Report on the joint project ”WarmWorld” bk1341

Project: **1341**
Project title: **WarmWorld**
Principal investigator: **Julia Duras**
Allocation period: **2025-01-01 to 2025-12-31**
Reporting period: **2025-01-01 to 2025-12-31**

1 Module Better

1.1 T1) Resolution and physics sensitivity studies and experimental validation *(MPI-M)*:

Granted for 2025:	**52,390 CPU node hours**	**185 GPU node hours**
Consumption by end of October:	**51,562 CPU node hours**	**132 GPU node hours**

Throughout the calendar year, we used the computing resources to carry two main types of simulations: 1) limited-area model of the ORCESTRA campaigns (curtains and re-runs of specific days at higher 3D sampling rates), and 2) global+nest simulations to enable them on GPU (part of ICON's development) and to pursue our study on the model's sensitivity to grid resolution.

Study 1) consisted of a series of targeted reruns of the ORCESTRA limited-area model[1] focused over the HALO flight days. Especially, 18 flights during which the EarthCARE satellite passed over the HALO plane. The simulations were run with the sole purpose of extracting so-called *curtains*, vertical slices of data along the satellite/plane tracks. These curtains were then fed to a forward-model, PAMTRA, which served as a transfer-function between model output (hydrometeor mass fractions and thermodynamics fields) and radar's output (cloud reflectivity, Doppler fall velocity, etc). This approach allowed for a direct comparison between ICON and HALO, EarthCARE's observations. An example of such analysis is shown in Figure 1, where the curtain visualisation and postprocessing process is illustrated in panel a). Panels b) and c) - built upon the 18 curtains - help us quantify the model's biases: an overestimated rainfall speed and a lack of high-altitude clouds. Though not shown in this report for the sake of conciseness, comparisons with the HALO radar observations appear consistent. These biases motivated us to overhaul the one-moment microphysics scheme, by deriving ab initio fall/evaporation/accretion rates assuming a log-normal cloud droplet size distribution, as opposed to the less realistic (though mathematically more practical) exponential distribution. This upcoming work is better described in the 2026 computing time proposal.

Further, 10k-15k node hours were used to rerun the campaign with a 4× higher 3D sampling rates, with additional variables needed to calculate radar reflectivity in any column (beyond the EarthCARE satellite tracks). This finalized dataset was archived in HEALPix format and catalogued accordingly, and is being used by the whole ORCESTRA community. An example of analysis enabled by this rerun is shown in Figure 2, where the ICON-Sapphire model is compared to the radiosonde soundings and the IFS-9km forecast. The good agreement over all soundings, irrespective of their launching station (METEOR ship, Cap Verde and Barbados weather stations), demonstrates the suitability of using ICON as a supplemental tool to field campaign, as it appears able to capture the wind patterns over a large range of scale, from large intraday variation to weekly/monthly oscillations.

As for 2), the development of both LAM and Nested simulation on GPU was completed early in the year. The subsequent runs were used to investigate the impact of spatial resolution on the vertical distribution of cloud condensate. As the amount of cloud ice and cloud liquid droplet largely governs the atmospheric radiative budget, it is a critical quantity that is finely-tuned to observations. Yet, it appears to strongly depend on the grid resolution, for both shallow and deep convection alike (see Figure 3). This result motivates the follow-up study for 2026 where we aim to derive a resolution-dependent cloud inhomogeneity factor, an important parameter (usually constant) that determines the amount of subgrid cloud cover variance, used in the radiative flux calculations. If we could derive a robust scaling over the 10 km to subkilometer range, which covers most of our current modelling efforts, we would eliminate one source of uncertainty.
Unfortunately, the cost of these GPU runs with 12 nodes have been quite high due to the memory demand of the LAM. Given the total number of GPU nodes at Levante, running these simulations on CPU nodes is more reliable.

[1] Documentation of the OCESTRA limited-area model: `https://orcestra-campaign.org/lam.html`

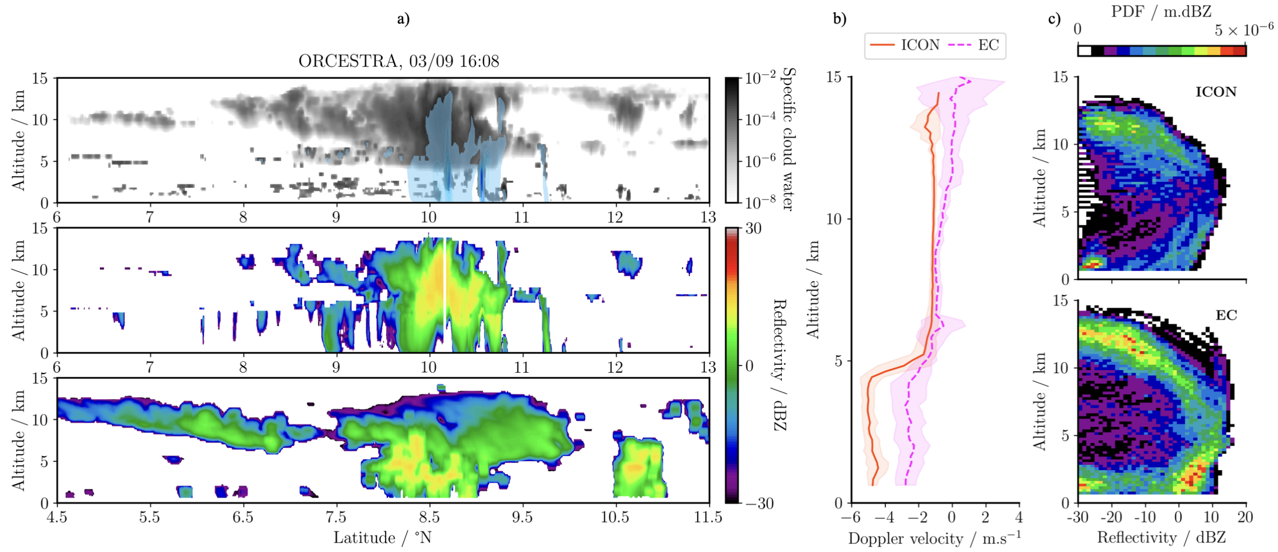

Figure 1: a) Instantaneous contours of (top) cloud and rain condensate, (middle) radar reflectivity evaluated from PAMTRA and (bottom) radar reflectivity measured by EarthCARE. b) Hydrometeors Doppler velocities for ICON (red) and EarthCARE (purple), shown with the $10^{th} - 90^{th}$ percentile envelopes. c) Two-dimensional probability density function of radar reflectivity for (top) ICON and (bottom) EarthCARE, sampled over the ORCESTRA campaign.

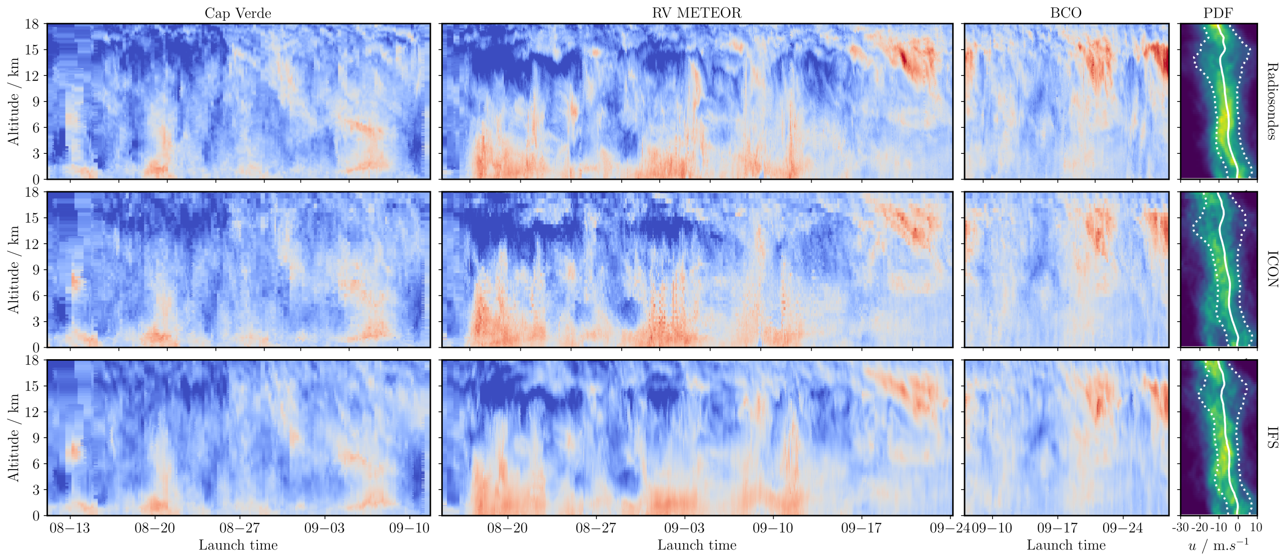

Figure 2: Meridional wind velocity contour along all radiosonde paths, ordered by launching station. The radiosonde locations are interpolated in time and space in the IFS (driving) and ICON simulations. Probability density function accumulated over all three stations are presented on the right panels, with (solid) 10-90^{th} percentile and mean profiles.

