

SOFTWARE METAPAPER

Xanthos – A Global Hydrologic Model

Xinya Li¹, Chris R. Vernon¹, Mohamad I. Hejazi², Robert P. Link², Leyang Feng², Yaling Liu² and Lynn T. Rauchenstein³

¹ Pacific Northwest National Laboratory, Richland, WA 99352, US

² Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, Maryland, US

³ (Formerly) Pacific Northwest National Laboratory, Richland, WA 99352, US

Corresponding author: Xinya Li (Xinya.Li@pnnl.gov)

Xanthos is an open-source hydrologic model, written in Python, designed to quantify and analyse global water availability. Xanthos simulates historical and future global water availability on a monthly time step at a spatial resolution of 0.5 geographic degrees. Xanthos was designed to be extensible and used by scientists that study global water supply and work with the Global Change Assessment Model (GCAM). Xanthos uses a user-defined configuration file to specify model inputs, outputs and parameters. Xanthos has been tested using actual global data sets and the model is able to provide historical observations and future estimates of renewable freshwater resources in the form of total runoff.

Keywords: Water Balance Model; Global Hydrologic Model; Runoff; River Routing; Potential Evapotranspiration (PET); Global Change Assessment Model (GCAM); Python

Funding statement: PNNL is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830.

(1) Overview

Introduction

Xanthos is an open-source hydrologic model written in Python,¹ designed to quantify and analyse global water availability. Xanthos was developed at the Joint Global Change Research Institute of the Pacific Northwest National Laboratory (<http://www.globalchange.umd.edu>). The main objective of the model is to quantify changes in future freshwater availability under various climate change regimes, and to serve as the freshwater supply component of the Global Change Assessment Model (GCAM) [1–3]. The model has been used in previous publications to explore different climate and socioeconomic scenarios over the 21st century and assess the influence and implications of the different scenarios on regional and global water availability [4–5].

The core of Xanthos is a hydrologic model that includes modules for calculating potential evapotranspiration (PET), runoff generation, and river routing (**Figure 1**). The underlying equations and algorithms for the hydrologic model are described in details in Hejazi et al. [4], and Zhou et al. [7]. The model requires inputs of gridded monthly temperature and precipitation, a maximum soil water storage capacity map (assumed static over time). Using those inputs, the model calculates gridded monthly runoff, PET, actual evapotranspiration, water storage in the soil column, channel water storage, and average channel flow.

The PET module uses the Hargreaves method [6], which requires average monthly temperature and average daily temperature range in degree Celsius. The stream routing module includes a cell-to-cell river routing scheme adopted from the modified river transport model (RTM), which uses a linear advection formula [7]. The aggregation module of Xanthos provides estimates of maximum naturally-available water fluxes (mm) or volumes (billion m³) monthly or annually (i.e., total monthly runoff) for each river basin. Xanthos includes an option to obtain the annual volumes (billion m³/year) of renewable water resources that are accessible for human use in each basin [8], which is calculated by the accessible-water module.

Xanthos can be applied as an independent model as well as serve as the water supply component in the GCAM. The primary goals of Xanthos are:

1. Be distributed as an open-source software that can be utilized and extended by researchers who study global hydrology.
2. Make estimates of global runoff, average channel flow, actual evapotranspiration, and water storage in the soil column on a monthly time step at a spatial resolution of 0.5 geographic degrees.
3. Employ a standardized configuration and simplified input structure that is easy for the user to update.
4. Direct diagnostics on the runoff results.

