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

Running title: Open web platforms facilitate data-driven learning in hydrology

Authors: Belize Lane[1*], Irene Garousi-Nejad[1], Melissa Gallagher[2], David Tarboton[1], Emad Habib[3]

[1] Utah Water Research Laboratory, Utah State University. Logan, Utah, USA

[2] University of Houston. Houston, Texas, USA

[3] University of Louisiana at Lafayette. Lafayette, Louisiana, USA

* Corresponding author, belize.lane@usu.edu

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

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 JupyterHub 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.

Key Words: HydroLearn, data-driven, learning, education, collaborative, web-based, HydroShare, open

Introduction

Hydrologists investigate the distribution and variation of water across a range of spatial and time scales. In the face of mounting water resources challenges - due to a growing population, climate and land use change, and shifting societal values - hydrology has evolved from a mainly applied engineering discipline to a fundamental underpinning of geo and environmental sciences (Eagleson 1991; National Research Council 1991; Wagener et al., 2007; Vogel et al., 2015). As an applied and interdisciplinary science, hydrology benefits from first-hand knowledge gained by working with many different datasets. Generic equations are not easily translated into local or regional solutions, and experience with specific systems and datasets is critical for hydrologic practice and research. Such data-driven analysis is often needed to conceptualize complex processes and to explore how well models or equations work in particular settings or to answer a specific problem.

As demands on hydrologists have grown, so have calls to enhance hydrology education at the upper division and graduate level to adequately prepare students for both research and industry (Merwade & Ruddell 2012; Ruddell & Wagener 2015). Enhancing students' ability to conceptualize, analyze and interpret complex hydrologic processes is an area of much research (Bourget, 2006; Wagener et al., 2007; Wagener et al., 2012; Ngambeki, Thompson, Troch, Sivapalan, & Evangelou, 2012; Merwade & Ruddell, 2012; Marshall, Castillo, & Cardenas, 2013; Ruddell & Wagener 2015; Habib, Deshotel, & Williams, 2018). Educators have recognized a need to augment traditional teacher-centered lectures centered on fundamental physical laws with student-centered, data-driven learning activities that enable students to explore the integrated spatio-temporal hydrological system using authentic datasets and modeling tools (Merwade & Ruddell, 2012). Problem-based learning activities that include the use of authentic, real-world problems and datasets have been shown to enhance engineering and hydrology learning outcomes and career preparation (Lyon & Teutschbein 2011; Litzinger, Lattuca, Hadgraft, & Newstetter, 2011; Habib, Ma, Williams, Sharif, & Hossain, 2012; Sanchez, Ruddell, Schiesser, & Merwade, 2016; Habib, Deshotel, Lai, & Miller, 2019; Gallagher et al., 2021; Merck et al., 2021). As a result, several web-based educational platforms that offer learning in an internet based environment have been developed to incorporate real-world data and modeling resources in hydrology learning activities (e.g., SERC, CSDMS, COMET, HydroViz, RWater).

At the same time, the sheer volume and access to hydrologic data has grown rapidly through breakthroughs in remote sensing and in situ data collection and data services. "Big data" promises to provide new hydrologic insights to address mounting water resources challenges, and collaborative open web-based platforms are being increasingly developed and adopted by the global hydrologic science community to harness these datasets (Slater et al., 2019; Goodall et al., 2017). “Open” in this case implies that data and computational resources can be openly shared, discovered, and accessed among the community (Chen et al., 2020) while the underlying software may, in some cases, be commercial (i.e., not open-source). Open-source software by contrast is developed by a non-profit community and is free to use, distribute, and modify. Open-source software provides unique opportunities in education for accessibility (Rajib et al., 2016) and in research for transparency and reproducibility as it reduces the financial and time costs for others to reproduce results (Rosenberg et al., 2020); however, it may not always have extensive technical support.

As in many fields, hydrology is trending toward a standardized open web-based structuring of data services, formats, and metadata to facilitate data management, analysis, and sharing needs. For example, the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) has developed an array of web-based data services and information systems specifically for the hydrologic science community (Horsburgh et al., 2008; Horsburgh et al., 2016; Goodall et al., 2017). Other open web-based platforms not specific to hydrology are also being increasingly adopted. For example, GoogleColab, Google Earth Engine, and Jupyter Notebooks all allow users to create and share documents that contain live code, equations, visualizations, narrative text and link to web-based data services. Collaborative platforms like these provide convenient, standard workspaces and tools for the hydrology community, but they also demand that hydrologists and hydrology instructors keep pace with the rapid advancements.

