

Better understanding of hydrologic process through data-driven learning facilitated by collaborative open web-based platforms

Belize Lane¹, Irene Garousi-Nejad¹, Melissa Gallagher², Dave Tarboton¹, and Emad Habib³

¹Utah State University

²University of Houston System

³University of Louisiana at Lafayette

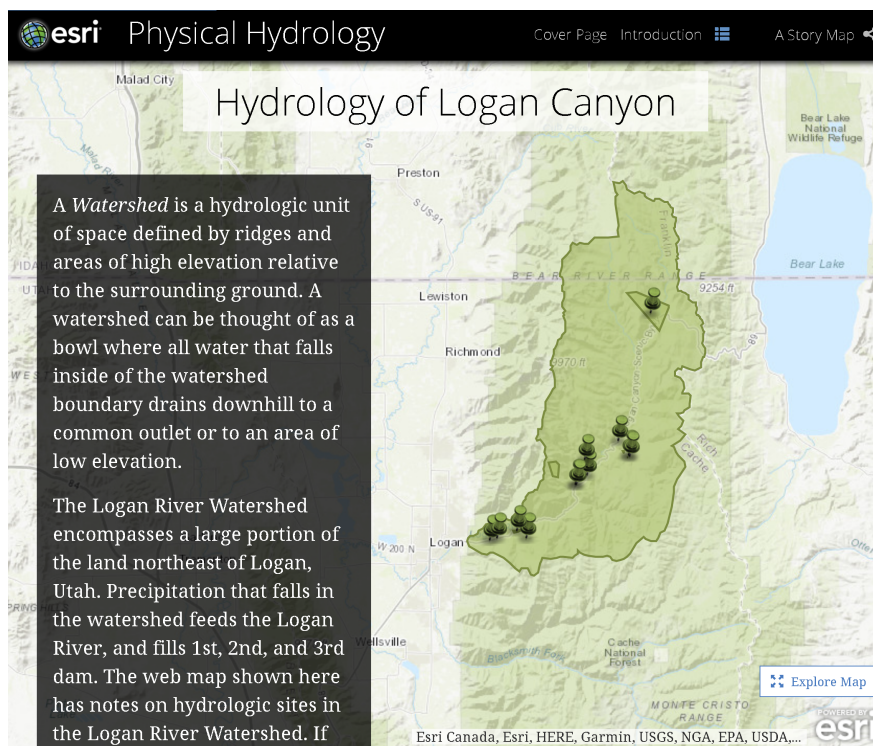
March 23, 2021

Abstract

The era of "big data" promises to provide new hydrologic insights, and open web-based platforms are being developed and adopted by the hydrologic science community to harness these datasets and data services. This shift accompanies advances in hydrology education and the growth of web-based hydrology learning modules, but their capacity to utilize emerging open platforms and data services to enhance student learning through data-driven activities remains largely untapped. Given that generic equations may not easily translate into local or regional solutions, teaching students to explore how well models or equations work in particular settings or to answer specific problems using real data is essential. This paper introduces an open web-based learning module developed to advance data-driven hydrologic process learning, targeting upper level undergraduate and early graduate students in hydrology and engineering. The module was developed and deployed on the HydroLearn open educational platform, which provides a formal pedagogical structure for developing effective problem-based learning activities. We found that data-driven learning activities utilizing collaborative open web platforms like HydroShare and CUAHSI Jupyter-Hub computational notebooks allowed students to access and work with datasets for systems of personal interest and promoted critical evaluation of results and assumptions. Initial student feedback was generally positive, but also highlights challenges including trouble-shooting and future-proofing difficulties and some resistance to open-source software and programming. Opportunities to further enhance hydrology learning include better articulating the myriad benefits of open web platforms upfront, incorporating additional user-support tools, and focusing methods and questions on implementing and adapting notebooks to explore fundamental processes rather than tools and syntax. The profound shift in the field of hydrology toward big data, open data services and reproducible research practices requires hydrology instructors to rethink traditional content delivery and focus instruction on harnessing these datasets and practices in the preparation of future hydrologists and engineers.