Publication: *Tropical convection and storm-resolving modeling: a case study based on the ORCESTRA campaign*, R. Fiévet, L. Linardakis, L. Kornblueh, M. Daniel-Lacombe, H. Gloeckner, L. Kluft, T. Fiolleau, C. Hohenegger and B. Stevens, Journal of Advances in Modeling Earth Systems (to be submitted soon)

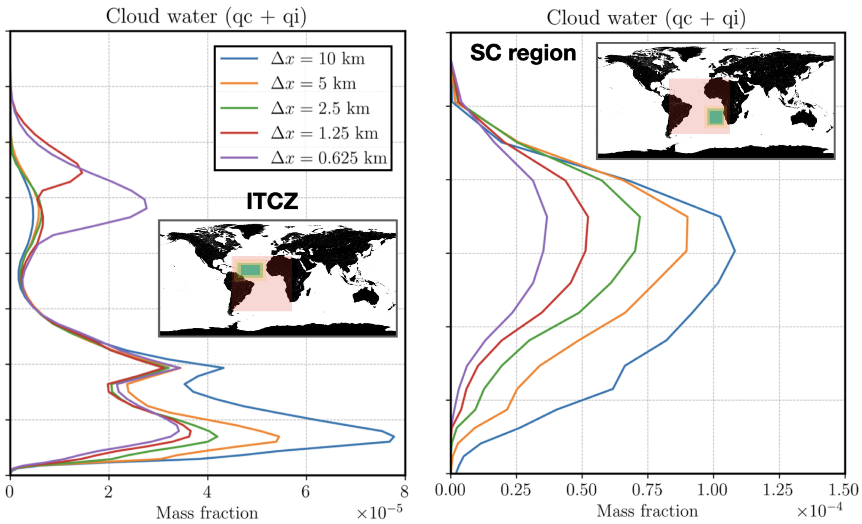

Figure 3: Vertical profiles of cloud condensates for (left) the deep-convective-rich region of the ITCZ and (right) the shallow-convection-rich region off the Namibian coast.

1.2 T2) Calibration of liquid-cloud microphysics in ICON *(University of Leipzig)*:

Granted for 2025:	**39,649 CPU node hours**
Consumption by end of October:	**20,000 CPU node hours**

We have used our computing resources to develop and test a new droplet parameterization that is now implemented in ICON-MPIM.

The entire procedure is as following:
External, daily resolved data of cloud condensation nuclei (CCN) are being read each simulation day, interpolated in time and in the vertical according to ICON pressure levels. The external CCN data consist of 6 variables reflecting the CCN spectrum over supersaturation. Instead of prescribing droplet numbers (CDNC), the 1-moment microphysical scheme is modified in such a way that CDNC are now computed via a new developed activation scheme using the CCN and the resolved vertical velocity. This scheme is a fitted version from (Kuba and Fujiyoshi, 2006), that has been tested on observational CCN data and the actually used CAMS CCN data. The CDNC that now vary in time and space are then further used in the autoconversion scheme as well as in the radiation scheme for consistency. This ensures that the microphysics and radiation of liquid clouds are coupled, unlike before.

First results look promising with distinct differences to the baseline simulation, using ICON-MPIM in Sapphire configuration (version 2025/04), see Figure 4 and Figure 5. However, we have noticed 2 deficiencies:

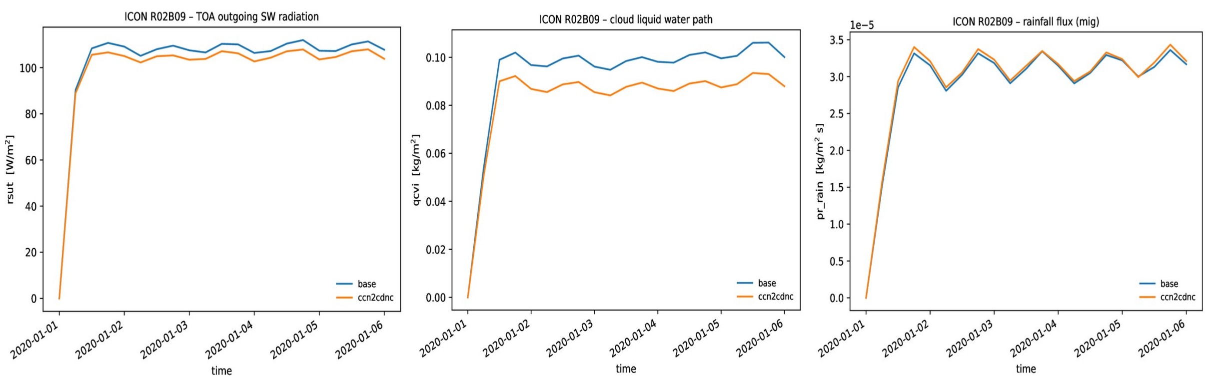

Figure 4: Comparison of relevant parameters from ICON-MPIM with droplet parameterization (ccn2cdnc) and without (base).

1. CDNC are too low compared to observations because resolved updrafts at 5km resolution are too low. Therefore, we plan to interact with tmx-developments to get information on subgrid-scale variability.

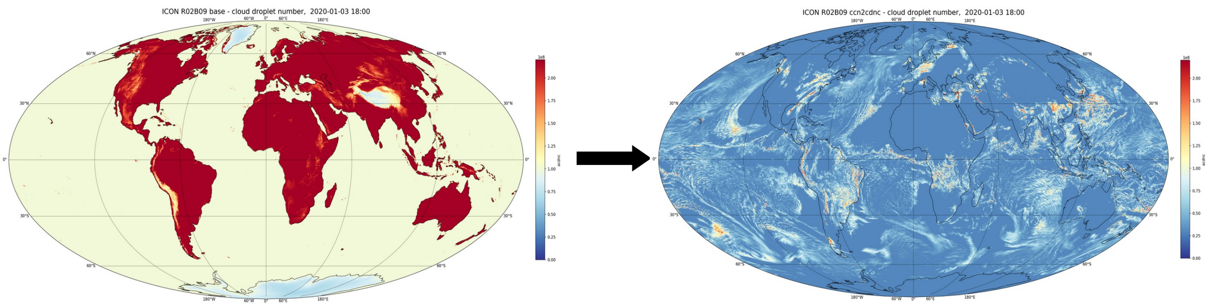

Figure 5: Cloud droplet number concentration (CDNC) as prescribed in ICON-MPIM (base) vs. CDNC as computed from CCN and resolved updraft at 5km resolution (ccn2cdnc).

2. The new implementations result in a runtime that is 4x that of the baseline simulation. That is ~ 10 simulated days/day instead of > 40 simulated days/day, or in other words we now require ~ 1000 node hours per simulated day instead of the 240 node hours we were aiming for. To reduce costs and increase efficiencies, we are currently testing a version of reduced complexity. Instead of reading 6 new variables, each with 4 dimensions (time,lev,lat,lon), we like to reduce this to one new variable with 3 dimensions (time,lat,lon), prescribing vertical distribution and supersaturation-spectrum within the model. These improvements are being tested at the moment.

Once the efficiency is increased and proper updraft variability is found for the activation parameterization, we can start simulating longer periods. We aim to do this towards the end of this year.

