Implementation of ICON global 1km atmosphere-only demonstrator and performance analysis

Deliverable D2.9

This project has received funding from the European Union's
Horizon 2020 Research and Innovation Programme
under Grant Agreement No 675191

About this document

Work package in charge: WP 2 Scalability
Actual delivery date for this deliverable: 1 August 2017
Dissemination level: PU (for public use)

Lead author:
Deutsches Klimarechenzentrum GmbH (DKRZ): Philipp Neumann

Other contributing authors:
Deutsches Klimarechenzentrum GmbH (DKRZ): Joachim Biercamp, Irina Fast
European Centre for Medium-Range Weather Forecasts (ECMWF): Peter Bauer
Max-Planck-Institut für Meteorologie (MPI-M): Matthias Brück, Thorsten Mauritsen, Leonidas Linardakis
Deutscher Wetterdienst (DWD): Daniel Klocke

Contacts: esiwace@dkrz.de
Visit us on: www.esiwace.eu
Follow us on Twitter: @esiwace

Disclaimer: This material reflects only the authors view and the Commission is not responsible for any use that may be made of the information it contains.

Index

1. Abstract /publishable summary

The ICON framework has been considered for the use in global high-resolution, atmosphere-only weather and climate simulations. Focus was put on two scenarios:

- an aqua-planet experiment using ECHAM physics and
- a realistic full-world simulation scenario in the context of numerical weather prediction.

Both scenarios were investigated at various grid resolutions and compute settings. The experiments—based on the partition compute2 of supercomputer Mistral—suggest that ICON is expected to deliver good scalability for global high-resolution simulations up to 1.2km when filling up to 70-100 horizontal grid cells (depending on the number of vertical grid levels used) per compute core. The main bottlenecks of the “computational workflow” for high-resolution simulations were found in the creation of initial data and external parameters for the large-scale runs, as well as in writing output; the latter is due to processing large amounts of data (200GB per time slice in a 2.5km globally resolved simulation) and is subject to reimplementation and improvement with regard to using asynchronous I/O on distributed memory systems.

2. Conclusion & Results

- For the first time, a global ICON grid with 1.2km resolution was created
- For the first time, a ICON atmosphere-only simulation at 1.2km global resolution (with ECHAM physics in an aqua-planet experiment) was run.
- For the first time, a ICON atmosphere-only simulation at 2.5km global resolution with realistic external parameters and initial data was run.
- We investigated scalability of ICON and performed extrapolation estimates for computability of global high-resolution scenarios on upcoming exa-scale systems

3. Project objectives

This deliverable contributes directly and indirectly to the achievement of all the macro-objectives and specific goals indicated in section 1.1 of the Description of the Action:

Macro-objectives	**Contribution of this deliverable?**
Improve the efficiency and productivity of numerical weather and climate simulation on high-performance computing platforms	Yes
Support the end-to-end workflow of global Earth system modelling for weather and climate simulation in high performance computing environments	Yes
The European weather and climate science community will drive the governance structure that defines the services to be provided by ESiWACE	No
Foster the interaction between industry and the weather and climate community on the exploitation of high-end computing systems, application codes and services.	Yes
Increase competitiveness and growth of the European HPC industry	No

Specific goals in the workplan	**Contribution of this deliverable?**
Provide **services** to the user community that will impact beyond the lifetime of the project.	Yes

Improve **scalability** and shorten the time-to-solution for climate and operational weather forecasts at increased resolution and complexity to be run on future extreme-scale HPC systems.	Yes
Foster **usability** of the available tools, software, computing and data handling infrastructures.	Yes
Pursue **exploitability** of climate and weather model results.	No
Establish governance of common software management to avoid unnecessary and redundant development and to deliver the best available solutions to the user community.	No
Provide **open access** to research results and **open source** software at international level.	Yes
Exploit **synergies** with other relevant activities and projects and also with the global weather and climate community	Yes

4. Detailed report on the deliverable

In a first step, potential simulation setups and simulation models were identified. These were chosen as follows:

1. Aqua-planet experiments (APE) using ICON in combination with ECHAM physics:
   This setup allows for a simple-to-start, yet effective analysis of the actual computational units (such as the dynamical core) for climate simulation since no topography or initial condition data are required (both are given analytically). The test setup was specified in cooperation with Thorsten Mauritsen, MPI-M.
2. Full-system simulations of the globe using ICON with NWP physics:
   This setup comprises realistic parameterisations (including topography, land use, etc.) for global weather forecasts. The test setup was specified in cooperation with Matthias Brück, MPI-M, and Daniel Klocke, DWD, so that it basically corresponds to analogous local numerical predictions carried out in the scope of the HD(CP)2 project.

