Report on the joint project "WarmWorld" bk1341

Project: **1341**
Project title: **WarmWorld**
Principal investigator: **Julia Duras / Elina Plesca**
Allocation period: **01 Jan 2026 to 31 Dec 2026**
Reporting period: **01 Jan 2026 to 31 Oct 2026**

The 2026 computing time proposal for WarmWorld (bk1341) covered a broad range of scientific experiments and technical developments across the four project modules: *Better*, *Faster*, *Easier*, and *Smarter*. The requested resources were intended to support the development and evaluation of storm-resolving Earth system models, including sensitivity studies of atmospheric physics, coupled atmosphere–ocean simulations, and improvements to turbulence and cloud microphysics. Further activities focused on performance portability and GPU optimisation of ICON and FESOM, the development of data-centric workflows for handling and accessing large simulation datasets, and the application of machine learning and advanced analysis methods. The proposal brought together activities from multiple partner institutions, with computing and storage requirements specified for individual tasks and experiments. Additional resources were requested and granted for Jul - Dec 2026 and the usage of these is also included in this report.

	CPU node hours	**GPU node hours**	**TiB in /work**	**TiB in /arch**
Granted for Jan - Dec 2026	130130	12620	544	101
Granted for Jul - Dec 2026	29170	15540	145	47
Consumption by 31 Oct 2026	128045	18451	557	6

1 Module Better

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

In 2025, the ICON-LAM hindcasts of the ORCESTRA campaign (1.25 km) showed that simulated rain falls about twice as fast as measured by the Doppler radars of EarthCARE and HALO. This bias was proposed as a possible common cause of the dry troposphere, the undersized convective systems and the lack of ice aloft found in these simulations (Fiévet et al., 2026, ESS Open Archive, in review). The 2026 allocation was used to correct this bias and to extend the evaluation to global configurations. As the allocation was smaller than requested, the resources were concentrated on *curtain* simulations, in which the model is sampled along the EarthCARE ground tracks at the overpass times and converted into radar quantities with the forward operator PAMTRA.

Regional curtains with new rain microphysics. The one-moment warm-rain rates of ICON (sedimentation, evaporation and accretion) were rederived using a log-normal drop size distribution (DSD), instead of the former exponential DSD, together with the terminal velocity law of Van Boxel (1998). The rates were then approximated by polynomials that the model evaluates cheaply. The ORCESTRA LAM campaign was rerun with the original exponential rates (ICON-Exp) and the new log-normal rates (ICON-LogN), with identical initial and boundary conditions, and compared with the 18 HALO–EarthCARE overpasses. Below the melting level, the mean Doppler velocity drops from -5.0 to $-2.9\,\mathrm{m\,s^{-1}}$, against $-2.5\,\mathrm{m\,s^{-1}}$ observed by EarthCARE. This removes 85% of the fall-speed bias against EarthCARE and 70% against HALO, at all convective intensities (Fig. 1). Above the melting level, the Doppler velocities are still about twice the observed ones. Further, the ice content are unchanged while the upper troposphere becomes drier, which points to the ice microphysics as the next target for model improvement. Finally, the new rain parameterization has since been implemented in ICON's default one-moment microphysical scheme.

Internal sensitivity studies. Additional LAM curtain runs were used for exploratory sensitivity tests. These compared the ICON-Sapphire 2.0 configuration with the previous Sapphire 1 setup and tested bounding choices in the warm-rain processes, such as a maximum rain evaporation rate. They guided the configuration choices for the new scheme and are not shown here.

Global curtains for model intercomparison (ECOMIP, ETCMIP). Global ICON simulations at 2.5 km (R2B10) were run for two intercomparison projects coordinated by the University of Tokyo (M. Satoh). For

ECOMIP[1], the ORCESTRA golden case of 3 September 2024 was simulated. For ETCMIP (its tropical cyclone counterpart), two runs of Hurricane Humberto were initialized on 26 and 27 September 2025. A two-month continuous simulation of the ORCESTRA period (global+nest or LAM) is now being discussed with the ECOMIP coordinators for the coming months or early 2027, as a contribution to the 2027 global km-scale hackathon.

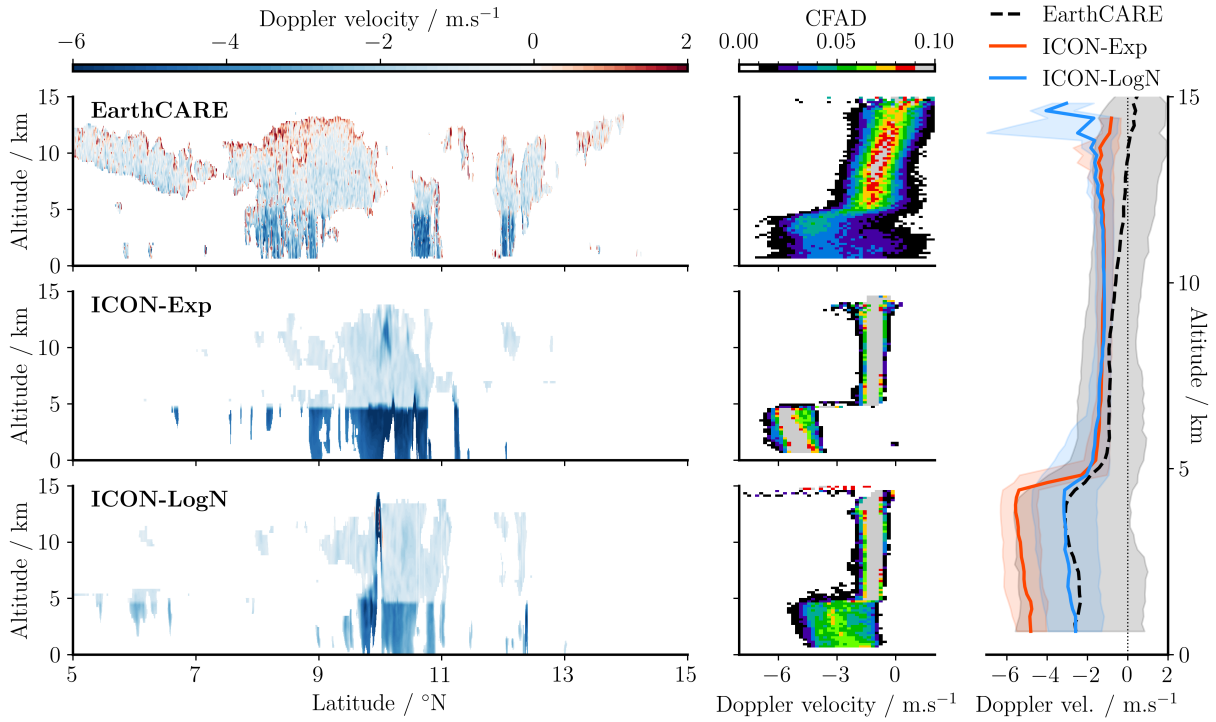


Figure 1: Doppler velocity measured by EarthCARE and simulated by ICON-LAM with the original exponential (ICON-Exp) and new log-normal (ICON-LogN) rain microphysics. Left, curtains of the 3 September 2024 overpass. Center, CFADs of the 18 overpasses, normalized at each altitude. Right, mean profiles of the 18 overpasses, with shading showing one standard deviation across overpasses.

Publications:

- Fiévet et al., 2026, '*Tropical convection and storm-resolving modeling: a case study based on the ORCESTRA campaign*' (in review, JAMES; preprint: ESS Open Archive, DOI: 10.22541/essoar.177315810.09022965/v1)
- Fiévet et al., '*From Observations to Correction: Rain Microphysics in a Storm-Resolving Model*' (in preparation)

