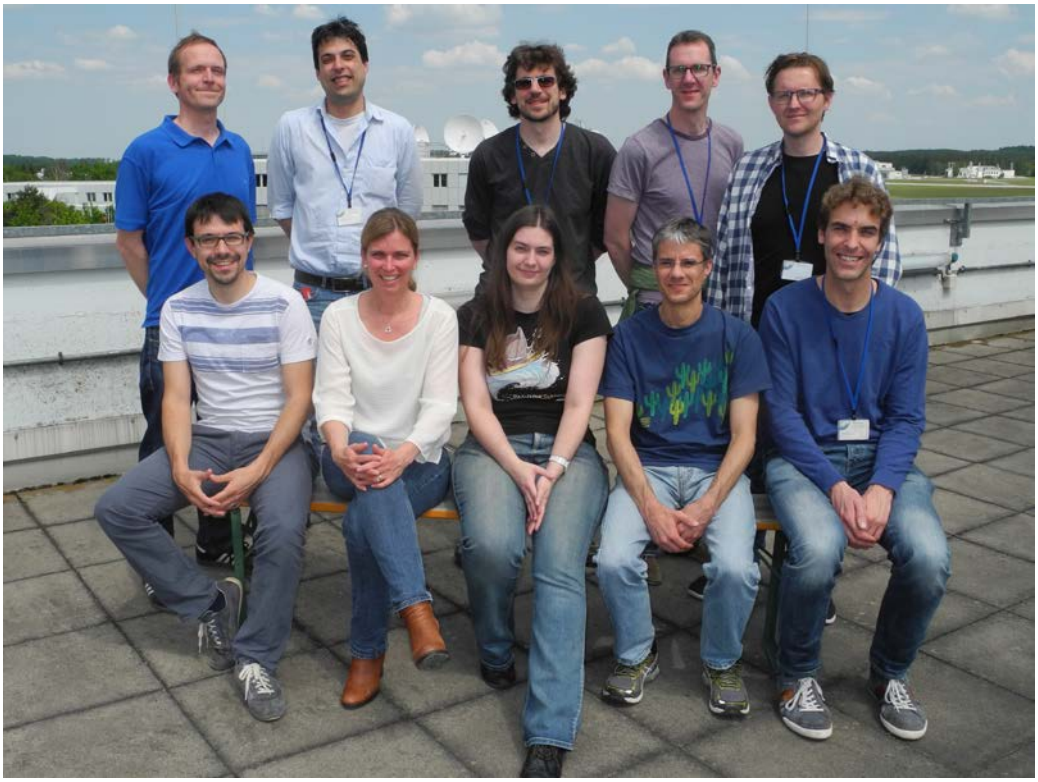
2nd ESMValTool Backend Coding Workshop

Workshop summary

Convener: Mattia Righi, Veronika Eyring, and Axel Lauer (DLR, Germany)



Workshop Participants, from left to right, front row: Mattia Righi (DLR, Germany), Veronika Eyring (DLR, Germany), Bettina Gier (DLR, Germany), Axel Lauer (DLR, Germany), Maarten Plieger (KNMI, Netherlands); back row: Björn Brötz (DLR, Germany), Valeriu Predoi (U Reading, UK), Javier Vegas (BSC, Spain), Bill Little (MetOffice, UK), Niels Drost (NLESC, Netherlands); Not on the photo: Grenville Lister (U Reading, UK), Alex Loew (LMU Munich, Germany), and David Hassell (U Reading, UK).

Venue: the 2nd ESMValTool Backend Coding Workshop was held at the Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institute of Atmospheric Physics in Oberpfaffenhofen, Germany, from 15-19 May 2017.

Workshop Goals: The coding workshop brought together experts from the ESMValTool and IRIS development teams with the goal to revise the current version of the ESMValTool backend towards improved efficiency and performance. The workshop was a follow-up workshop of the '1st ESMValTool Backend Coding Workshop' that was held at the MetOffice in Exeter (UK) from 6-10 February 2017.

The ESMValTool applies diagnostics and metrics scripts to a wide range of input data, including models from CMIP5 and other model intercomparison projects (MIPs), selected models in their native format (EMAC, GFDL, EC-Earth and, in the future, the MetOffice model), and observations. Before the analysis and diagnostic codes can be applied, a number of common operations such as extraction, reformatting, regridding, masking, etc., need to be applied to the input data. These operations are collectively named pre-processing and are (or should be) performed by the backend. The current version of the ESMValTool performs only a few of these operations in a centralized way

in the backend, while others are performed at a lower level, for example within specific diagnostics. This results in several drawbacks, such as slow performance, code duplication, lack of consistency among the approaches, and limited documentation.

The workshop's main goal is to rewrite the backend based on IRIS in order to allow an efficient pre-processing of the input data (models and observations), covering common operations such as format check, variable derivation, regridding, masking, and temporal and spatial subsetting. Given the huge amount of data that CMIP6 will produce (40-60 Pb), an efficient processing of such data represent a significant challenge for the diagnostic tools. The revised ESMValTool backend will allow for a better separation of pre-processing tasks from the diagnostic analysis, centralizing the most common operations and implementing new features to increase efficiency, including parallelization. The revised ESMValTool backend will be largely based on the IRIS python package developed at the MetOffice, thus also enhancing the collaboration across institutions and within the community.

Achievements:

The workshop was organized in four groups focusing on each of the pre-processing tasks mentioned above and in the backend report, as well as on the higher level requirements for coupling the backend with the ESMValTool:

- *Variable reading and CMORization*: functions for checking data compliance have been implemented. These check variable's name, units and coordinates, and report and fix the most common errors in input data.
- *Regridding*: several regridding routines have been programmed and tested. They cover horizontal regridding of regular and irregular grids to a user-defined target grid. The first allows for several regridding schemes (bilinear, area-conserving, nearest-neighbour), whereas for the latter only an unstructured-nearest scheme is available in IRIS. Further method will be explored for this case. Vertical regridding of both atmospheric pressure and ocean depth coordinates is also available, using a linear scheme. Python tests have been implemented for all the regridding routines.
- *Masking*: a first version of a masking routine for input data has been developed, allowing for efficient masking of land and ocean areas, and for other user-specified regions. A concept has been discussed for a consistent masking of missing values across the input models and observations: the corresponding code will be implemented soon.
- *New ESMValTool namelist format*: as discussed in Exeter, a new namelist format will be adopted to allow for a better control of the backend operations by the user. The yaml format will replace the current xml format. A first version of the corresponding Python parser has been developed and the conversion of the xml namelists to yaml is ongoing, with about 70% of the namelist already converted. Further testing and refining of the new namelist defitions will be required.
- *Higher level requirements*: the possibility to run multiple instances of the ESMValTool namelists at the same time has been included. This is realized by writing the temporary files in a different temporary folder for each instance. The overall processing chain of the backend operations has been defined and a first version has been coded.

The workshop was held under the auspices of the ESA Climate Model User Group (CMUG), the EU Horizon 2020 Coordinated Research in Earth Systems and Climate: Experiments, kNowledge, Dissemination and Outreach (CRESCENDO) project, the EU Horizon 2020 PRocess-based climate sIMulation: AdVances in high-resolution modelling and European climate Risk Assessment (PRIMAVERA) project, and the Institute of Atmospheric Physics of the German Aerospace Center (DLR).