Hosted file

HydroLearn manuscript_2021.pdf available at <https://authorea.com/users/387348/articles/514898-better-understanding-of-hydrologic-process-through-data-driven-learning-facilitated-by-collaborative-open-web-based-platforms>



For this exercise, we will use the HydroShare Jupyter Notebook environment. The notebook is created within [this](#) resource, which is accessed using your HydroShare login information. You will need to create a HydroShare account at [HydroShare](#) if you haven't already. You can scroll down and read the abstract. To use Jupyter Notebook for this assignment, you need to click on the **Open** with on the top right of the resource. Select **CUAHSI JupyterHub** and for Server Options select "**R [version] Scientific**."

Note that you will need to become a member of the [CUAHSI JupyterHub group](#) to use this free service. This account must be approved, so setting this up *ahead of time* is recommended.

It is recommended that you open the Notebook from this page in a new window and then continue with HydroLearn for guidance on how to work through the notebook. That is best done with the HydroLearn instructions and the Jupyter Notebook open side-by-side.

http://www.hydroshare.org/resource/bd0t is available that replaces this version.' Below this, the authors (Irene Garousi-Nejad, Belize Lane), owners (Belize Lane), resource type (Composite Resource), storage (45.5 KB), creation date (Jul 17, 2020 at 7:05 p.m.), last updated date (Dec 02, 2020 at 12:44 a.m.), and DOI (10.4211/hs.873b5490f986428bbd23e37c81906402) are listed. On the right side, there is a 'Open with...' dropdown menu showing options: Jupyter (CUAHSI JupyterHub), OpenDAP, CyberGIS-Jupyter for Water, and MATLAB Online."/>

jupyterhub
Physical_Hydrology_Stats_STU...
Last Checkpoint: 11/28/2020 (autosaved)
Logout
Control Panel

File
Edit
View
Insert
Cell
Kernel
Help

Run
Markdown
Memory: 154 / 4096 MB

Problem 4. Daily Streamflow

Perform your calculations below.

To download the streamflow data for an observation gage, we need to use `dataRetrieval` library in R. This library is a collection of functions to help retrieve U.S. Geological Survey (USGS) and U.S. Environmental Protection Agency (EPA) water quality and hydrology data from web services.

```
In [ ]: # First you need to install the package.
install.packages('dataRetrieval')
```

```
In [ ]: # Now, load the package
library(dataRetrieval)
```

The `dataRetrieval` library has two functions that let you retrieve the daily values and instantaneous values. Those functions are introduced below. For this exercise, we will use the function that retrieves daily values. This function takes site number, what variable to download, what type of data, and the duration. The parameters are defined in below. You can find this information for your selected gage on <https://waterwatch.usgs.gov>.

`readNWISdv`: daily value USGS NWIS data retrieval
`readNWISsv`: Instantaneous value data from USGS (NWIS)

```
In [ ]: # Select a USGS gage ID. Here, I select the Blacksmith Fork USGS Gaging Station.
SiteNumber = "10113500" # Site ID
ParamCode = "00060" # The code for the discharge
StatCode = "00003" # The code for the average values
Start_Date = "1972-10-01" # Start date, format: YYYY-MM-DD
End_Date = "2015-09-30" # End date, format: YYYY-MM-DD
```

```
In [ ]: Daily_Q = readNWISdv(SiteNumber, ParamCode, Start_Date, End_Date, StatCode)
```

Let's take a look at the data we have retrieved from the USGS using `readNWISdv` function of `dataRetrieval` library in R. If you

Student Response to Question 4:

The plot shows daily mean discharge observed at the Blacksmith Fork USGS gage over 43 years (1972-2015). The minimum and maximum values are 26 and 1530 ft³/s. The data was retrieved using `dataRetrieval` in R.