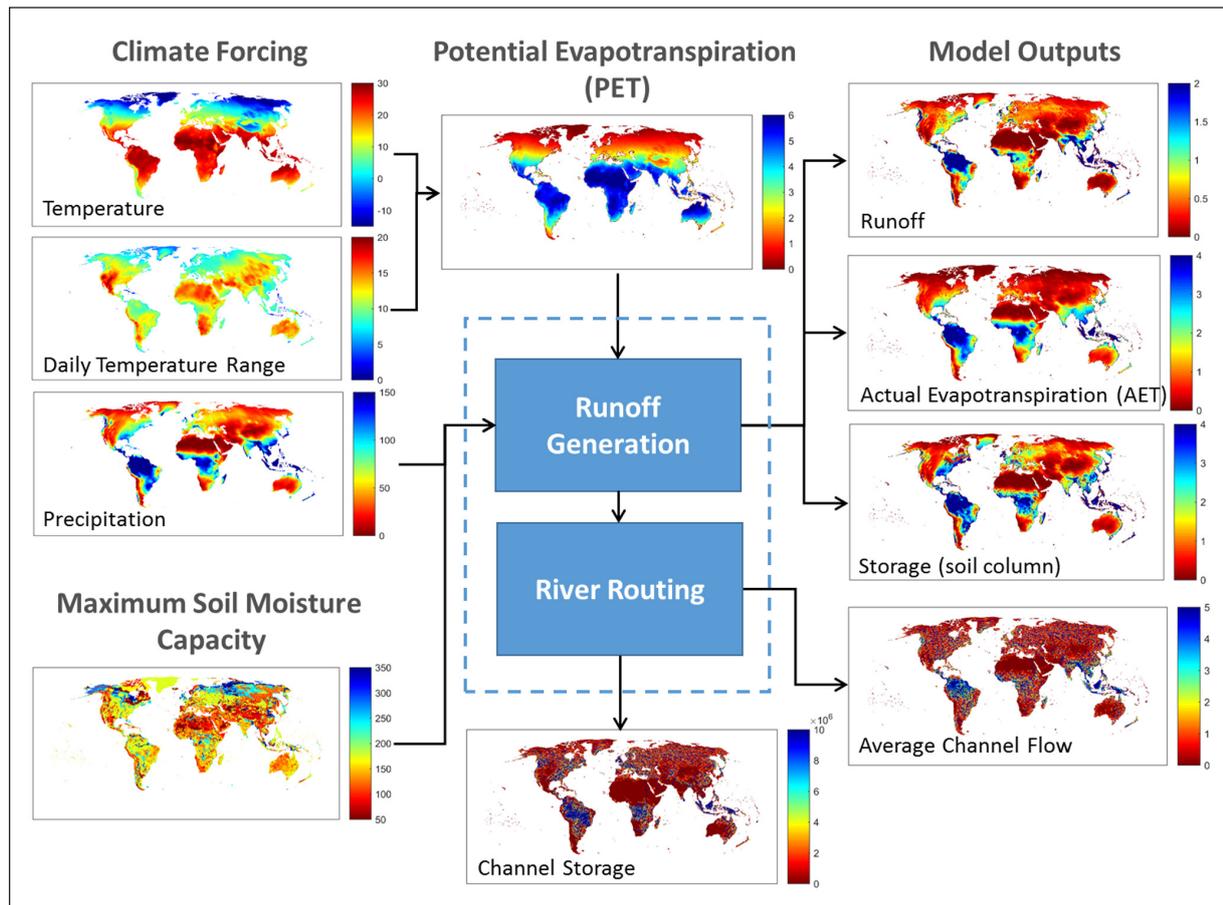


Figure 1: Main inputs and outputs of Xanthos hydrologic model (Runoff Generation & River Routing).

To achieve the above goals, Python was selected as the programming language because of its ease of use and extensibility.

Implementation and Architecture

The main inputs used by Xanthos are global gridded maps of monthly climate data and common map data including basin, country and region information. The spatial resolution of the global gridded data maps is 0.5 geographic degrees. These global gridded data maps contain a total of 259,200 grid cells (360 × 720) of which 67,420 grid cells are categorized at “land grids” and are considered valid for modelling purposes. The valid grid cells are used to define a “gridded” map according to the coordinates listed in the coordinate file in the input folder. The coordinate file, including the coordinates and the indexes of the 67,420 cells on the 360 × 720 grid, is used to convert a 2D map data into a 1D vector and vice versa. To aggregate the gridded data into basin/country/region scale for outputs and diagnostics, certain commonly used global data maps such as IDs of basins/countries/regions are harmonized into the gridded format required by Xanthos. The inputs converted using the 67,420 grid cells according to the coordinate data file are called harmonized inputs.

