Observatory for Gridded Hydrometeorology (OGH) automated retrieval, preprocessing, and visualization for spatial-temporal analysis

Python in GeoScience Seminar
University of Washington eScience Institute
December 4, 2018
Presented by Christina Bandaragoda

Content for presentation from: Phuong, J., C. Bandaragoda, E. Istanbulluoglu, C. Beveridge, R. Strauch; L. Setiawan; S. D. Mooney. (2018) Automated retrieval, preprocessing, and visualization of gridded hydrometeorology data products for spatial-temporal exploratory analysis and intercomparison, Env. Modeling & Software, (submitted, in revision).

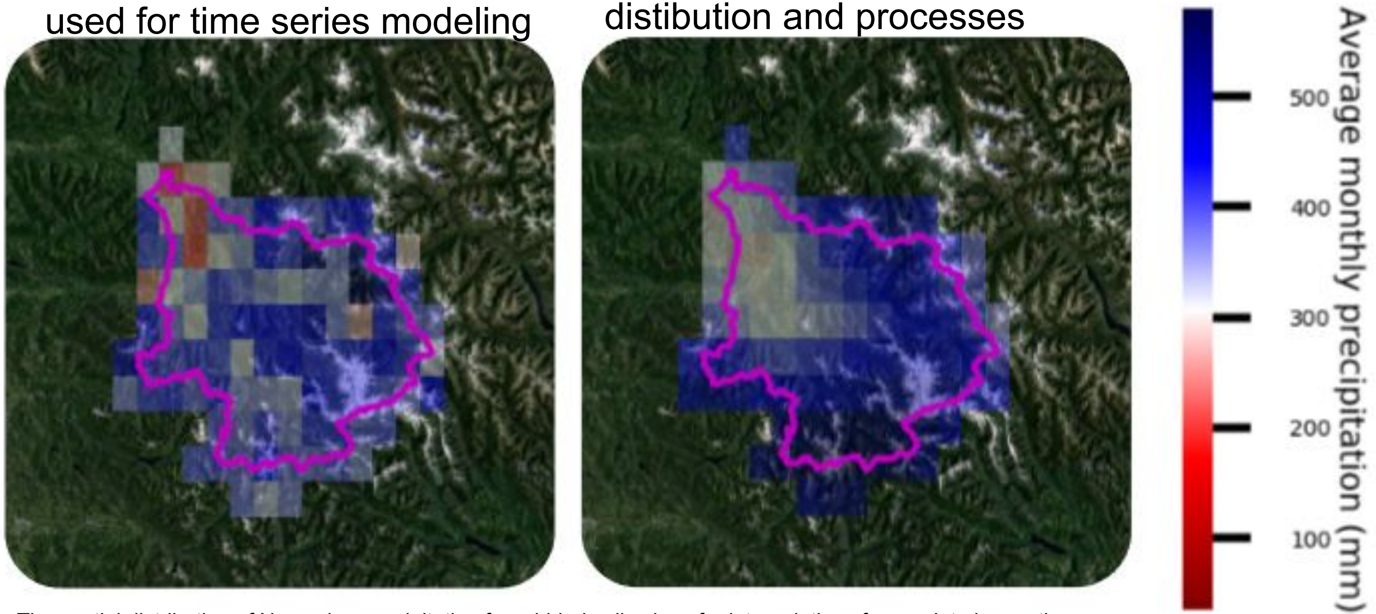
Acknowledgements

This work benefited from the contributions from **University of Washington (UW) Watershed Dynamics group** who helped test and develop OGH, and members **UW eScience Institute** that helped with developing and using the Python toolkits. This project was supported in part by National Science Foundation **HydroShare** Cyberinfrastructure project (ACI 1148453), SI2:SSI **Landlab** project (ACI-1450412), **UW Civil & Environmental Engineering** Department in collaboration with researchers and scientists of the **Sauk-Suiattle Indian Tribe** and the **Skagit Climate Consortium**. The project uses the HydroShare platform, which is supported by the **Consortium of Universities for the Advancement of Hydrologic Sciences, Inc. (CUAHSI)**, a research organization supported by NSF cooperative agreement (EAR 1338606). Special thanks to early reviewers: Drs. Dan Ames, Emilio Mayorga, and Nicoleta Cristea.

Gridded hydrometeorology

Interpolation from
COOP Station observation:
used for time series modeling

Atmospheric Model:
Used for modeling spatial
distibution and processes



The spatial distribution of November precipitation for gridded cell values for interpolations from point observations (Livneh *et al.* (2013)) and from an atmospheric model (Salathé *et al.* (2014) WRF model outputs).

Gridded climate data products are an alternative to instrumental measurements as inputs to earth surface models

BUT

you don't know which gridded data works best for your study site and research question.

Ledesma JLJ, Futter MN. Gridded climate data products are an alternative to instrumental measurements as inputs to rainfall - runoff models. Hydrological Processes 2017;31:3283 - 3293. https://doi.org/10.1002/Hyphttps://doi.org/10.1002/Hyp.1126