To carry out simulations, it became clear that the respective setups would exceed the allocation for computing time of the original DKRZ development project. Therefore, additional compute time was applied for and granted (50 000 nodes hours in the first phase).

APE simulations with ICON

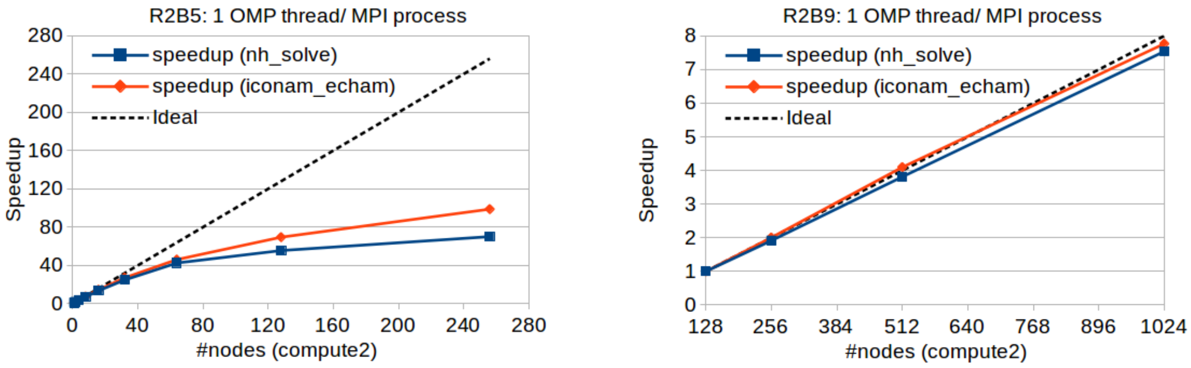
The APE setup was used to reconsider compile time optimisations as well as to evaluate various hybrid MPI/OpenMP configurations for different setups. The findings were in line with previous studies carried out in the scope of ICON benchmarking and suggest that the current settings are already effective.

A major ingredient to large-scale global simulations is the generation of the underlying unstructured ICON grid. The existing grid generators are, however, not designed for parallel grid generation yet. For this reason, global grids at resolutions finer than 2.5km had not been possible at the start of the project. In collaboration with Leonidas Linardakis (MPI-M) the MPI-M ICON grid generator could be used on a fat memory hardware hosted at DKRZ to construct a 1.2km grid (R2B11; 272GB). With the grid at hand, we were able to investigate scalability of ICON in the APE framework at various resolutions, ranging from R2B4 (160km res.) to R2B11 (1.2km res.). Input and selected measurement data for configurations up to R2B9 are published on the DKRZ websites in form of a performance benchmark[1]. The speedup in strong scalability experiments for R2B5 and R2B9, each consisting of 90 vertical levels and running in MPI-only

[1] https://redmine.dkrz.de/projects/icon-benchmark, ICON Benchmark v16.0

mode on the DKRZ supercomputer Mistral, partition compute2 (nodes equipped with 2x18 Broadwell cores), are shown in Figure 1. Similar results have been obtained in hybrid OpenMP/MPI settings.

The graphs suggest that the setup R2B9 (corresponding to a resolution of 5km) exhibits close to ideal scaling up to 1024 Broadwell compute nodes (each node consists of 36 cores; this corresponds to 36 864 MPI processes). The limits in scalability can be seen from the R2B5 case. Here, we reach ca. 50% in parallel efficiency on 1152 MPI processes, corresponding to ca. 71 horizontal cells per core or 2560 horizontal cells per compute node. These numbers are comparable to IS-ENES2 performance benchmarks carried out by Panagiotis Adamidis, considering ICON-APE scalability of R2B7 grids with 96 levels on Mistral, partition compute (parallel efficiency of 74% on 12 288 cores, corresponding to 107 horizontal cells per core).



(a) (b)

Figure 1: Scalability of dynamical core (nh_solve) and ECHAM physics kernel (iconam_echam) in MPI-only (a) R2B5- and (b) R2B9-global APE simulations using ICON.

Finally, we successfully ran a 1.2km APE experiment on Mistral, partition compute2. The simulation with 45 vertical levels was executed on 1408 nodes using 2 MPI processes per node, 18 OpenMP threads per MPI process. The run time for 200 time steps (time step was chosen as 4 seconds) was measured as ca. 453 seconds. Considering the 50%-efficiency mark as still acceptable (i.e. 71 horizontal grid cells with 90 vertical levels per core) and taking into account the difference in vertical levels in the R2B5, R2B9 and R2B11 runs, we extrapolate that a 1.2km APE simulation is affordable to be run on up to 70 000 nodes of type Mistral, compute2 (Broadwell), corresponding to ca. 2.5 million cores.