With the promises of "big data" in hydrology come new challenges related to data management and reproducibility. Reproducibility is a critical requisite to advancing hydrologic discovery and innovation, and to subsequent integration and reuse of findings by the community (Wilkinson et al., 2016; Hutton et al., 2016; Stagge et al., 2019; Essawy et al., 2020; Choi et al., 2021). The complexity and diversity of hydrologic systems reflected in emerging data requires that scientists can reproduce methods developed in specific settings more broadly across a range of scales and locations to robustly evaluate hypotheses and assumptions (Clark et al., 2016; Hutton et al., 2016). Particularly as datasets and models become more complex, analysis procedure and code need to be transparent and well-documented to allow for reproduction (Stagge et al., 2019; Rosenberg et al., 2020). The increasing use of open and open-source software by the hydrologic science community underpins these dual aims of accessibility and reproducibility.

The shift in data availability and analysis capabilities offered by open web-based platforms and the call for reproducible research has fundamentally transformed the role of hydrology instructors from disseminators of knowledge to guides in learning, critical thinking, and good research practices. However, these changes have not yet fully translated into changes in the education of future hydrologists. While educational platforms are emerging to support authentic, problem-based learning, as described above, they are mostly static and lack mechanisms for harnessing the emerging open data services and practices being adopted by the professional community. They also generally lack a formalized pedagogical structure to help instructors develop their own learning activities with these aims in mind. One exception is HydroLearn, a web-based collaborative hydrology education platform that provides a formalized and validated pedagogical structure - including tools to support instructors in creating learning objectives, formative assessment questions - to develop authentic, problem-based learning activities. Student learning of concepts and technical skills has been found to increase after using HydroLearn modules (Merck et al., 2021). However, HydroLearn's capacity to harness emerging open platforms and data services to enhance conceptual understanding in hydrology through data-driven learning remains largely untapped.

Advancing understanding in hydrological processes requires a workforce trained in working with data and learning from data, and learning platforms and modules designed to facilitate data-driven learning have the potential to change the way hydrologists do research that advances hydrological processes. This paper describes a HydroLearn physical hydrology learning module targeting advanced undergraduate and early graduate students in hydrology and engineering. The objectives of the learning module are (1) to harness emerging open web-based platforms in order to (2) develop data-driven learning skills whereby students actively explore key concepts using real data that is relevant and meaningful to them thereby (3) enhancing student learning of fundamental hydrology concepts while (4) training students in good data management and reproducibility practices. The paper briefly describes several open web-based platforms for hydrology and their potential educational utility, introduces the learning module including integration of these platforms, and offers initial student perceptions and instructor reflections on the module.

Open web-based platforms for hydrologic analysis

Collaborative open web-based platforms and tools are being increasingly adopted by the hydrologic science community. A comprehensive review of available resources is beyond the scope of this paper. Instead, here we briefly summarize the platforms utilized in the HydroLearn physical hydrology learning module and their potential educational utility, including HydroShare, CUAHSI's JupyterHub (henceforth JH), and ESRI Story Maps. HydroShare is a web-based collaborative platform for hydrology data storage, retrieval, sharing, and processing (Tarboton et al., 2014; Horsburgh et al., 2016; Essawy et al., 2020). Hydrology instructors and students increasingly use

HydroShare to access free cloud-based versions of several software programs and hydrologic models and use them for various research and learning applications, or to access previously uploaded static teaching resources (Ward et al., 2020). CUAHSI's JH is an open cloud-based environment for computational notebooks that allows users to create and share documents that contain live code, equations, visualizations and narrative text (Choi et al., 2021). JH computation notebooks are used to write, build, and run codes as well as run pre-installed software (e.g., TauDEM, Tesfa et al., 2011; Tarboton, 2018), but can also be used as teaching tools to build programming and data management skills. Finally, ESRI Story Maps combine narrative text with immersive content that fills the screen with maps, images, or videos for an engaging learning experience. While the code for ESRI Story Maps is not open-source, these cloud-accessed resources harness ArcGIS's analysis tools and GIS platforms, and can be hosted and made publicly available directly through ArcGIS Online. Instructors may choose to develop Story Maps to allow students to directly interact with the data through a personalized hands-on experience (e.g., Kerski 2019). Alternatively, students could be assigned to create their own Story Maps to dynamically communicate project results (e.g., Battersby & Remington 2013).