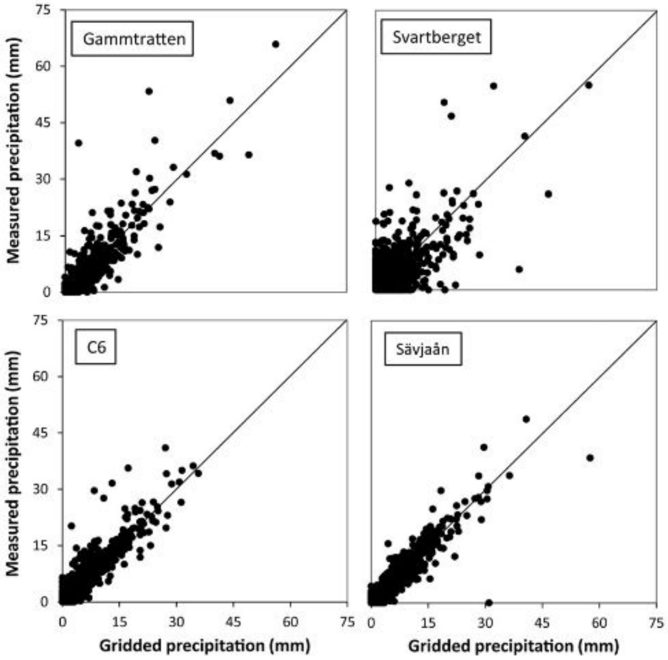


FIGURE 1 Regression plots of instrumental measured versus gridded daily precipitation for the study sites

Don't Use

Probably Use

"Readily available gridded climate products are an underutilized source of temperature and precipitation time series for rainfall – runoff modelling, which may overcome some of the performance issues associated with poor quality instrumental data in small headwater monitoring catchments."

One gridded produce (E-OBS)
Two rainfall – runoff models (PERSiST and HBV)
Six small Swedish catchments.

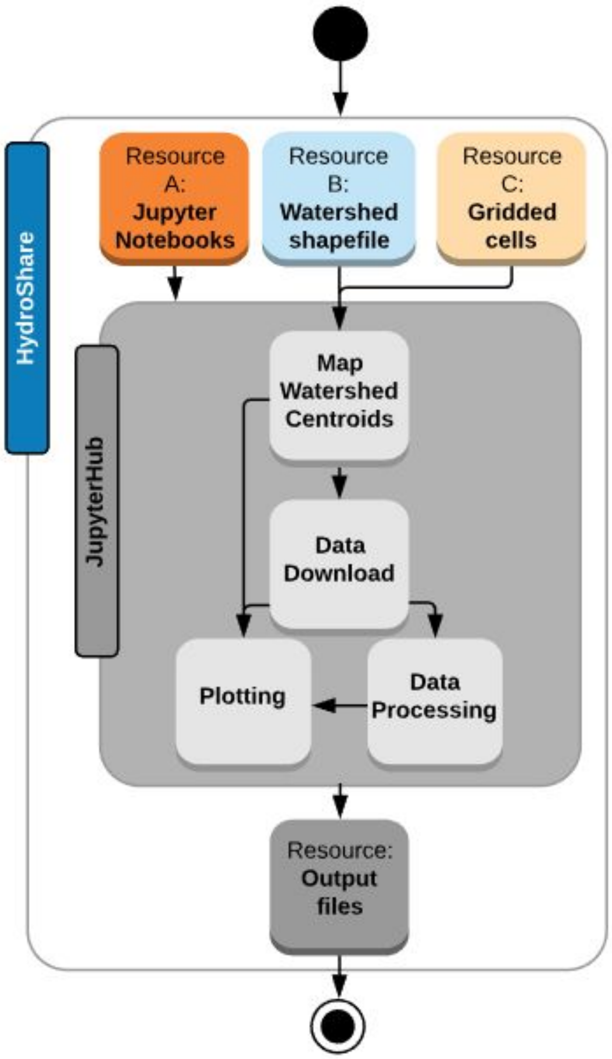
How do you access, download, process, and analyze all the new data for **usability with your research project**?

How do you access, download, process, and analyze all the new data for **usability with your research project**?

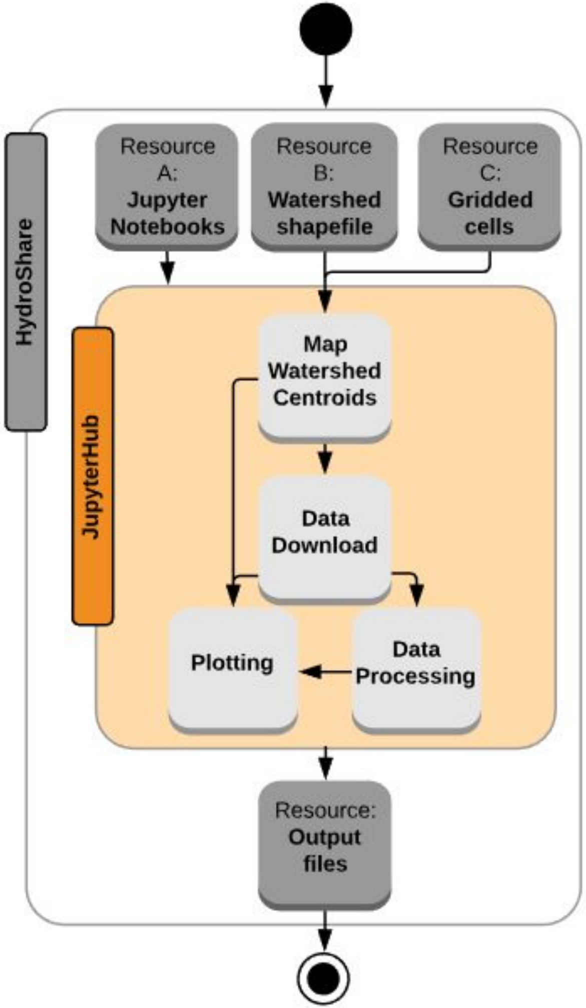
1. Switch your PhD topic to atmospheric science?
2. Spend 1-2 years of your PhD becoming friendly with climate data?
3. Use Observatory for Gridded Hydrometeorology (OGH)

How do you access, download, process, and analyze all the new data for **usability with your research project**?