However, some issues were encountered. Due to an unresolved memory bottleneck with the ICON-ECHAM configuration, the 1.2km experiment could not be executed on less nodes, or with more vertical levels. This needs to be investigated in more detail in future. Besides, output was disabled in this scenario; see next Section for information on output generation.

Numerical weather prediction with realistic input data with ICON

In a second test suite, we considered NWP simulations. The input configurations were chosen in line with realistic settings that had been used in the past for production runs for local numerical weather predictions. A major issue we encountered was the generation of input data for

- initial conditions: given IFS simulation data from ECMWF, we made use of the parallelized DWD ICON tools to convert the data into netcdf-compatible ICON input. While the generation was successful up to 2.5km (R2B10), we could not generate input for ICON at 1.2km resolution (R2B11) due to memory issues,
- external parameters such as topographic data, land use data etc.: We made use of the tool EXTPAR, V3.0, for this purpose. Since this version is not parallelized for distributed memory systems, we could—similarly as for the initial conditions—successfully generate the external

parameters for ICON runs for resolutions up to 2.5km resolution. For 1.2km, all parameters could be generated; however, the final consistency check of EXTPAR, that removes inconsistencies among the different data and aggregates all data, failed due to memory issues.

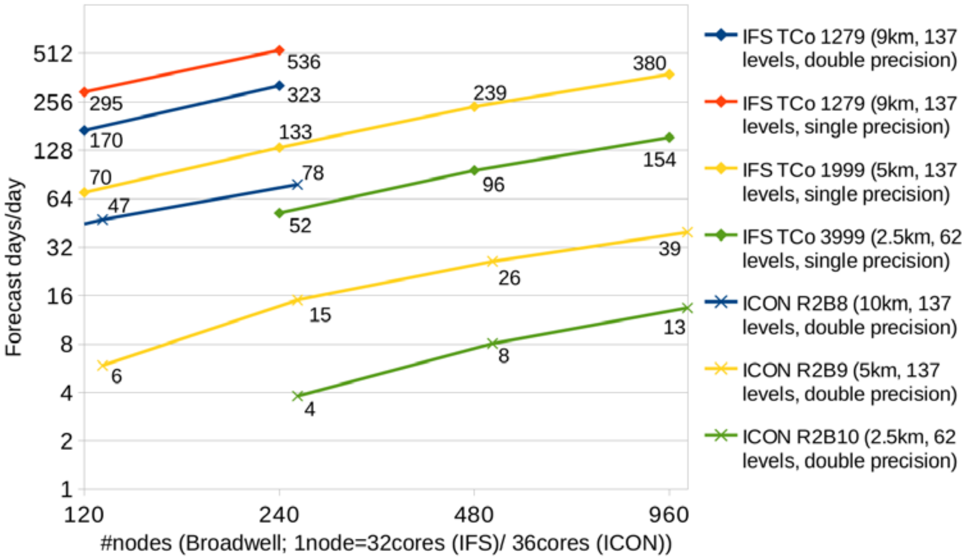


Figure 2: Scalability of ICON and IFS for different NWP simulation scenarios

We were able to perform various scalability and performance measurements for grid resolutions in the range of 10km (R2B8) to 2.5km (R2B10) with and without I/O. The resolutions and number of vertical grid levels were chosen in accordance with corresponding experiments at ECMWF based on the simulation model IFS. The performance (forecast days/day) neglecting I/O is shown in Figure 2. ICON performs ca. 10x slower than IFS with regard to compute forecast days per day. This can be explained by the following:

- Most of the IFS runs were based on single precision mode. According to ECMWF, this amounts to a reduction in compute time for IFS of up to 40%, compared to double precision runs.
- IFS was run in hydrostatic mode, whereas ICON is executed in non-hydrostatic mode by default. Within IFS, this yields speedup factor of ca. 2.
- The remaining factor of 3 is expected to originate from the different time stepping schemes. While ICON is purely explicit, IFS relies on a semi-implicit formulation. The latter results in higher computational work per time step, but enables the use of significantly longer time steps; that is, a 2.5km resolution in IFS can run with time steps of 120s whereas a time step of ca. 20s is currently estimated for ICON. Preliminary results suggest potentially the need for smaller time steps.

The measurements in Figure 2 did not consider I/O. However, generating output is essential and may become a bottleneck for large-scale scenarios. We therefore ran experiments including I/O on the R2B10 setup (62 vertical levels). The cloud distribution after 26 hours of simulation using ICON (R2B10) is shown in Figure 3.