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

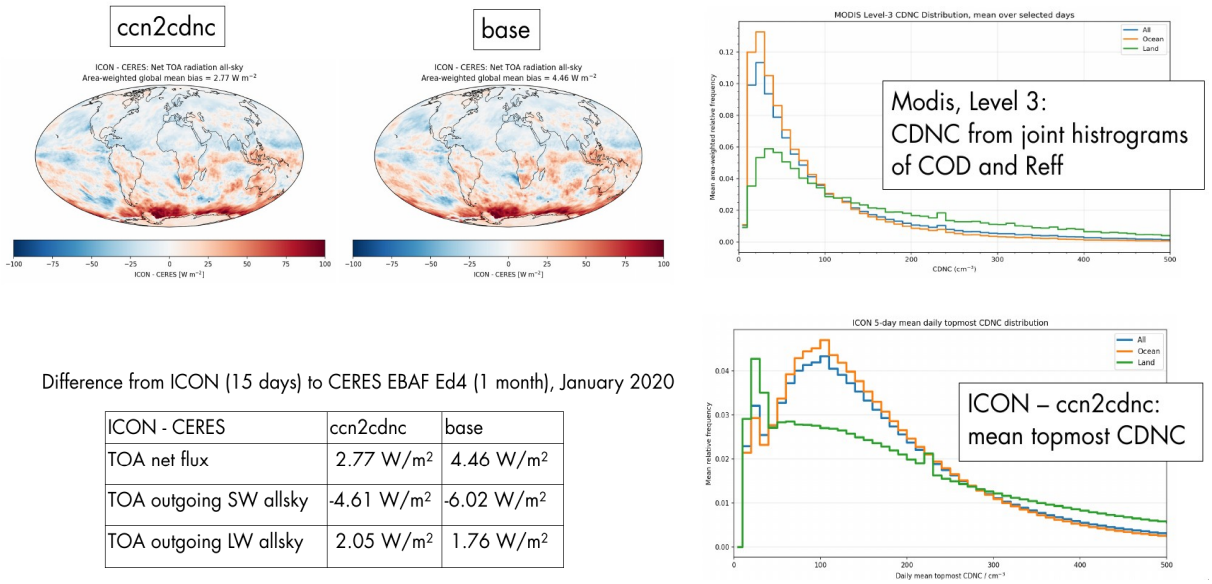
We have used our computing resources to thoroughly test and improve the new ccn2cdnc scheme in ICON. We were able to resolve the existing deficiencies of too low droplet numbers and most importantly, we dramatically decreased model runtime by restructuring the workflow and make use of complexity-reduced input data. The 2025 runtime of 4x that of the baseline simulation have thus been reduced to 1.05x the baseline runtime in 2026; in other words: running ICON with the new ccn2cdnc scheme costs only 5% more node hours than running it without it.

Based on our scientific achievements in 2025, the number of activated cloud droplets is computed from cloud condensation nuclei (CCN) at specific supersaturations (Ssat) and updraft velocity at cloud base. The activation parameterization is a fitted version of the one from Kuba and Fujiyoshi (2006), CCN data are taken from CAMS reanalysis (Block et al., 2024). In order to safe computational costs only surface CCN at the required Ssat are being read into ICON. Additional preprocessing handles the data reduction and ensured a proper Ssat-scaling as is recommended by Anders et al (2026). The data can now be read in on daily or monthly frequencies, depending on the purpose of the simulation; the time interpolation is done within ICON. CCN are prescribed over height within ICON (with various options of vertical profiles), depending on the simulated height of the planetary boundary layer (PBL) which is computed in the tmx interface.

[1] `https://www.earthcarescience.net/projects/ecomip`

The adjusted processing of CCN ensures the best fit to observations over all required Ssat while increasing simulation speed due to a reduced complexity of CCN. Besides the changes on CCN processing, we updated the cloud base updraft velocity used for activation by including a scalable subgrid-scale vertical velocity which is computed from parameters in the tmx scheme. Together with the PBL-CCN feature, this creates a nice connection between tmx and the mig scheme.

We compared the performance and stability of the base simulation (1-moment and 2-moment) with the ccn2cdnc simulation (only 1-moment) using several updraft scaling factors and CCN input frequencies. Furthermore we compared the simulated CDNC with CDNC retrieved from Modis and compared radiative fluxes of both schemes (base and ccn2cdnc) with CERES retrievals. The statistical distributions of CDNC (global as well as over land and ocean) feature similar behavior as the retrieved ones, although they tend to show somewhat larger number concentrations (Fig. 2) The bias in radiative fluxes decrease when the ccn2cdnc scheme is used, especially in outgoing TOA SW radiation, however longer runs are needed in order to present statistically robust numbers. This is currently in progress and we plan to finish the long global AMIP simulations by the end of 2026.



ICON - CERES	ccn2cdnc	base
TOA net flux	2.77 W/m²	4.46 W/m²
TOA outgoing SW allsky	-4.61 W/m²	-6.02 W/m²
TOA outgoing LW allsky	2.05 W/m²	1.76 W/m²

Figure 2: Preliminary evaluation of ICON ccn2cdnc feature Left panel: Top of atmosphere (TOA) radiative flux biases from 15 days of ICON simulation (with ccn2cdnc and without (base), with maps of net flux bias shown on top; Right panel: CDNC distributions from MODIs retrievals (top) and ICON ccn2cdnc simulations pooled from several days.

Publications:

- Anders et al., 2026, '*Evaluation and improvement of CAMS-derived CCN number concentrations using in-situ measurements*' (Submitted to and accepted for publication by GMD, https://doi.org/10.5194/egusphere-2026-1404)

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

Developments for the coupled FESOM2 configurations during the period focused on addressing the limitations of FESOM2 itself that affect all of them: reproducibility in hybrid MPI + OpenMP mode, performance, mixed precision and GPU portability. This work was therefore done on the stand-alone FESOM2 ocean model, in close collaboration with T12 (see Sect. 3.3). As such resorces are also shared across them.

A notable advance in a coupled configuration, ICON–FESOM, is reported below. As was noted in the 2025 compute report, it was agreed during the last WarmWorld General Assembly that the focus of *WarmWorld–Better* should shift more towards developing and evaluating the **ICON–FESOM coupled configuration** for climate simulations across spatial and temporal scales.

During the current allocation period, substantial progress has been achieved in the development, calibration, and evaluation of the coupled ICON–FESOM system. Building on earlier sensitivity experiments, the R2B4–CORE2

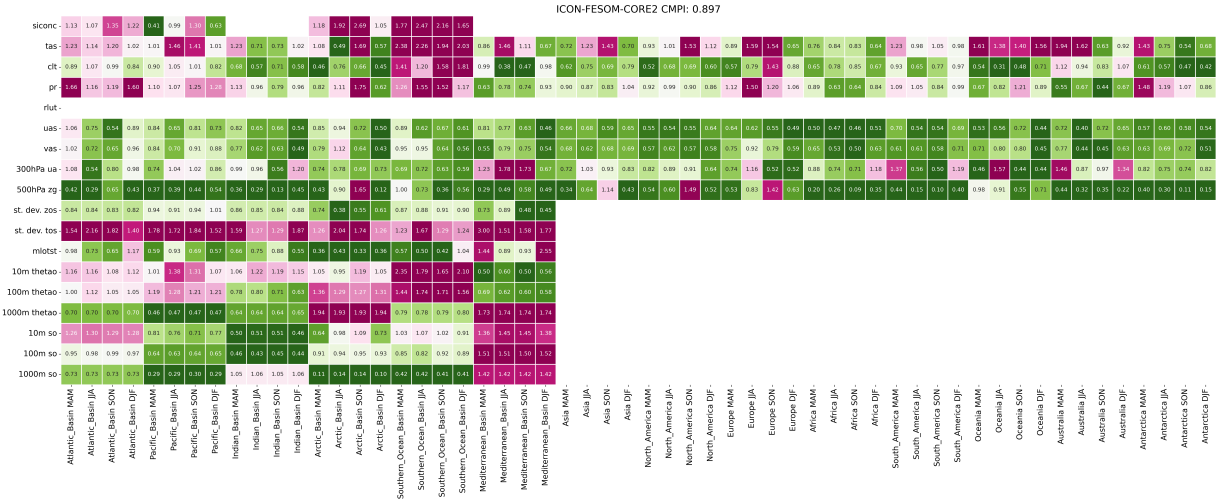


Figure 3: Performance indices for the R2B5-CORE2 historical simulations, presented as a fraction of the absolute error averaged over CMIP6 models. Values below/above 1 correspond to below (green)/above (pink) CMIP6 average biases.

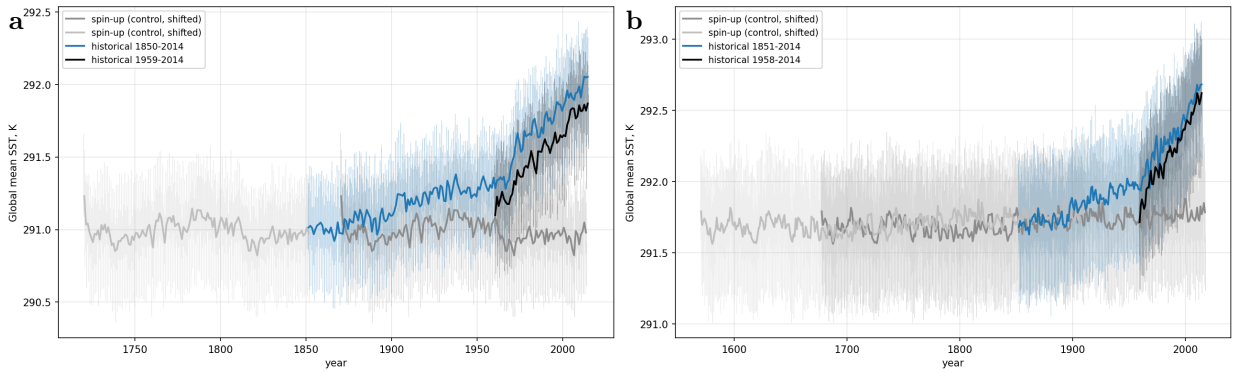


Figure 4: Global-mean sea surface temperature (K) in the R2B5–CORE2 configuration for the R2B5-CORE2 (a) and R2B4–CORE2 (b) configurations: pre-industrial control spin-up (grey, shifted) and the historical branches covering 1850–2014 (blue) and 1959–2014 (black).

and R2B5–CORE2 configurations have been further evaluated, with particular emphasis on the simulated mean climate state, systematic model biases, and response to historical forcing.