Inputs


HydroShare
Resource A: Jupyter Notebooks
Resource B: Watershed shapefile
Resource C: Gridded cells
JupyterHub
Map Watershed Centroids
Data Download
Plotting
Data Processing
Resource: Output files

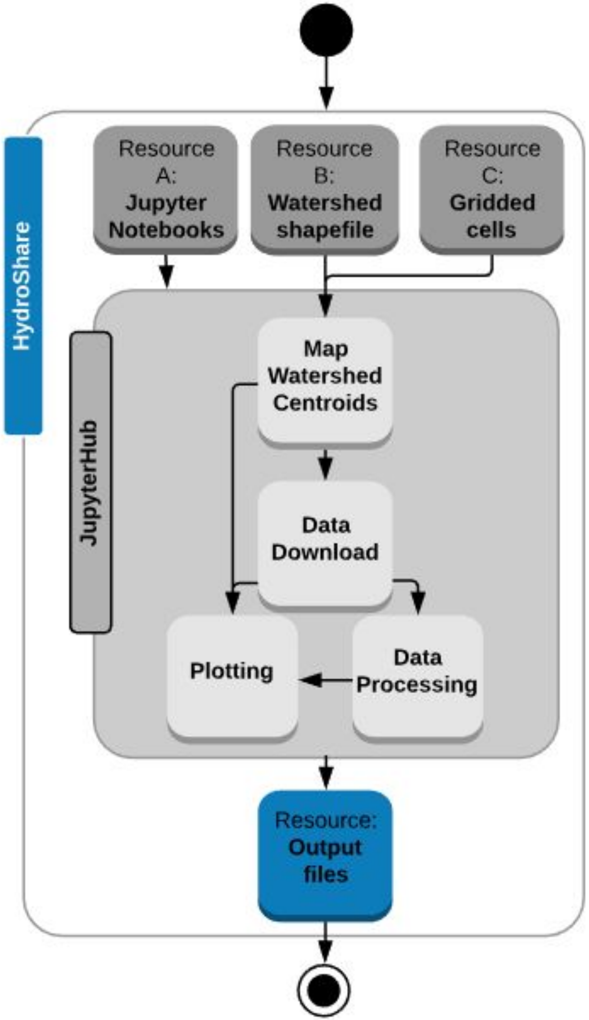
Work



HydroShare
JupyterHub
Resource A: Jupyter Notebooks
Resource B: Watershed shapefile
Resource C: Gridded cells
Map Watershed Centroids
Data Download
Plotting
Data Processing
Resource: Output files

UNIVERSITY of WASHINGTON

Outputs



Inputs

Table 1: Summary of seven daily, 1/16° gridded data product. Descriptions in this summary compare data products by variable, date range, analysis type, intended spatial coverage, and their source publication.

Data set	Features and variables (in order)	Start date	End date	Analysis type	Spatial coverage	Publication
Climate station meteorology						
dailymet_livneh2013	PRECIP, TMIN, TMAX, WINDSPEED	1915-01-01	2011-12-31	raw	CONUS	[Livneh et al., 2013]
dailymet_bclivneh2013	PRECIP, TMIN, TMAX, WINDSPEED	1915-01-01	2011-12-31	bias-corrected	Columbia river	[Livneh et al., 2013]
dailymet_livneh2015	PRECIP, TMIN, TMAX, WINDSPEED	1950-01-01	2013-12-31	raw	CONUS	[Livneh et al., 2015]
WRF-NNRP model meteorology						
dailywrf_salathe2014	PRECIP, TMIN, TMAX, WINDSPEED	1950-01-01	2010-12-31	raw	Columbia river	[Salathé et al., 2014]
dailywrf_bcsalathe2014	PRECIP, TMIN, TMAX, WINDSPEED	1950-01-01	2010-12-31	bias-corrected	Columbia river	[Salathé et al., 2014]
Variable Infiltration Capacity						
dailyvic_livneh2013	YEAR, MONTH, DAY, EVAP, RUNOFF, BASEFLOW, SMTOP, SMMID, SMBOT, SWE, WDEW, SENSIBLE, LATENT, GRNDFLUX, RNET, RADTEMP, PREC	1915-01-01	2011-12-31	Physics-based model	CONUS	[Livneh et al., 2013]
dailyvic_livneh2015	YEAR, MONTH, DAY, EVAP, RUNOFF, BASEFLOW, SMTOP, SMMID, SMBOT, SWE, WDEW, SENSIBLE, LATENT, GRNDFLUX, RNET, PETTALL, PETSHORT, PETNATVEG	1950-01-01	2013-12-31	Physics-based model	CONUS	[Livneh et al., 2015]