To process larger sets of data, asynchronous output is enabled in ICON using separate MPI I/O processes. These processes can be fed with chunks of data such that a predefined number of vertical levels is bundled per chunk and sent to the I/O process. In our experiments, the chunk size needed to be chosen smaller than or equal to 16 vertical levels to fit into memory. While time integration resulted in a run time of ca. 2.7 seconds on 515 compute2 nodes (512 nodes for computing, 3 nodes for output processing), writing output required ca. 400 seconds per written time step. In the current conditions, even writing data in 30min intervals would already yield a slowdown of the simulation. Improvements with regard to increasing the parallelism in the asynchronous I/O operations are work in progress. Besides, storing data in the current form may also become a bottleneck. The (minimal) three- and two-dimensional output of one time step of the R2B10 (62 levels) setup already resulted in ca. 202GB of data.

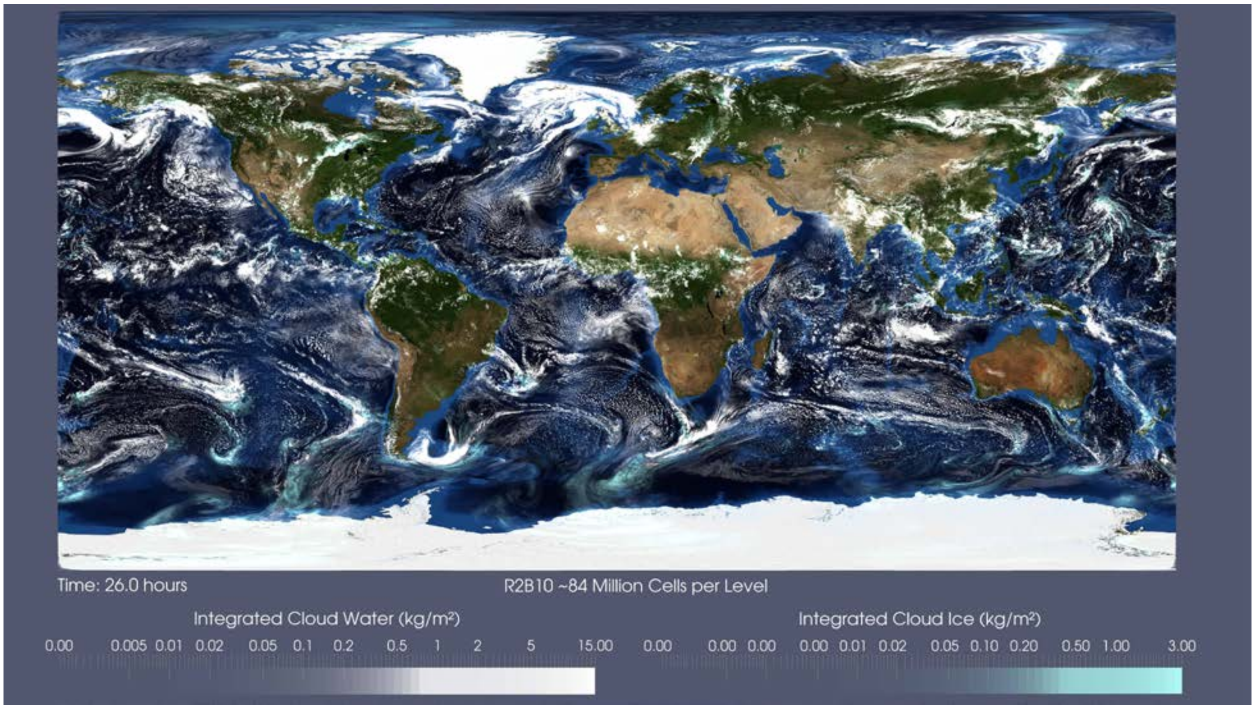


Figure 3: Cloud distribution after 26h of a global 2.5km ICON NWP simulation.

5. References *(Bibliography)*

- Zängl, G., Reinert, D., Prill, F., Giorgetta, M., Kornblueh, L., Linardakis, L., Müller, S., Rast, S. ICON Users's Guide, 2016
- Zängl, G., Reinert, D., Rípodas, P., Baldauf, M. The ICON (Icosahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. Q.J.R. Meteorol. Soc. 141: 563-579, 2015
- Dipankar, A., Stevens, B., Heinze, R., Moseley, C., Zängl, G., Giorgetta, M., Brdar, S. Large eddy simulation using the general circulation model ICON. Journal of Advances in Modeling Earth Systems 7(3): 963-986, 2015

6. Dissemination and uptake

6.1 Dissemination

The outcome of the ICON high-resolution global simulations have been presented at the European Geosciences Union General Assembly 2017.

