and the use of github copilot for AI-assisted coding and debugging of the workflow toolkit.

References

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities, Mon. Wea. Rev., 139, 3887–3905, https://doi.org/10.1175/mwr-d-10-05013.1, 2011.

Bechtold, P., Köhler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M. J., Vitart, F., and Balsamo, G.: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales, Q. J. R. Meteorolog. Soc., 134, 1337–1351, https://doi.org/10.1002/qj.289, 2008.

Bender, M. A. and Ginis, I.: Real-Case Simulations of Hurricane–Ocean Interaction Using A High-Resolution Coupled Model: Effects on Hurricane Intensity, Mon. Weather Rev., 128, 917 – 946, https://doi.org/10.1175/1520-0493(2000)128<0917:RCSOHO>2.0.CO;2, 2000.

Davis, C., Wang, W., Chen, S. S., Chen, Y., Corbosiero, K., DeMaria, M., Dudhia, J., Holland, G., Klemp, J., Michalakes, J., Reeves, H., Rotunno, R., Snyder, C., and Xiao, Q.: Prediction of Landfalling Hurricanes with the Advanced Hurricane WRF Model, Mon. Weather Rev., 136, 1990 – 2005, https://doi.org/10.1175/2007MWR2085.1, 2008.

Davis, C., Wang, W., Cavallo, S., Done, J., Dudhia, J., Fredrick, S., Michalakes, J., Caldwell, G., Engel, T., and Torn, R.: High-Resolution Hurricane Forecasts, Comput. Sci. Eng., 13, 22–30, https://doi.org/10.1109/MCSE.2010.74, 2011.

Dipankar, A., Stevens, B., Heinze, R., Moseley, C., Zängl, G., Giorgetta, M., and Brdar, S.: Large eddy simulation using the general circulation model ICON, J. Adv. Model. Earth Syst., 7, 963–986, https://doi.org/10.1002/2015MS000431, 2015.

Dong, J., Liu, B., Zhang, Z., Wang, W., Mehra, A., Hazelton, A. T., Winterbottom, H. R., Zhu, L., Wu, K., Zhang, C., Tallapragada, V., Zhang, X., Gopalakrishnan, S., and Marks, F.: The Evaluation of Real-Time Hurricane Analysis and Forecast System (HAFS) Stand-Alone Regional (SAR) Model Performance for the 2019 Atlantic Hurricane Season, Atmosphere, 11, https://doi.org/10.3390/atmos11060617, 2020.

Enz, B. M., Engelmann, J. P., and Lohmann, U.: Use of threshold parameter variation for tropical cyclone tracking, Geosci. Model Dev., 16, 5093–5112, https://doi.org/10.5194/gmd-16-5093-2023, 2023.

Gao, K., Harris, L., Bender, M., Chen, J.-H., Zhou, L., and Knutson, T.: Regulating Fine-Scale Resolved Convection in High-Resolution Models for Better Hurricane Track Prediction, Geophys. Res. Lett., 50, e2023GL103 329, https://doi.org/https://doi.org/10.1029/2023GL103329, e2023GL103329 2023GL103329, 2023.

Gopalakrishnan, S. G., Bacon, D. P., Ahmad, N. N., Boybeyi, Z., Dunn, T. J., Hall, M. S., Jin, Y., Lee, P. C. S., Mays, D. E., Madala, R. V., Sarma, A., Turner, M. D., and Wait, T. R.: An Operational Multiscale Hurricane Forecasting System, Mon. Weather Rev., 130, 1830 – 1847, https://doi.org/10.1175/1520-0493(2002)130<1830:AOMHFS>2.0.CO;2, 2002.

Gutmann, E. D., Rasmussen, R. M., Liu, C., Ikeda, K., Bruyere, C. L., Done, J. M., Garrè, L., Friis-Hansen, P., and Veldore, V.: Changes in Hurricanes from a 13-Yr Convection-Permitting Pseudo–Global Warming Simulation, J. Climate, 31, 3643 – 3657, https://doi.org/10.1175/JCLI-D-17-0391.1, 2018.