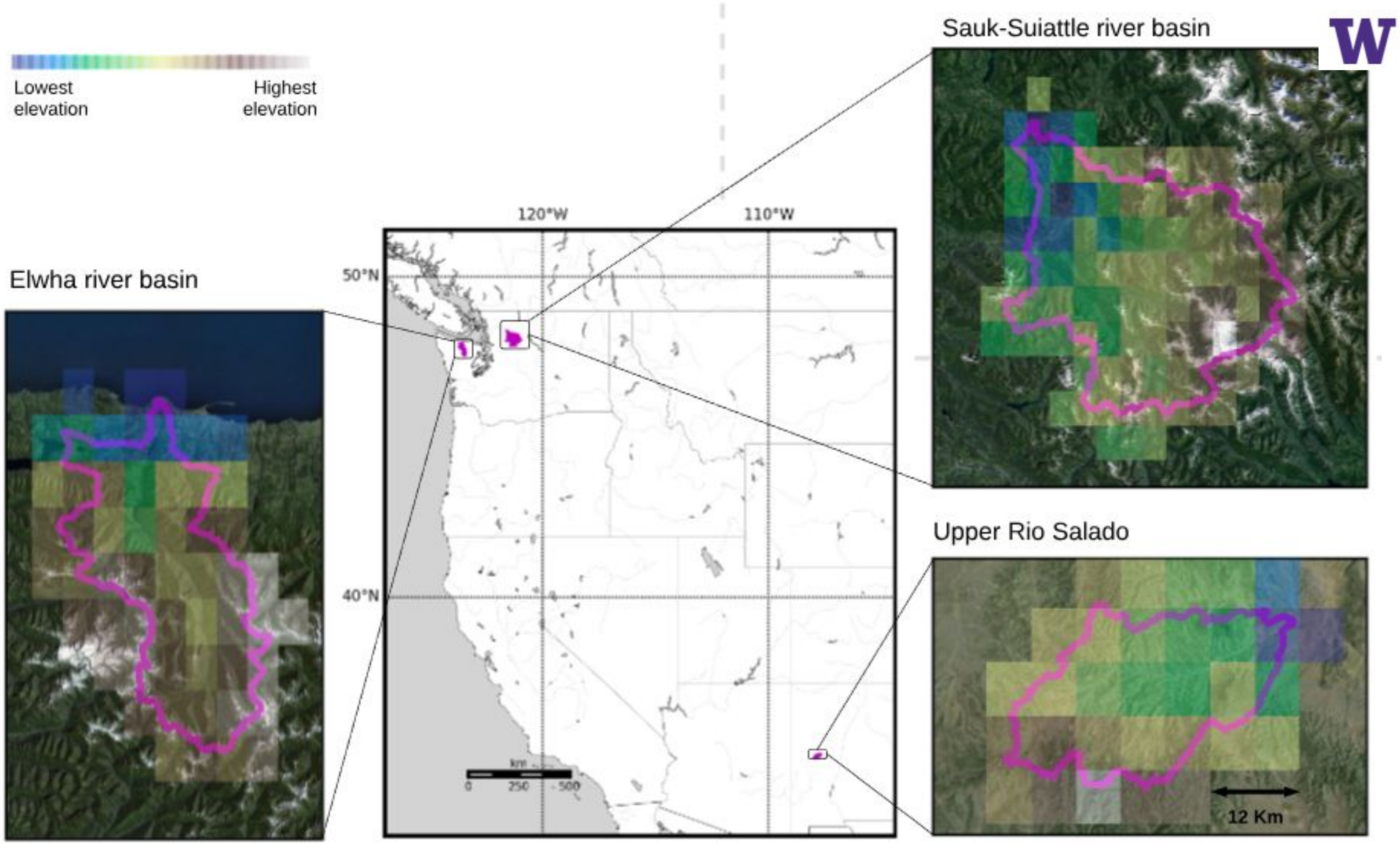


Figure 4. Aerial view of watersheds and gridded cells. The Sauk-Suiattle (164-2216 m), Elwha (36-1642 m), and Upper Rio Salado (1962-2669 m) watersheds located in western United States were visualized using the *multisiteVisual* function with an EPSG:3857 geospatial projection. Each watershed (outlined in magenta) and their gridded cells are visualized using the *griddedCellGradient* function at the 1/16^{o} spatial-resolution (~6 km). In the Sauk-Suiattle watershed, five reference markers denote the highest elevation gridded cell (gridded cell 69, elevation: 2216 m), the lowest elevation gridded cells (gridded cells 3 and 24, elevation: 164 m), the Darrington Ranger Station site (*; COOP station 451992, elevation: 167 m) for observed meteorology data, and the Sauk River Near Sauk, WA streamflow gauge (§, USGS-12189500, elevation: 81 m) for observed streamflow discharge measured at the downstream-most tip of the watershed.