1.3 T3) Coupled IFS-FESOM ghost simulations for Warmworld simulations *(AWI)*:

Granted for 2025:	**68,296 CPU node hours**
Consumption by end of October:	**37,798 CPU node hours**

During the 2025 WarmWorld general assembly, the WarmWorld principal investigators (Thomas Jung and Bjorn Stevens) agreed to refrain from conducting additional very high–resolution (HR) IFS–FESOM simulations within WarmWorld-Better. Instead, focus towards the development of IFS–FESOM and ICON–FESOM high–resolution modeling capabilities, including configurations, performance optimizations, coupling interfaces, and supporting utilities such as ensemble generation frameworks that are essential for HR experiments.
This is in consequence to several of the previously proposed HR coupled (nudged) simulations—such as the 9 km and 4 km IFS–FESOM (nudged) configurations—are also being performed in the related EU projects DestinE and EERIE. This strategic decision ensures a responsible and frugal use of DKRZ computing resources while maintaining a complementarity synergy across these initiatives. This revised focus of WarmWorld–Better toward capability development and will be reflected in the upcoming compute proposal, which will request significantly reduced compute resources compared to previous years.

These development efforts have already delivered tangible results. Multiple IFS–FESOM configurations were developed and optimized using WarmWorld resources, including the 1 km setup for NextGEMS and several configurations such as `TCO159--CORE2`, `TCO319--DARS`, and `TCO399--DARS` used in DestinE and EERIE. For instance the climate model ensemble generation method created under WarmWorld–Better with IFS–FESOM has been adopted in EERIE Phase 2 to complement perform a 100-year historical simulation (TCO1279-NG5) (see Figure 6, for simulations performed for EERIE project, which partially used WarmWorld compute resources and model developments). Similarly, configurations and ensemble members generated through this work are contributing to DestinE nudged–storyline simulations conducted at MareNostrum5, which may additionally serve as ghost simulations for WarmWorld if required.

In line with above goals, compute resources allocated under WarmWorld–Better for 2025 have been used to further develop and test the new coupled ICON–FESOM configuration for climate simulations, extending beyond what was originally planned in the previous compute proposal. This work is being carried out in close collaboration with WarmWorld–Faster (Task T12), DKRZ (Nils-Arne Drier), and GEOMAR–Kiel (Stephan Juricke's group). As part of this effort, we successfully conducted 100-year coupled-countol simulations using the ICON–FESOM (`R2B4--CORE2`) setup (ICON–atm with NWP physics) and the ICON–FESOM (`R2B5--DART(DARS)`) configuration (ICON–atm with NWP physics), and plan to test configurations employing ICON–XPP physics and higher spatial resolutions in the upcoming compute proposal. While both runs demonstrated long-term stability, they exhibited notable temperature and precipitation biases, among others. For instance, Figure 7 shows temperature biases and precipitation biases in climatologies w.r to ERA5. ICON-FESOM is too cold in

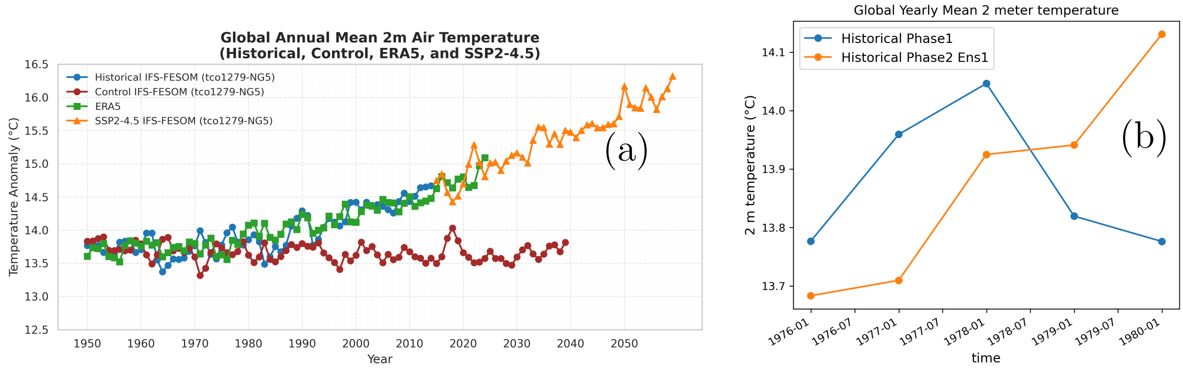

Figure 6: Global annual mean 2 m air temperature (T_{2m}) from IFS–FESOM simulations for the EERIE project: IFS–FESOM `TCO1279--NG5` at approximately 9 km. (a) shows results from the Phase 1 EERIE experiments, while (b) presents ensembles for the historical simulation (ongoing) for Phase 2 using model developments from WarmWorld.

the Arctic and warm in the Antarctic. Among the options to tune the model, the coupling interface, integrating hydrological discharge, assumed assymetry factor for liquid cloud cover are currently being investigated through targeted model tuning experiments, each of these tuning-experiment simulations are conducted for about 100 years to produce statistically-significant climatologies for evaluation.

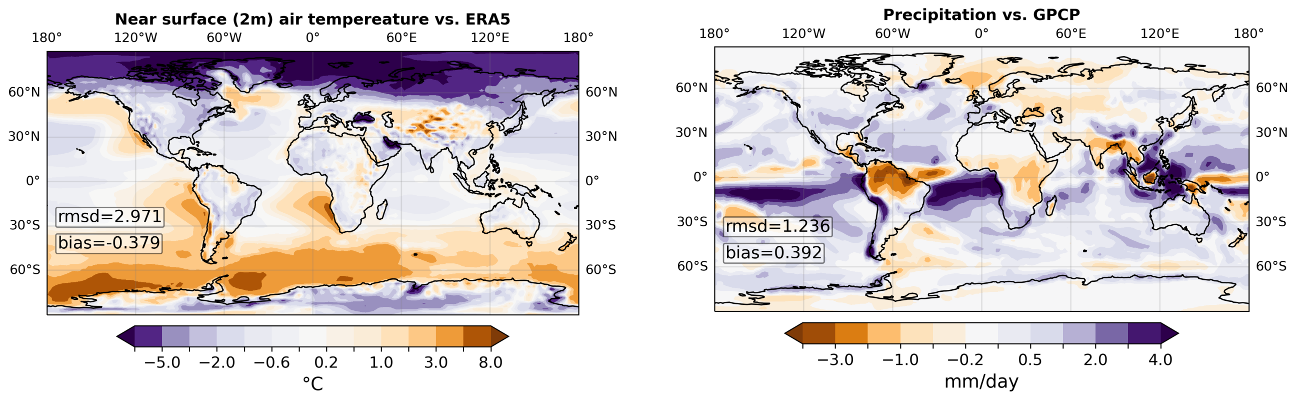

Figure 7: Annual mean biases of 2 m air temperature (T_{2m}, left) and precipitation (right) from the coupled ICON–FESOM (`R2B4--CORE2`) simulation relative to ERA5 climatology.

Publications:

- **ICON–FESOM: Coupled Climate Model Description and Evaluation.** Svetlana Loza *et al.* In preparation, to be submitted once the model tuning is complete.
- **Ocean Eddy-Rich Kilometre-Scale Global Climate Simulation with IFS–FESOM.** Rohit Ghosh *et al.* In preparation for submission to *Geoscentific Model Development*.
- While the members from related HR projects: DestinE, EERIE and NextGEMS were very appreciative of IFS-FESOM model developments and resources arising from WarmWorld-Better, every effort will be made to acknowledge WarmWorld and DKRZ computing resources in relavant publications arising from those projects.

1.4 T4) Evaluating the impact from mixed-phase and ice-cloud processes on cloud glaciation *(KIT)*:

Granted for 2025:	**8,624 CPU node hours**
Consumption by end of October:	**5,253 CPU node hours**

To assess the impact of secondary ice production (SIP) on cloud microphysics, dynamics, and radiation, five deep convective cloud (DCC) cases were simulated using ICON with 2-moment microphysics in limited-area mode (LAM). These included four aircraft-observed continental cases (CAIPEEX, DCMEX, MC3E, STEPS) and one satellite-observed marine case (ORCESTRA) (??). Simulations used online nesting with grids R3B7 to R3B10 (2000–500 km domains). SIP processes such as rime-splintering (Hallett and Mossop, 1974), raindrop shattering (Sullivan et al., 2018), and ice-ice collision (Phillips et al., 2017) were already in ICON, while sublimation breakup (Deshmukh et al., 2022) was newly added. A detailed raindrop shattering scheme from Phillips et al. (2018) was also implemented and compared to the existing one.

