Geophysical Research Abstracts
Vol. 20, EGU2018-12322, 2018
EGU General Assembly 2018

Towards a Framework for Parallelized Post-Processing and Evaluation of Ensemble Forecasts

Robert Redl (1), Christian Keil (1), George Craig (1), Sebastian Lerch (2), and Joachim Eichhorn (3)

(1) LMU Munich, Meteorological Institute, Munich, Germany (robert.redl@lmu.de), (2) Heidelberg Institute for Theoretical Studies (HITS), and Institute for Stochastics, Karlsruhe Institute of Technology, HITS gGmbH, Heidelberg, Germany, (3) Institute for Atmospheric Physics, University of Mainz, Mainz, Germany

From the computational point of view, post-processing and evaluation of model output is an increasingly challenging task, even though the calculations themselves are often relatively simple. Model resolutions continue to get higher, as does the number of ensemble members. Both in combination creates datasets that are not anymore easily processable on individual workstations or compute-nodes which are limited by their amount of memory and their network connection to the storage system. Even the most efficient algorithm has to read the whole dataset once. The limitations can be addressed by parallelized and distributed computations. With modern parallel file systems, the available bandwidth scales with the number of nodes, which is also the case for the available memory. However, exploiting hardware available for parallel processing still requires a (too) high level of programming skills, often in low-level languages.

Within the Transregional Collaborative Research Center "Waves to Weather (W2W)", which is focused on the limits of predictability in numerical weather prediction, we concentrate on the Python programming language. It is popular in the scientific community for its interactive character and ease of learning. Our aim is to create a user-friendly and portable open source Python package for ensemble forecast evaluation and post-processing called "Ensemble Tools". The foundation of the package is a unified interface to the most common model output formats netCDF, grib1, and grib2. This interface enables parallelized and distributed I/O based on the increasingly popular xarray and dask libraries. On top of that standard methods of statistical model evaluation like the continuous ranked probability score (CRPS), advanced process-oriented diagnostics like the analysis of Rossby Wave Packets as well as auxiliary utility functions like interpolation between different grids are implemented.

The intention of the development is on the one hand the support of project scientists within W2W but on the other hand also the publication of developed methods for improved reproducibility of scientific results. The current early stage of development is mainly focused on the I/O framework and output from the numerical models COSMO, ICON, and IFS. Future developments will build on-top that implemented functionality.