Publications in preparation OR submitted

In preparation OR submitted?	Title	All authors	Title of the periodical or the series	Is/Will ***open access*** be provided to this publication?

In preparation	Computational Horizons for Earth System Modelling	Biercamp, J., Stevens, B., Schulthess, T., Bauer, P., Wedi, N., Schär, C., Fuhrer, O., Adamidis, P., Neumann, P.	n.n.	Yes

6.2 Uptake by the targeted audience

As indicated in the Description of the Action, the audience for this deliverable is:

x	The general public (PU)
	The project partners, including the Commission services (PP)
	A group specified by the consortium, including the Commission services (RE)
	This reports is confidential, only for members of the consortium, including the Commission services (CO)

This is how we are going to ensure the uptake of the deliverables by the targeted audience

The deliverable will be provided online on the website of ESiWACE. Benchmark data have been made available for global ICON aqua planet experiments which is expected to increase interaction with vendors.

7. The delivery is delayed: ☐ Yes ☒ No

8. Changes made and/or difficulties encountered, if any

None.

9. Sustainability

9.1. Lessons learnt: both positive and negative that can be drawn from the experiences of the work to date

We obtained many insights from scalability considerations and extrapolations towards upcoming exa-scale systems. The arising predictions are mostly in line with actual expectations, based on analytical and thought experiments. However, the efforts for setting up the simulations (initial phase) were underestimated.

9.2 Links built with other deliverables, WPs, and synergies created with other projects

Links with ECMWF (D2.8) were established in terms of a joint scalability study of comparable scenarios. We further established links to other international initiatives, targeting high-resolution global simulations in the scope of a birds-of-a-feather session at ISC 2017, and discussions beyond.

10. Dissemination activities

Type of dissemination and communication activities	Number	Details	Total funding amount	Type of audience reached In the context of all dissemination & communication activities	Estimated number of persons reached
Participation to a conference	1	Talk at EGU 2017 "ESiWACE A Center of Excellence for HPC applications to support cloud resolving earth system modelling", Presenter: Philipp Neumann (DKRZ) http://meetingorganizer.copernicus.org/EGU2017/EGU2017-2367.pdf	See costs declared in form C of DKRZ	Scientific Community (higher education, Research)	80
Organisation of a workshop	1	Organisation of the EGU 2017 session AS4.10/CL5.12/ESSI1.14/OS4.15 "Recent developments in numerical atmospheric, oceanic and sea-ice models: towards global cloud and eddy resolving simulations on exascale supercomputers (co-organized)", 26 April 2017, Vienna (AT) Convener: Peter Dueben (ECMWF). Co-convener: Pier Luigi Vidale (UREAD) /PRIMAVERA Project. http://meetingorganizer.copernicus.org/EGU2017/session/23945	See costs declared in form C of ECMWF	Scientific Community (higher education, Research)	80
Organisation of a workshop	1	Minisymposium on "ESiWACE: The Center of Excellence in Simulation of Climate and Weather in Europe" organised at Platform for Advanced Scientific Computing (PASC) Conference,28 June 2017, Lugano, Switzerland. Presenters: Joachim Biercamp and Phillip Neumann (DKRZ) https://pasc17.pasc-conference.org/	See costs declared in form C of DKRZ	Scientific Community (higher education, Research)	80
Organisation of a workshop	1	Birds-of-a-feather session "Cloud Resolving Global Earth-System Model: HPC at its Extremes" on 20 June 2017, organised at ISC 2017. Speakers: See the full list here https://www.esiwace.eu/events/bof-cloud-resolving-global-earth-system-models-hpc-at-its-extreme-	See costs declared in form C of DKRZ	Scientific Community (higher education, Research)	80

		at-isc-2017			
Poster	1	Poster in Project Posters Track of ISC 2017 https://www.esiwace.eu/results/misc/esiwace-poster-2017/view	See costs declared in form C of DKRZ	Scientific Community (higher education, Research)	100
Flyers	150	ESiWACE (ad hoc) flyers distributed at the conference PASC 2017 https://www.esiwace.eu/results/misc/esiwace-at-a-glance-flyer-2017/view	See costs declared in form C of DKRZ	Scientific Community (higher education, Research)	150
Exhibition	1	Exhibition booth at PASC 2017: https://pasc17.pasc-conference.org/program/exhibitors/	See costs declared in form C of DKRZ	Scientific Community (higher education, Research)	200

Intellectual property rights resulting from this deliverable

Not applicable.