To implement and test two new SIP schemes (Phillips et al., 2018; Deshmukh et al., 2022), 24 short simulations were run for ORCESTRA (5×48 h), MC3E (5×48 h), and STEPS (14×24 h). Six additional runs per DCC case assessed the impact of individual SIP processes, including configurations with all or specific SIP processes disabled. One extra simulation per case evaluated the effect of a detailed raindrop shattering scheme (Phillips et al., 2018). Each run (24–48 h) used 8 CPU nodes, 40 NH per simulated day, and 0.6 TB of storage. For ORCESTRA and DCMEX, domain size was refined during initial tests (3 runs each) to better match EarthCARE and aircraft data, requiring 750 NH total. Production runs (7 per case) used 2000 NH, while testing, debugging, and tuning consumed 1200 NH.

The simulated total ice number concentrations (INCs) agree well with the EarthCARE-cloud Profiling Radar (CPR) observations at all vertical levels (Figure 8a). However, between −20 and −40 °C levels, ICON predicts the total INC that is about an order of magnitude higher than those from the CPR. This discrepancy is likely due to ICON overpredicting INC at these levels, as well as the limited sensitivity of the CPR to large concentrations of small ice particles, which, despite being numerous, contribute little to radar reflectivity. This results in an underestimation of the total INC in cloudy regions dominated by small ice-crystals.
SIP exhibits the most pronounced impact on Longwave (LW) cloud radiative heating (CRH) (Figure 8b), which decreases by up to 15% (in continental DCCs, not shown here) and 50% (in ORCESTRA marine DCCs) due to reduced absorption of outgoing LW radiation. In continental DCCs, excluding SIP causes minimal change (< 5%) in shortwave (SW) CRH. In contrast, excluding SIP from ORCESTRA marine DCCs exhibits enhanced SW radiative warming (from strong cooling to weak warming), primarily due to increased incoming SW radiation through the cloud column. This is because of a significant reduction in INC (by a factor of 10^3, Fig. 8a), which relatively forms optically thinner mixed-phase clouds and thereby allows more SW radiation to reach and absorb at lower levels.
In summary, along with changes in latent heating and convective relative humidity (CRH), the absence of SIP leads to a net reduction of up to 10% in the mean vertical motion within the simulated deep convective clouds (DCCs), resulting in cloud tops that are shallower by approximately 1–3 km.

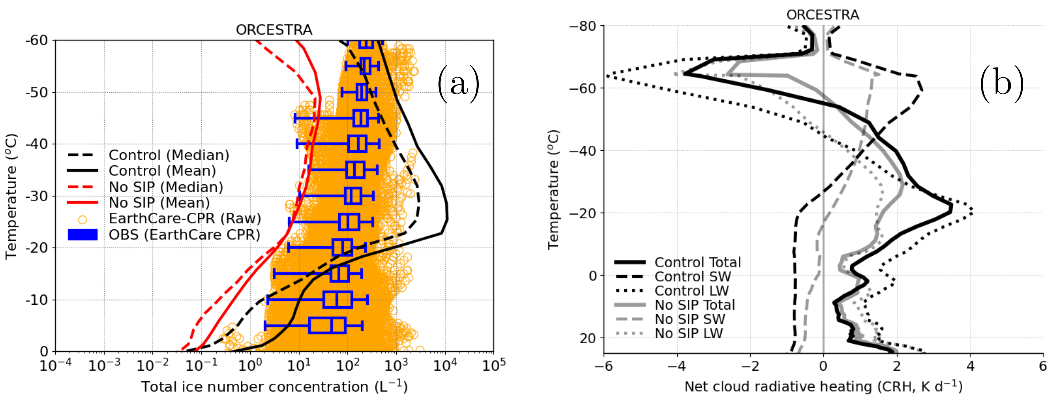

Figure 8: (a) Comparison of the simulated total ice number concentration (solid lines are median and dashed lines are mean) from the 4-SIP (Control, solid lines) and No SIP (No SIP, dashed lines) ICON-NWP runs with the observations from Cloud Profiling Radar (CPR) onboard the EarthCARE satellite. (b) Simulated domain-averaged net cloud radiative heating (CRH, K d1) from the control (solid black lines) and No SIP (solid gray lines) simulations of ORCESTRA clouds. Also shown are the SW (dashed lines) and LW (dotted lines) components of CRH.

Publication: A research article titled '*Impacts of Secondary Ice Production on the Microphysics and Dynamics of Deep Convective Clouds in Different Environments*', which explores the impact of SIP on the properties of these DCCs, will be submitted to the Journal of the Atmospheric Sciences.

1.5 T5) Development of ICON turbulence parameterization *(University of Hamburg)*:

Granted for 2025:	**11,904 CPU node hours**
Consumption by end of October:	**7,282 CPU node hours**

For the development of the turbulence parametrization in the software package turbulent mixing (TMX), 7,300 CPU node hours were used for development and validation tests. A new test case was added for a convective boundary layer (CBL) in an idealized torus domain, to validate the current turbulence schemes in TMX with analytical solution for the laminar CBL. Furthermore, the current implementation of the surface exchange coefficients was revisited in collaboration with sub-project TurbO-Arctic (AWI and University of Cologne) and tests have revealed a bug in the current implementation that reduces the surface heat flux abruptly for specific stable stratification regimes, see subsection 4.3. In addition, a new turbulence parametrization based on the Leonard term for the subgrid-scale stress tensor was implemented and is checked by laminar CBL test cases.

1.6 T6) Turbulent mixing (tmx) simulations *(DWD, MPI-M)*:

Granted for 2025:	**742 CPU node hours**	**174 GPU node hours**
Consumption by end of October:	**48 CPU node hours**	**1,455 GPU node hours**

Of the 742 CPU node hours granted for project T6), only 48 node hours have so far been used for visualization work. The reason for this is that during the integration of the TURBDIFF and TURBTRAN modules into TMX, it became apparent that a global memory manager significantly simplifies implementation. This task was therefore postponed to the end of the current and the beginning of the upcoming computing time period. The requested GPU node hours were instead used for reference simulations to investigate ParFlow's influence on cloud distribution and precipitation, particularly in regions with complex orography, such as the Alps, see task T7). The simulations are performed in a limited-area configuration covering the European continent with a horizontal grid spacing of 3.3 km (R03B09 mesh) and 90 vertical grid layers. To exploit synergies, the stability analysis described in the compute time proposal was performed on the same mesh rather than the original, smaller mesh covering only the D2-domain. The results are currently still being analyzed.

1.7 T7) Latitude-belt simulations with ParFlow *(FZ Jülich)*:

Granted for 2025:	**13,330 CPU node hours**	**19,240 GPU node hours**
Consumption by end of October:	**28,814 CPU node hours**	**9,786 GPU node hours**

A first publication on the coupling between ICON/ICON-Land and the ParFlow hydrological model was submitted in March 2025 serving as a proof-of-concept and presenting a first analysis of global-scale coupled simulations. Approximately 15,000 CPU node hours were used during the review process as several simulations needed to be redone before completing the revision. The GPU resources intended for high-resolution latitude-belt simulations of the tropics have not been fully used so far due to technical problems with ParFlow on GPU, which could be resolved, and performance issues with the high-resolution setup. Meanwhile, additional CPU resources were used to work around the problems on GPU. Approximately 10,000 GPU node hours were used instead for production runs of coupled simulations of the months April to August of 2024 on a European domain using a limited area configuration (3 km ParFlow horizontal resolution). This work was conducted in cooperation with Moritz Waldmann, see task T6). These simulations are currently evaluated against a stand-alone ICON reference and observations. Special focus is placed on differences in soil moisture distribution, surface quantities and fluxes (see Figure 9), and possible feedbacks on precipitation and the evolution of the atmosphere on a daily, weekly, and monthly basis.

Publication: A research article titled '*The ICON-ParFlow Coupling: Integrating a Continental-Scale Hydrological Model into a Global Atmospheric Model*', has been submittes to the Journal of Advances in Modeling Earth Systems and is currently under review.