In addition to the platforms described above, numerous programming packages can be used in learning activities to facilitate hydrologic data retrieval (e.g., *DataRetrieval*; *rnrfa),* analysis (e.g., *TauDEM*, Tarboton, 2018), modeling (e.g., Choi et al., 2021), and visualization. For example, the R *waterData* package allows a user to import daily hydrological data from the USGS web services and plot time-series data (Ryberg & Vecchia, 2012; R Core Team 2020). A detailed description of R packages relevant for hydrologic analysis is provided in Slater et al. (2019).

HydroLearn Physical Hydrology Learning Module

HydroLearn is itself an open web-based platform that aims to help hydrology instructors develop, share, and adapt learning modules. It combines research-based active learning methods with authentic online learning modules. The modular nature of HydroLearn and the dynamic computational notebooks allows instructors to use, combine, or adapt content, datasets and scripts from existing learning modules to their specific instructional needs and geographic settings. Active learning is supported through the ability to embed video and questions, iFrames, and learning activity templates. Common elements of HydroLearn modules include Check-Your-Understanding (CYU) questions, quantitative problems, and authentic learning activities. CYU question formats include multiple choice, checkbox, drag-and-drop questions, and open response to higher-level questions related to process interpretation. By contrast, authentic tasks are high cognitive-demand tasks built to reflect how knowledge is used in real life and to simulate the type of problems that a professional might tackle. Each authentic task has a rubric, an assessment tool intended to set clear expectations for students and make grading more objective. the platform provides wizards and templates to help instructors develop strong learning objectives and align the teaching activities, learning outcomes, and assessments, a process referred to in the learning literature as constructive alignment (Biggs & Tang, 2011; Kandlbinder 2014).

The HydroLearn Physical Hydrology learning module incorporates the above elements and consists of six sections: (1) data analysis and statistics in hydrology, (2) geographical information systems in hydrology, (3) runoff generation, (4) water in the soil, (5) infiltration modeling, and (6) calculating runoff using TOPMODEL concepts. Below we briefly discuss sections 1, 2, 3, and 6 of the module to highlight the use of active and authentic learning, open web-based platforms, and data-driven learning skills. Sections 4 and 5 of the module are not covered here because they are similar to other sections in format and learning elements used. Table 1 lists key learning objectives, learning activities, open web-based platforms and data sources for each section.

Section 1- Data Analysis and Statistics in Hydrology

Here, students are introduced to fundamental concepts in hydrology while learning basic data analysis and management skills through a set of problems and an authentic learning activity. Key terms and concepts are introduced using an ESRI Story Map. The problems and authentic learning activity are performed in a JH computational notebook accessed through HydroLearn.

Through a link on the module Introduction page, students are first directed to an *Introduction to Physical Hydrology* Story Map (Figure 1) that provides a map-based virtual tour of physical hydrology in the Logan River watershed, Utah, USA. This Story Map introduces key terms and concepts related to hydrologic processes and provides place-based examples and illustrative images and figures to help students connect with the landscape and concepts personally. As part of the development process, open repositories of videos and images of hydrologic processes in different settings were assembled from existing web products to visualize fundamental processes and make this media available to other educators.

Several quantitative problems ask students to make basic calculations related to water storage, fluxes, and uncertainty in key components of the hydrologic cycle. These calculations are performed in a JH computational notebook stored in a HydroShare resource embedded within the learning module using an iFrame for easy access (Figure 2). Students are required to create an account in HydroShare in order to run the notebook. The Notebook uses the *R* programming language and is intended for students who have had no or limited programming experience. The packages and code needed to perform the calculations are provided and well notated to familiarize students with basic programming notation and key functions. This prepares them for the next section in which they are required to modify the code slightly to their selected dataset. The notebook used here provides a gentle and context-based introduction to R programming recognizing that learners without programming knowledge are more likely to be interested and see its value when it is applied in the context of an authentic problem (Kalelioglu 2015).