HYDROLEARN
jupyterhub

2. Basic grid analysis using TauDEM functions for watershed delineation

The first step in computing the topographic wetness index for Spawn Creek Watershed is to **delineate the Spawn Creek Watershed** from National Elevation Dataset (NED) 30 arc-second dataset (10 m of the Logan River watershed). Here, we use the TauDEM basic grid analysis functions (i.e., `Pitremove` to fill sinks, `D8 Flow Direction` to compute the downslope flow directions, `D8 Flow Accumulation` to calculate the contributing area at each grid cell, and `gagedwatershed` function that takes the Spawn outlet and delineates the watershed) to delineate the Spawn Creek watershed. You have learned how to work with these functions in homework 3.

The following shows the command-line functions that need to be used. You can find more details on the inputs and outputs to each function in the Python Jupyter Notebook on HydroLearn resource.

```
# pitremove -o logan.tif -d logan.tif
# d8flowdir -f logan.tif -p logan.tif -d8 logan.tif
# areaout -p logan.tif -o spawnout.tif -d8 spawn.tif
# gagedwatershed -o spawnout.tif -p logan.tif -p spawn.tif -d8 spawn.tif
```

After using `gdal_polygonize.py` script from the GDAL library, the shapefile of the Spawn Creek watershed is created. Follow the instruction and codes in the Jupyter Notebook to visualize the Spawn Creek Watershed shapefile as shown in Figure 3.

2.2 Delineating the Spawn Creek Watershed

Having the D8 contributing area result, we can create the Spawn Creek watershed using the TauDEM `gagedwatershed` function. We need the Spawn Creek watershed to mask the pit removed DEM file and only focus on Spawn Creek area, not the entire Logan River watershed. This function takes `p.tif` and `SpawnOutlet.shp` as the inputs and calculates the gaged watershed grid, i.e. `spawn_watershed.tif`.

```
gagedwatershed -o SpawnOutlet.shp -p logan.tif -p spawn_watershed.tif
```

VISUALIZATION: See the zoomed-in Spawn Creek Watershed.

```
sp = MultiPolygon(shapefile['geometry']) for pol in fiona.open("../spawn_watershed.shp"):
    cx = pol.get('centroid')
    num_colours = len(sp)

fig = plt.figure(figsize=(10, 10))
ax = fig.add_subplot(111)
xmin, xmin2, xmax, xmax2 = sp.bbox
ax.set_xlim(xmin - 0.2 * w, xmax + 0.2 * w) # use for the zoom in
ax.set_ylim(ymin - 0.2 * h, ymax + 0.2 * h) # use for the zoom in
ax.set_aspect(1)
```

Expected results shown in HydroLearn are generated in JupyterHub computational notebook

Clarify learning objectives and requirements

What State University CEE6400
Physical Hydrology

Discover New

Help

balance

Learning Objectives

At the end of this section, the following is expected:

1. Given a digital elevation model (DEM), the student will be able to **compute** the topographic wetness index and **describe** its role and use in TOPMODEL runoff calculations.
2. Using CUAHSI JupyterHub and TauDEM and given necessary catchment and storm characteristics, the student will be able to **apply** TOPMODEL principles and equations to **calculate** runoff.
3. Given information about a catchment, the student will be able to **critically assess** TOPMODEL assumptions and determine if and why TOPMODEL is an appropriate tool in that setting.

Introduction

The focus of this assignment is the simulation of runoff generation using a semi-distributed conceptual hydrologic model called TOPMODEL. In this activity, the student will learn how to (1) calculate the topographic wetness index using digital elevation models (DEMs) following up on a previous module on DEMs and GIS in Hydrology (2) apply TOPMODEL concepts and equations to estimate soil moisture deficit and runoff generation across a watershed given necessary watershed and storm characteristics; and (3) critically assess concepts and assumptions to determine if and why TOPMODEL is an appropriate tool given information about a specific watershed.

Requirements

Before completing this exercise, please read the following sources

- <https://tinyurl.com/rainfall-cunoff> (Rainfall Runoff Processes StoryMap, Chapter 6)

Provide content, including images & equations

II. Topographic Wetness Index

Given the tendency of water to flow downhill, **specific catchment area** serves as a useful surrogate for the **subsurface lateral moisture flux** comprised of interflow and groundwater flow at a location (Figure 3).

