Towards Reproducible Hydrological Modelling with eWaterCycle

Niels Drost[1], Jerom Aerts[2], Fakhereh Alidoost[1], Bouwe Andela[1], Jaro Camphuijsen[1], Yifat Dzigan[1], Nick Van De Giesen[2], Rolf Hut[2], Eric Hutton[3], Peter Kalverla[1], Maarten van Meersbergen[1], Gijs van den Oord[1], Inti Pelupessy[1], Stefan Verhoeven[1], Berend Weel[1], and Ben van Werkhoven[1]

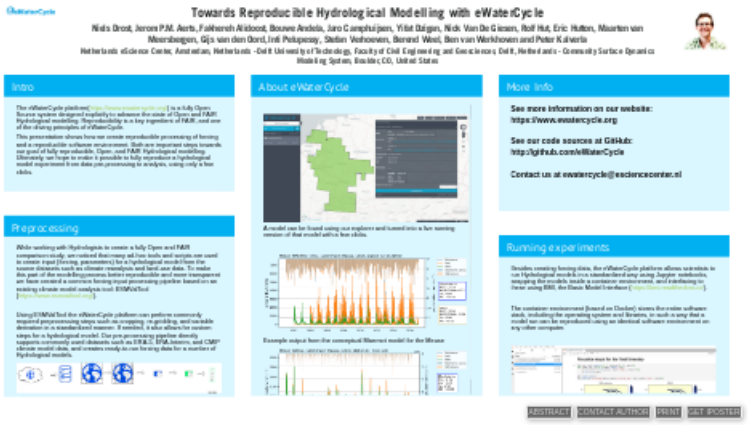
[1]Netherlands eScience Center
[2]Delft University of Technology
[3]Community Surface Dynamics Modeling System

November 23, 2022

Abstract

The eWaterCycle platform(https://www.ewatercycle.org/) is a fully Open Source system designed explicitly to advance the state of Open and FAIR Hydrological modelling. Reproducibility is a key ingredient of FAIR, and one of the driving principles of eWaterCycle. While working with Hydrologists to create a fully Open and FAIR comparison study, we noticed that many ad-hoc tools and scripts are used to create input (forcing, parameters) for a hydrological model from the source datasets such as climate reanalysis and land-use data. To make this part of the modelling process better reproducible and more transparent we have created a common forcing input processing pipeline based on an existing climate model analysis tool: ESMValTool (https://www.esmvaltool.org/). Using ESMValTool the eWaterCycle platform can perform commonly required pre-processing steps such as cropping, re-gridding, and variable derivation in a standardized manner. If needed, it also allows for custom steps for a Hydrological model. Our pre-processing pipeline directly supports commonly used datasets such as ERA-5, ERA-Interim, and CMIP climate model data, and creates ready-to-run forcing data for a number of Hydrological models. Besides creating forcing data, the eWaterCycle platform allows scientists to run Hydrological models in a standardized way using Jupyter notebooks, wrapping the models inside a container environment, and interfacing to these using BMI, the Basic Model Interface (https://bmi.readthedocs.io/). The container environment (based on Docker) stores the entire software stack, including the operating system and libraries, in such a way that a model run can be reproduced using an identical software environment on any other computer. The reproducible processing of forcing and a reproducible software environment are important steps towards our goal of fully reproducible, Open, and FAIR Hydrological modelling. Ultimately, we hope to make it possible to fully reproduce a Hydrological model experiment from data pre-processing to analysis, using only a few clicks.

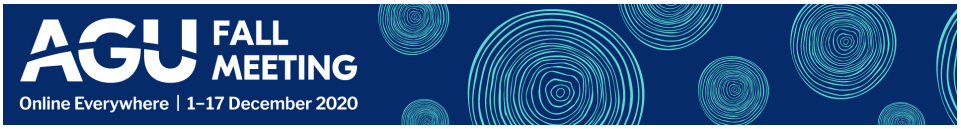
Towards Reproducible Hydrological Modelling with eWaterCycle



Niels Drost, Jerom P.M. Aerts, Fakhereh Alidoost, Bouwe Andela, Jaro Camphuijsen, Yifat Dzigan, Nick Van De Giesen, Rolf Hut, Eric Hutton, Maarten van Meersbergen, Gijs van den Oord, Inti Pelupessy, Stefan Verhoeven, Berend Weel, Ben van Werkhoven and Peter Kalverla

Netherlands eScience Center, Amsterdam, Netherlands - Delft University of Technology, Faculty of Civil Engineering and Geosciences, Delft, Netherlands - Community Surface Dynamics Modeling System, Boulder, CO, United States