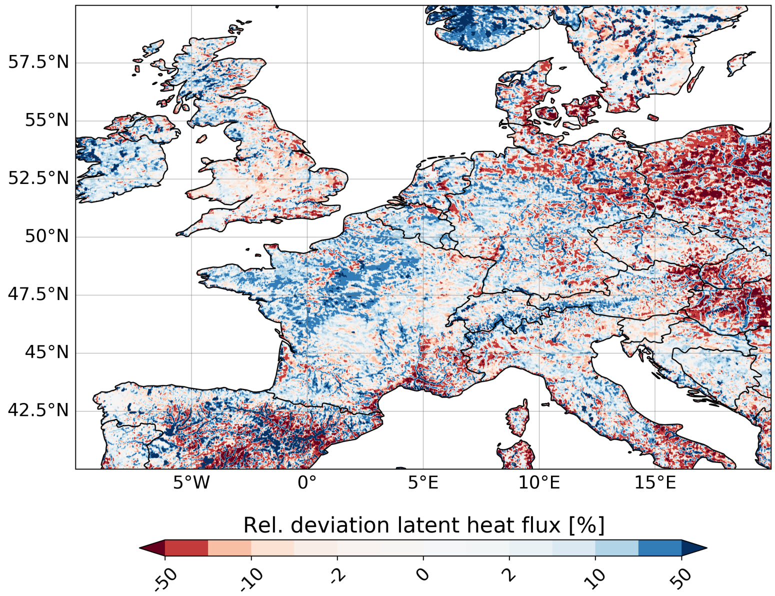

Figure 9: Relative deviation in the surface latent heat flux between a coupled ICON-ParFlow simulation and a stand-alone ICON reference simulation averaged from July 1 to July 5 2024 revealing river corridors, topographic features, and different soil hydraulic parameters.

2 Module Easier

2.1 T8) Development of a data-centric ICON workflow *(DKRZ)*:

Granted for 2025:	**11,699 CPU node hours**	**4 GPU node hours**
Consumption by end of October:	**8,000 CPU node hours**	**27 GPU node hours**

For the development of storage and access methods for km-scale simulations across storage tiers, about 1000 CPU node hours were use in this reporting period. They allowed development, testing and application of a toolchain to handle the transfer and retrieval of data located across different storage backends, corresponding to the metadata queries of the user, when exploring km-scale ESM simulations. The toolchain consists of a *zarranalyzer* to transform the zarr-stores to tarballs and a *reference file generator* for the contents of the tarballs in kerchunk as well as parquet format.
For the zarranalyser development, the data-sets of the km-scale ESM simulations `ngc3028` and `ngc4008`, generated by the nextGEMS projects[2], were used. The zarr-stores were reorganized, so that they can be optimally fit into the tape archives, keeping in view the size and data transfer rate constraints imposed by the HSM system.

The remaining 7,000 CPU node hours and 27 GPU node hours were used to develop and test the Hiopy output module.[3] This output module is designed to write large datasets in zarr-format on a hierarchy of grids. The use of zarr provides coherent datasets at the petabyte scale. As for each dataset, a full resolution hierarchy from 12 cells covering the globe to the full model resolution is written, users can chose the coarsest resolution that still allows an analysis to be performed, thus minimizing the amount of data needing to be processed, and enabling efficient analysis of the model output across scales.

Simulations used resolutions up to HEALPix zoom level 12 (kilometer-scale). In addition to validating functionality and workflow, performance metrics were measured and optimized. As the output needs to be performance-optimized for massive high-resolution simulations, also the tests need to be run on up to hundreds of nodes, so 7,000 node hours are the result of very conservative testing.

The WCRP Digital Earths Global Hackathon demonstrated the possibility of working with remote petabyte-scale datasets, as DKRZ, University of Tokyo, and the JASMIN supercomputer provided large datasets online to the global community. At the same time, it showed the need for running proxy servers at the compute centers to cache the access to these datasets, as we experienced difficulties especially at less-connected nodes. To circumvent these difficulties we performed first tests with running a proxy-server locally. They show a significant speed-up even under single-user scenarios. We will expand on these tests in the next year.

[2] nextGEMS simulations overview: `https://easy.gems.dkrz.de/DYAMOND/NextGEMS/simulation_overview.html`
[3] `https://nils.gitlab-pages.dkrz.de/coyote/hiopy.html`

3 Module Faster

3.1 T9) ICON performance benchmarks *(DKRZ)*:

Granted for 2025:	**40 CPU node hours**	**163 GPU node hours**
Consumption by end of October:	**2,858 CPU node hours**	**589 GPU node hours**

To ensure the performance of the ICON model during the code developments within the project, a test setup has been defined. It is an AMIP atmosphere-only setup, similar to production setups, with horizontal grid resolutions of R2B4, R2B6, R2B8, R2B9 and R2B10. These resolutions allow to test the same setup on different architectures (CPUs and GPUs) using several hundred of nodes down to only a fraction of a node. This setup has been converted to a mkexp configuration file to enable run script generation with the mkexp tool, allowing all WarmWorld developers as well as all ICON users at DKRZ to make use of it.
In addition to short test runs during configuration file development, some benchmarks were conducted to evaluate the performance with respect to the build wrapper configure option `--enable-realloc-buf` and environment variables for GPU setup. Although the currently used GPU software stack is not yet able to utilise the `gdr_copy` feature, performance at an R2B8 resolution could be improved by 2 %.
The overall performance development of this setup (R2B8-L90 resolution, 5 simulated days) is as follows:

- 2024: 15.3 SH/NH at 32 GPU nodes
- 2025: 13.0 SH/NH at 32 GPU nodes (March 2025)

Further work is currently being carried out to enable mkexp to generate run scripts for coupled setups running on the *gpu* or the *dolpung* partition, with the atmosphere conducted on the GPUs and the ocean and I/O processes on the CPUs of the nodes.

3.2 T10) Testing and developing granules *(DKRZ)*:

Granted for 2025:	**403 CPU node hours**	**222 GPU node hours**
Consumption by end of October:	**112 CPU node hours**	**320 GPU node hours**

The work on setting the radiation code in ICON to run concurrently with the other components of an ICON atmosphere run has not been attempted due to a shift in priority. Instead, we successfully rewrote 41 performance-critical Fortran routines from ICON's math library (libiconmath) using C++ and Kokkos to achieve performance portability across GPU architectures. The computing hours were spent on building and testing this rewrite and checking the performance gain thereafter.
Performance testing was conducted using the ICON Testbed, where we ran 100 executions of each of the routines we ported after initialising ICON with a particular grid structure. These tests were performed at two grid resolutions: R2B6 (single-GPU node) and R2B8 (multi-node GPU), testing only double-precision arithmetic for the iconmath-interpolation module. The comparison of performance between the original Fortran version and the newly implemented C++/Kokkos version is presented in Figure 10.

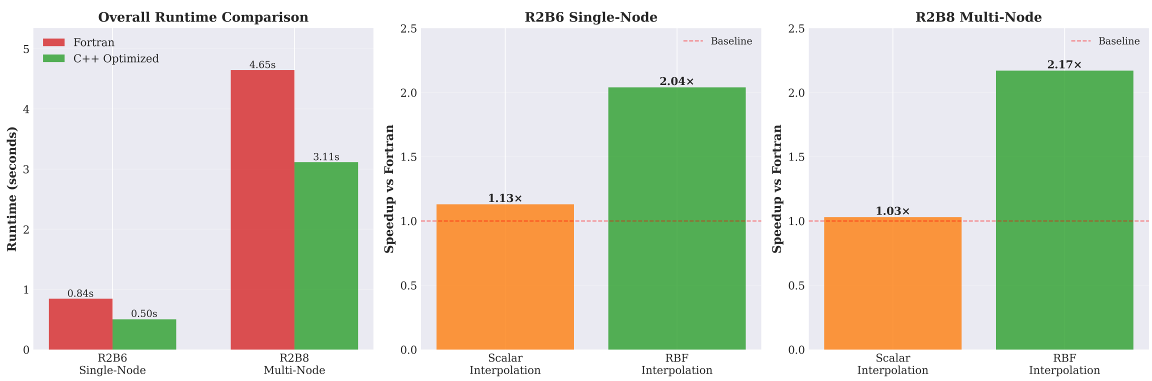

Figure 10: Performance comparison showing overall runtime improvements (left) and module-level speedups (middle, right) for R2B6 single-node and R2B8 multi-node configurations. The module RBF interpolation demonstrates the strongest gains (2.04-2.17×) with excellent scaling characteristics.