The evaluation of the coupled system using the performance index of Reichler and Kim (2008), following the methodology of Streffing et al. (2022), reveals differences in model performance across atmospheric and oceanic variables and geographical regions (Fig. 3). While several aspects of the simulated climate are reasonably well represented, notable biases remain, particularly in ocean temperature and some regional atmospheric fields. These diagnostics provide a basis for further model calibration and assessment of the coupled system.

An important achievement is the successful extension of the coupled simulations to the historical period (1850–2014) using both ICON R2B4–CORE2 and R2B5–CORE2 configurations. The evolution of global mean sea-surface temperature (SST) demonstrates a clear warming response to historical forcing in both configurations, particularly during the second half of the twentieth century (Figs. 4a and 4b). In contrast, the corresponding preindustrial control integrations exhibit relatively stable global mean SSTs, without a comparable long-term warming trend. These results demonstrate the ability of ICON–FESOM to reproduce a physically plausible global SST response to historical forcing, although further evaluation against observations is required to assess the magnitude and regional distribution of the simulated changes.

Overall, the simulations completed so far demonstrate the robustness and scientific potential of the coupled ICON–FESOM framework for climate applications. Further model optimization and evaluation are ongoing, with particular attention to improving the representation of Southern Ocean climate, including ocean temperature biases, air–sea heat exchanges, and sea-ice conditions. The previously proposed extensions towards higher-resolution configurations, including km-scale ICON–FESOM simulations, remain planned for subsequent development and testing.

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

During January–October 2026, two major tasks were performed:

Implementing aerosol-aware INP activation schemes in ICON-AES: To improve the representation of ice microphysics and cloud–radiation interactions, aerosol-aware INP parameterizations following DeMott et al. (2015) and Ullrich et al. (2017) were implemented in the ICON-AES one-moment microphysics scheme, replacing the default temperature-based formulation of Cooper (1986). The implementation is partially complete, with testing and validation currently underway. We estimated INP concentrations using mineral dust and black carbon mass mixing ratios from the CAMS aerosol climatology (Inness et al., 2019). The implementation was tested through global simulations and regional sensitivity experiments over the Tropical Atlantic during ORCESTRA-2024 at 10 km resolution. Ten regional experiments were performed using the three INP schemes, including aerosol perturbations with dust and black carbon concentrations scaled by factors of 100 and 1000, to assess their influence on cloud and radiative properties (Fig. 5). For this task, a total of 20 global simulations (R2B4 grid) and 10 regional simulations (ORCESTRA, 26–29 August 2024) were performed, consuming about 600 node hours (NH) and 0.5 TB of working storage.

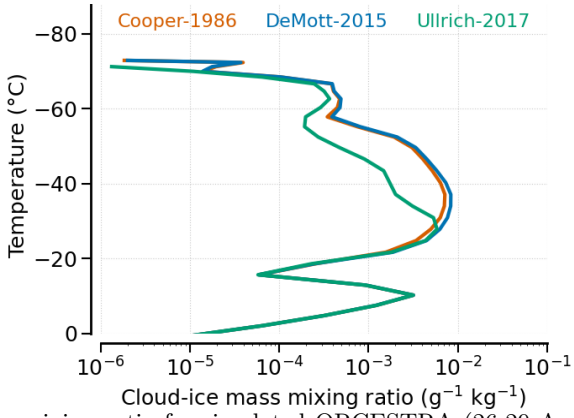


Figure 5: Mean cloud-ice mass mixing ratio for simulated ORCESTRA (26-29 August, 10 km resolution) convection from simulations with different INP schemes. The reference run uses the Cooper (1986) scheme (Orange line), while blue and green lines show analysis from the simulations with DeMott et al. (2015) and Ullrich et al. (2017) INP activation schemes, respectively.

Data-driven approach for ice multiplication: Secondary ice production (SIP) can influence cloud microphysics and cloud–radiation interactions (Zhao and Liu, 2021; Waman et al., 2026), but is not explicitly represented in simpler one-moment microphysics schemes. To develop a computationally efficient formulation for ice multiplication, the analysis was extended from four to seven deep convective cloud (DCC) cases, including two continental cases from STEPS and one marine case from ORCESTRA. The Ice Enhancement Factor (IEF), defined as the ratio of ice particle number to activated ice nuclei at cloudy points (Waman et al., 2022), was used to quantify ice multiplication. Each case was simulated under four contrasting CCN conditions (marine, intermediate, continental, and polluted continental), resulting in 28 simulations. An additional 14 simulations examined the sensitivity of IEF to INP concentrations by perturbing initial INPs within observational uncertainty. All simulations have been completed, and the relationships between IEF and storm properties (temperature, vertical velocity, and supercooled liquid and ice mass contents) are currently being analyzed. A total of 42 simulations (28 CCN and 14 INP sensitivity experiments) consumed approximately 1,680 NH, with each simulation requiring about 0.2 TB (total 10 TB) of working storage.

Publication: '*Impacts of Secondary Ice Production on the Microphysics and Dynamics of Deep Convective Clouds in Different Environments*' (Published in Atmospheric Chemistry and Physics (ACP), Waman et al., 2026)

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

A study analyzing the impact of ParFlow-calculated subsurface hydrology on the atmosphere began in the previous compute period and continued in the current one. These simulations, along with subsequent data processing and analysis, consumed all granted GPU resources. Additionally, some simulations were shifted to CPUs so that more of the granted resources (up to the requested amount) were used. Compared with the

previous study described in the mid-report of June 2026, we altered the setup slightly. Instead of running a month continuously, the atmosphere was reinitialized daily using DWD's global initialized analyses. The land was restarted (cycled) from the previous day's data. This lets the land variables develop freely while forcing the atmosphere into a realistic state, comparable to observations.

The simulation results confirm the findings reported in June 2026. Mountains show a wet bias in ICON-standalone, as cooler high-altitude conditions reduce evapotranspiration. As a result, these cells are even moister than other regions in ICON-standalone. In ICON-ParFlow, this changes dramatically. As water is transported laterally, it flows downhill and dries the mountain ridges. In addition, water fluxes converge in valleys, increasing soil moisture. Grid cells in flatlands or on hillslopes do not show such strong differences, suggesting that lateral water exchange drives these (see Fig. 6). However, the results also underline that changes appear especially in dry, water-limited regions. In northern Europe, where soil tends to be closer to saturation or contains frozen water, lateral transport becomes less influential than natural variability.

The effects can be traced along the process chain from soil moisture to atmospheric variables. For instance, on dry mountain ridges, energy partitioning favors sensible heat transfer over latent heat transfer because evapotranspiration is reduced. This increases the near-surface temperature and decreases the near-surface humidity (see Fig.7). Curtain plots of the instantaneous specific water vapor and temperature fields show how these surface differences propagate into the atmosphere. The boundary-layer height also responds to surface changes. However, comparing with EarthCARE data (e.g., cloud cover, cloud base height, or cloud top height) is not trivial, as the white-background noise is too strong to distinguish model-induced differences from natural atmospheric variability. We are currently summarizing the study's findings and will publish them in a peer-reviewed journal next month.

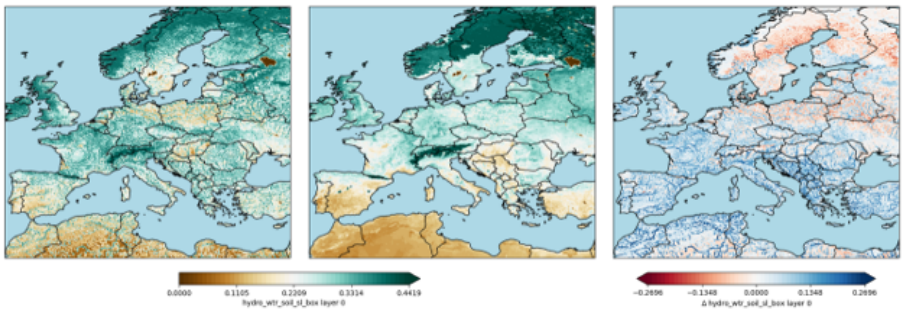


Figure 6: Volumetric soil water content in the upper soil layer for ICON-ParFlow (l.) and ICON-standalone (m.), and their difference, ICON-ParFlow - ICON-standalone (r.).