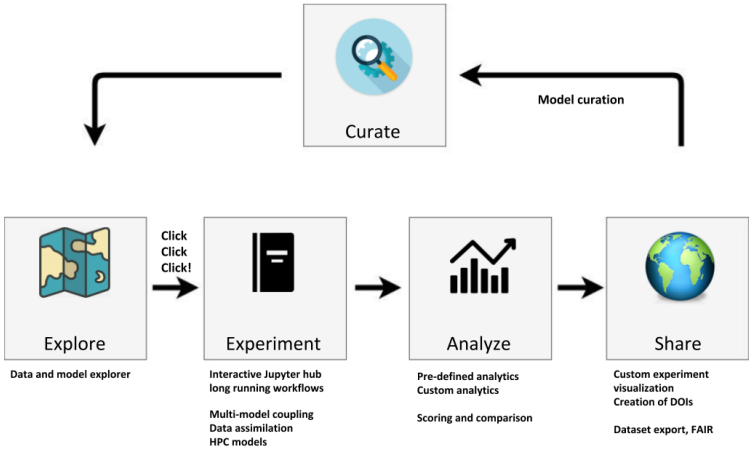
PRESENTED AT:



INTRO

The eWaterCycle platform(https://www.ewatercycle.org/ (https://www.ewatercycle.org/)) is a fully Open Source system designed explicitly to advance the state of Open and FAIR Hydrological modelling. Reproducibility is a key ingredient of FAIR, and one of the driving principles of eWaterCycle.

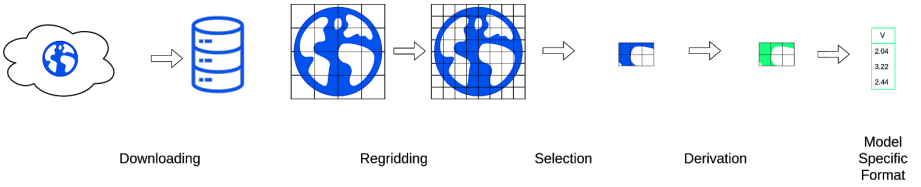
This presentation shows how we create reproducible processing of forcing and a reproducible software environment. Both are important steps towards our goal of fully reproducible, Open, and FAIR Hydrological modelling. Ultimately, we hope to make it possible to fully reproduce a hydrological model experiment from data pre-processing to analysis, using only a few clicks.



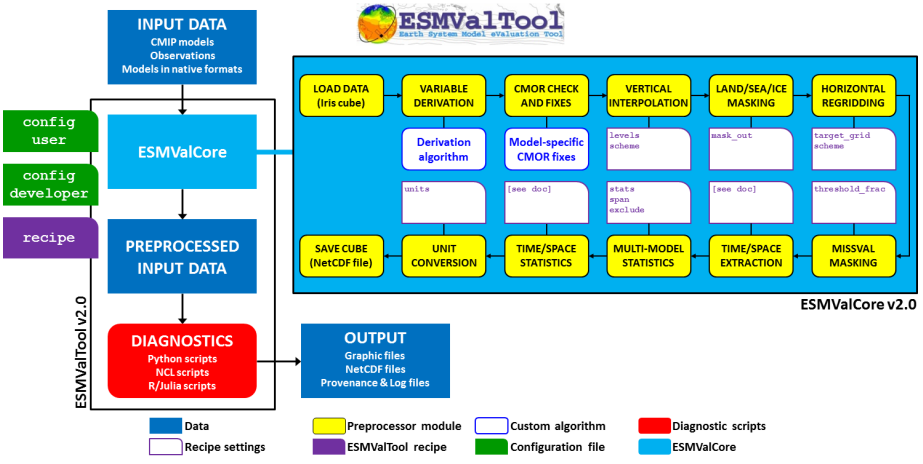
PREPROCESSING

While working with Hydrologists to create a fully Open and FAIR comparison study, we noticed that many ad-hoc tools and scripts are used to create input (forcing, parameters) for a hydrological model from the source datasets such as climate reanalysis and land-use data. To make this part of the modelling process better reproducible and more transparent we have created a common forcing input processing pipeline based on an existing climate model analysis tool: ESMValTool (https://www.esmvaltool.org/ (https://www.esmvaltool.org/)).

Using ESMValTool the eWaterCycle platform can perform commonly required preprocessing steps such as cropping, re-gridding, and variable derivation in a standardized manner. If needed, it also allows for custom steps for a hydrological model. Our pre-processing pipeline directly supports commonly used datasets such as ERA-5, ERA-Interim, and CMIP climate model data, and creates ready-to-run forcing data for a number of Hydrological models.

