The Pilot Lab Exascale Earth System Modelling

Catrin I. Meyer (1), Sabine Attinger (2), Achim Basermann (3), Stefano Bianco (6), Arne Biastoch (4), Peter Braesicke (5), Jean Braun (6), Lars Buntemeyer (7), Daniel Caviedes-Voullieme (1), Martin Claus (4), Charlotte Debus (3), Doris Dransch (6), Thomas Fischer (2), Stephan Frickenhaus (8), Bernadette Fritsch (8), Sabine Grießbach (1), Jens-Uwe Grooß (9), Kerstin Hartung (3), Jan Hegewald (8), Harrie-Jan Hendricks-Franssen (10), Lars Hoffmann (1), Katharina Höflich (4), Johannes Holke (3), Daniela Jacob (7), Patrick Jöckel (3), Thomas Jung (8), Astrid Kerkweg (11), Timo Kirfel (11), Ole Kirner (5), Margrit Klitz (3), Olaf Kolditz (2), Nikolay Koldunov (8), Stefan Kollet (10), Daniel Kreyling (8), Tobias Meisel (2), Jörg Meyer (5), Helge Mohn (8), Rolf Müller (9), Zbigniew Piotrowski (1), Felix Plöger (9), Natalja Rakowsky (8), Willi Rath (4), Luis Samaniego (2), Rober Sausen (3), Magdalena Scheck-Wenderoth (6), Simon Schmitz (3), Yuri Shprits (6), Dmitry Sidorenko (8), Kevin Sieck (7), Olaf Stein (1), Achim Streit (5), Domenico Taraborrelli (11), Martin Thomas (6), Kay Thust (1), Stefan Versick (5), Benedikt von St. Vieth (1), Ingo Wohltmann (8), Sourena Yadegari (2), and Martin Schultz (1)

(1): Jülich Supercomputing Centre, Forschungszentrum Jülich
(2): Helmholtz-Zentrum für Umweltforschung
(3): Deutsches Zentrum für Luft- und Raumfahrt
(4): GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel
(5): Karlsruher Institut für Technologie
(6): GFZ Potsdam
(7): Helmholtz-Zentrum Hereon
(8): Alfred-Wegener-Institut
(9): Institut für Energie- und Klimaforschung Stratosphäre, Forschungszentrum Jülich
(10): Institute for Bio- and Geosciences, Forschungszentrum Jülich
(11): Institut für Energie- und Klimaforschung Troposphäre, Forschungszentrum Jülich

The Pilot Lab Exascale Earth System Modelling (PL-ExaESM) is a “Helmholtz-Inkubator Information & Data Science” project and explores specific concepts to enable exascale readiness of Earth System models and associated work flows in Earth System science. PL-ExaESM provides a new platform for scientists of the Helmholtz Association to develop scientific and technological concepts for future generation Earth System models and data analysis systems. Even though extreme events can lead to disruptive changes in society and the environment, current generation models have limited skills particularly with respect to the simulation of these events. Reliable quantification of extreme events requires models with unprecedentedly high resolution and timely analysis of huge volumes of observational and simulation data, which drastically increase the demand on computing power as well as data storage and analysis capacities. At the same time, the unprecedented complexity and heterogeneity of exascale systems, will require new software paradigms for next generation Earth System models as well as fundamentally new concepts for the integration of models and data. Specifically, novel solutions for the parallelisation and scheduling of model components, the handling and staging of huge data volumes and a seamless integration of information management strategies throughout the entire process-value chain from global Earth System simulations to local scale impact models will be developed in PL-ExaESM. The potential of machine learning to optimize these tasks is investigated. At the end of the project, several program libraries and workflows will be available, which provide the basis for the development of next generation Earth System models.
In the PL-ExaESM, scientists from 9 Helmholtz institutions work together to address 5 specific problems of exascale Earth system modelling:

- Scalability: models are being ported to next-generation GPU processor technology and the codes are modularized so that computer scientists can better help to optimize the models on new hardware.
- Load balancing: asynchronous workflows are being developed to allow for more efficient orchestration of the increasing model output while preserving the necessary flexibility to control the simulation output according to the scientific needs.
- Data staging: new emerging dense memory technologies allow new ways of optimizing I/O operations of data-intensive applications running on HPC clusters and future Exascale systems.
- System design: the results of dedicated performance tests of Earth system models and Earth system data workflows are analysed in light of potential improvements of the future exascale supercomputer system design.
- Machine learning: modern machine learning approaches are tested for their suitability to replace computationally expensive model calculations and speed up the model simulations or make better use of available observation data.