P, precipitation
SOF, Saturation overland flow
SSSF, Subsurface Storm flow, $q = ra$
S, slope
a, contributing area
 a_{sat} , saturated area (variable source area)
r, recharge

Figure 3. Definition of the upslope area draining through a point within a catchment. Streamflow is separated into surface runoff (SOF) generated by surface water input on saturated contributing areas and subsurface downhill flow (SSSF) comprising baseflow and return flow.

Check Your Understanding (CYU) questions

Runoff calculation based on TOPMODEL is widely used. However, the ... excess surface runoff assumption employed by this model was originally applied and verified in ... areas; thus, this assumption may not be met in ... areas where overland flow dominates due to ... excess.

☐ infiltration, dry, humid, saturation

☒ saturation, humid, dry, infiltration ✓

☐ infiltration, humid, dry, saturation

☐ saturation, dry, humid, infiltration

Learning activity grading rubric

No Evidence	Level #1 (Does Not Meet Expectations)	Level #2 (Meets Expectations)
Q1. Absent, no evidence, not applicable, or not addressed 0 point	Some parts are not presented. Units are not presented or not correct. 1-5 points	D8 contributing area in both the number of cells and in square kilometers are reported. The area of a single grid cell is reported. Units are correctly mentioned. 6 points
Q2. Absent, no evidence, not applicable, or not addressed 0 point	Some parts are not presented. Units are not correct. No discussion on the differences. 1-6 points	D-infinity contributing area is reported with the correct unit. The difference between the D8 and D-infinity contributing areas are presented in terms of grid cell size and discussed. 7 points

Module Section	Learning Objectives student will be able to... (The)	Learning Activities	Open web-based platforms and datasets
(1) Data Analysis and Statistics	Calculate water storage, fluxes, and uncertainty in components of the hydrologic cycle	Problems: Water balance, uncertainty	ESRI Story Map- Introduction to Physical Hydrology
	Navigate public websites to extract key hydrologic information	Problems: streamflow time series analysis	StreamStats: USGS NWIS
	Perform basic hydrologic data analysis for a watershed of interest	Authentic Task: For user-selected stream gage: (i) describe watershed attributes, (ii) retrieve streamflow data, (iii) perform statistical analyses including daily and seasonal plots, flow duration curve, exceedance, etc.	JH Notebook (R), including packages: DataRetrieval, zoo, ggplot
	Assess and interpret hydrologic trends in the context of a specific watershed	CYU: Multiple Choice and Open Response	
(2) Geographical Information Systems	Derive hydrologically useful information from digital elevation models (DEMs)	Problems, CYU: Open Response	USGS National Elevation Dataset
	Describe the sequence of steps involved in mapping stream networks, catchments, and watersheds from DEMs	Problems: Basic terrain analysis for hydrologic research	JH Notebook (Python), including packages: TauDEM, gdal, geopandas, rasterio, rasterstats
(3) Runoff Generation	Compute an approximate water balance for a watershed using open data	Problems: Water Balance, Runoff Ratio	
	Use appropriate terms to describe the processes involved in runoff generation	CYU: Concepts and definitions of runoff generation mechanisms	
	Differentiate between runoff generation mechanism and when and where each is likely to occur	Problems: Multiple Choice	ESRI Story Map- Rainfall-Runoff Processes
(6) Calculating Runoff using TOPMODEL	Justify why and how specific changes in physical and climate attributes will influence dominant runoff mechanism and storm hydrograph shape	CYU: Open Response	
	Compute the topographic wetness index and describe its role and use in runoff calculations, given a watershed DEM	CYU: Topographic wetness index, variable source area, hydrologic models	ESRI Story Map- Rainfall-Runoff Processes
	Apply TOPMODEL principles and equations to calculate runoff using CUAHSI JupyterHub and TauDEM and given necessary catchment and storm characteristics	Authentic Learning Activity: Simulate runoff using semi-distributed hydrologic model	JH Notebook (Python); ESRI Story Map - Rainfall-Runoff Processes
	Critically assess assumptions and determine if and why an appropriate tool, given information about specific catchment and storm attributes	Authentic Learning Activity: Estimate runoff across a watershed	