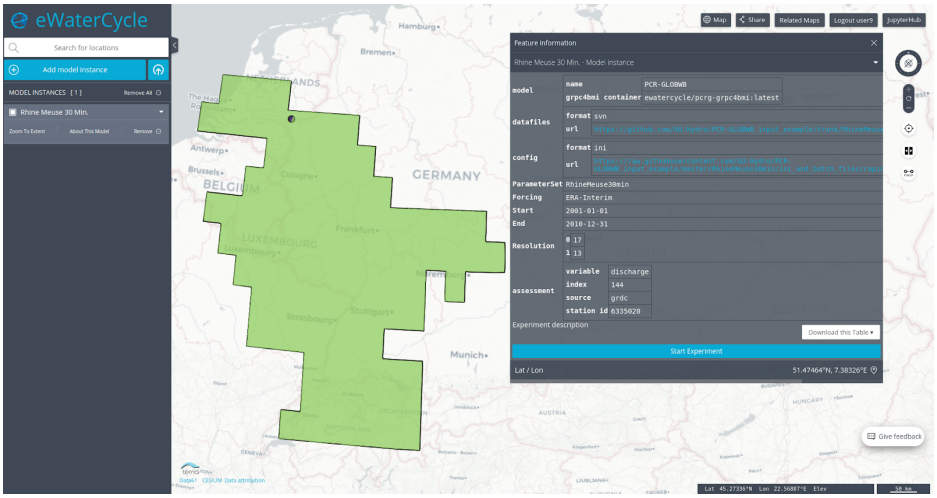


Example of a preprocessing pipeline for a certain model

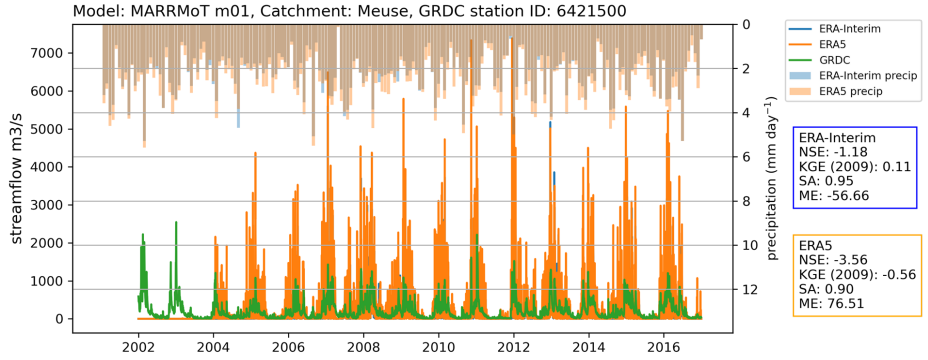


The full pipeline of preprocessing functionality in ESMValTool

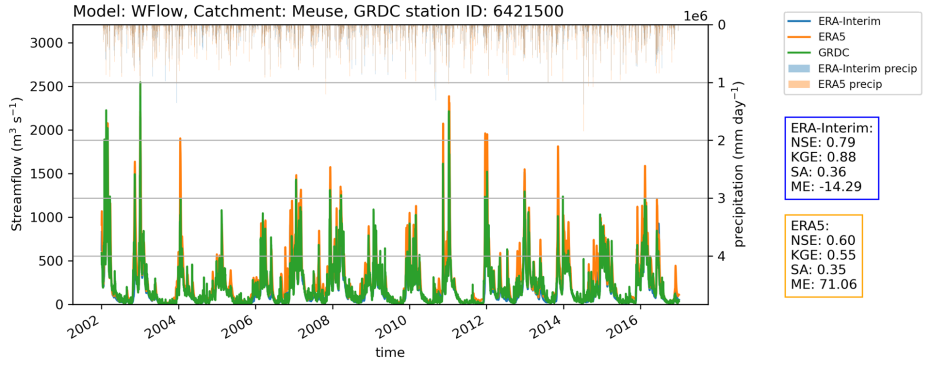
ABOUT EWATERCYCLE



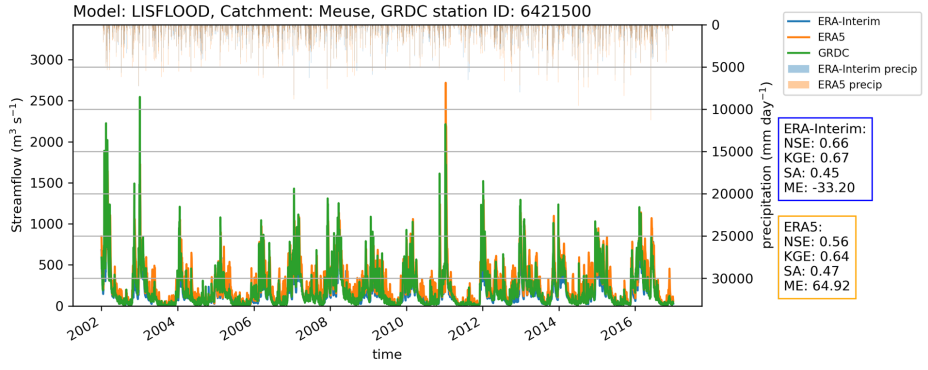
A model can be found using our explorer and turned into a live running version of that model with a few clicks.



