



## ESMValCore: analyzing CMIP data made easy

**Bouwe Andela**<sup>1</sup>, Lisa Bock<sup>2</sup>, Björn Brötz<sup>2</sup>, Faruk Diblen<sup>1</sup>, Laura Dreyer<sup>7</sup>, Niels Drost<sup>1</sup>, Paul Earnshaw<sup>7</sup>, Veronika Eyring<sup>2,3</sup>, Birgit Hassler<sup>2</sup>, Nikolay Koldunov<sup>8,9</sup>, Axel Lauer<sup>2</sup>, Bill Little<sup>7</sup>, Saskia Loosveldt-Tomas<sup>5</sup>, Lee de Mora<sup>6</sup>, Valeriu Predoi<sup>4</sup>, Mattia Righi<sup>2</sup>, Manuel Schlund<sup>2</sup>, Javier Vegas-Regidor<sup>5</sup>, and Klaus Zimmermann<sup>10</sup>

<sup>1</sup>Netherlands eScience Center, Amsterdam, Netherlands

<sup>2</sup>Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, Germany

<sup>3</sup>Institute of Environmental Physics (IUP), University of Bremen, Bremen, Germany

<sup>4</sup>National Centre for Atmospheric Science, University of Reading, Reading, United Kingdom

<sup>5</sup>Barcelona Supercomputing Center, Barcelona, Spain

<sup>6</sup>Plymouth Marine Laboratory, Plymouth, United Kingdom

<sup>7</sup>Met Office, Exeter, United Kingdom

<sup>8</sup>Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

<sup>9</sup>MARUM - Center for Marine Environmental Sciences, Bremen, Germany

<sup>10</sup>Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden

The Earth System Model Evaluation Tool (ESMValTool) is a free and open-source community diagnostic and performance metrics tool for the evaluation of Earth system models participating in the Coupled Model Intercomparison Project (CMIP). Version 2 of the tool (Righi et al. 2019, [www.esmvaltool.org](http://www.esmvaltool.org)) features a brand new design, consisting of ESMValCore (<https://github.com/esmvalgroup/esmvalcore>), a package for working with CMIP data and ESMValTool (<https://github.com/esmvalgroup/esmvaltool>), a package containing the scientific analysis scripts. This new version has been specifically developed to handle the increased data volume of CMIP Phase 6 (CMIP6) and the related challenges posed by the analysis and the evaluation of output from multiple high-resolution or complex Earth system models. The tool also supports CMIP5 and CMIP3 datasets, as well as a large number of re-analysis and observational datasets that can be formatted according to the same standards (CMOR) on-the-fly or through scripts currently included in the ESMValTool package.

At the heart of this new version is the ESMValCore software package, which provides a configurable framework for finding CMIP files using a “data reference syntax”, applying commonly used pre-processing functions to them, running analysis scripts, and recording provenance. Numerous pre-processing functions, e.g. for data selection, regridding, and statistics are readily available and the modular design makes it easy to add more. The ESMValCore package is easy to install with relatively few dependencies, written in Python 3, based on state-of-the-art open-source libraries such as Iris and Dask, and widely used standards such as YAML, NetCDF, CF-Conventions, and W3C PROV. An extensive set of automated tests and code quality checks ensure the reliability of the package. Documentation is available at <https://esmvaltool.readthedocs.io>.

The ESMValCore package uses human-readable recipes to define which variables and datasets to use, how to pre-process that data, and what scientific analysis scripts to run. The package provides convenient interfaces, based on the YAML and NetCDF/CF-convention file formats, for running diagnostic scripts written in any programming language. Because the ESMValCore framework takes care of running the workflow defined in the recipe in parallel, most analyses run much faster, with no additional programming effort required from the authors of the analysis scripts. For example, benchmarks show a factor of 30 speedup with respect to version 1 of the tool for a representative recipe on a 24 core machine. A large collection of standard recipes and associated analysis scripts is available in the ESMValTool package for reproducing selected peer-reviewed analyses. The ESMValCore package can also be used with any other script that implements its easy to use interface. All pre-processing functions of the ESMValCore can also be used directly from any Python program. These features allow for use by a wide community of scientific users and developers with different levels of programming skills and experience.

Future plans involve extending the public Python API (application programming interface) from just preprocessor functions to include all functionality, including finding the data and running diagnostic scripts. This would make ESMValCore suitable for interactive data exploration from a Jupyter Notebook.