Project: **XYZ**

Project title: **DYAMOND Data Intercomparison**

Principal investigator: **Daniel Klocke, Tomoki Miyakawa, Florian Ziemen**

Allocation period: **2020-01-01 to 2020-12-31**

Project overview

The DYAMOND (DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains) project is an **international framework for the intercomparison of an emerging new class of atmospheric circulation models**, that, through their resolution of the major modes of atmospheric heat transport, endeavor to represent the most important scales of the full three-dimensional fluid dynamics of the atmospheric circulation. The first phase of DYAMOND, which was completed in 2019 had nine participating models from three continents which ran successfully at globally storm resolving scales (horizontal mesh size <5km) for 40 days and nights from the 1st August 2016 (Stevens et al., 2019). Scientific results from all participants are currently collected in a special issue of the Journal of the Meteorological Society of Japan[1]. The success of the first phase motivated the participants of DYAMOND to complement the first phase simulations by a **second set of simulations, this time modeling a boreal winter period and also adding coupled models with the goal to: (i) compare the representation of the Madden-Julian-Oscillation in this class of models; (ii) investigate the effect of the atmosphere-ocean coupling at storm and ocean-eddy resolving scales on convection and the general circulation; and (iii) link to the EUREC^{4}A [2] campaign, which targets meso-scale convection patterns and the coupling to the upper ocean processes.**

With the centre of excellence ESiWACE considering 1km-simulations, but rather with regard to scalability and performance, there are close links that have been established between DYAMOND and ESiWACE. Amongst others, The ESiWACE compute project (bk1040) supports and comprises the DYAMOND experiment runs with the ICON model that will be a joint undertaking of DKRZ and MPI-M. These numerical experiments will comprise a 40-day period of simulations, running at 2.5 km and 5 km resolution.

In this project, we apply for storage on the data servers at DKRZ to host the results of all participating DYAMOND groups and so enable the exchange of data and corresponding data intercomparison. Based on the first phase of DYAMOND and the experiments defined in the DYAMOND protocol[3], we estimate the data storage requirements, as follows: **we expect 11 participating models**, each submitting a high and a low resolution run with fixed sea-surface temperatures, which amounts to about 100 TB per model. An estimated 7 models will also perform coupled runs, again an estimated 100 TB per model. For post-processing data for the intercomparison (interpolation of some data on common grid) we estimate in total 300 Tb. This will result in a total memory requirement of 2100 TB. We are in contact with the DKRZ data management group and will build on the experience from the first phase to optimally share the data among participants and make it available to the international research community.

Range of planned work from the scientific view

The actual DYAMOND simulations that yield scientific insight are carried out by every participant separately, and DKRZ/MPI-M have applied for corresponding compute time in the scope of the ESiWACE project bk1040 for the contribution with ICON to DYAMOND. This project's focus is to collect the data in one place and make it available to participants. It is a necessary prerequisite for all participating groups and of utmost importance for a successful intercomparison in the scope of DYAMOND.

---

[1] http://jmsj.metsoc.jp/special_issues_editions/DYAMOND_info.html

[2] http://eurec4a.eu/, Field Campaign to Elucidate the Couplings Between Clouds, Convection and Circulation

[3] https://www.esiwace.eu/services/dyamond/dyamond-2nd-phase-the-winter-experimental-protocol/view

Mathematical and/or computational aspects

The aspects depend on the actual models and do not yield any restrictions or impacts on the proposed project.

Algorithmic/mathematical/numerical methods and solution procedures

The procedures depend on the actual models and do not yield any restrictions or impacts on the proposed project. All groups rely on same data formats, that is Grib or NetCDF. During the first phase of DYAMOND, the climate data operators (cdo) were further developed to handle all data formats and grid structures of the DYAMOND models.

Particular suitability to solve the problem with help of HLRE-3

DKRZ has one of the largest data archives and is therefore the perfect partner to host, administrate and handle the large amounts of data that will be produced by the DYAMOND initiative.

Performance benefits depending on the number of used CPUs (scalability)

Not applicable.

Required computing time and amount of storage space

We currently apply for 11 groups x 2 experiments, plus 7 groups with coupled experiments, plus space for post-processed data = 2100 TB of storage space in the work directory.

Long-time archival of data is envisaged and will be realized in close collaboration with DKRZ's data department. For this, we request 3.6 PB of space on /docu/ (1.5 PB for phase 1 and 2.1 PB for phase 2). For intermediate storage and preparation of the final archiving, we request 2.1 PB of space on /arch/.

For compute-intensive post-processing, we request 10 000 node hours.

Additional value compared to other projects

Other intercomparison projects exist, such as CMIP. DYAMOND is different as it tackles significantly higher resolution at global scale which is expected to be essential for future climate and weather science, as indicated in the project overview. It is the first intercomparison project of this new class of models, and is of wide interest to the internationally leading climate modeling groups.

Data Storage Usage Plan

Currently **users from 18 institutions around the world have direct disk access to the DYAMOND data on mistral** for shared analysis. To provide continued good access, we are **closely collaborating with DKRZ's data department to make key data available via the ESGF**. This will ensure a convenient long-time access to this unique data set for the worldwide scientific community. We furthermore plan to provide **long-time access to the full datasets (1.5 PB) via /docu/**. We **envisage providing long-time data access to the results of the second phase** as well. For this, we **request an additional 2.1 PB, totalling at 3.6 PB of space on /docu/**. For **intermediate storage** of data during the preparations for the final archiving, we request **2.1 PB on /arch/**.

Literature

Stevens, B.; Satoh, M.; Auger, L.; Biercamp, J.; Bretherton, C. S.; Chen, X.; Düben, P.; Judt, F.; Khairoutdinov, M.; Klocke, D.; Kodama, C.; Kornblueh, L.; Lin, S.-J.; Neumann, P.; Putman, W. M.; Röber, N.; Shibuya, R.; Vanniere, B.; Vidale, P. L.; Wedi, N. & Zhou, L.: DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains, *Progress in Earth and Planetary Science,* **2019***, 6*, 61