Analysis

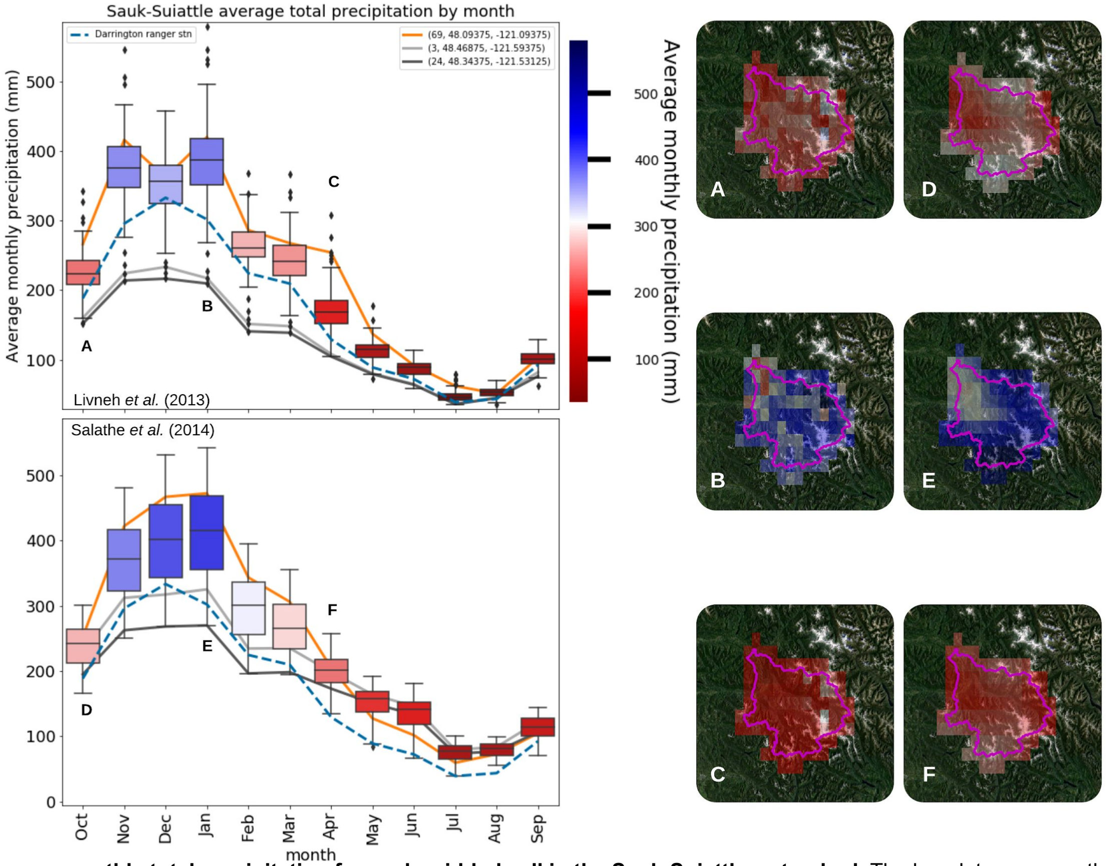


Figure 5: Comparison of the average monthly total precipitation for each gridded cell in the Sauk-Suiattle watershed. The boxplots compare the statistical distribution of the average monthly total precipitation (inches) to the spatial distribution of precipitation in each gridded cell (A-F), using data from Jan 1 1950 through Dec 31 2010. Created using the *renderValueInBoxplot* function, the boxplot colors represent the median value of the gridded cell distributions. Reference trend lines were included to illustrate Sauk-Suiattle's highest elevation gridded cell (#69; orange) and the lowest elevation gridded cells (#3 and #24; light and dark gray), found using the *findCentroidCode* function. The gridded cell distributions are rendered spatially with a basemap using the *renderValueInPoints* function. The spatial distribution of gridded cell values are rendered using the *renderValueInPoints* function for Livneh *et al.* (2013) interpolated meteorology for A) October, B) January, and C) April, compared with to Salathé *et al.* (2014) WRF model outputs for D) October, E) January, and F) April. All spatial maps and boxplots use the same colorbar legend and numerical distribution shown in the top-left.

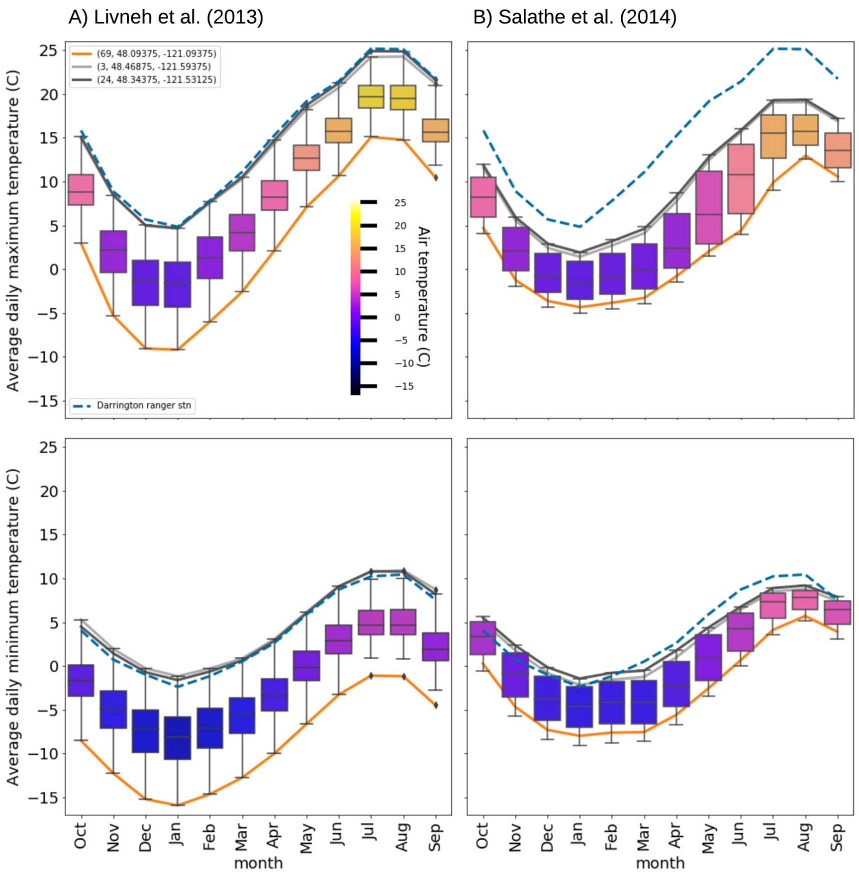


Figure 6: Comparison of monthly mean of daily minimum and maximum temperature. The monthly mean of daily maximum (top) and minimum (bottom) temperatures (in Celsius) were computed for each of the 99 Sauk-Suiattle gridded cells. The boxplots represent the observations from Livneh *et al.,* (2013) meteorology (left) and Salathé *et al.,* (2014) WRF model outputs (right). Reference trend lines were included to represent the highest elevation gridded cell (orange) and the lowest elevation gridded cells (light and dark gray) in Sauk-Suiattle. The field observations (blue dashed line) measured at Darrington Ranger Station (elevation: 167 m) indicates that maximum daily temperature (top) are more closely represented by Livneh *et al.,* (2013) in the Sauk-Suiattle watershed, while there are no remarkable differences observable for minimum daily temperature (bottom).

Metadata

Table 2: Minimum annotation criteria for gridded data products.