Heikenfeld, M., Marinescu, P. J., Christensen, M., Watson-Parris, D., Senf, F., van den Heever, S. C., and Stier, P.: tobac 1.2: towards a flexible framework for tracking and analysis of clouds in diverse datasets, Geosci. Model Dev., 12, 4551–4570, https://doi.org/10.5194/gmd-12-4551-2019, 2019.

Heinze, R., Dipankar, A., Henken, C. C., Moseley, C., Sourdeval, O., Trömel, S., Xie, X., Adamidis, P., Ament, F., Baars, H., Barthlott, C., Behrendt, A., Blahak, U., Bley, S., Brdar, S., Brueck, M., Crewell, S., Deneke, H., Girolamo, P. D., Evaristo, R., Fischer, J., Frank, C., Friederichs, P., Göcke, T., Gorges, K., Hande, L., Hanke, M., Hansen, A., Hege, H.-C., Hoose, C., Jahns, T., Kalthoff, N., Klocke,

D., Kneifel, S., Knippertz, P., Kuhn, A., van Laar, T., Macke, A., Maurer, V., Mayer, B., Meyer, C. I., Muppa, S. K., Neggers, R. A. J., Orlandi, E., Pantillon, F., Pospichal, B., Röber, N., Scheck, L., Seifert, A., Seifert, P., Senf, F., Siligam, P., Simmer, C., Steinke, S., Stevens, B., Wapler, K., Weniger, M., Wulfmeyer, V., Zängl, G., Zhang, D., and Quaas, J.: Large-eddy simulations over Germany using ICON: a comprehensive evaluation, Quart. J. Roy. Meteor. Soc., 143, 69–100, https://doi.org/10.1002/qj.2947, 2017.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. R. Meteorolog. Soc., 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Hogan, R. J. and Bozzo, A.: A Flexible and Efficient Radiation Scheme for the ECMWF Model, J. Adv. Model. Earth Syst., 10, 1990–2008, https://doi.org/https://doi.org/10.1029/2018MS001364, 2018.

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., Giorgetta, M., Gutjahr, O., Haak, H., Hanke, M., Ilyina, T., Jahns, T., Jungclaus, J., Kern, M., Klocke, D., Kluft, L., Kölling, T., Kornblueh, L., Kosukhin, S., Kroll, C., Lee, J., Mauritsen, T., Mehlmann, C., Mieslinger, T., Naumann, A. K., Paccini, L., Peinado, A., Praturi, D. S., Putrasahan, D., Rast, S., Riddick, T., Roeber, N., Schmidt, H., Schulzweida, U., Schütte, F., Segura, H., Shevchenko, R., Singh, V., Specht, M., Stephan, C. C., von Storch, J.-S., Vogel, R., Wengel, C., Winkler, M., Ziemen, F., Marotzke, J., and Stevens, B.: ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales, Geosci. Model Dev., 16, 779–811, https://doi.org/10.5194/gmd-16-779-2023, 2023.

ICON Documentation: Build and Run: Input Data — Lateral Boundary Conditions, https://docs.icon-model.org/documentation/buildrun/buildrun_input_data.html, last accessed: 1 July 2026, 2026.

ICON partnership: ICON release 2025.04, https://doi.org/10.35089/wdcc/iconrelease2025.04, dWD; MPI-M; DKRZ; KIT; C2SM, 2025.

Kleppek, S., Muccione, V., Raible, C. C., Bresch, D. N., Koellner-Heck, P., and Stocker, T. F.: Tropical cyclones in ERA-40: A detection and tracking method, Geophys. Res. Lett., 35, https://doi.org/https://doi.org/10.1029/2008GL033880, 2008.

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., Mazumder, P., Copik, M., Weber, B., Herten, A., Bockelmann, H., Hoefler, T., Hohenegger, C., and Stevens, B.: Computing the Full Earth System at 1km Resolution, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC '25, p. 125–136, Association for Computing Machinery, New York, NY, USA, ISBN 9798400714665, https://doi.org/10.1145/3712285.3771789, 2025.

Kurihara, Y., Tuleya, R. E., and Bender, M. A.: The GFDL Hurricane Prediction System and Its Performance in the 1995 Hurricane Season, Mon. Weather Rev., 126, 1306 – 1322, https://doi.org/10.1175/1520-0493(1998)126<1306:TGHPSA>2.0.CO;2, 1998.