The flowchart of Xanthos is illustrated in **Figure 2** and **Figure 3** shows the overall architecture of Xanthos. The

PET calculation, runoff generation, and river routing are the core components of the hydrologic model. The gridded input files required by Xanthos can be divided into two classes, 1) monthly climate data and 2) static global maps and hydrologically related data. Xanthos majorly consists of four module packages:

1. DataReader: Configuration (class “ConfigSettings”) and import inputs (“DataLoad” module) according to user defined INI¹¹ file (“IniReader” module).
2. DataWriter: Export outputs (“OUTWriter” module) by options.
3. Diagnostics: Aggregate gridded runoff results at basin/country/region scales (“Aggregation” module), perform diagnostics (“Diagnostics” module), or create time series plots (“TimeSeries” module) by options.
4. RunoffStreamGen: Calculate PET (“PETCalculation” module), calculate runoff (“RunoffGen” module) and river routing (“StreamRouting” module) using monthly water balance model, calculate accessible water (“Accessible” module).

Xanthos can be executed in either historical or future mode. For the historical mode, the climate forcing data is taken from a historical time period (e.g. data before year 2005). As for the future mode, initial soil moisture and channel storage conditions can be taken either from

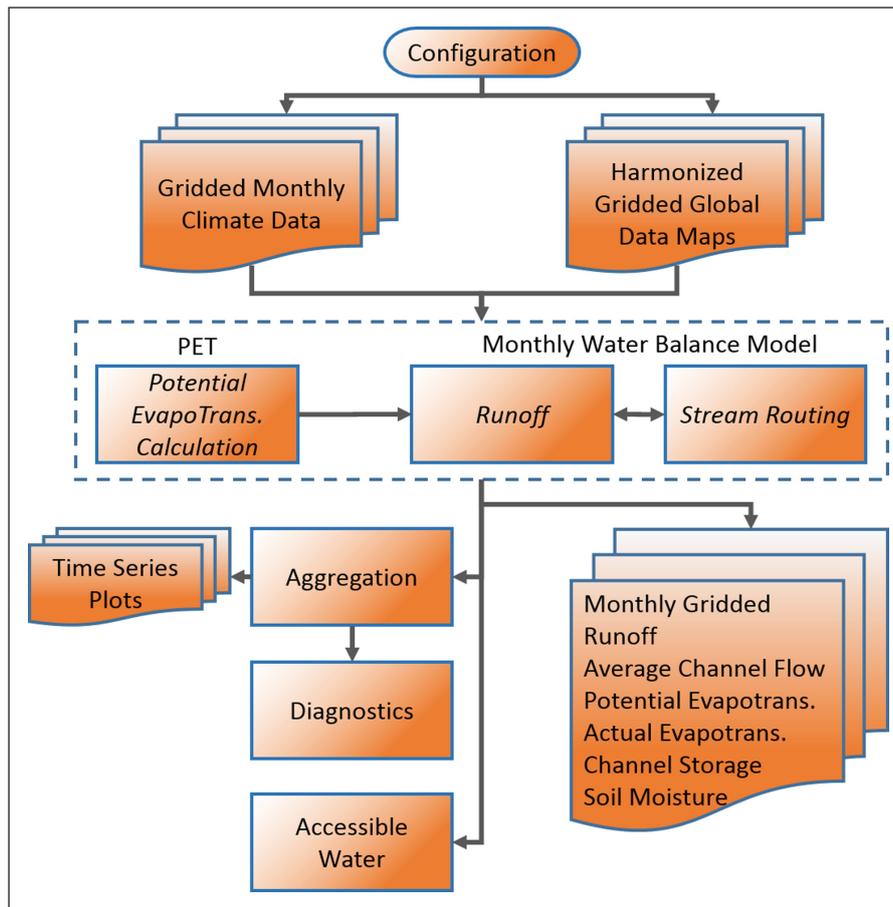


Figure 2: Flowchart of Xanthos.

a historical case calculated by Xanthos or from other models, and the climate forcing data is for future years (e.g. 2006–2100). Xanthos allows the spin-up (model initiation by using the results from the first few years) phase by setting the years for the spin-up time.