Metadata	Metadata descriptions
File location	
1. Dataset	name of the gridded data product
2. Spatial resolution	the distance between gridded cell centroids
3. Web protocol	the data transfer protocol
4. Domain	the web domain
5. Subdomain	the subdomain path
6. Decision steps	the file organization for locating data files
7. Filename structure	the standard components to the filename
8. File format	the file type at download
File structure	
9. Start date	the start date of the time-series
10. End date	the end date of the time-series
11. Temporal resolution	the unit increment for time-steps
12. Delimiter	the column separator within each line of data
13. Variable_list	the list of variables in order of appearance
14. Reference	the sources of metadata
Variable structure	
15. Variable_info	
• desc	the long name of the variable
• dtypes	the expected data type
• units	the unit increment of the data

Results

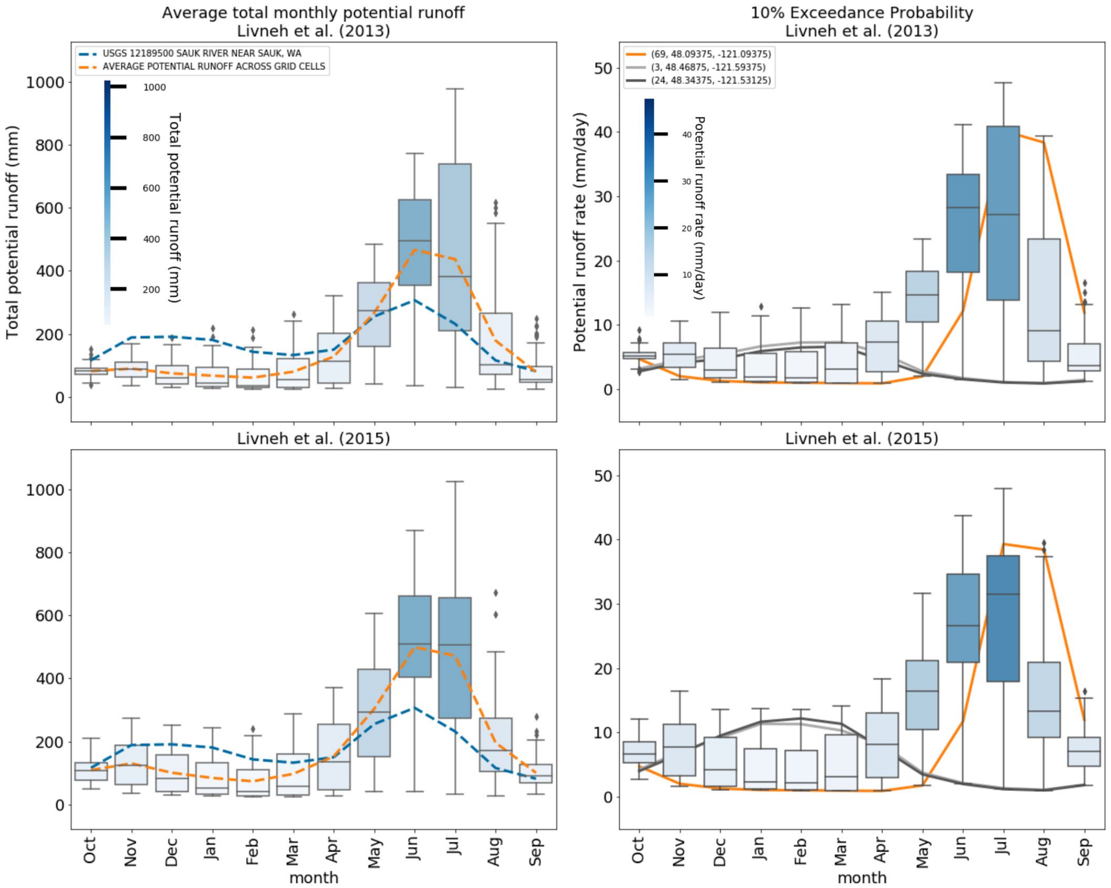
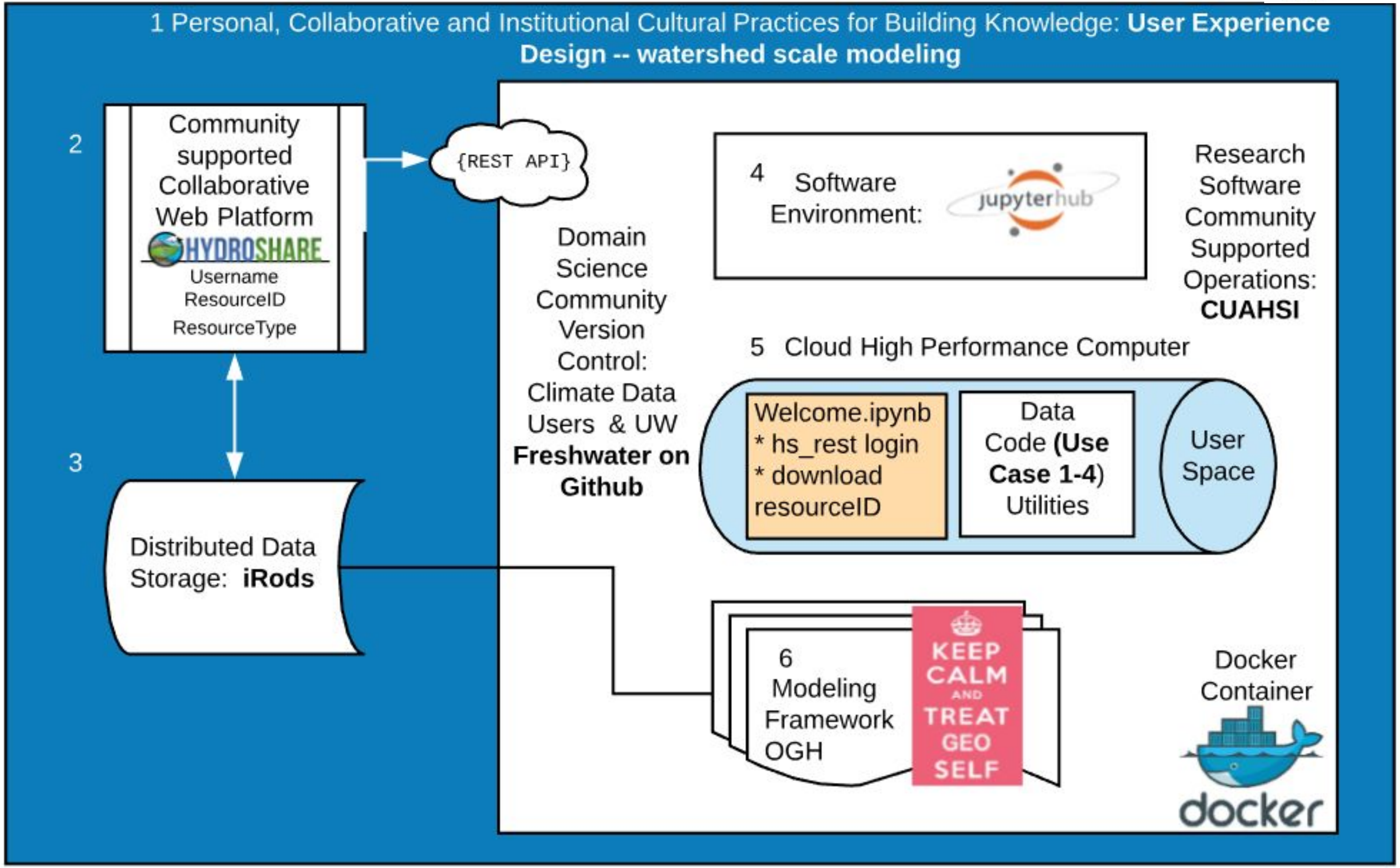