Landsea, C. W. and Franklin, J. L.: Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format, Mon. Weather Rev., 141, 3576 – 3592, https://doi.org/10.1175/MWR-D-12-00254.1, 2013.

Mauritsen, T., Redler, R., Esch, M., Stevens, B., Hohenegger, C., Klocke, D., Brokopf, R., Haak, H., Linardakis, L., Röber, N., et al.: Early development and tuning of a global coupled cloud resolving model, and its fast response to increasing CO2, Tellus Series A: Dynamic Meteorology and Oceanography, 74, 346–363, https://doi.org/https://doi.org/10.16993/tellusa.54, 2022.

Müller, W. A., Früh, B., Korn, P., Potthast, R., Baehr, J., Bettems, J.-M., Bölöni, G., Brienen, S., Fröhlich, K., Helmert, J., Jungclaus, J., Köhler, M., Lorenz, S., Schneidereit, A., Schnur, R., Schulz, J.-P., Schlemmer, L., Sgoff, C., Pham, T. V., Pohlmann, H., Vogel, B., Vogel, H.,

Wirth, R., Zaehle, S., Zängl, G., Stevens, B., and Marotzke, J.: ICON: Toward Vertically Integrated Model Configurations for Numerical Weather Prediction, Climate Predictions, and Projections, Bull. Am. Meteorol. Soc., 106, E1017 – E1031, https://doi.org/10.1175/BAMS-D-24-0042.1, 2025.

National Hurricane Center: Tropical Cyclone Report: Hurricane Paulette (AL172020), Tech. rep., National Hurricane Center, Miami, FL, USA, https://www.nhc.noaa.gov/data/tcr/AL172020_Paulette.pdf, last access: 5 February 2026, 2021.

Satoh, M., Noda, A. T., Seiki, T., Chen, Y.-W., Kodama, C., Yamada, Y., Kuba, N., and Sato, Y.: Toward reduction of the uncertainties in climate sensitivity due to cloud processes using a global non-hydrostatic atmospheric model, Prog. Earth Planet. Sci., 5, 67, https://doi.org/10.1186/s40645-018-0226-1, 2018.

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., and Düben, P.: Global Cloud-Resolving Models, Curr. Clim. Change Rep., https://doi.org/10.1007/s40641-019-00131-0, 2019.

Schulzweida, U.: CDO User Guide, https://doi.org/10.5281/zenodo.10020800, 2023.

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., Brunner, L., Cheedela, S. K., Das, S., Denissen, J., Dragaud, I., Dziekan, P., Ekblom, M., Engels, J. F., Esch, M., Forbes, R., Frauen, C., Freischem, L., García-Maroto, D., Geier, P., Gierz, P., González-Cervera, A., Grayson, K., Griffith, M., Gutjahr, O., Haak, H., Hadade, I., Haslehner, K., ul Hasson, S., Hegewald, J., Kluft, L., Koldunov, A., Koldunov, N., Kölling, T., Koseki, S., Kosukhin, S., Kousal, J., Kuma, P., Kumar, A. U., Li, R., Maury, N., Meindl, M., Milinski, S., Mogensen, K., Niraula, B., Nowak, J., Praturi, D. S., Proske, U., Putrasahan, D., Redler, R., Santuy, D., Sármány, D., Schnur, R., Scholz, P., Sidorenko, D., Spät, D., Sützl, B., Takasuka, D., Tompkins, A., Uribe, A., Valentini, M., Veerman, M., Voigt, A., Warnau, S., Wachsmann, F., Wacławczyk, M., Wedi, N., Wieners, K.-H., Wille, J., Winkler, M., Wu, Y., Ziemen, F., Zimmermann, J., Bender, F. A.-M., Bojovic, D., Bony, S., Bordoni, S., Brehmer, P., Dengler, M., Dutra, E., Faye, S., Fischer, E., van Heerwaarden, C., Hohenegger, C., Järvinen, H., Jochum, M., Jung, T., Jungclaus, J. H., Keenlyside, N. S., Klocke, D., Konow, H., Klose, M., Malinowski, S., Martius, O., Mauritsen, T., Mellado, J. P., Mieslinger, T., Mohino, E., Pawłowska, H., Peters-von Gehlen, K., Sarré, A., Sobhani, P., Stier, P., Tuppi, L., Vidale, P. L., Sandu, I., and Stevens, B.: nextGEMS: entering the era of kilometer-scale Earth system modeling, Geosci. Model Dev., 18, 7735–7761, https://doi.org/10.5194/gmd-18-7735-2025, 2025.