The second part, the planed implementation of the ICON-A advection has not been started yet. Instead, time was spent on the ground work for implementing the C++ infrastructure of Ragnarok that provides access to

ICONs communication and domain data (p_patch). This will simplify further C++/Kokkos implementation of the aes-physics including TMX and will also support future implementations of C++/Kokkos advection kernels. The memory manager has been developed further to provide a simpler Fortran interface for CPU and GPU usage. The usability and correctness have been demonstrated within the current Fortran version of TMX. This development was done using single node resources outside the project contingent.

3.3 T11) Further development of ICON-C components using different programming languages *(DKRZ)*:

Granted for 2025:	**310 CPU node hours**	**740 GPU node hours**
Consumption by end of October:	**1485 CPU node hours**	**324 GPU node hours**

The compute hours were spent on (a) building and testing the C++/Kokkos infrastructure and project dependencies, (b) porting code from Fortran to C++/Kokkos and (c) preliminary performance measurements of the newly integrated components into ICON.
The porting of several components of the AES physics package started with the thermodynamics, the graupel microphyiscs, the turbulence (tmx) and the mathematical libraries (libiconmath). Preliminary results from running an AMIP experiment with the first two ICON Kokkos components (thermodynamics and graupel microphysics) of the AES physics are shown in Figure 11. This experiment is one of the standard test experiments of the AES pyhsics package.

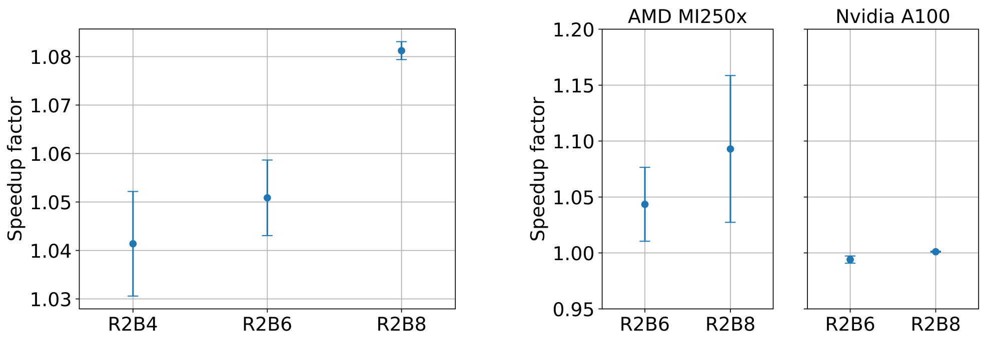

Figure 11: Speedup of ICON Kokkos versus ICON Fortran for an AMIP experiment with AES pyhsics, at different resolutions, for 3 simulated hours, on different hardware platforms.

3.4 T12) Development of next-version of FESOM and its coupled configurations *(AWI)*:

Granted for 2025:	**1,550 CPU node hours**	**1,110 GPU node hours**
Consumption by end of October:	**53 CPU node hours**	**0 GPU node hours**

During this reporting period, all testing of new coupled model configurations—specifically IFS–FESOM and ICON–FESOM—was transferred to *WarmWorld–Better (Task T3)*. Accordingly, the compute resources originally allocated under *WarmWorld–Faster* for this activity will instead be utilized within *WarmWorld–Better*. From this point onward, the focus of *WarmWorld–Faster* shifts more strongly toward technical developments related to FESOM, in close synergy with the overarching goals of both *WarmWorld* and *WarmWorld–Better*.

As outlined in the previous compute proposal, and analogous to the modernization of the ICON model, a new generation of the FESOM model—**FESOM3(x)**—is being developed in Python, with **JAX** serving as the initial computational backend. The solver core of FESOM2 has been modularized into a *dwarf* configuration and successfully ported to Python, reproducing *bit-identical* results to the original Fortran-based FESOM2 for the benchmark Munk–Gyre test case.
Due to the experimental nature of this early-stage development, most tests have been performed locally on AWI-procured workstations equipped with **Intel i9-13950HX (13th Gen, 32 CPU threads)** and **NVIDIA RTX5000 (Ada)** GPUs. This setup allows for rapid prototyping and verification, reducing the need for large-scale CPU or GPU allocations compared to the previous proposal. However, this local workflow also implies that scalability tests of the JAX-based Python implementation—particularly on multi-node CPU and GPU systems—have not yet been carried out.
By the end of the calendar year, we plan to utilize a portion of the remaining compute resources to conduct

these scalability tests on distributed architectures, assessing the performance of JAX for high-resolution and multi-node configurations for unstructured meshes. These tests will be crucial in evaluating the feasibility of using Python/JAX for large-scale, performance-portable implementations of **FESOM3(x)** for the future.

3.5 T13) WarmWorld Summer School 2025 *(DKRZ)*:

Granted for 2025:	**310 CPU node hours**	**370 GPU node hours**
Consumption by end of October:	**585 CPU node hours**	**1 GPU node hours**

The Summer School took place between 28.07.2025 and 7.08.2025 in Lauenburg, Germany. 30 Master and PhD Students of 14 nationalities took part in the extensive programme, which contained lectures, hands-on exercises and coding challenges. The lectures are publicly available online, at `https://warmworld.gitlab-pages.dkrz.de/summerschool2025/lectures/`.

The allocated node hours were used for running ICON experiments with the purpose of extracting and post-processing data, visualisation and development of new ICON plugings, written in python, connected via ComIn. The students also ported the energy diagnostic module of the AES physics to Kokkos and evaluated the performance of different programming approaches by running the Bubble experiment.

4 Module Smarter

4.1 T14) ELPHE *(GEOMAR)*:

Granted for 2025:	**15,500 CPU node hours**	
Consumption by end of October:	**5.318 CPU node hours**	**2073 GPU node hours**

Due to insufficient CPU allocation, we were unable to execute the planned FESOM2 production simulations. Our efforts were therefore mostly focused on large OceanParcels runs originally planned for later project stages using high-resolution CMEMS and NEMO datasets. This simulations will be later used for direct comparison with FESOM2 and ICON simulations.

We ran OceanParcels on CMEMS GLORYS12 (1/12°) and NEMO outputs. Lagrangian experiments comprised 4-year integrations with two release strategies: 10,000 particles every 5 days and 100,000 particles monthly. We tracked particle position and hydrographic properties. For the Amazon River plume, a seasonal mean particle concentration, averaged over 3 years, can be seen in Figure 12. The Amazon River runoff into the Atlantic Ocean accounts 17% of global river input to the oceans (Dippner et al., 2024) and plays a crucial role in near-field offshore stratification, nutrients delivery and CO_2 sequestration (Louchard et al., 2023; Subramaniam et al., 2008). This particle trajectories will now be investigated to better understand seasonal versus mesoscale patterns in the ARP.

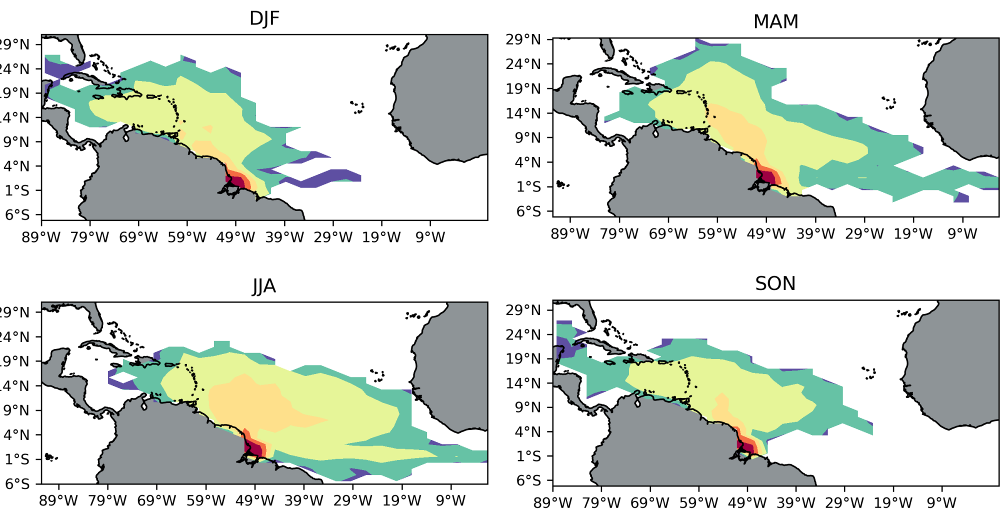

Figure 12: Grand-ensemble seasonal mean particle concentration (log10 fractional presence), averaged over June 2022–July 2025 and binned at 1° × 1°. Seasons follow DJF, MAM, JJA, SON