Example output from the conceptual Marrmot model for the Meuse



Example output from the WFLOW SBM model for the Meuse



Example output from the LISFLOOD model for the Meuse

MORE INFO

See more information on our website:

https://www.ewatercycle.org

See our code sources at GitHub:

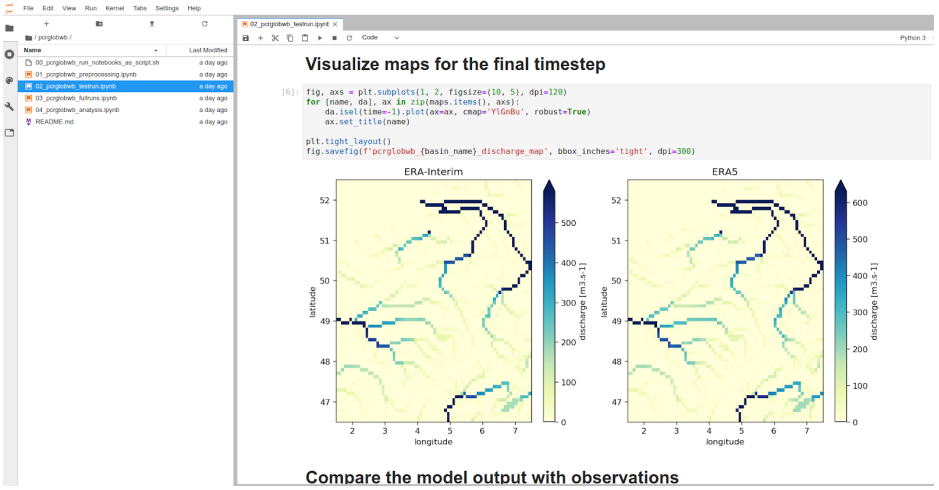
http://github.com/eWaterCycle

Contact us at ewatercycle@esciencecenter.nl

RUNNING EXPERIMENTS

Besides creating forcing data, the eWaterCycle platform allows scientists to run Hydrological models in a standardized way using Jupyter notebooks, wrapping the models inside a container environment, and interfacing to these using BMI, the Basic Model Interface (https://bmi.readthedocs.io/ (https://bmi.readthedocs.io/)).

The container environment (based on Docker) stores the entire software stack, including the operating system and libraries, in such a way that a model run can be reproduced using an identical software environment on any other computer.



A Jupyter notebook running a model and visualizing model output.

ABSTRACT

The eWaterCycle platform(https://www.ewatercycle.org/ (https://www.ewatercycle.org/)) is a fully Open Source system designed explicitly to advance the state of Open and FAIR Hydrological modelling. Reproducibility is a key ingredient of FAIR, and one of the driving principles of eWaterCycle.

While working with Hydrologists to create a fully Open and FAIR comparison study, we noticed that many ad-hoc tools and scripts are used to create input (forcing, parameters) for a hydrological model from the source datasets such as climate reanalysis and land-use data. To make this part of the modelling process better reproducible and more transparent we have created a common forcing input processing pipeline based on an existing climate model analysis tool: ESMValTool (https://www.esmvaltool.org/ (https://www.esmvaltool.org/)).

Using ESMValTool the eWaterCycle platform can perform commonly required pre-processing steps such as cropping, re-gridding, and variable derivation in a standardized manner. If needed, it also allows for custom steps for a Hydrological model. Our pre-processing pipeline directly supports commonly used datasets such as ERA-5, ERA-Interim, and CMIP climate model data, and creates ready-to-run forcing data for a number of Hydrological models.

Besides creating forcing data, the eWaterCycle platform allows scientists to run Hydrological models in a standardized way using Jupyter notebooks, wrapping the models inside a container environment, and interfacing to these using BMI, the Basic Model Interface (https://bmi.readthedocs.io/ (https://bmi.readthedocs.io/)). The container environment (based on Docker) stores the entire software stack, including the operating system and libraries, in such a way that a model run can be reproduced using an identical software environment on any other computer.

The reproducible processing of forcing and a reproducible software environment are important steps towards our goal of fully reproducible, Open, and FAIR Hydrological modelling. Ultimately, we hope to make it possible to fully reproduce a Hydrological model experiment from data pre-processing to analysis, using only a few clicks.