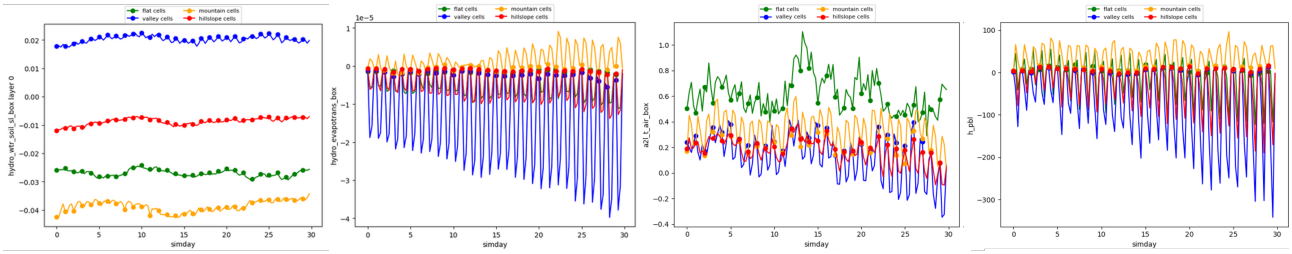


Figure 7: Temporal evolution of spatial category means for differences (ICON-ParFlow - ICON-standalone) in volumetric soil water content (l.), evapotranspiration (m.l.), near-surface temperature (m.r.), and boundary layer height (r.). The categories are obtained by grouping cells based on their topographic position index (TPI) and slope. A strongly negative TPI indicates valleys (blue), while mountain ridges (yellow) have a strong positive TPI. TPI. Significant slopes indicate hillslopes (red), and flatland cells (green) have small slopes and TPI values close to 0.

1.6 T19) 2-moment cloud microphysics *(KIT, U Cologne, U Leipzig, DWD, MPI-M)*

Development of an ICON model configuration to study aerosol-cloud interactions with 2-moment cloud microphysics: As efforts are made by various countries across the world to curb their emission of anthropogenic aerosols, it is becoming increasingly important to understand what impacts such reductions may have on the radiative properties of clouds, and therefore, the net radiative budget of the planet in a future cleaner world. As a step towards gaining a more robust quantitative understanding of aerosol-cloud interactions, a configuration of ICON was developed in which the interactive aerosol module, HAM-lite (Weiss et al. (2025)), was coupled with the 2-moment cloud microphysics scheme of Seifert and Beheng (2006). This can facilitate

experiments with perturbed aerosol concentrations, thereby allowing detection-attribution based studies of the impacts of anthropogenic air pollution on climate. Month-long simulations at R2B8 ($\sim$ 10 km) resolution, initialized on August 01, with present-day (PD) and pre-industrial (PI) aerosol emissions demonstrate changes in global-mean longwave ($\Delta LWCRE$) and shortwave ($\Delta SWCRE$) cloud radiative effects of 0.3 W m^{-2} and -2.7 W m^{-2}, respectively (see Fig. 8), yielding a net change of -2.4 W m^{-2} which is higher than current estimates. Further tuning, testing and development at higher resolutions would be undertaken with additional resources in future.

Resources used: 35000 NH and 40 TB work storage.

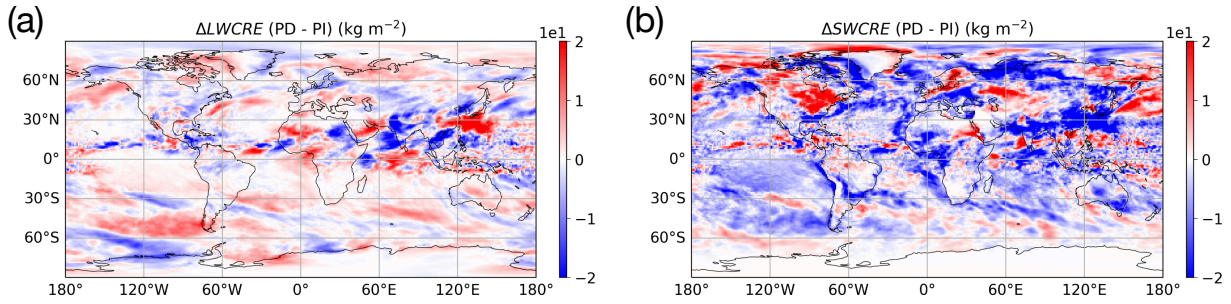


Figure 8: Changes in time-averaged top-of-the-atmosphere (a) longwave cloud radiative effect ($\Delta LWCRE$) and (b) shortwave cloud radiative effect ($\Delta SWCRE$); the differences have the convention of PD minus PI, and fluxes are positive downwards such that a postive change reflects warming and negative implies cooling.

Representing Energy Conservation during Precipitation Sedimentation in ICON Precipitation sedimentation transports hydrometeors across model levels, carrying enthalpy with them. Without an explicit correction, this transport can introduce a thermodynamic inconsistency, as temperature is not updated to account for the energy deposited or removed by falling hydrometeors. In ICON's two-moment microphysics scheme, hydrometeor sedimentation is treated using explicit and semi-implicit solvers. To account for the associated energy transport, we implemented an enthalpy correction in the explicit sedimentation solver, which has received scientific approval through a merge request to the ICON-NWP. The correction was tested using an idealized Weisman–Klemp supercell case (6–24 h) on Horeka HPC at KIT, showing approximately 0.1 K cooling in the upper troposphere (Fig. 9). Implementation in the semi-implicit solver is ongoing, for which eleven limited-area simulations were performed for a marine convection case observed during ORCESTRA over the Tropical Atlantic. In total, these simulations consumed approximately $11 \times 40 = 440$ NH.

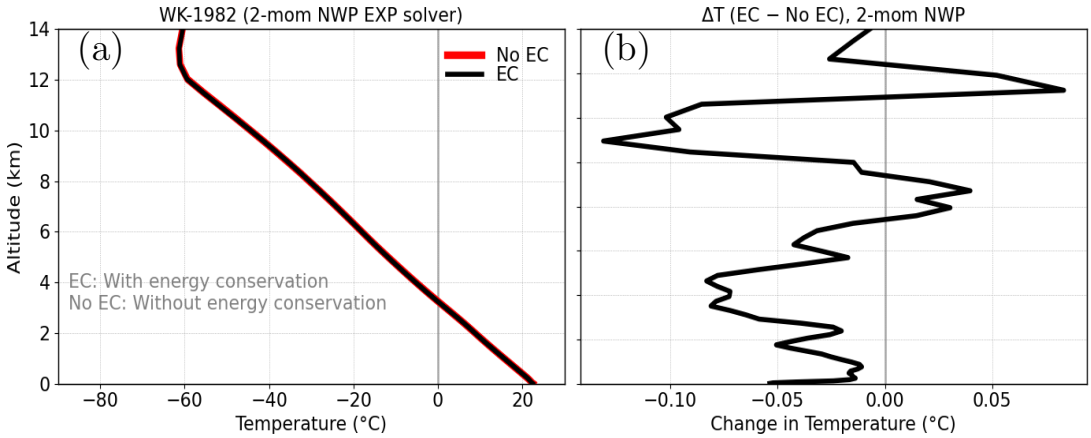


Figure 9: Simulated mean (a) temperatures with (black line) and without (red line) energy conservation (EC) applied, and (b) temperature change due to energy conservation during precipitation sedimentation. The simulations are performed with the Weisman-Klemp idealized test case.

2 Module Easier

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

Upgrading zarrcurator (formerly zarranalyzer) to handle zarr v2 as well as zarr v3 including data trees The ICON model output for km scale ESM simulations has been adapted to the latest format provided

by the version 3 of the zarr package. Now the entire output of a single km scale ESM simulation is represented as a data tree with each node corresponding to a combination of freqeuncy(P1D,PT1H etc.) and the method (inst or mean). Each of these nodes further contain childeren, with each child corresponding to a zoom level of the HEALPix grid hierarchy. Furthermore, the data is now stored in shards(collection of chunks) which result in a lesser number of files in the file system and thereby reducing the 'inode' burden on the operating system. To curate these data trees into the tape drive, the zarrcurator was refactored and tested with different test cases consisting of synthetic zarr stores as well as stores from real km scale simulations running into hundreds of TB.

Resources used: 3026 NH and 117.80 TB work storage.

During the reporting period, we conducted debugging and testing using Hiopy in large-scale environments. This process required 1600 CPU node hours. Specifically, Hiopy was utilized to generate output for the XSPIES run on JUPITER. To optimize performance and resolve issues, we replicated these settings within levante, conducting several large yet brief runs.

3 Module Faster

3.1 T9) ICON performance benchmarks *(DKRZ)*

The mkexp tool was successfully extended to also enable the generation of hybrid coupled runscripts with the atmosphere being computed on the GPUs and the ocean and output writing on the CPUs of the same nodes on both the Levante GPU and Dolpung partitions. Since production setups are in general coupled setups and in the hybrid configuration the ocean computations create almost no overhead, we decided to change our benchmark configuration to a coupled setup with a R2B8 (10 km) atmosphere and a R2B9 (5 km) ocean. Instead of running the benchmarks manually every two months, they were set up to run as weekly gitlab CI jobs and the results are stored in a monitoring dashboard. Since these CI jobs are running on the Dolpung partition no further benchmarks were done on Levante GPUs. Over the past 3 months the throughput for this configuration on 11 Dolpung nodes was between 274 and 283 simulated days per day.