Seifert, A. and Beheng, K. D.: A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description, Meteor. Atmos. Phys., 92, 45–66, https://doi.org/https://doi.org/10.1007/s00703-005-0112-4, 2006.

Senf, F.: Data for Nested Hurricane-Centric ICON Runs for Hurricane Paulette (2020), https://www.wdc-climate.de/ui/entry?acronym=DKRZ_LTA_1376_dsg0001, 2025.

Senf, F.: ICON v2.6.6 Patch: for flexible SSTs in nested domains, https://doi.org/10.5281/zenodo.18385022, 2026a.

Senf, F.: Jupyter Notebooks for Plotting Analysis of "Advancing Hurricane-Centric Simulations with ICON", Submission Release, https://doi.org/10.5281/zenodo.19002039, 2026b.

Senf, F.: Hurricane Centric Toolset Version v2026.01 - Common Initialization and Perturbation Experiment Capabilities, https://doi.org/10.5281/zenodo.18271898, 2026c.

Senf, F.: ICON v2025-04-2 Patch: for handling vertical interpolation in init of two-moment cloud microphysics, https://doi.org/10.5281/zenodo.18387047, 2026d.

Senf, F., Voigt, A., Clerbaux, N., Hünerbein, A., and Deneke, H.: Increasing Resolution and Resolving Convection Improve the Simulation of Cloud-Radiative Effects Over the North Atlantic, J. Geophys. Res. Atmos., 125, e2020JD032 667, https://doi.org/10.1029/2020JD032667, e2020JD032667 10.1029/2020JD032667, 2020.

Shukla, J., Palmer, T. N., Hagedorn, R., Hoskins, B., Kinter, J., Marotzke, J., Miller, M., and Slingo, J.: Toward a New Generation of World Climate Research and Computing Facilities, Bull. Am. Meteorol. Soc., 91, 1407 – 1412, https://doi.org/10.1175/2010BAMS2900.1, 2010.

Sokolowsky, G. A., Freeman, S. W., Jones, W. K., Kukulies, J., Senf, F., Marinescu, P. J., Heikenfeld, M., Brunner, K. N., Bruning, E. C., Collis, S. M., Jackson, R. C., Leung, G. R., Pfeifer, N., Raut, B. A., Saleeby, S. M., Stier, P., and van den Heever, S. C.: *tobac* v1.5: introducing fast 3D tracking, splits and mergers, and other enhancements for identifying and analysing meteorological phenomena, Geosci. Model Dev., 17, 5309–5330, https://doi.org/10.5194/gmd-17-5309-2024, 2024.

Stevens, B., Acquistapace, C., Hansen, A., , and Coauthors incl. Senf, F.: The Added Value of Large-eddy and Storm-resolving Models for Simulating Clouds and Precipitation, J. Meteor. Soc. Japan, 98, 395–435, https://doi.org/10.2151/jmsj.2020-021, 2020.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Mon. Wea. Rev., 117, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:acmfsf>2.0.co;2, 1989.

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., and Reed, K. A.: TempestExtremes v2.1: a community framework for feature detection, tracking, and analysis in large datasets, Geosci. Model Dev., 14, 5023–5048, https://doi.org/10.5194/gmd-14-5023-2021, 2021.

Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core, Quart. J. Roy. Meteor. Soc., 141, 563–579, https://doi.org/10.1002/qj.2378, 2014.

Zängl, G., Reinert, D., and Prill, F.: Grid refinement in ICON v2.6.4, Geosci. Model Dev., 15, 7153–7176, https://doi.org/10.5194/gmd-15-7153-2022, 2022.