The default outputs from the model are monthly gridded total runoff, average channel flow, potential evapotranspiration, actual evapotranspiration, and channel storage, and water storage in soil column. There are several options to output the gridded results. The default unit for these five outputs is mm/month formatted as classic NetCDF⁶ file. The alternative output format is CSV (comma-separated values). Another optional unit is billion m³/month. In addition to monthly output results, there is an option for annually output results. In historical mode, two more outputs will be saved, monthly soil column moisture (mm/month) and monthly channel storage (m³/month). These two files will be served as initialization data for future mode. Optional outputs include: 1) aggregated runoff by basin, country or region; 2) time series plots of runoff and average channel flow by basin, country or region; 3) runoff comparisons with other models (diagnostics, described in “Quality Control”); and 4) accessible water.

“GCAM_Hydro.py” defines the main function that integrates all the modules shown in **Figure 2**. “test.py” is the executable Python file connecting “config.ini” and

“GCAM_Hydro.py”. A model simulation can be executed with a single command:

```
$ python test.py
```

To simplify the model run process, all the input data files are located in a separate folder from the code folders. The input folder contains:

1. Climate data files are the major user controlled inputs. There are three classic NetCDF files; 1) precipitation, 2) temperature and 3) daily maximum temperature range. Each climate data file has a size of 67,420 rows and NM (total number of months, for example, for year 1996–2005, NM = 10*12 = 120) columns.
2. Harmonized gridded global data maps and other required parameter data files required, which can be modified or replaced by the user.
3. Example data files of other models for comparison in diagnostics.
4. Example data files for accessible water.

The metadata (data source, format, related pre-processing, etc.) of all the input files are described in a document called “ReadMe_Input_Data.txt” included in the input folder.

The interface between the user and Xanthos is a configuration file, i.e., “config.ini”. One simulation requires

a unique configuration file. The configuration file is an INI file that is organized into six separate sections, each of which contains a few of setting parameters. The six sections are:

1. Project (Required): defines the major settings such as the paths of input and output folders, the output formatting, options for aggregated runoff results, an option for diagnostics, an option for time series plot, and an option for calculation of accessible water.
2. Climate (Required): defines the historical/future mode, paths of input climate forcing data – precipitation, temperature and daily temperature range files, and an option for initialization.
3. GriddedMap (Required): defines the required data maps, such as coordinates of the grids, maximum soil moisture of each grid, ID of basin/country/region for each grid.
4. Diagnostics (Required only if “PerformDiagnostics = 1” in section 1): defines the paths of the data files from other models (example models are described in “Quality Control” section), and the options for comparison at basin, country or region scale.
5. TimeSeriesPlot (Required only if “CreateTimeSeriesPlot = 1” in section 1): defines the scale for plots and the selected basin/country/region to be used for generating the plots.
6. AccessibleWater (Required only if “CalculateAccessibleWater = 1” in section 1): defines all the related parameters to the calculation of accessible water, such as the total reservoir capacity at basin level, and a baseflow index file.

Section 4, 5 and 6 are optional responding to the flags defined in Section 1. Default data files for inputs are included in the input folder. An output folder will be automatically generated at the path defined in the configuration file.