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

The node hours were used for a) extending the C++/Kokkos infrastructure of ICON by defining new abstractions which encapsulate CPU/GPU parallelism concepts, like `icon::block_loop`, b) to fully port the "Cloud migration scheme" which uses the already ported graupel microphysics and c) to port 2/3 of the mathematical library, iconmath. We also note progress on porting the turbulence mixing framework. Recent performance results from running the R2B8 AES AMIP experiment show a speedup of the ported code over the legacy Fortran and are depicted in Table 1.

Table 1: Fortran vs. C++/Kokkos Runtimes, AES AMIP R2B8, 3 simulated hours (total avg over ranks, s), `timers_level`=12, mean of 10 runs each; Levante CPU: Intel 2025.3.2 compiler stack, 32 nodes / 1024 MPI ranks. Levante GPU: NVHPC 24.7 compiler stack, Fortran with OpenACC, 8 nodes / 32 MPI ranks

	Levante CPU				Levante GPU			
Timer	Fortran	C++/Kokkos	Δ	Speed-up	Fortran	C++/Kokkos	Δ	Speed-up
total	313.661	303.271	−3.3 %	(noise)	117.471	117.062	−0.3 %	(noise)
integrate_nh	302.293	291.098	−3.7 %	(noise)	110.572	109.875	−0.6 %	(noise)
nh_solve	101.229	101.285	+0.1 %	1.00×	27.917	27.801	−0.4 %	1.00×
aes_bcs	22.857	14.402	−37.0 %	(noise)	20.547	20.863	+1.5 %	(noise)
aes_phy	108.935	106.039	−2.7 %	1.03×	43.081	42.192	−2.1 %	1.02×
interface_aes_rad	45.847	45.875	+0.1 %	1.00×	15.078	15.068	−0.1 %	1.00×
interface_aes_tmx	53.443	52.688	−1.4 %	1.01×	25.186	24.727	−1.8 %	1.02×
interface_cloud_mig	**7.062**	**5.613**	**−20.5 %**	**1.26×**	**2.362**	**1.936**	**−18.0 %**	**1.22×**
cloud_mig	**6.888**	**5.368**	**−22.1 %**	**1.28×**	**1.990**	**1.468**	**−26.2 %**	**1.36×**
satad	**0.895**	**0.807**	**−9.8 %**	**1.11×**	**0.165**	**0.157**	**−4.8 %**	**1.05×**
graupel	**5.676**	**4.364**	**−23.1 %**	**1.30×**	**1.477**	**0.930**	**−37.0 %**	**1.59×**
interface_aes_wmo	0.515	0.502	−2.5 %	1.03×	0.125	0.125	0.0 %	1.00×

The performance is credited to using a local memory pool which reuses memory allocations. This is especially useful for GPU memory and avoids allocating/deallocating the arrays in each time step. The impact on the memory footprint is summarized in Table 2.

Table 2: Memory use of the runs in Table 1 (mean ± sd over runs). Host: SLURM MaxRSS of the largest task (rank 0). Device: used memory of rank 0's GPU (80 GB), as printed by ICON every time step; identical in all runs of a build.

	Levante CPU			Levante GPU		
Metric	Fortran	C++/Kokkos	Δ	Fortran	C++/Kokkos	Δ
Host MaxRSS (GiB)	2.481 ± 0.082	2.471 ± 0.094	−0.4 %	47.074 ± 0.095	47.217 ± 0.071	+0.3 %
Device used, time step 1 (GB)	–	–	–	36.475	36.477	+0.0 %
Device used, peak (GB)	–	–	–	**69.262**	**70.568**	**+1.9 %**

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

During 2026 the allocation was used mainly to make FESOM2 faster and reproducible across configuration, on CPUs and, for the first time, on GPUs. Most of this work is merged and part of the FESOM2 releases 2.8.1–2.8.3, which are also used for upcoming the CMIP7 simulations.

Highlights of how the compute time was used:

1. **OpenMP coverage and a better SSH-solver preconditioner improve scaling.** Most of the time step now runs threaded, including lock-free sea-ice (mEVP/aEVP) and split-explicit solvers. A per-rank scatter replaced the global broadcast of atmospheric forcing. Together these let CORE2 scale further. On 27 nodes (3456 cores) throughput rose from 93 SYPD (pure MPI) or 224 SYPD (best hybrid layout) to 286 SYPD. Before, pure MPI stopped scaling at 864 cores; now throughput keeps rising to 3456 cores and holds up to 6912 (Fig. 10a). A symmetric preconditioner for the sea-surface-height (SSH) solver is now in the release. A new Chebyshev preconditioner with a single-reduction CG, on a development branch, cuts solver iterations from 88 to 33. At 3456 cores it shortens the whole time step by 25 % (Fig. 10b).
2. **Mixed precision.** FESOM2 now runs in single precision, with the accumulations and layer thicknesses that matter for conservation kept in double precision. On one CORE2 node this is 1.58× faster than double precision. Heat is conserved as well as in double precision, and volume and salt 3–4 orders of magnitude better than in plain single precision. The gain is 1.3–1.6× up to 512 cores. It falls to about 1.08× beyond 1728 cores, where communication latency dominates; the best throughput is 309 SYPD at 3456 cores (Fig. 11).
3. **GPU portability.** An OpenACC port (merged in 2.8.3) runs the full CORE2 model on NVIDIA A100 GPUs. Every porting step was verified against the previous one (bit-identical, or round-off where a math function moved to the GPU). Throughput rose from 1.2 to 17.8 SYPD on one A100 and from 3.4 to 32.1 SYPD on one 4-GPU node. One GPU now beats a 128-core CPU node with the Intel compiler (14.7 SYPD), and a GPU node matches a CPU node with the NVIDIA compiler (34.5 SYPD). The first runs on the 7.4-million-node NG5 mesh scale from 2 to 8 GPU nodes (15.7 to 3.8 s per step) but are not yet competitive with CPUs (32 nodes: 0.53 s per step). They set the targets for 2027.
4. **Modularity and testing.** In the released code, step diagnostics run in a separate post-step phase, and new budget diagnostics check volume, heat and salt conservation online. Module dependencies were cleaned up, and CI now runs thread-invariance, conservation, single-precision and idealised-configuration tests on every pull request. On a development branch the sea ice runs as a library with its own stand-alone driver, needing 50 instead of 95 module files. It can also run as a separate executable next to the ocean, bit-identical to the single executable. New sea-ice and tracer "dwarfs" serve as small testbeds, including for FESOM3(x).

4 Module Smarter

4.1 T14) ELPHE *(GEOMAR)*

In ELPHE, we are extending the ocean analysis toolbox "Parcels v4" (https://parcels-code.org/blog/parcels-v4) (Lange and van Sebille, 2017; Delandmeter and van Sebille, 2019) to accurately estimate particle transports on unstructured ocean meshes such as those used by FESOM2 and ICON. We repeated particle tracking experiments for the Amazon river plume dispersion on Glorys12v1 reanalysis (in native C grid, 1/12 degrees resolution) with 15.2 million particles after discovering a major bug in the initial run leading to large amounts of particle losses. To fix it, we implemented a kernel with adaptive time stepping that correctly handles particles near the mesh boundaries. This reduced particle loss to 0.1%.

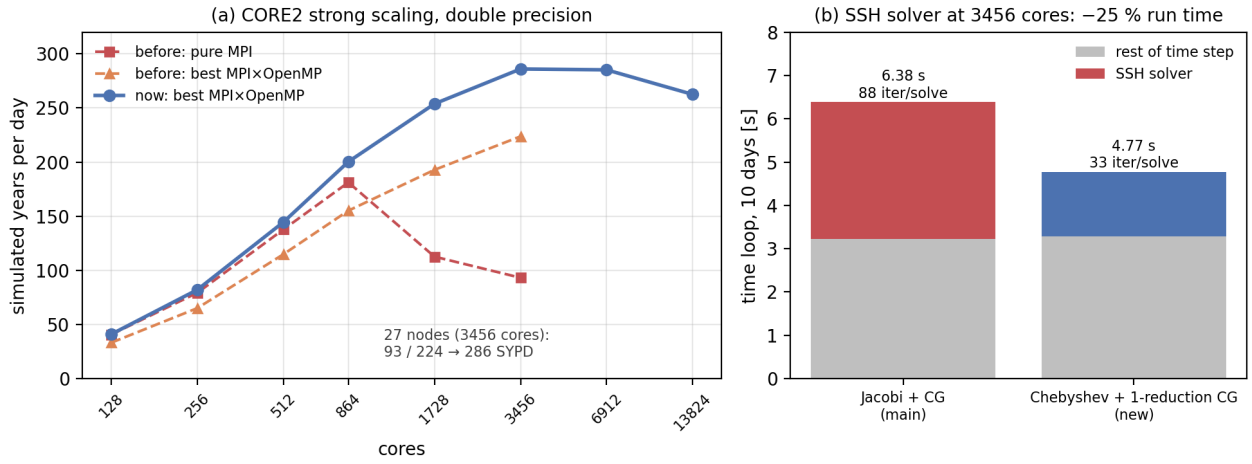


Figure 10: FESOM2 CORE2 mesh on Levante (20 model days, no output). (a) Strong scaling in double precision, simulated years per day for the best MPI×OpenMP layout at each core count, before and after the OpenMP and forcing-scatter work. (b) Time loop for 10 days at 3456 cores with the released Jacobi-preconditioned CG and with the new Chebyshev preconditioner and single-reduction CG; the SSH-solver share is coloured.

