EGU2020-14745, updated on 18 Jan 2022
https://doi.org/10.5194/egusphere-egu2020-14745
EGU General Assembly 2020

ESMValTool pre-processing functions for eWaterCycle

Fakhereh Alidoost[1], Jerom Aerts[2], Bouwe Andela[1], Jaro Camphuijsen[1], Nick van De Giesen[2], Gijs van Den Oord[1], Niels Drost[1], Yifat Dzigan[1], Ronald van Haren[3], Rolf Hut[2], Peter C. Kalverla[1], Inti Pelupessy[1], Stefan Verhoeven[1], Berend Weel[1], and Ben van Werkhoven[1]

[1]Netherlands eScience Center, Amsterdam, the Netherlands
[2]Delft University of Technology, Faculty of Civil Engineering and Geoscience, Delft, the Netherlands
[3]Rijkswaterstaat, Utrecht, the Netherlands

eWaterCycle is a framework in which hydrological modelers can work together in a collaborative environment. In this environment, they can, for example, compare and analyze the results of models that use different sources of (meteorological) forcing data. The final goal of eWaterCycle is to advance the state of FAIR (Findable, Accessible, Interoperable, and Reusable) and open science in hydrological modeling.

Comparing hydrological models has always been a challenging task. Hydrological models exhibit great complexity and diversity in the exact methodologies applied, competing for hypotheses of hydrologic behavior, technology stacks, and programming languages used in those models. Pre-processing of forcing data is one of the roadblocks that was identified during the FAIR Hydrological Modelling workshop organized by the Lorentz Center in April 2019. Forcing data can be retrieved from a wide variety of sources with discrepant variable names and frequencies, and spatial and temporal resolutions. Moreover, some hydrological models make specific assumptions about the definition of the forcing variables. The pre-processing is often performed by various sets of scripts that may or may not be included with model source codes, making it hard to reproduce results. Generally, there are common steps in the data preparation among different models. Therefore, it would be a valuable asset to the hydrological community if the pre-processing of FAIR input data could also be done in a FAIR manner.

Within the context of the eWaterCycle II project, a common pre-processing system has been created for hydrological modeling based on ESMValTool (Earth System Model Evaluation Tool). ESMValTool is a community diagnostic and performance metrics tool developed for the evaluation of Earth system models. The ESMValTool pre-processing functions cover a broad range of operations on data before diagnostics or metrics are applied; for example, vertical interpolation, land-sea masking, re-gridding, multi-model statistics, temporal and spatial manipulations, variable derivation and unit conversion. The pre-processor performs these operations in a centralized, documented and efficient way. The current pre-processing pipeline of the eWaterCycle using ESMValTool consists of hydrological model-specific recipes and supports ERA5 and ERA-Interim data provided by the ECMWF (European Centre for Medium-Range Weather Forecasts). The pipeline starts with the downloading and CMORization (Climate Model Output Rewriter) of input

data. Then a recipe is prepared to find the data and run the preprocessors. When ESMValTool runs a recipe, it will also run the diagnostic script that contains model-specific analysis to derive required forcing variables, and it will store provenance information to ensure transparency and reproducibility. In the near future, the pipeline is extended to include Earth observation data, as these data are paramount to the data assimilation in eWaterCycle.

In this presentation we will show how using the pre-processor from ESMValTool for Hydrological modeling leads to connecting Hydrology and Climate sciences, and increase the impact and sustainability of ESMValTool.