In one problem, students first estimate long-term average evapotranspiration rates for several watersheds using a simple water balance model, and then compute the 95% relative and absolute uncertainties in these estimates. The Notebook begins with text descriptions and equations, followed by a section with necessary scripts needed to perform the calculations. Students must run the script to generate results, and are then prompted to add a new section and switch from calculation to text response to describe their results. They are then asked to check their understanding by comparing and contrasting the water balance and uncertainty results across watersheds in their own words. Specifically, they are prompted to indicate what catchment and climate conditions may be influencing the long-term water balances of these watersheds and then describe how specific hydrologic processes may be affected by these conditions.

In the authentic learning task, students identify a USGS stream gage of interest and work through several steps to explore the streamflow patterns and statistics at that location, including to: obtain streamflow data, delineate the upstream catchment, generate time series plots and calculate summary statistics, interpret the hydrological behavior of the river using a flow duration curve, evaluate seasonality trends, and use histograms and probability distributions to describe characteristics of the flow (Figure 3). For this activity, the Notebook walks students through each step in detail, paralleling background information and summary questions in the HydroLearn module. Complete code is provided to the students but a few parameters (e.g., stream gage ID, start date) must be updated to customize the script to their specific dataset. First, they are provided with a tutorial video in the learning module demonstrating how to select a stream gage, delineate a watershed, and describe basin characteristics. Then, they are directed how to use the *dataRetrieval* package in R to bring streamflow data from the USGS National Water Information System website into their cloud-based workspace for subsequent analysis without the need to download data. Finally, they explore plotting and visualization tools to generate a flow duration curve and a single plot including annual peak flow, average flow, and seven-day minimum flow. Students are also shown how to manage streamflow data using data-frames, the definition and use of Water Years (October to September in the U.S.), and how to aggregate data by year or month to calculate summary statistics (e.g. average June flow). The section concludes with CYU multiple choice questions to reinforce key concepts. Having students select a USGS stream gage of interest to perform these analyses provides students with experience in reproducible research practices and opens up opportunities for students to evaluate and compare hydrological responses in distinct climate and catchment settings. It also allows students to assess results in the context of a system they are familiar with to promote conceptual understanding.

Section 2- Digital Elevation Models and GIS in Hydrology

This section introduces the use of geospatial processing tools and basic terrain analysis to derive hydrologically useful information from digital elevation models (DEMs) for an example watershed, including watershed delineation and stream network generation. Similar to Section 1, HydroLearn links to a JH computational notebook that is used to complete this data-driven learning activity. Here, the Notebook employs the Python programming language (i.e. iPython Notebook) so students can explore the distinct utility of Python-based spatial analysis packages and visualization tools. The full set of learning activities in Section 2 could be adapted for another watershed, and provide students with experience and training in open web-based tools and reproducible research practices. The specific activities are summarized below to illustrate the authentic, data-driven approach.

1- Preparation, Libraries, and Getting Oriented Key concepts and input files are presented, followed by an overview of Python libraries and functions that will be used to extract hydrologic information about the study watershed and a figure illustrating expected results. Students are then prompted to open the iPython Notebook and step through the individual sections, following instructions provided throughout the document. They are also encouraged to keep the HydroLearn module open on their browser alongside the dynamic notebook to guide the activity (Figure 4). Questions are posed periodically throughout the notebook, corresponding with questions in the HydroLearn module, to clarify key ideas for students and what they should be able to calculate or describe at any point in the notebook. Students are directed to the USGS National Water Information System (NWIS) website embedded as an iFrame within the module to explore the stream gage location and extract key information including drainage area. To simplify the exercise, the USGS 10-m DEM for the Logan River watershed is provided for the students in a linked HydroShare resource. Guidance is also provided in the module for obtaining DEM datasets for other locations of interest to facilitate the use of these analysis tools and concepts more broadly by both students and instructors.

2- Basic DEM Hydrologic Analysis After a brief discussion of the conceptual underpinnings of the geospatial processing toolset, TauDEM, and links to useful resources, students are introduced to the Basic Grid Analysis toolset including *Pit Removal, D8 Flow Direction,* and *D8 Contributing Area.* The learning module again provides static code snippets and figures of expected results (Figure 4), as well as summary questions associated with the learning activity.

3- Stream Network Analysis A network analysis function within TauDEM is used to create and analyze a small section of stream network in the Logan River watershed given a DEM. The student generates a stream network and prepares an attribute table of a subset of reaches (including the link number, downstream and upstream linked reaches, contributing area, length, and corresponding watershed ID) to describe the properties of the stream network and subwatersheds. Next, the student identifies the number of grid cells in each subwatersheds, calculates the area of each subwatershed, and reconciles the values with the contributing area values from the stream network.