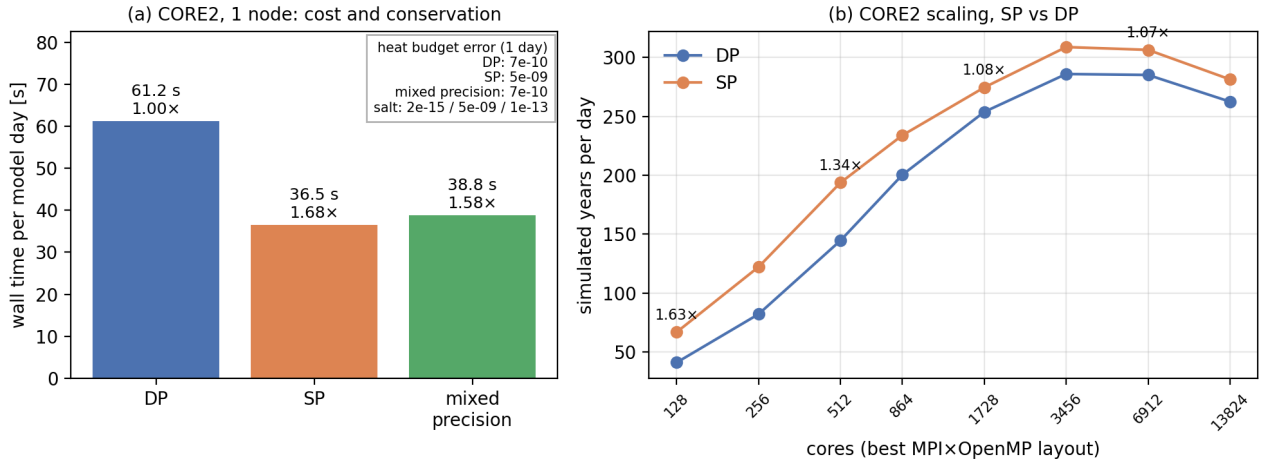


Figure 11: Mixed precision in FESOM2, CORE2 mesh. (a) Wall time per model day on one node for double precision (DP), plain single precision (SP) and mixed precision, with the heat and salt budget errors after one day. (b) Strong scaling of SP and DP (best layout per core count); labels give the SP/DP speed-up.

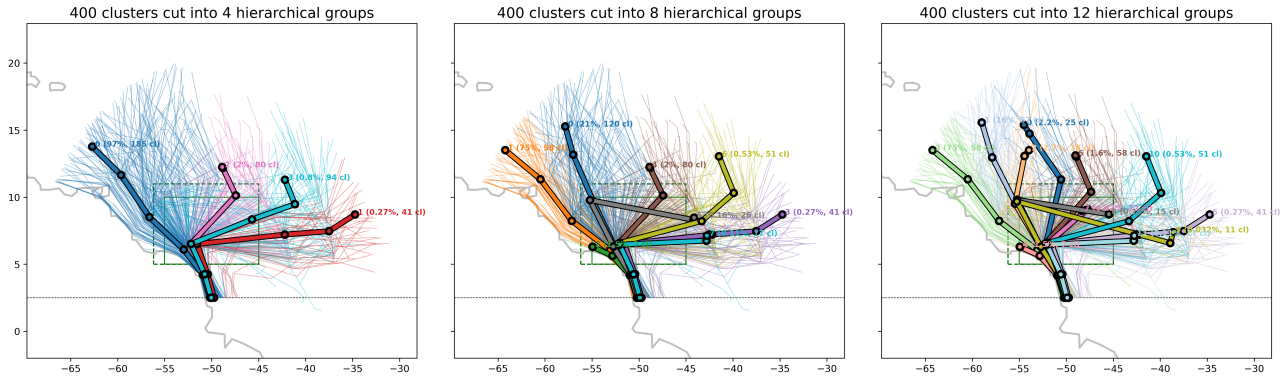


Figure 12: Freshwater pathways from the Amazon outflow identified by clustering particle trajectories. The 15.2 million particle trajectories were grouped into 400 k-means clusters, which were then merged by hierarchical clustering into 4 (left), 8 (middle) and 12 (right) groups. Thin lines show the mean trajectories of the 400 clusters, coloured by group; thick lines show the mean pathway of each group. Labels give the group number, the percentage of particles in the group, and the number of clusters it contains (cl). The dotted line marks 2.5° N, where the particles were regrouped. The solid green box marks the North Brazil Current (NBC) retroflection region.

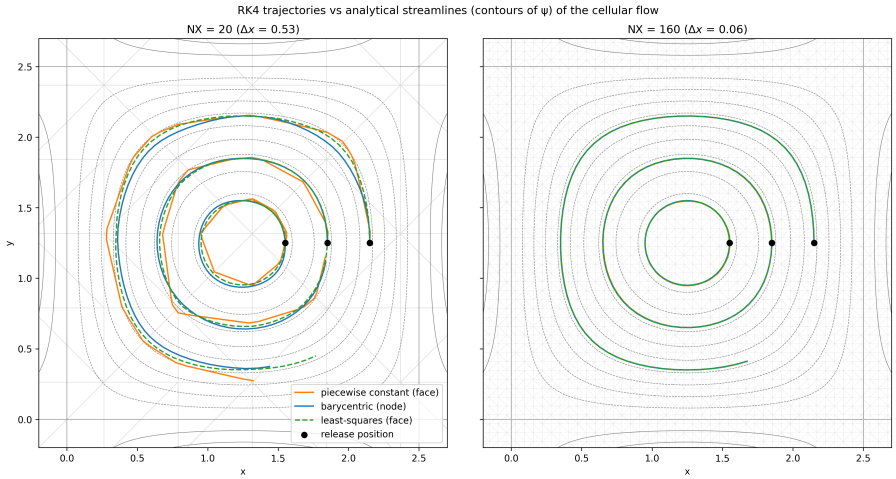


Figure 13: Test of the three unstructured-grid interpolators implemented in Parcels v4 (barycentric, piecewise constant and least-squares) in an idealized cellular flow, $u = U_0 \sin kx \cos ky$, $v = -U_0 \cos kx \sin ky$, for which every particle should circle along a closed streamline. Coloured lines show particle trajectories computed with Parcels (fourth-order Runge–Kutta); grey contours show the analytical streamlines, and black dots mark the release positions. Left: coarse mesh (20×20 nodes, $\Delta x \approx 0.53$); right: fine mesh (160×160 nodes, $\Delta x \approx 0.06$). On the coarse mesh, the piecewise-constant scheme produces trajectories that drift off their streamlines, while the barycentric and least-squares schemes follow them more closely. On the fine mesh, all three schemes follow the analytical solution.

Furthermore, we trained a k-means clustering algorithm with 400 clusters on a subset of 500,000 particles, drawn from the 15.2 million particles that cross 2.5° N in the tropical Atlantic after being released at the Amazon river mouth. The aim was to understand how mesoscale processes shape freshwater pathways from the Amazon river plume. The analysis revealed 12 distinct pathways (Figure 12). Publication of the results is in progress.

To isolate mesoscale variability, we developed and tested several filters employed to the input data for Parcels: spatial filters (Gaussian and implicit) and temporal filters. We ran them on GPUs and selected temporal filtering because it handles coastal boundaries better. This allows us to assess the role of the mesoscale during extreme events in years with anomalous Amazon river runoff.

We have also implemented three interpolation schemes for unstructured grids in Parcels v4: piecewise constant, barycentric, and least-squares. We are currently testing them in idealized configurations, comparing the results against analytical solutions for idealized flows (Figure 13). In addition, we are testing an Eulerian tracer as a reference solution for a double-gyre model intercomparison (ICON, FESOM2 and NEMO), which will serve as a benchmark for evaluating the Parcels v4 interpolators.

