hddtools: Hydrological Data Discovery Tools

27 December 2016

Summary

The hddtools (Vitolo, 2017) (**h**ydrological **d**ata **d**iscovery **tools**) is an R package (R Core Team, 2016) designed to facilitate access to a variety of online open data sources relevant for hydrologists and, in general, environmental scientists and practitioners. This typically implies the download of a metadata catalogue, selection of information needed, formal request for dataset(s), de-compression, conversion, manual filtering and parsing. All those operation are made more efficient by re-usable functions.

Depending on the data license, functions can provide offline and/or online modes. When redistribution is allowed, for instance, a copy of the dataset is cached within the package and updated twice a year. This is the fastest option and also allows offline use of package's functions. When re-distribution is not allowed, only online mode is provided.

Datasets for which functions are provided include: the Global Runoff Data Center (GRDC), the Scottish Environment Protection Agency (SEPA), the Top-Down modelling Working Group (Data60UK and MOPEX), Met Office Hadley Centre Observation Data (HadUKP Data) and NASA's Tropical Rainfall Measuring Mission (TRMM).

This package follows a logic similar to other packages such as rdefra (Vitolo et al., 2016b) and rnrfa (Vitolo et al., 2015): sites are first identified through a catalogue (if available), data are imported via the station identification number, then data are visualised and/or used in analyses. The metadata related to the monitoring stations are accessible through the functions: `catalogueGRDC()`, `catalogueSEPA()`, `catalogueData60UK()` and `catalogueMOPEX()`. Time series data can be obtained using the functions: `tsGRDC()`, `tsSEPA()`, `tsData60UK()`, `tsMOPEX()` and `HadDAILY()`. Geospatial information can be retrieved using the functions: `KGClimateClass()` returning the Koppen-Greiger climate zone and `TRMM()` which retrieves global historical rainfall estimations.

The retrieved hydrological time series (e.g. using `tsData60UK()`) can be used to feed hydrological models such as fuse (Vitolo et al., 2012, 2016a), topmodel (Buytaert, 2011) and hydromad (Andrews and Guillaume, 2016; Andrews et al., 2011).

For more details and examples, please refer to the help pages and vignette.

References

Andrews, F. and Guillaume, J.: hydromad: Hydrological Model Assessment and Development. [online] Available from: http://hydromad.catchment.org/, 2016.

Andrews, F., Croke, B. and Jakeman, A.: An open software environment for hydrological model assessment and development, Environmental Modelling & Software, 26(10), 1171–1185, doi:http://dx.doi.org/10.1016/j.envsoft.2011.04.006, 2011.

Buytaert, W.: topmodel: Implementation of the hydrological model TOPMODEL in R. [online] Available from: https://CRAN.R-project.org/package=topmodel, 2011.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. [online] Available from: https://www.R-project.org/, 2016.

Vitolo, C.: hddtools: Hydrological Data Discovery Tools. [online] Available from: https://CRAN.R-project.

org/package=hddtools, 2017.

Vitolo, C., Wells, P., Dobias, M. and Buytaert, W.: fuse: Framework for Understanding Structural Errors. [online] Available from: https://github.com/cvitolo/fuse, 2012.

Vitolo, C., Fry, M. and Buytaert, W.: rnrfa: UK National River Flow Archive Data from R. [online] Available from: https://CRAN.R-project.org/package=rnrfa, 2015.

Vitolo, C., Wells, P., Dobias, M. and Buytaert, W.: fuse: An R package for ensemble Hydrological Modelling, The Journal of Open Source Software, 1(8), doi:10.21105/joss.00052, 2016a.

Vitolo, C., Russell, A. and Tucker, A.: rdefra: Interact with the UK AIR Pollution Database from DEFRA, The Journal of Open Source Software, 1(4), doi:10.21105/joss.00051, 2016b.