4- Water Balance In the final learning activity of this section, the student delineates the watershed that was analyzed in the previous steps and then uses the delineated watershed to calculate key water balance components. Streamflow data is downloaded from the USGS NWIS website for the Logan River stream gage. A web client is used to retrieve the annual 800-m precipitation for 30-year normals (1981-2010) from PRISM (Daly et al., 2000). This dataset is visualized and then clipped to the watershed extent in the Notebook. Finally, the student calculates mean annual precipitation over the watershed and reports this value along with watershed area, mean annual streamflow, and the runoff ratio.

Section 3- Runoff Generation

Section 3 is unique in that it focuses on building conceptual understanding of runoff generation processes and therefore only addresses the first two aims considered in this paper: promoting active learning and harnessing open web-based software. The bulk of the content is provided through a linked standalone ESRI Story Map called *Introduction to Rainfall-Runoff Processes*. This resource is an adaptation of an online physical hydrology workbook developed for COMET in 2003 (Tarboton, 2003). We reconceived the workbook as an interactive Story Map including updated and compatible images, videos, and animations. The HydroLearn section solidifies key concepts discussed in the Story Map and promotes active learning through targeted CYU questions.

Section 6- Simulating Runoff using TOPMODEL

The culminating section of the module builds on data-driven learning skills developed in Section 2 to simulate semi-distributed variable source area runoff generation in a tributary to the Logan River. TOPMODEL is a conceptual hydrologic model that uses basic topographic and soils information to estimate runoff from the saturated and unsaturated zones (Beven 1989). The location of the interface between the two zones, quantified by the water table elevation, corresponds to the soil water saturation deficit and controls the types and amounts of flow simulated by the model. Following the established HydroLearn structure, the section starts by delineating the learning objectives and provides key background information, a detailed grading rubric, learning activity instructions and summary questions (Figure 5). Section 6 first provides background information on TOPMODEL, including key equations, terms, and assumptions, which are covered in more depth in a linked Story Map. CYU questions promote active learning and prepare students for the Authentic Learning Activity. Similar to Section 2, an iPython computational notebook guides the students through the specific activity steps outlined in the module. Similar to Section 2, this exercise could be adapted for another watershed of interest by instructors or students, which again provides training in research reproducibility practices. Finally, the module provides several summary questions to solidify concepts applied in the learning activity.

Learning data management and reproducibility practices

The HydroLearn Physical Hydrology learning module educates students in data management and reproducible research practices through multiple means. It provides direct training for students with findable, accessible, interoperable, and reusable resources accessed through a series of learning activities. Students gain experience working with open web-based tools such as HydroShare and CUAHSI JH that are explicitly designed to share hydrology data and code to facilitate transparency and reproducibility. The authentic learning activities are also designed to be readily adapted and reproduced at other locations including specific guidance on how to obtain other datasets. This approach aims to empower students to reproduce analyses and reuse platforms and data services introduced in the learning module.

Implementation

We piloted our learning module during the Fall 2019 and Fall 2020 semesters with 7 and 15 students, respectively, enrolled in a first-year graduate-level physical hydrology course taught through the Department of Civil and Environmental Engineering at Utah State University. The majority of participants were in Civil Engineering or Watershed Sciences graduate programs. None of the students had used computational notebooks before, and most had limited to no programming experience. At the end of each module section, students were asked to provide open-ended feedback on their experience. In particular, we were interested to understand student perceptions of the HydroLearn module and the utility of open web-based tools, what worked and didn't work for them, and how their conceptual understanding of the material improved after participating.

Results & Discussion

The HydroLearn physical hydrology learning module described in this paper was intended to provide distinct benefits for hydrology students, the larger research community, and hydrology instructors. For students, the learning module aims to enhance data-driven learning through student-centered learning activities that harness emerging open data services. It also provides experience and training in open web-based tools and reproducible research practices. For the research community more broadly, this type of learning module explicitly addresses the call in hydrology for open and reproducible research and provides training in data-driven, process-oriented thinking needed to advance hydrological research. For instructors, the collaborative, modular nature of the HydroLearn platform and computational notebooks allows content, datasets, and scripts to be readily shared, combined, and adapted to specific instructional needs and geographic settings.