Finally, we did some preliminary testing in a series of one-year sensitivity experiments with a global FESOM2 setup. Refinement was focussed on the tropical Atlantic (locally 5 km, 120 levels) tuning the constant air–sea transfer coefficients for momentum, latent heat and sensible heat (cd, Ce, Ch) to reduce the global and tropical Atlantic SST biases for future use in Amazon river runoff studies.

4.2 T16) TurbO-Arctic *(AWI and U Cologne)*

During the current allocation period, CPU resources were primarily dedicated to:

- Running nested ICON-LAM simulations centered around the research vessel *Polarstern* during the Arctic MOSAiC campaign in winter 2019/2020
- Running a baseline pan-Arctic LAM simulation with the default ICON configuration for the entire MOSAiC winter

Early this year, the 5 km ICON-LAM simulations completed in the previous allocation period were extended to a nested configuration with a 2.5 km inner domain (Fig. 14a). ICON tends to generate excessively stable temperature inversions over sea ice in polar winter (Fig. 14b). In particular, ICON exhibits a pronounced cold bias in surface temperature, including episodes of runaway surface cooling lasting several days (Fig. 14c). These are triggered by excessive radiative surface cooling resulting from deficiencies in the representation of supercooled liquid-bearing clouds.

Over the past few weeks, a pan-Arctic baseline simulation (domain north of 65°N) at 5 km horizontal resolution (R2B9) was successfully set up and run for the entire MOSAiC winter (Fig. 15).

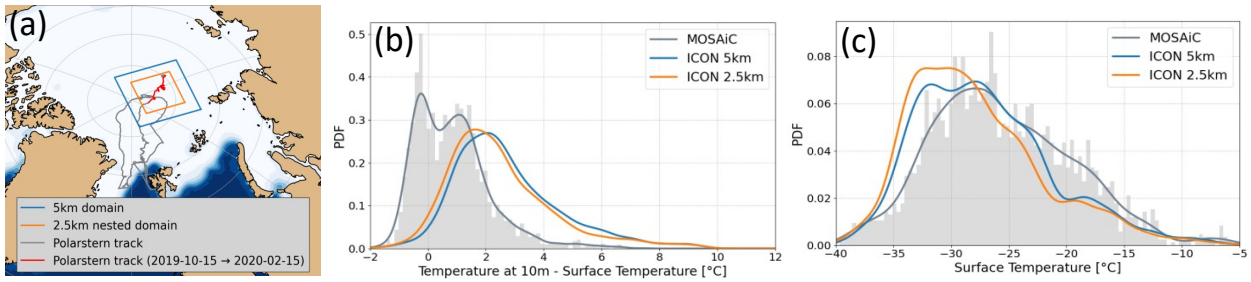


Figure 14: (a) Track of research vessel *Polarstern* during MOSAiC (gray) and the nested LAM domains (blue, orange) centered around the ship's position from October 2019 to February 2020. Histograms of near-surface stability (b) and sea ice surface temperature (c) for MOSAiC observations and the ICON domains.

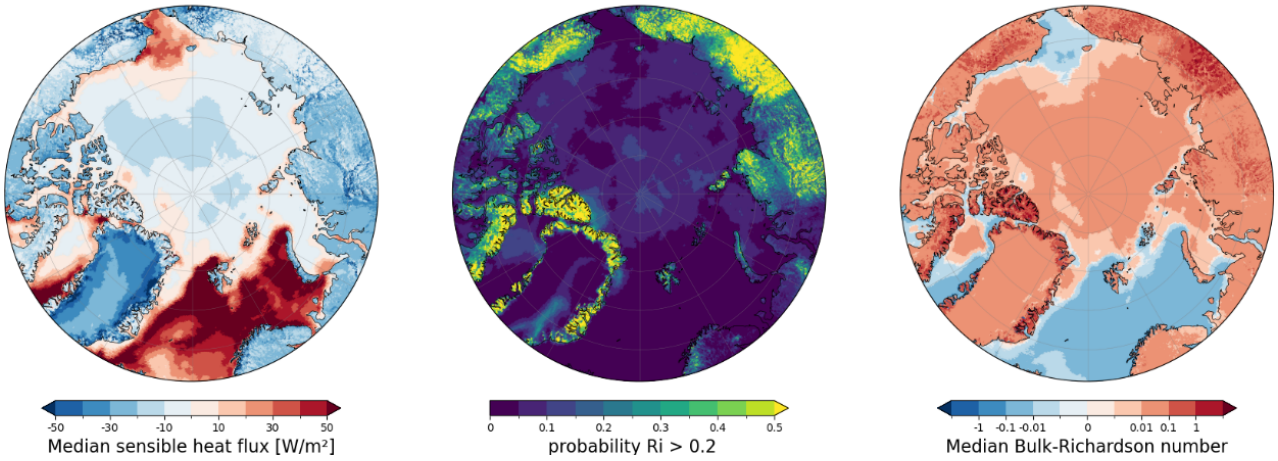


Figure 15: First results of a baseline pan-Arctic LAM simulation at 5 km horizontal resolution (R2B9) covering the MOSAiC winter, 1 October 2019 to 10 March 2020. Shown are the spatial distributions of the median sensible heat flux (left), the occurrence frequency of strongly stable conditions with Bulk Richardson number greater 0.2 (middle), and the median surface Bulk Richardson number (right).

Furthermore, simulations of MOSAiC days with a spatial resolution of 20 m, available at the university of Cologne, were post-processed using coarse-graining to different target-resolutions ranging from 80m to 5 km to extract turbulence closure coefficients and study their variability to inform ICON parameterisation development.

For the remainder of the current allocation period, resources will be further used to test the stochastic extension of the turbulent surface fluxes within this experimental setup. Over the course of the year, this extension was adapted to the surface bulk scheme in ICON using MOSAiC data and has just been implemented in the model. It allows a more realistic representation of turbulent surface mixing under stable and strongly stable conditions, which are common over polar regions and for which the current scheme produces very weak surface fluxes (Fig. 15).

4.3 T17) ProImpact *(GERICS and KIT)*

During the current allocation period, CPU resources were mainly used to develop and test ICON/ComIn plugins, data processing workflows, and the ProImpact web-based platform. In the previous reporting period, we established the basis for combining runtime diagnostics, data streaming, and catalog based data access within a unified web platform consisting of multiple dashboards for data access, scientific analysis, and visualization. During the current period, we focused on extending these components and improving their integration. For the ComIn plugins, we continued development high frequency precipitation diagnostics, wind interpolation, and thermal comfort indices. We also tested their storage requirements on global and regional grids. To reduce the amount of stored data, we worked on reusable tools that allow users to select specific regions of interest and define thresholds below which data are not retained. In addition, we simplified the configuration of ComIn plugins using YAML files and intergated the setup of ComIn and Hiopy into the Started Package for ICON-CLM Experiments (SPICE).

In parallel, we made substantial changes to the ProImpact web interface and its software architecture. We reorganized the code to make it easier to maintain, extend, and integrated new components. We introduced a centralized logging system, improved the management of computational tasks, and separated user sessions to support multiple users working at the same time. The *DataHub* was extended to support local, online, and server hosted datasets, including ICON/ComIn outputs in different formats (e.g., zarr, nc, ncz). We also introduced a new workspace that allows users to manage datasets and reuse them across different scientific applications. In addition, we expanded the scientific functionality with new plugins, including UrbanPlot and Radiosondes, and imporved existing climate diagnostics, visualization tools, automated tests, and the user interface. The web application was deployed on a DKRZ virtual machine and made available at https://www.proimpact.dkrz.de/. User authentication was integrated with the centralized DKRZ authentication system through the Freva login page, so access to the platform requires a valid DKRZ account.

One of the main achievements during this reporting period was the closer integration of the ICON/ComIn developments with the ProImpact web interface through the *DataHub*. We added new functionality to inspect ComIn datasets and generate configuration files (namelists) for the ComIn plugins. We also introduced a dedicated LOCO component that allows users to access streaming data directly through the web interface. In parallel, we made progress in transferring model derived statistics, such as histograms and mean values, from ICON simulations to the platfrom. This functionality was succesfully tested, although further development and integration were still ongoing. Together, these developments strengthened the connection between model configuration, runtime diagnostics, and interactive climate data analysis.