Xanthos was developed using Python¹ (version 2.7) and related scientific libraries. NumPy² and SciPy³ are the fundamental packages for scientific computing data processing. Matplotlib⁴ is a library used for 2D plotting of timer series plots and scatter plots from diagnostics. Pandas⁵ was used in accessible water module for data analysis. The results are able to be saved into two formats: CSV and classic NetCDF (Version 3). Classic NetCDF⁶ is a self-describing, machine-independent data formats for array-oriented scientific data which requires much smaller storage space [14]. By default, the input climate data files should be stored as classic NetCDF files, but they can also be stored as mat (MATLAB⁷ formatted) files if needed. The mat files can be read and write by input/output modules of SciPy.

Quality Control

This hydrologic model was tested by using different climate forcing datasets. Its ability to reproduce historical values is evaluated against observations, statistical assessments and other global hydrologic models (Figures S3 and S4 in the Supplement of [4]). The testing was completed using

different operating systems: Linux Windows and Mac. To help the user get familiar with the features and functions of Xanthos, two example test cases were included with the input climate data files:

1. Test case for a historical run of 10 years (1996–2005) with options for aggregation, diagnostics, accessible water and time series plots, turned on. The configuration file is “config_hist.ini” and the outputs are saved in the folder of “Test_Case_Hist” (Figure 3).
2. Test case for a future run of 15 years (2006–2020) with the initial channel storage and soil moisture data taken from the ending values of test case A and with basic options. The configuration file is “config_futu.ini” and the outputs are saved in the folder of “Test_Case_Futu” (Figure 3).

By changing the name of the configuration file (line 19 in “test.py”), the use is able to run a certain case.

An automatically generated log file provides details of each model run. The log file lists model settings, the processing steps, and CPU cost and warnings if applicable. The log file will record each step and indicate if the model run successfully. For example, if the hydrologic simulation finished successfully, “-----Simulation has finished successfully.....” will be printed into log file. When the model runs through to the end, a message will be printed. An example is:

```
ProjectName : Test_Case_Futu
.....
# Hydrologic Model total time: 361.753000021
seconds
End of Test_Case_Futu
```

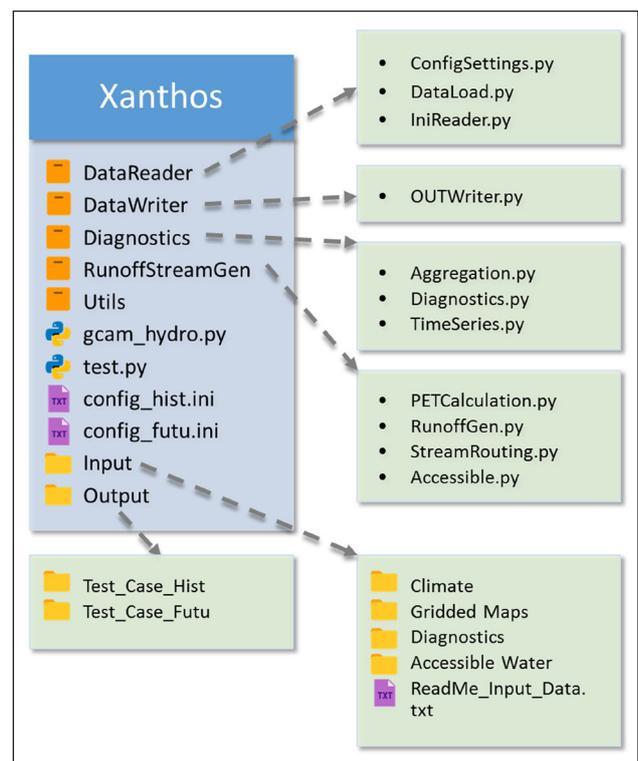


Figure 3: The architecture of Xanthos in application.