Figure 7: Average total monthly potential runoff (mm) and 10% exceedance probability for each monthly unrouted potential runoff (mm/day) within the Sauk-Suiattle watershed. The boxplots are comprised of 99 gridded cell values for each month. Peak of average total monthly potential runoff (left) occurs in November and June months shown by the observed USGS streamflow discharge (blue dash), and observable by the spatial average of the gridded cell (orange dash). The 10% exceedance probability for each gridded cell (right) is a function of the spatial average of peak flow occurs in November and July. The snowmelt season is the major period for expected runoff for highest elevation gridded cell (orange line), while the rainfall season is the major period contributing to runoff for lowest elevation gridded cells (light and dark gray lines).

Cyberinfrastructure



Adapted from Bandaragoda et al., 2018 (submitted)

Interactive Modules: Use Case 1-4

Go to HydroShare, Discover, Search: [treatgeoself]

Observatory for Gridded Hydrometeorology: TreatGeoSelf to gridded data

Open with...

Authors: Jimmy Phuong · Christina Bandaragoda · Claire Beveridge · Ronda Strauch · Landung Setiawan · Erkan Istanbulluoglu
Owners: Jimmy Phuong · Christina Bandaragoda
Resource type: Composite Resource
Created: Sep 06, 2017 at 8:29 p.m.
Last updated: Dec 04, 2018 at 9:09 p.m. by Christina Bandaragoda

Abstract

This iPython notebook demonstrates the workflow for obtaining and processing gridded meteorology data files with the Observatory for Gridded Hydrometeorology Python library.

Using the Sauk-Suiattle, Elwha, and Upper Rio Salado watersheds as the study sites of interest, each Jupyter notebook will guide the user through assembling the datasets and analyses from each of seven gridded data product.

In Usecase 1, users may inspect their study site of interest given in the form of summary spatial visualizations. The treatgeoself() function will yield a mapping file per study site, which reduces the gridded cell centroids to the subset that intersects with the study area (i.e., within the watershed). Within treatgeoself(), the user may determine the amount of buffer space to include outside of the study site (default is 0.06-degree buffer region).

In Usecase 2, each of the mapping files are used to guide data retrieval from each of the gridded data products. A series of _get_ functions then downloads the files to designated subfolders. The resulting file paths are cataloged into the mapping file, which can be summarized for data availability according to the elevation gradient using the mappingfileSummary() function. These downloaded files are compressed into tar.gz files, then migrated with their respective mapping files as content files within a new HydroShare resource, for ease of collaborative use.

In Usecase 3, the downloaded files from Usecase 2 are processed in to spatial and temporal summary statistics. The gridclim_dict() function compiles and computes daily, monthly, annual, and monthly-yearly average values for each variable described in the gridded data product metadata (e.g., the ogh_meta class dictionary). Monthly averages are then visualized as time-series plots, while spatial averages are visualized as spatial heatmaps. Finally, the dictionary of dataframes (the product of the spatial-temporal analyses) is saved into a json file and migrated out as a content file within a new HydroShare resource.