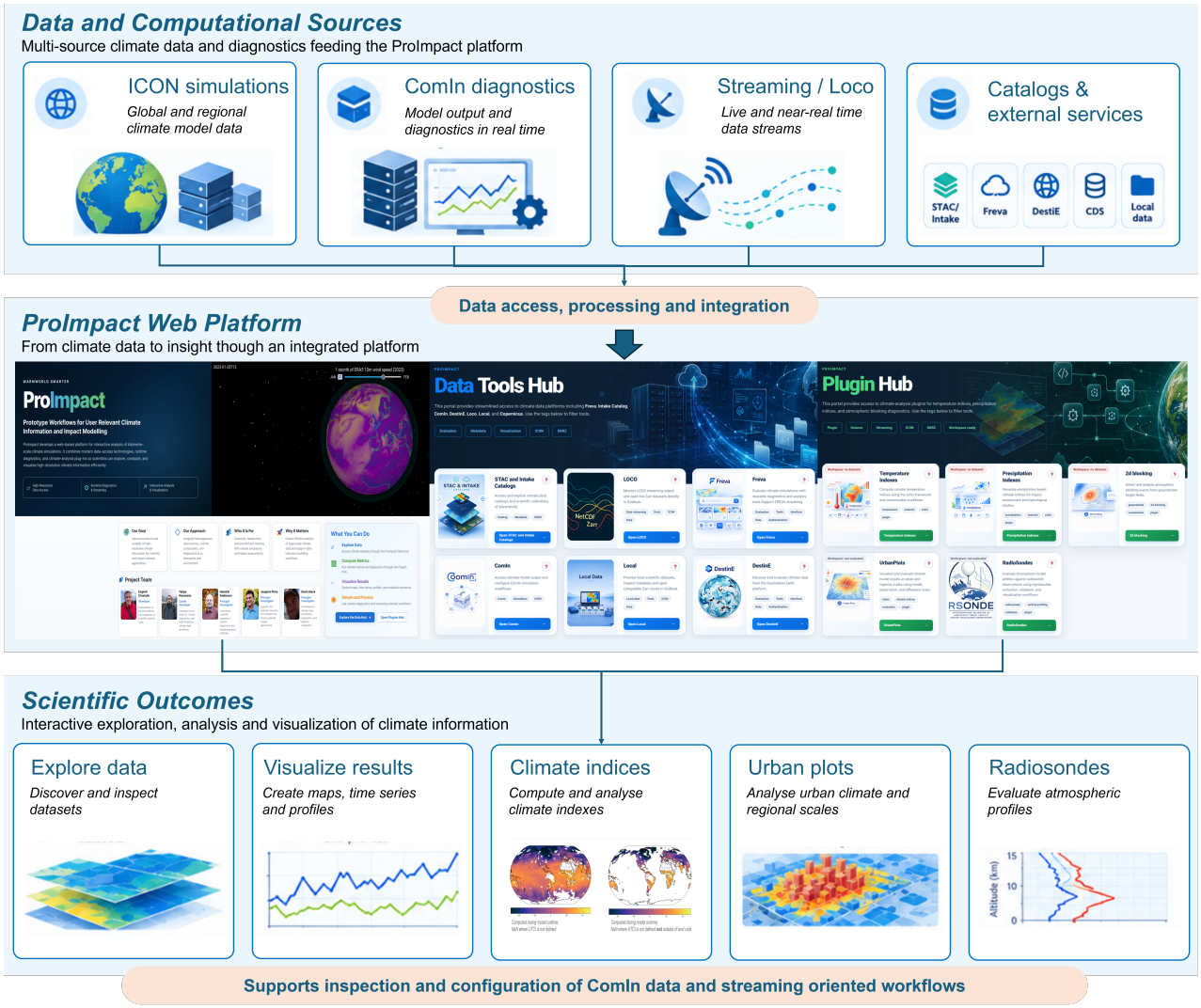


Figure 16: Overview of the ProImpact web platform, integrating climate data sources, ICON/ComIn diagnostics, streaming services, and scientific analysis workflows

4.4 T18) IconRep *(ECMWF and FZJ)*

Under the WarmWorld project and the granted compute resources, we extend the capabilities of the WeatherGenerator model to correctly respond to climate forcings and provide well-calibrated, multi-decadal forecast ensembles. To do so, we train the WeatherGenerator model with data from the free-running EERIE climate simulations. The EU-funded EERIE project (European Eddy-RIch Earth System Models, https://eerie-project.eu) simulates climate scenario over century-long time scales at high spatial resolution (i.e. 9km atmosphere and 4.4km ocean). We are converting this climate data into machine learning-ready datasets on which we can train our model under varying scenarios.

The native high spatial resolution is interpolated to 1-degree to pre-train the model initially. We can leverage the WeatherGenerator's native capability to work on multiple input-/output-datasets at different resolutions to test different configurations in feeding the model with data from atmosphere, ocean, and sea ice. Furthermore, different approaches to condition the model on various climate forcings are tested.

We also used the granted compute resources to train and evaluate a generative model under sea-surface-temperature and sea-ice-cover forcings. We participated in the AIMIP initiative where different machine-learning based climate emulators are compared following the Atmospheric Model Intercomparison Project (AMIP). We tested the capabilities of the same generative model by removing the monthly-mean forcings in a free-running simulation and found that the unforced configuration also performs well compared against numerical climate models, e.g. MPI-ESM1.2-HR.

References

W. A. Cooper. Ice initiation in natural clouds. *Meteorological Monographs*, 21(43):29–32, 1986. doi: 10.1175/0065-9401-21.43.29.

Philippe Delandmeter and Erik van Sebille. The Parcels v2.0 Lagrangian framework: new field interpolation schemes. *Geoscientific Model Development*, 12(8):3571–3584, August 2019. ISSN 1991-959X. doi: 10.5194/gmd-12-3571-2019. URL `https://gmd.copernicus.org/articles/12/3571/2019/`. Publisher: Copernicus GmbH.

P. J. DeMott, A. J. Prenni, and coauthors. Integrating laboratory and field data to quantify the immersion freezing ice nucleation activity of mineral dust particles. *Atmospheric Chemistry and Physics*, 15(1):393–409, 2015. doi: 10.5194/acp-15-393-2015.

Antje Inness, Melanie Ades, Anna Agustí-Panareda, Jérôme Barré, Anna Benedictow, Anne-Marlene Blechschmidt, Juan Jose Dominguez, Richard Engelen, Henk Eskes, Johannes Flemming, et al. The cams reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, 19(6):3515–3556, 2019. doi: 10.5194/acp-19-3515-2019.

Michael Lange and Erik van Sebille. Parcels v0.9: prototyping a Lagrangian ocean analysis framework for the petascale age. *Geoscientific Model Development*, 10(11):4175–4186, November 2017. ISSN 1991-959X. doi: 10.5194/gmd-10-4175-2017. Publisher: Copernicus GmbH.

Thomas Reichler and Junsu Kim. How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, 89(3):303–312, 2008. doi: 10.1175/BAMS-89-3-303.

A. Seifert and K. Beheng. A two-moment cloud microphysics parameterization for mixed-phase clouds. part 1: Model description. *Meteorology and Atmospheric Physics*, 92:45–66, 2006. doi: 10.1007/s00703-005-0112-4. URL `https://doi.org/10.1007/s00703-005-0112-4`.

J. Streffing, D. Sidorenko, T. Semmler, L. Zampieri, P. Scholz, M. Andrés-Martínez, N. Koldunov, T. Rackow, J. Kjellsson, H. Goessling, M. Athanase, Q. Wang, J. Hegewald, D. V. Sein, L. Mu, U. Fladrich, D. Barbi, P. Gierz, S. Danilov, S. Juricke, G. Lohmann, and T. Jung. Awi-cm3 coupled climate model: description and evaluation experiments for a prototype post-cmip6 model. *Geoscientific Model Development*, 15(16):6399–6427, 2022. doi: 10.5194/gmd-15-6399-2022. URL `https://gmd.copernicus.org/articles/15/6399/2022/`.

R. Ullrich, C. Hoose, and coauthors. A new ice nucleation active site parameterization for desert dust and soot. *Journal of Atmospheric Sciences*, 74(3):699–717, 2017. doi: 10.1175/JAS-D-16-0074.1.

D. Waman, S. Patade, A. Jadav, A. Deshmukh, A. Gupta, V. T. J. Phillips, A. Bansemer, and P. J. DeMott. Dependencies of four mechanisms of secondary ice production on cloud-top temperature in a continental convective storm. *Journal of the Atmospheric Sciences*, 79(12):3375–3404, 2022. doi: 10.1175/JAS-D-21-0278.1.

D. Waman, J. Muesel, B. Keshtgar, G. Wallentin, C. Barthlott, S. Patade, S. Shete, T. Prabhakaran, R. Fievet, D. Finney, A. Blyth, and C. Hoose. Impacts of secondary ice production on the microphysics and dynamics of deep convective clouds in different environments. *Atmospheric Chemistry and Physics*, 26(15):11371–11407, 2026. doi: 10.5194/acp-26-11371-2026.

P. Weiss, R. Herbert, and P. Stier. Icon-ham-lite 1.0: simulating the earth system with interactive aerosols at kilometer scales. *Geoscientific Model Development*, 18(12):3877–3894, 2025. doi: 10.5194/gmd-18-3877-2025. URL `https://gmd.copernicus.org/articles/18/3877/2025/`.

X. Zhao and X. Liu. Global importance of secondary ice production. *Geophysical Research Letters*, 48(11): e2021GL092581, 2021. doi: 10.1029/2021GL092581.