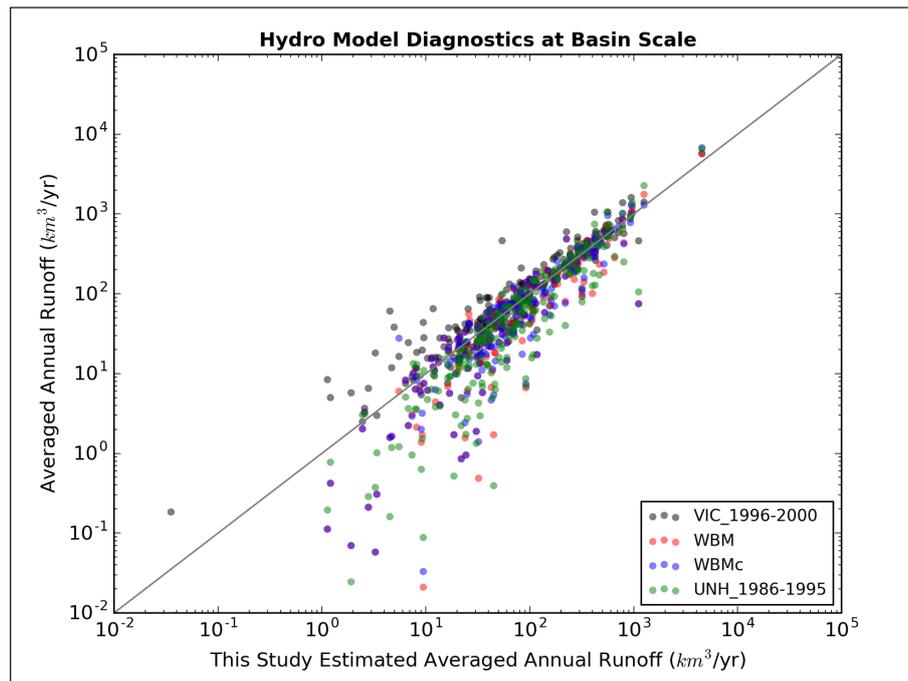


Figure 4: Example of diagnostics plot for comparison of averaged annual runoff with results from four other models at basin scale.

When the model does not run successfully, errors and warnings will be recorded in the log file. Same information will also be printed in the Python console.

The diagnostics module provides another option for the user to examine the quality of the model results of runoff. It compares the runoff results by basin scale, country scale or region scale to data sets of other models, such as the Variable Infiltration Capacity (VIC) model [9], Water Balance Model (WBM) and Water Balance Model Composite (WBMc) [10], UNH/GRDC [11–12]. **Figure 4** is a scatter plot that shows an example of the plots created by the diagnostics module. If the points are below the “x = y” line, the model overestimated the runoff compared to the four models (VIC, WBM, WBMc and UNH/GRDC), and if the points are above the “x = y” line then the model underestimated the model runoff. The gridded streamflow outputs are not regulated streamflow results. They are not suitable for direct comparison with other model results that are regulated streamflow (e.g. Dai’s results [13]), and are not included in diagnostics.

(2) Availability

Operating system

Xanthos has been tested successfully on Linux (64-bit), Windows 7 and Mac OS X Sierra.

Programming language

Python (2.7.11)

Additional system requirements

The core modules are designed to process large data sets. Thus, a minimum memory size of 8 GB is recommended to run the model and memory capacity also determines how fast the model can run according to the large size of the data sets.

Xanthos is a pure python code. The user is recommended to use other tools to manage the run of Xanthos.

The modules were tested with Anaconda 2.3.0⁸ (64-bit) on Linux.

The modules were tested using an open-source desktop Python IDE (PyDev⁹ for Eclipse¹⁰) on Windows.

Dependencies

All the listed dependencies can be obtained by using the Python’s tool for installing packages called pip (<https://pypi.python.org/pypi/pip>).

- NumPy (version 1.11)
- Scipy (version 0.18)
- Matplotlib (version 2.0)
- Pandas (version 0.19)
- configobj (version 5.0.6)

Software location

Archive

Name: GitHub

Persistent identifier: <https://github.com/JGCRI/xanthos/releases/tag/v1.0.0>

Licence: BSD 2-Clause

Publisher: Chris R Vernon

Version published: v1.0.0

Date published: 02/08/2017

Code repository

Name: GitHub

Identifier: <https://github.com/JGCRI/xanthos>

Licence: BSD 2-Clause (<https://github.com/JGCRI/xanthos/blob/master/LICENSE>)

Date published: 30/05/2017

Language

English.

