



Contribution ID: 19

Type: **Oral presentation**

# Toward a petascale data hackathon for exploring the 1-km IFS nature run

*Wednesday, 8 June 2022 14:20 (20 minutes)*

During 2020-2021, the European Center for Medium-Range Weather Forecasts (ECMWF) and the Oak Ridge National Laboratory (ORNL) received an award from the U.S. Department of Energy to simulate and develop a baseline for weather and climate simulations at 1-km resolution, using the Summit supercomputer, hosted at the Oak Ridge Leadership Computing Facility (OLCF). The ECMWF hydrostatic Integrated Forecasting System (IFS), with explicit deep convection on an average grid spacing of 1.4 km, was used to perform the simulations in two phases, as a set of two nature runs (NR) for the periods corresponding to 2018 November to 2019 February (four months) and August - October in 2019 (three months). The simulations revealed unprecedented detail of the earth's atmosphere. These baseline simulations will provide further guidance toward future model developments and satellite mission planning. The project won the 2020 *HPCwire Readers Choice Award* for Best Use of HPC in Physical Sciences.

The simulation output amounts to over 500 TB in total volume, much of it stored as compressed and bit-packed grib files. After the spectral and grid point conversion the volume inflates to over 2 PB. Once our seminal manuscript was published, we received a few inquiries regarding access to the data. All of the interested research teams lacked the necessary infrastructure and/or bandwidth to download and manage the large volume of data. So, we are facilitating direct access to the data and computational resources via a data hackathon, jointly sponsored by the OLCF and the ECMWF.

A formal announcement for the event will solicit proposals from teams that are interested in relevant research topics, such as extreme weather events, including hurricanes and severe storms. At 1-km resolution, many of the smaller scale processes are resolved, such as most of the gravity wave spectrum. Hence, this dataset provides a unique opportunity to not only smaller scale processes but also to inform the development of their parameterization in order to improve numerical weather and climate prediction. We are also emphasizing and encouraging the development of AI and ML applications for surrogate models, emulators for data assimilation systems, satellite retrievals of earth observations, etc. Visualization of the global 1-km simulations are also of special interest for science investigations as well as for outreach activities to inform and educate.

Initially the model output was saved every 3 hours, with the anticipation that we could rerun short periods as necessary with higher temporal frequency output for specific science cases. Both model level data at all 137 levels and pressure level fields at 31 levels have been archived, along with key single level fields. Since then with feedback from the community, we identified a set of special cases, including a tropical cyclone and three severe weather episodes. We used checkpoint files to restart these simulations for the special cases but with data output every 15 minutes. The high temporal frequency output has added an additional 500 TB of grid point data. Observing System Simulation Experiments (OSSEs) can be performed for designing, configuring and evaluating future satellite instruments for prediction of severe weather events.

We are currently in the process of finalizing the plans for the data hackathon, scheduled to launch on 1 June 2022. The announcement of opportunity, soliciting proposals, will be disseminated in early May 2022. The selection of projects will be completed toward the end of May. The hackathon will be for a duration of 6 months. The teams will have access to Andes at OLCF, a Linux cluster with 704 nodes, each with 256 GB memory per node. For machine learning and AI workloads, we will also facilitate access to Ascent, a stand-alone 18-node system with an architecture similar to Summit. The users can also deploy jupyter notebook instances for interactive analytics. The approved users and teams will be able to get familiar with the OLCF systems and the data via tutorial and training sessions. Mentors will be available to provide support related

to the systems and the data collection. We have already engaged several external stakeholders with diverse research interests. During our presentation, we will discuss our experiences and the lessons learned from planning and preparation for the hackathon, the nature of various science investigations, and the roadmap for the event.

**Primary authors:** ANANTHARAJ, Valentine (Oak Ridge National Laboratory); WEDI, Nils (ECMWF); POLICHTCHOUK, Inna; DUEBEN, Peter (ECMWF); PARETE-KOON, Suzanne (Oak Ridge National Laboratory); PAPTHERODORE, Thomas (Oak Ridge National Laboratory)

**Presenter:** ANANTHARAJ, Valentine (Oak Ridge National Laboratory)

**Session Classification:** User Engagement and Feedback

**Track Classification:** UEF2022