Report on Project: **1040**

Project title: **ESiWACE: Scalability of Earth System Models**

Principal investigator: **Joachim Biercamp**

Report period: **2022-11-01 to 2023-10-31**

Data provision for DYAMOND

The initiative *DYAMOND* (DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains) sets up a framework for the intercomparison of global storm-resolving models, i.e. an emerging class of atmospheric circulation models. Through their explicit resolution of the major modes of atmospheric heat transport, they endeavour to represent the most important scales of the full three-dimensional fluid dynamics of the atmospheric circulation. DYAMOND is coordinated by the Max Planck Institute for Meteorology (MPI-M) and the German Climate Computing Center (DKRZ) and supported by the Centre of Excellence in Simulation of Weather and Climate in Europe (*ESiWACE*[1]). It consists of two experiment phases: *DYAMOND Summer*, closed in 2018 and *DYAMOND Winter*, starting in 2019.

The project has been used to provide access to the two both data collections. While bk1040 contains the data and regulates permissions, the user project bb1153, has been used to give users access to compute and storage resources without risking them accidentally burning the resources of this project. The provided easy access to global high-resolution data sets especially attracts young scientists from all over the world, to work with these data and learn how to make efficient use of future pre-exascale simulation. But also Nvidia and Microsoft Research have chosen this data for their ML projects.
The DYAMOND Summer data-sets are archived under project mh1113 and are provided at disk within bk1040. About 30 data retrieval requests in order of a TB or larger were carried out. The data requests were mostly for DYAMOND Summer output. Between July and September, about 120 TB of the 263 TB data has been used.
For the second phase of DYAMOND, bk1040 is used as archival and disk storage. Here about 130 TB out of the 809 TB have been in use during July and September 2023.
Beside data access, the scientific value of DYAMOND can be seen by its publication list provided at https://www.esiwace.eu/services/dyamond-initiative/dyamond-related-publications.

Standardizing and archiving DYAMOND Winter data-sets

To ease multi-model comparisons, we are standardising all contributed DYAMOND Winter simulation data sets. In this process, we assign common variable names based on the CMIP6-standard and rearrange the data into files with one file per variable and day. We see this granularity as the ideal for

[1] https://www.esiwace.eu/

providing data on user-request, and allowing for efficient analysis. For a simplified multi-model comparisons with semantic data access we provide an intake-esm catalogue.
In 2023 we could standardise and achieve 5 additional data sets:

- GRIST-5km uncoupled and coupled runs (CAMS)
- IFS-4km and IFS-9km coupled runs (ECMWF)
- ICON-SAP-5km uncoupled and coupled runs
- GEOS-1km and GEOS-6km coupled runs (NASA)
- MPAS-3km run (NCAR)

The total size of the processed files is more than 300TB.

Development of YAVIZ

In the ESiWACE project we have developed YAVIZ (Yet Another VIzualiZation Component), which connects the ICON model with Catalyst, the in-situ interface of the visualization tool paraview. For that we used the coupler YAC to get access to the data during runtime. The development and testing of this effort required a parallel setup. A detailed description can be found at https://doi.org/10.5281/zenodo.7584067.
Furthermore, we also tested the application on heterogeneous configurations, running the model on CPU nodes and the visualization on GPUs. A further part of the compute time was spent for the creation of demonstration videos, e.g. https://youtu.be/sMY0nzzHA4Q.

Post-processing of Aquaplanet experiments

MPi-M has used bk1040 for post-processing and analysis of global Aquaplanet experiments using ICON-Sapphire with horizontal resolutions between 160 and 1.25 km. These experiments are quite heavy if we let them be in the native grid and ICON output format. So some of the resources have been used for bitwise compression and usage of Zarr format rather than netcdf. This reduced the memory footprint by factor ⅛.
From the science perspective, a paper by Angel Peinadeo Bravo, Daniel Klocke and Bjorn Stevens is going to be finished that indicates apparent physical and numerical convergence of the atmosphere component of ICON using the Sapphire configuration. One of the main results is the required horizontal resolution estimation for different atmospheric characteristics as the Inter-Tropical Convergence Zone to have a good representation of them.