The Journal of Open Source Software

Hydrodata: A portal to hydrology and climatology data through Python

Taher Chegini[1] **and Hong-Yi Li**[1]

1 University of Houston

DOI: DOIunavailable

Software

- Review
- Repository
- Archive

Editor: Pending Editor

Reviewers:

- @Pending Reviewers

Submitted: N/A
Published: N/A

Summary

Over the last decade, the increasing availability of web services that offer hydrology and climatology data has facilitated publishing reproducible scientific researches in these fields. Such web services allow researchers to subset big databases and perform some of the common data processing operations on the server-side. However, implementing such services increases the technical complexity of code development as it requires sufficient understanding of their underlying protocols to generate valid queries and filters. `Hydrodata` tries to bridge this gap by providing a unified and simple Application Programming Interface (API) to web services that are based on three of the most commonly used protocols for geo-spatial/temporal data publication: REpresentational State Transfer (RESTful), Web Feature Services (WFS), and Web Map Services (WMS). `Hydrodata` is a software stack and includes the following Python packages:

- Hydrodata: Provides access to NWIS (National Water Information System), NID (National Inventory of Dams), HCDN-2009 (Hydro-Climatic Data Network), NLCD (National Land Cover Database), and SSEBop (operational Simplified Surface Energy Balance) databases. Moreover, it can generate an interactive map for exploring NWIS stations within a bounding box, compute categorical statistics of land use/land cover data, and plot five hydrologic signature graphs. There are several helper functions one of which returns a roughness coefficients lookup table for each NLCD land cover type. These coefficients can be useful for overland flow routing among other applications
- PyGeoOGC: Generates valid queries for retrieving data from supported RESTful-, WMS-, and WFS-based services. Although these web services limit the number of features in a single query, under-the-hood, `PyGeoOGC` takes care of breaking down a large query into smaller queries according to the service specifications. Additionally, this package offers several notable utilities: data re-projection, asynchronous data retrieval, and traversing a JSON (JavaScript Object Notation) object.
- PyGeoUtils: Converts responses from PyGeoOGC's supported web services to geo-dataframes (vector data type) or datasets (raster data type). Moreover, for gridded data, it can mask the output dataset based on any given geometry.
- PyNHD: Provides the ability to navigate and subset NHDPlus (National Hydrography Database), at medium- and high-resolution, using NLDI (Hydro Network-Linked Data Index), WaterData, and TNM (The National Map) web services. Additionally, it can retrieve over 30 catchment-scale attributes from ScienceBase that are linked to the NHDPlus database via Common Identifiers (ComIDs). `PyNHD` has some additional river network tools that use NHDPlus data for routing through a river network. These tools sort the river network topologically from upstream to downstream, then based on a user defined function transport the specified attribute through the network.
- Py3DEP: Gives access to topographic data through 3DEP (3D Elevation Program) service. This package can pull 12 types of topographic data from the 3DEP service such as Digital Elevation Model, slope, aspect, and hillshade.
- PyDaymet: Retrieves daily climate data as well as their monthly and annual summaries from the Daymet dataset. It is possible to request data for a single location as well as

a grid (any valid geometrical shape) at 1-km spatial resolution.

Furthermore, PyGeoOGC and PyGeoUtils are low-level engines of this software stack that the other four packages utilize for providing access to some of the most popular databases in the hydrology community. These two low-level packages are generic and developers can use them for connecting and sending queries to any other web services that are based on the protocols that Hydrodata supports.

Statement of need

Preparing input data for conducting hydrology and climatology studies is often one of the most time-consuming steps in such studies. The difficulties for processing such input data stem from diverse data sources and types as well as their sizes. For example, hydrological modeling of watersheds might require climate data such as precipitation and temperature, topology data such as Digital Elevation Model, and river network. Climate and topology data are available in raster format, and river network could be from a vector data type. Additionally, these data are available from different sources, for example, we can retrieve climate data from Daymet (Thornton et al., 2020), topology data from 3D Elevation Program (Thatcher et al., 2020), and river network from National Hydrography Database (Buto & Anderson, 2020). The diversity in data sources and their large size may hinder reproducible publication of such studies. Hydrodata software stack provides access to such databases through plethora of web services that public and private entities have made available.

There are several open-source packages that offer similar functionalities. For example, hydrofunctions is a Python package that retrieves streamflow data from NWIS and ulmo is another Python package that provides access to several public hydrology and climatology data. Dataretrieval gives access to some of the USGS (United States Geological Survey) databases and has two versions in R and Python. There is also a R Package with the same name, HydroData, that provides access to 15 earth system datasets. Although these packages offer similar functionalities to Hydrodata, none of the Python packages offer access to datasets from diverse sources and offer post-processing functionalities that Hydrodata provides.

Acknowledgements

We acknowledge contributions from Austin Raney and Emilio Mayorga to this project.

References

Buto, S. G., & Anderson, R. D. (2020). *NHDPlus high resolution (NHDPlus HR)—a hydrography framework for the nation*. US Geological Survey. https://doi.org/10.3133/fs20203033

Thatcher, C. A., Lukas, V., & Stoker, J. M. (2020). *The 3D elevation program and energy for the nation*. US Geological Survey. https://doi.org/10.3133/fs20193051

Thornton, M. M., Shrestha, R., Wei, Y., Thornton, P. E., Kao, S., & Wilson, B. E. (2020). *Daymet: Daily surface weather data on a 1-km grid for north america, version 4*. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1840