

EGU23-9782, updated on 04 Mar 2024
<https://doi.org/10.5194/egusphere-egu23-9782>
EGU General Assembly 2023
© Author(s) 2024. This work is distributed under
the Creative Commons Attribution 4.0 License.



User-friendly climate data discovery and analysis with ESMValCore

Bouwe Andela¹, Peter Kalverla¹, Remi Kazeroni², Saskia Loosveldt Tomas³, Valeriu Predoi⁴, Manuel Schlund², Stef Smeets¹, and Klaus Zimmermann⁵

¹Netherlands eScience Center, Amsterdam, Netherlands (b.andela@esciencecenter.nl)

²Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

³Barcelona Supercomputing Center, Barcelona, Spain

⁴NCAS-CMS, University of Reading, Reading, UK

⁵Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

The Earth System Grid Federation (ESGF) offers a wealth of climate data that can be used to do interesting research. For example, the latest edition of the Coupled Model Intercomparison Project (CMIP6) output features 20 petabytes of data. However, the heterogeneity of the data can make it difficult to find and work with. Here, we present new features of ESMValCore, a Python package designed to work with large climate datasets available from ESGF and beyond. ESMValCore now provides a Python interface that makes it easy to discover what data is available on ESGF and locally, download it if necessary, and make it analysis-ready. The analysis-ready data can then be used as input to the ESMValCore preprocessor functions, a collection of functions that can be used to perform commonly used analysis steps such as regridding and statistics. When searching for data on ESGF as well as when loading the NetCDF files, the software intelligently corrects small issues in the metadata that otherwise make working with this data a time-consuming, manual effort. Data and metadata issues are fixed in memory for fast performance. The search and download features are user-friendly and will automatically use a different server if one of the ESGF servers is unavailable for some reason. Several Jupyter notebooks demonstrating these new features are available at <https://github.com/ESMValGroup/ESMValCore/tree/main/notebooks>.

ESMValCore has been designed for use on computing systems that are typically used by researchers: it works well on a laptop or desktop computer, but also comes with example configuration files for use on large compute clusters attached to ESGF nodes. For reliable computations, ESMValCore makes use of the Iris library developed by the UK Met Office. This in turn is built on top of Dask, a library for efficient parallel computations with a low memory footprint. In 2023, we aim to improve our use of Dask in collaboration with the Iris developers, for even better computational performance.

For easy reproducibility, ESMValCore also offers “recipes” in which standard analyses can be saved. A large collection of such recipes is available in the Earth System Model Evaluation Tool

(ESMValTool). ESMValTool started out as a set of community-developed diagnostics and performance metrics for the evaluation of Earth system models. Recently it has also turned out to be useful for other users of climate data, such as hydrologists and climate change impact researchers. Both ESMValCore and ESMValTool are developed by and for researchers working with climate data, with the support of several research software engineers. An important recent achievement is the use of these packages to produce the figures for several chapters of the IPCC AR6 report. Documentation for both ESMValCore and ESMValTool is available at <https://docs.esmvaltool.org>.