Initial FESOM2 work focused on configuration and trial tests in the Western Tropical Atlantic/Amazon River Plume (ARP) region, and on a coarse-resolution global proof-of-concept for workflow validation. We performed tuning across multiple meshes, evaluating configurations with and without the so-called backscatter parametrization scheme (Juricke et al., 2020) to inform mesh-specific parameters and assess subgrid backscatter impacts on solution stability and energy spectra. Furthermore, GPUs were used to explore a transformer-based ocean emulator (adapted from ArchesWeather) on NextGEMS simulations regridded to 1.5°, with emphasis on the Atlantic region which may be used in later project stages. The emulator reproduces key features (e.g., tropical and extratropical Atlantic dynamics) and exhibits improved long-term stability.

4.2 T15) SmartWeather21 *(KIT)*:

Granted for 2025:	**22,455 GPU node hours**
Consumption by end of October:	**1,153 GPU node hours**

We implemented a version of the Pangu-Weather model, and trained it on different resolutions (0.25°, 1.5° and 3°) of the ERA-5 datasets. We performed an architecture study regarding patch embeddings to improve model predictive performance. This included, among others, non-overlapping and overlapping convolutions, and patch embeddings with and without direct association of certain embedded features with certain input variables. We further conducted experiments on the effects of constant masks on model predictive performance as well as on the impact of hidden dimension size. Towards testing the transferability of the architecture, we conducted some training runs on the task of tracer advection and subseasonal mean temperature and precipitation prediction.

Based on this optimized model, we started work on a probabilistic model in the form of a Bayesian neural network (BNN), trained with variational inference. Training BNNs is computationally very challenging due to the increased number of model parameters from using parameterized distributions as weights, and the necessity to sample from these distributions. Hence, we conducted initial exploration runs using the 1.5° resolution model. Figure 13 shows an exemplary prediction of the global 2 m temperature, achieved with initial training runs. We are working on parallelization strategies to extend this work to higher resolutions, using domain and sample parallelization approaches. These developments are currently ongoing.

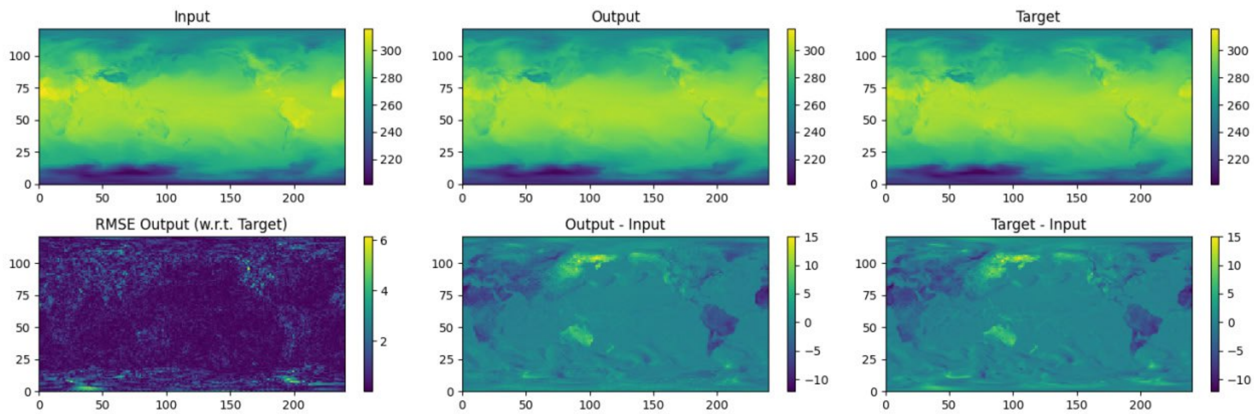

Figure 13: Prediction of global 2 m temperature with a Bayesian Transformer model. Left: input state; middle: neural network prediction; Right: ground truth

We further explored training of our Bayesian model with data parallelism; however, we were not able to scale beyond a single node so far. In combination with the wall-time limits for running jobs on the system, this hindered us so far from making full use of the granted GPU resources

In line with attempts to enable initialization with kilometer-scale climate projections, requiring models to handle much larger inputs than previously possible, we conducted scaling experiments on our newly developed algorithm for domain and tensor parallelism. However, like in the case on BNNs, larger scaling experiments would require us to scale beyond a single node, which we were not able to realize from a technical point of view.

4.3 T16) TurbO-Arctic *(AWI and Uni Köln)*:

Granted for 2025:	**403 CPU node hours**
Consumption by end of October:	**2,574 CPU node hours**

Over the past year, CPU resources were primarily dedicated to:

- Running ICON-LAM simulations centered around the research vessel Polarstern during the Arctic MOSAiC campaign in winter 2019/2020
- Postprocessing and analysis of model output data
- Preprocessing of high-resolution MOSAiC raw data and turbulence intermittency studies.

The ICON-LAM experiments were performed at a horizontal resolution of 5 km over a small domain of approximately 600 × 600 km around Polarstern. Following the successful setup, a series of short test runs were performed to assess and better understand the model's behavior. Following this, a winter-long simulation (October 2019 — February 2020) was conducted with an Local Area Mode (LAM) of the ICON WarmWorld configuration around the track of Polarstern during the MOSAiC expedition (Figure 14). This allowed for a first comparisons between model output at the Polarstern location and 10-minute postprocessed MOSAiC observational data.

The statistical analysis revealed that ICON tends to generate excessively stable temperature inversions over

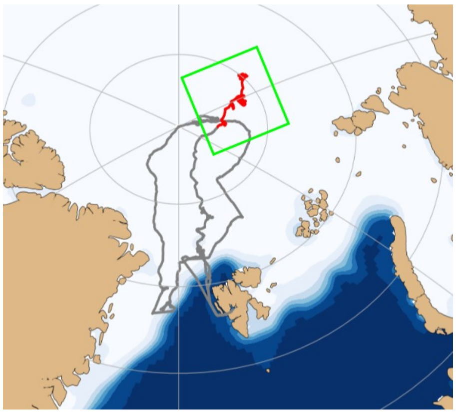

Figure 14: Track of research vessel Polarstern during MOSAiC (gray) and a Local Area modelling domain (green) centered around the ship's position in from October 2019-February 2020 (red).

sea ice in polar winter. In particular, episodes of runaway surface cooling lasting several days were observed, triggered by overly strong radiative surface cooling. An upcoming more detailed investigation of the model output and surface coefficient computations identified a bug in the surface scheme of the turbulent mixing (tmx) package, which effectively suppressed turbulent fluxes under very stable conditions (Figure 15, middle panel). In strong collaboration with University Hamburg, task T5), this issue has since been fixed. The winter-long ICON-LAM simulations were subsequently rerun using both the corrected surface stability scheme and an alternative long-tailed stability function derived from the earlier Arctic SHEBA campaign (Gryanik and Lüpkes, 2018). Analyses indicate weaker temperature inversions and a reduced surface cold bias over Arctic sea ice in winter, resulting from enhanced turbulent fluxes under stable conditions (Figure 15, right panel).