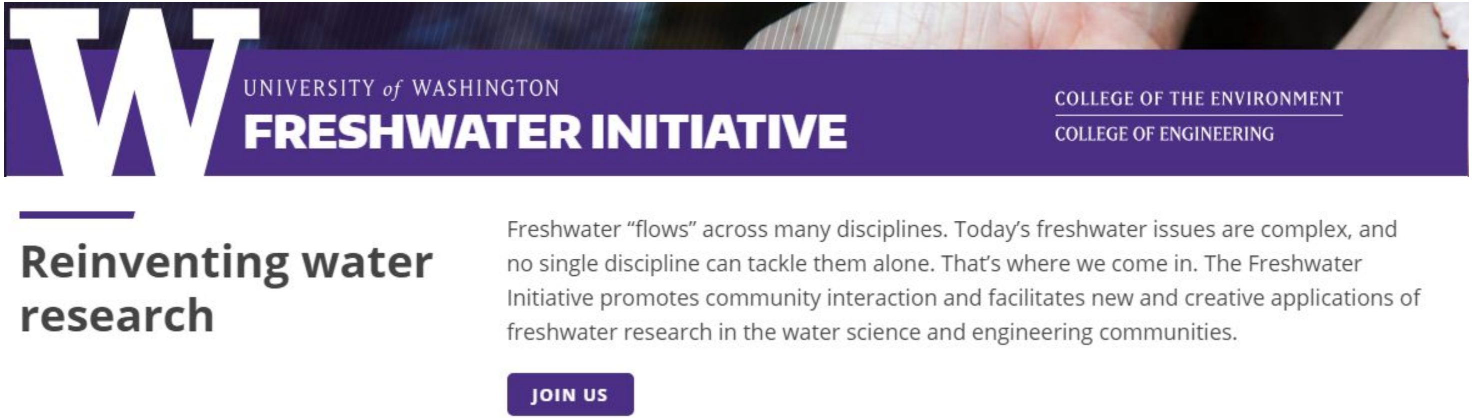
Access and Next Development Steps

The tutorial notebooks are also hosted from the University of Washington Freshwater Initiative Observatory github.
https://github.com/freshwater-Initiative/Observatory
Currently, pip and conda installation will get you OGH v.0.1.11.

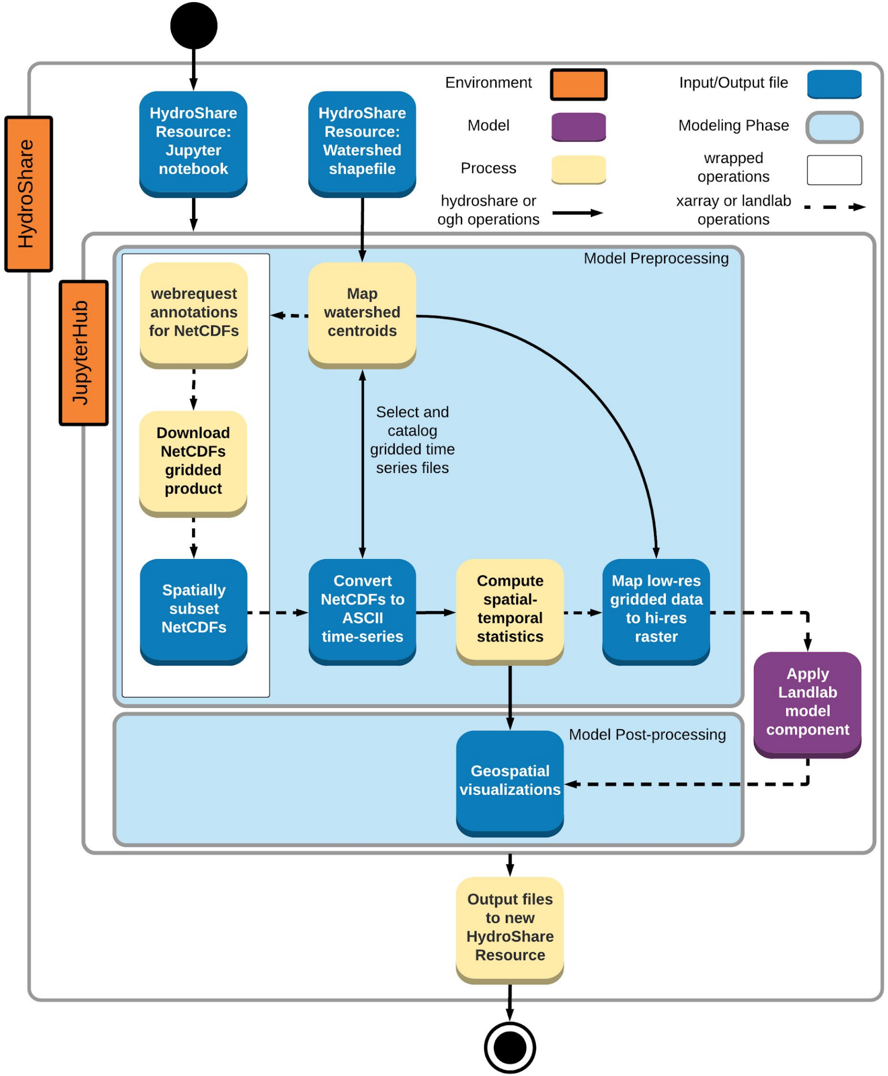
ogh_xarray_landlab (OXL) work with Jeff Keck will be part of v.0.1.12 (or V1).

ogh_geohydro_dhsvm (OGD) work with Zhuoran Duan will be part of v.0.1.13 (or V2).

Contribute to this open source project!



use-case 8


Environment
Input/Output file
Model
Modeling Phase
Process
wrapped operations
hydroshare or ogh operations
xarray or landlab operations
HydroShare
JupyterHub
HydroShare Resource: Jupyter notebook
HydroShare Resource: Watershed shapefile
Model Preprocessing
webrequest annotations for NetCDFs
Map watershed centroids
Download NetCDFs gridded product
Select and catalog gridded time series files
Spatially subset NetCDFs
Convert NetCDFs to ASCII time-series
Compute spatial-temporal statistics
Map low-res gridded data to hi-res raster
Apply Landlab model component
Model Post-processing
Geospatial visualizations
Output files to new HydroShare Resource

landlab integration