Contact

For questions, technical supporting and user contribution, please contact:

Li, Xinya Xinya.Li@pnnl.gov
 Vernon, Chris R Chris.Vernon@pnnl.gov
 Link, Robert P Robert.Link@pnnl.gov
 Hejazi, Mohamad I Mohamad.Hejazi@pnnl.gov

Installation

The “InstallationRequirements” file on the repository is to help the user set up the Python environment for a proper run. It explains the steps required for a user to download and install the software with all its dependencies. Execution of the python code, “test_environment”, will be helpful for the user to testify if all the required packages are installed.

(3) Reuse potential

The Python language and the dependent library packages used are all open-source. Xanthos is highly modularized and most of the modules can be used independently by the user. The independence of the modules ensures the future development and feasibility of user contribution. For example, the current PET calculation is based on Hargreaves method. “PETCalculation” as a module can be imported into other programs to calculate monthly PET or can be replaced by the user using other methods to calculate monthly PET. Another example is the module package “DataReader”, that can be utilized by the user independently to import the included input data file for other applications.

Modification of a certain step could be restricted to the corresponding module. Extension of the model is achievable by adding a new module to an existing sub-folder or a new sub-folder. Documentation is organized through intensive comments inside the python code and the example configuration file. Execution will also produce a detailed log file lists model settings, the processing steps, CPU cost and warnings if applicable. The users can get support by contacting the authors when issues/bus are found. The users may also contact the authors for contributions to the code base. The contact information for support is included in the “README” file of the GitHub repository. To help the user about the installation and requirements of Xanthos, a file named “InstallationRequirements.pdf” is provided as part of the repository.

Xanthos is built as a part of an integrated modelling software for global water demand, supply, and scarcity, which the authors’ team is continuing to develop.

Notes

- ¹ Python <https://www.python.org/>.
- ² NumPy <http://www.numpy.org/>.
- ³ SciPy <https://www.scipy.org/>.
- ⁴ Matplotlib <https://matplotlib.org/>.
- ⁵ Pandas <http://pandas.pydata.org/>.

- ⁶ NetCDF classic http://www.unidata.ucar.edu/software/netcdf/docs/netcdf_introduction.html.
- ⁷ MATLAB <https://www.mathworks.com/>.
- ⁸ Anaconda <https://www.continuum.io/Anaconda-Overview>.
- ⁹ PyDev <http://www.pydev.org/>.
- ¹⁰ Eclipse <https://www.eclipse.org/ide/>.
- ¹¹ INI https://en.wikipedia.org/wiki/INI_file#Format.

Acknowledgements

This research was supported by the Office of Science of the U.S. Department of Energy through the Integrated Assessment Research Program.

Competing Interests

The authors have no competing interests to declare.