In addition, a first prototype of a stochastic surface flux parameterization was successfully implemented and

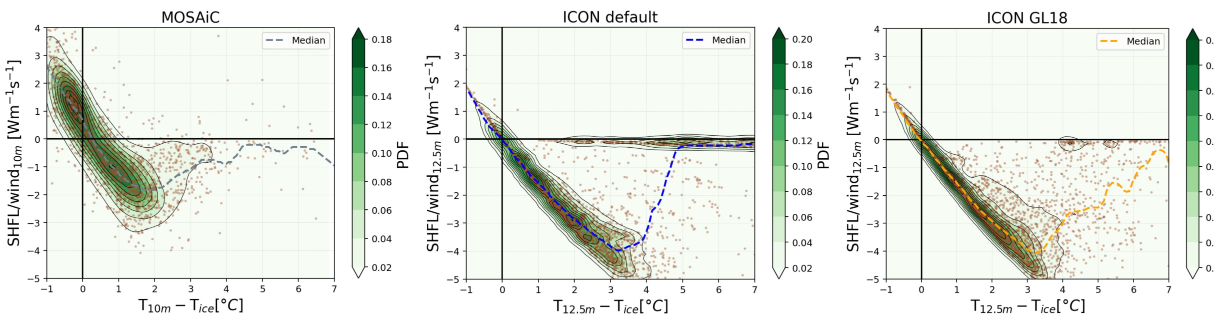

Figure 15: Scatterplot (dots) of 10-min averaged near-surface thermal stratification versus scaled sensible heat fluxes overlaid with the probability density funtion (contours) and median (blue dashed line). Left: MOSAiC measurements from October 2019-February 2020. Middle: ICON default results which allowed to identify the bug in the tmx package for stable stratification. Right: ICON results with alternative long-tailed stability function according to Gryanik and Lüpkes (2018).

tested in ICON. This scheme computes turbulent surface exchange coefficients by solving a stochastic differential equation, thereby accounting for intermittency and turbulent bursts under strongly stable stratification. The implementation will be further investigated and refined in the next stages of development.

4.4 T17) ProImpact *(GERICS and KIT)*:

Granted for 2025:	**1,544 CPU node hours**
Consumption by end of October:	**3,705 CPU node hours**

During the first phase of the project, our main objective was to familiarize ourselves with the technologies developed and used within the WarmWorld project and especially the Easier group. This included gaining hands-on experience with several key components of the ecosystem, such as STAC catalogs, ComIn, COYOTE, YAC, Zarr data formats, and HEALPix. Through this process, we established a solid technical foundation that enabled us to begin developing our own tools and instruments based on these technologies with focus on "streaming" analysis.

After the initial familiarization stage, we focused on integrating, testing, and extending selected components within our workflow. The main achievements of this phase are summarized below:

1. **Development and testing of the ComIn precipitation plug-in**
   We developed and validated a new ComIn plug-in for the computation of 5-minute accumulated precipitation. The implementation is compatible with both AES and NWP physics configurations within ICON, ensuring flexibility and interoperability across different model setups.

2. **First prototype of 5-minute precipitation data streaming**
   By combining the ICON–ComIn–YAC interface with the COYOTE framework, we created our first functional emulator for data streaming of 5-minute precipitation results. This achievement demonstrates the feasibility of near-real-time data exchange and processing within ICON.

3. **Prototype web portal for visualization and analysis**
   We developed the first version of the web portal, available at `https://www.proimpact.dkrz.de`, which integrates the official **DKRZ authentication** system. The portal currently features three main sections (all currently under development and can be modified):

   - **Page 1:** Project overview and data animator for interactive visualization of model output.
   - **Page 2:** Data hub providing unified access to multiple data sources and repositories.
   - **Page 3:** Plug-in hub for scientific evaluation tools and diagnostics.

4. **Adaptation of the 2D blocking tracing algorithm for streaming data**
   Ongoing work focuses on adapting the **2D blocking tracing algorithm** to operate with **streaming data**. This development aims to enable online detection and analysis of atmospheric blocking events during model runtime, leveraging the streaming capabilities established through the COYOTE–ComIn integration.

4.5 T18) IconRep *(ECMWF and FZJ))*:

Granted for 2025:	**5,165 GPU node hours**
Consumption by end of October:	**4,780 GPU node hours**

The ICON-Rep project aims to demonstrate large-scale representation learning of physical atmospheric processes using a transformer-based neural network on climate data, with applications in climate model data compression and rapid interpolation between climate scenarios. This year has been dedicated to preparing the core model of the WeatherGenerator for training on ICON data, with significant computational resources invested in developing the infrastructure necessary for this goal.

The primary focus this year has been establishing the foundational capabilities required to train the WeatherGenerator on ICON climate simulation data. A substantial portion of computational resources was consumed in developing an ICON CMIP6 data loader, which is essential for efficiently processing the climate model output required for training. This data loader enables the model to handle ICON's native grid structure and multi-variable datasets across various temporal resolutions. In parallel with infrastructure development, we have conducted experiments involving atmospheric data to contribute to core model development. These experiments have provided valuable insights into the model's behavior with real atmospheric variables.

With the data infrastructure now established and initial atmospheric experiments completed, the project is positioned to transition to intensive model training phases. Future work will focus on implementing climate forcing handling, conducting full-scale training runs on multiple ICON scenario runs (equilibrium and non-equilibrium states with varying CO_2 concentrations), and evaluating the model's interpolation and extrapolation capabilities across different climate scenarios. The infrastructure developed this year provides a solid foundation for achieving the project's ultimate goal of creating a prototype climate emulator with well-documented capabilities and limitations.

References

A. Deshmukh, V. T. J. Phillips, A. Bansemer, S. Patade, and D. Waman. New empirical formulation for the sublimational breakup of graupel and dendritic snow. *Journal of Atmospheric Sciences*, 79(1):317–336, 2022. doi: 10.1175/JAS-D-20-0275.1.

Joachim W. Dippner, Joseph P. Montoya, Ajit Subramaniam, Jacqueline Umbricht, and Maren Voss. The Amazon River plume—a Lagrangian view. *Limnology and Oceanography: Methods*, 22(8):572–589, 2024. ISSN 1541-5856. doi: 10.1002/lom3.10626.

Vladimir M Gryanik and Christof Lüpkes. An efficient non-iterative bulk parametrization of surface fluxes for stable atmospheric conditions over polar sea-ice. *Boundary-Layer Meteorology*, 166(2):301–325, 2018. doi: 10.1007/s10546-017-0302-x.

J. Hallett and S. Mossop. Production of secondary ice particles during the riming process. *Nature*, 249:26–28, 1974. doi: 10.1038/249026a0.

Stephan Juricke, Sergey Danilov, Nikolay Koldunov, Marcel Oliver, and Dmitry Sidorenko. Ocean kinetic energy backscatter parametrization on unstructured grids: Impact on global eddy-permitting simulations. *Journal of Advances in Modeling Earth Systems*, 12(1):e2019MS001855, 2020.

N. Kuba and Y. Fujiyoshi. Development of a cloud microphysical model and parameterizations to describe the effect of ccn on warm cloud. *Atmospheric Chemistry and Physics*, 6(10):2793–2810, 2006. doi: 10.5194/acp-6-2793-2006. URL `https://acp.copernicus.org/articles/6/2793/2006/`.

Domitille Louchard, Matthias Münnich, and Nicolas Gruber. On the role of the amazon river for n2 fixation in the western tropical atlantic. *Global Biogeochemical Cycles*, 37(2):e2022GB007537, 2023.

V. T. J. Phillips, J. J. Yano, and A. Khain. Ice multiplication by breakup in ice–ice collisions. part i: Theoretical formulation. *Journal of Atmospheric Sciences*, 74(6):1705–1719, 2017. doi: 10.1175/JAS-D-16-0224.1.

V. T. J. Phillips, S. Patade, J. Gutierrez, and A. Bansemer. Secondary ice production by fragmentation of freezing drops: Formulation and theory possible high ice particle production during graupel–graupel collisions. *Journal of Atmospheric Sciences*, 75(9):3031–3070, 2018. doi: 10.1175/JAS-D-17-0190.1.

Ajit Subramaniam, PL Yager, EJ Carpenter, C Mahaffey, K Björkman, S Cooley, AB Kustka, JP Montoya, SA Sañudo-Wilhelmy, R Shipe, et al. Amazon river enhances diazotrophy and carbon sequestration in the tropical north atlantic ocean. *Proceedings of the National Academy of Sciences*, 105(30):10460–10465, 2008.

S. C. Sullivan, C. Barthlott, J. Crosier, I. Zhukov, A. Nenes, and C. Hoose. The effect of secondary ice production parameterization on the simulation of a cold frontal rainband. *Atmospheric chemistry and physics*, 18(22): 16461–16480, 2018. doi: 10.5194/acp-18-16461-2018.