Geophysical Research Abstracts
Vol. 17, EGU2015-11033, 2015
EGU General Assembly 2015

Facilitating hydrological data analysis workflows in R: the RHydro package

Wouter Buytaert (1), Simon Moulds (1), Jon Skoien (2), Edzer Pebesma (3), and Dominik Reusser (4)

(1) Imperial College London, Civil and Environmental Engineering, London, United Kingdom (w.buytaert@imperial.ac.uk), (2) Institute for Environment and Sustainability, Joint Research Centre - European Commission, Ispra, Italy, (3) Institute for Geoinformatics, University of Münster, Münster, Germany, (4) Potsdam Institute for Climate Impact Research, Potsdam, Germany

The advent of new technologies such as web-services and big data analytics holds great promise for hydrological data analysis and simulation. Driven by the need for better water management tools, it allows for the construction of much more complex workflows, that integrate more and potentially more heterogeneous data sources with longer tool chains of algorithms and models. With the scientific challenge of designing the most adequate processing workflow comes the technical challenge of implementing the workflow with a minimal risk for errors.

A wide variety of new workbench technologies and other data handling systems are being developed. At the same time, the functionality of available data processing languages such as R and Python is increasing at an accelerating pace.

Because of the large diversity of scientific questions and simulation needs in hydrology, it is unlikely that one single optimal method for constructing hydrological data analysis workflows will emerge. Nevertheless, languages such as R and Python are quickly gaining popularity because they combine a wide array of functionality with high flexibility and versatility. The object-oriented nature of high-level data processing languages makes them particularly suited for the handling of complex and potentially large datasets.

In this paper, we explore how handling and processing of hydrological data in R can be facilitated further by designing and implementing a set of relevant classes and methods in the experimental R package RHydro. We build upon existing efforts such as the sp and raster packages for spatial data and the spacetime package for spatiotemporal data to define classes for hydrological data (HydroST). In order to handle simulation data from hydrological models conveniently, a HM class is defined. Relevant methods are implemented to allow for an optimal integration of the HM class with existing model fitting and simulation functionality in R.

Lastly, we discuss some of the design challenges of the RHydro package, including integration with big data technologies, web technologies, and emerging data models in hydrology.