HPC-based Data Science and Learning Environment for the Analysis of Unstructured Grid-data in the Climate-change Domain

Desta Gebre Gebremedhin (desta.Gebremedhin@unitn.it)

DISI, University of Trento

Introduction

The latest Intergovernmental Panel on Climate Change report, IPCC 2022, indicates that climate change is causing substantial damages some of which are increasingly irreversible in terrestrial, coastal, and open ocean marine ecosystems.

To tackle these ramifications, extensive study of internal variability of climate and weather behaviors as well as observed climate change patterns are being conducted in many multinational research centers. Climate models that predict the state of the next time step by iteratively running simulations on current climate states have been built mostly based on numerical weather prediction (NWP) models. These kind of models try to numerically solve the fundamental governing equations of processes contributing to climate change.

In this regard, intergovernmental collaborative initiatives and projects including the Earth System Grid Federation (ESGF), coupled model intercomparison projects (CMIP5 and CMIP6) and the World Climate Research Programme (WCRP) are working towards a shared goal.

Data-driven Algorithms

Data-driven algorithms being investigated, optimized and scaled up for a research on unstructured grid are:

- ✓ Graph neural network models (GNN-based models)
- ✓ Convolutional neural network models (CNN-based models)
- ✓ Continuous GNN (a kind of LSTM-CNN). A new proposal

So far, we have implemented and trained GNN-based algorithm for large volumes of structured grid data.

Undelying Infrastructure

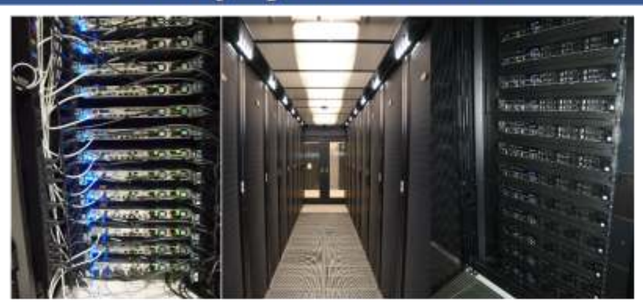

Figure 1: HPC@unitin

The @UniTrento HPC cluster with calculation capacity currently equal to 478 TFLOPs (theoretical peak performance) and 7,674 cores

Results so far

At this stage of the research the following results have been achived:

- Release of first prototypal release of the HPC-based environment at the University of Trento, with a full Python-base data science and learning software ecosystem.
- Design and implementation of GNN based algorithm for unstructured mesh grids in the climate domain
- Preliminary results of the data-driven algorithm on two unstructured datasets

Motivation

Over the last years, increased availability of computational power has pushed research boundaries in climate change and weather research towards higher spatio-temporal resolution models, larger ensembles as well as longer time-scale simulations.

Simulations at such scale have led to an unprecedented growth in data production which in turn calls for new data intelligent approaches and proper infrastructural solutions to enable scientific discovery. In such a context, advances in machine learning jointly with big data and high-performance computing convergence make migration to data-driven approaches logical and feasible. In this research an HPC-based data science and learning environment for climate change is being developed, with a specific focus on challenges regarding unstructured grid models which are raising a great interest in the community as they provide stronger flexibility to climate modelers and a better balance between computation and accuracy.

Datasets

A preliminary set of data from the AWI Coupled Model (AWI-CM) retrieved from the Earth System Grid Federation (ESGF.

Prominent projects hosted by the ESGF (CMIP5 and CMIP6) which delivers data relevant to the major activities of the World Climate Research Programme (WCRP) extensively used in the IPCC Assessment Reports (Ars).

Unstructured grid models from AWI-CM in the CMIP6 project are used and in this research. These data are populated in the data space of the HPC-based environment at the University of Trento

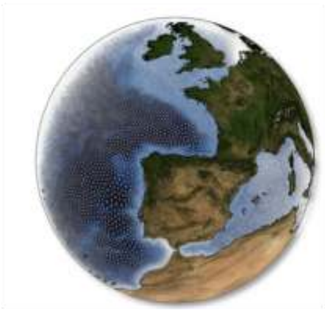

Figure 2: FESOM Data

Conclusion

A data driven algorithm based on graph neural network seems to handle irregularities in the unstructured grid. Although much further research is required and is being conducted and it could be premature to draw a definitive conclusion, promising results on modeling weather prediction using native unstructured grids are observed. Moreover, taking advantage of the available computing power, better results seem to be achieved without interpolation of variables.

Research Goals

This research overall focuses on achieving two-fold goals.

- ✓ Design, implementation, integration, parallelization and performance analysis of data-driven algorithms.
- ✓ Design of the HPC-based data science and learning infrastructure and software stack ecosystem
  - ✓ Tackling research challenges at the intersection of HPC, big data and data science for unstructured grid climate data

In general, the goal of this research is to realize a seamless integration of big data and data science tools into the same HPC-based problem-solving environment by providing high-level application programing interface and extensible software ecosystems.

Environment Architecture

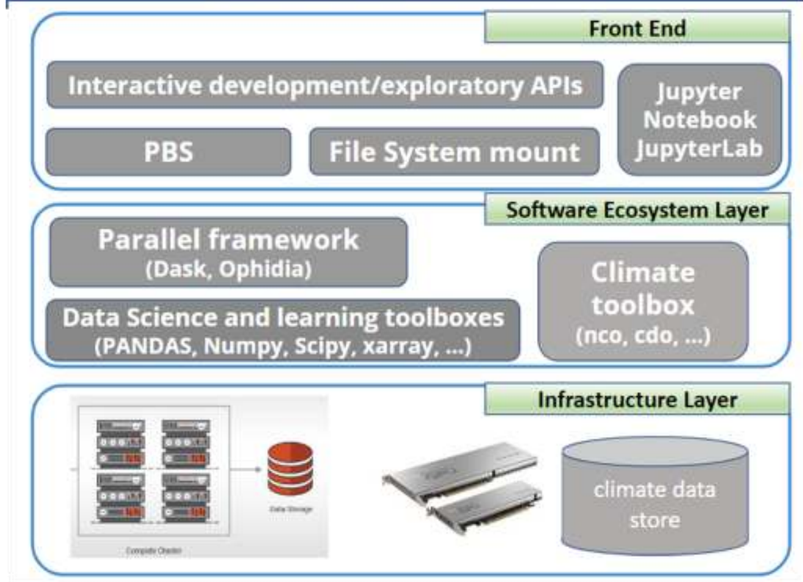

Figure 1: Environment Architecture

References

1. Ryan Keisler (15 Feb 2022) Forecasting Global Weather with Graph Neural Networks
2. Stephan Rasp, Nils Thuerey (02 February 2021) Data-Driven Medium-Range Weather Prediction With a Resnet Pretrained on Climate Simulations: A New Model for WeatherBench
3. Peter Bauer et al. (01 February 2021) **A digital twin of Earth for the green transition**
4. Yang et al. (2022l) Learning to Simulate Unseen Physical Systems with Graph Neural Networks
5. M Asch et al. (2018) Big data and extreme-scale computing: Pathways to Convergence-Toward a shaping strategy for a future software and data ecosystem for scientific inquiry
6. Nicola Jones (2017) How Machine Learning Could Help to Improve Climate Forecasts
7. https://www.ipcc.ch/report/ar6/wg2/
8. https://www.mooc.org/