References

1. **Edmonds, J and Reilly, J M** 1985 Global Energy: Assessing the Future, 317. Oxford University Press, New York.
2. **Edmonds, J, Wise, M, Pitcher, H, Richels, R, Wigley, T and Maccracken, C** 1997 An integrated assessment of climate change and the accelerated introduction of advanced energy technologies-an application of MiniCAM 1.0. *Mitigation and adaptation strategies for global change*, 1(4): 311–339. DOI: <https://doi.org/10.1023/B:MITI.0000027386.34214.60>
3. **Kim, S H, Edmonds, J, Lurz, J, Smith, S J and Wise, M** 2006 The objECTS framework for integrated assessment: Hybrid modeling of transportation. *The Energy Journal*, Special Issue #2: 51–80. DOI: <https://doi.org/10.5547/ISSN0195-6574-EJ-VolSI2006-NoSI2-4>
4. **Hejazi, M I, Edmonds, J, Clarke, L, Kyle, P, Davies, E, Chaturvedi, V, Wise, M, Patel, P, Eom, J and Calvin, K** 2014 Integrated assessment of global water scarcity over the 21st century under multiple climate change mitigation policies. *Hydrology and Earth System Sciences*, 18: 2859–2883. DOI: <https://doi.org/10.5194/hess-18-2859-2014>
5. **Hejazi, M I, Edmonds, J, Chaturvedi, V, Davies, E and Eom, J** 2013 Scenarios of global municipal water-use demand projections over the 21st century. *Hydrological Sciences Journal*, 58(3): 519–538. DOI: <https://doi.org/10.1080/02626667.2013.772301>
6. **Hargreaves, G L, Hargreaves, G H and Riley, J P** 1985 Irrigation water requirements for Senegal River basin. *Journal of Irrigation and Drainage Engineering*, 111(3): 265–275. DOI: [https://doi.org/10.1061/\(ASCE\)0733-9437\(1985\)111:3\(265\)](https://doi.org/10.1061/(ASCE)0733-9437(1985)111:3(265))
7. **Zhou, Y, Hejazi, M, Smith, S, Edmonds, J, Li, H, Clarke, L, Calvin, K and Thomson, A** 2015 A comprehensive view of global potential for hydro-generated electricity. *Energy & Environmental Science*, 8: 2622–2633. DOI: <https://doi.org/10.1039/C5EE00888C>
8. **Kim, S H, Hejazi, M, Liu, L, Calvin, K, Clarke, L, Edmonds, J, Kyle, P, Patel, P, Wise, M and Davies, E** 2016 Balancing global water availability and use at basin scale in an integrated assessment model.

- Climatic Change*, 136(2): 217–231. DOI: <https://doi.org/10.1007/s10584-016-1604-6>
9. **Leng, G, Tang, Q** and **Rayburg, S** 2015 Climate change impacts on meteorological, agricultural and hydrological droughts in China. *Global and Planetary Change*, 126: 23–34. DOI: <https://doi.org/10.1016/j.gloplacha.2015.01.003>
 10. **Fekete, B M, Vörösmarty, C J** and **Grabs, W** 1999 Global, composite runoff fields based on observed river discharge and simulated water balances. *Volume 22 of GRDC-Report*. Global Runoff Data Centre, Federal Institute of Hydrology, Koblenz, Germany.
 11. **Fekete, B M, Vörösmarty, C J** and **Grabs, W** 2002 High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochemical Cycles*, 16(3): 15-1–15-10. DOI: <https://doi.org/10.1029/1999GB001254>
 12. **Fekete, B M** and **Vörösmarty, C J** 2011 ISLSCP II UNH/GRDC Composite Monthly Runoff. In: Hall, F G, Collatz, G, Meeson, B, Los, S, Brown de Colstoun, E and Landis, D (eds.), *ISLSCP Initiative II Collection*. Data set. Available on-line <http://daac.ornl.gov/> from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. DOI: <https://doi.org/10.3334/ORNLDAAC/994>
 13. **Dai, A, Qian, T, Trenberth, KE** and **Milliman, JD** 2009 Changes in continental freshwater discharge from 1948 to 2004. *Journal of Climate*, 22(10): 2773–2792. DOI: <https://doi.org/10.1175/2008JCLI2592.1>
 14. **Rew, R, Davis, G, Emmerson, S, Davies, H, Hartnett, E, Heimbigner, D** and **Fisher, W** (2016, November 21) NetCDF. Retrieved August 01, 2017, from: <http://www.unidata.ucar.edu/software/netcdf/docs/index.html>.

How to cite this article: Li, X, Vernon, C R, Hejazi, M I, Link, R P, Feng, L, Liu, Y and Rauchenstein, L T 2017 Web Tool for Navigating and Plotting SomaLogic ADAT Files. *Journal of Open Research Software* 5: 21, DOI: <https://doi.org/10.5334/jors.181>

Submitted: 02 June 2017 **Accepted:** 04 August 2017 **Published:** 11 September 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.