While this learning module has only been implemented in two graduate courses to date, initial student perceptions and the instructor's reflections are summarized below. This discussion is intended to provide some guidance regarding implementation and customization by other instructors and highlight opportunities to more effectively harness open web-based platforms for upper-level data-driven hydrology education. Based on the student perceptions and the instructor's reflections, we found that: (a) the utility of JH Notebooks could have been more clearly communicated, (b) harnessing web-based platforms facilitated data-driven learning, and (c) opportunities remain to enhance student learning. We discuss each of these themes below, first with evidence from the student questionnaires and then the instructor's feedback.

Utility of open web-based platforms

The students found the linked ESRI Story Maps very effective at delivering information interactively and appreciated the combination of text, videos, figures, and hyperlinks to other resources. One student indicated that they wished all of their textbooks were Story Maps, and that the 'highly accessible content makes sharing with others who may be interested in these topics much easier than information out of a textbook.' Another indicated that having videos and figures interspersed with text 'helped to not only explain but show the time and space variability of the processes.' Several students noted that the interactive aspects helped break up the text and drive home concepts.

Perceptions on the utility of HydroShare and JH computational notebooks were variable. In Section 1, students with even a small amount of programming experience were initially far more receptive to these tools than students with no prior experience, but this discrepancy diminished over the semester as students established more familiarity with the platforms and programming syntax. A subset of students with no programming experience indicated early on that they thought they would appreciate these tools more once they developed basic programming skills and were now interested to do so. Others said that they appreciated knowing that Jupyter Notebooks exist, even if they were "still unable to replicate or augment the code so far." In terms of the structure of the notebooks themselves, most students were grateful for the amount of code that was already provided for them, but some indicated that doing more of the coding themselves would improve learning outcomes. Some expressed frustration about technological challenges such as losing server connection to CUAHSI JH and needing to log out and start over.

Several students were frustrated by the amount of time it took to work through the programming scripts. For instance, "it can be frustrating when you understand what you're trying to do but can't find the code to do it. I think those types of frustrations take away some of the benefits of the lab and cause students to worry more about coding than what we are doing with the code." Furthermore, "sometimes I spend more time looking and thinking at and about the code rather than the concepts." Even by the third learning activity, some students were still unsure of the utility of the computational notebooks as learning tools. "I still don't really like this format in general. It's cool to have the open source tools but I feel like I haven't gained the skill to apply this code in any other context… I would prefer to build these together in class so we could actually learn how to do it ourselves."

In early applications of this learning module, the intense focus by the instructor on familiarizing students with the tools distracted from clearly conveying the value of these tools. Since this was not clearly articulated, several students questioned the need to learn how to use programming and computational notebooks to complete learning activities given other common software programs and GUIs already available to accomplish similar things. For example, "training us to use open software programs is great but … can hurt the learning process when compared with a program that has a user-friendly GUI and is easy to create maps like ArcGIS." Given that students may prefer the use of tools they are already familiar with or have heard of, instructors must clearly articulate the anticipated value associated with using open web-based tools.

Harnessing web-based platforms facilitates data-driven learning

Students generally had positive feedback about the use of HydroLearn to lay out learning objectives, activities and expectations, as well as its integration with HydroShare, CUAHSI JH, and ESRI Story Maps. One student said HydroLearn was "easy and straightforward to use, and provided all relevant links making it convenient to access everything… the layout was such that I easily followed instructions for the learning activities and found the questions I needed to answer." Other students noted that the "variety of ways in which the material was presented allowed for better understanding" and "allowed for more focus on principles rather than just coding." The CYU questions in particular appeared to help students solidify key ideas and support higher learning levels. For instance, in Section 1, students computed water balance uncertainty in several watersheds through a series of calculations and were then asked to check their understanding by comparing the different watershed results in their own words in the context of their physical catchment and climate settings. The JH computational notebooks allowed students to easily switch between calculations and text-based response in the same document. One student indicated that the CYU questions throughout the module "helped me focus on what was particularly important within smaller blocks of information" and "were really helpful for developing my understanding of rainfall-runoff processes." Another student articulated that "one of the things I enjoyed most about the module was that it really tested your understanding.The CYU questions in particular had relatively simple answers, but they did a good job of testing actual understanding of the concepts- especially the CYU question hints and explanations as to why the answer was correct."

The effectiveness of integrating multiple open web-based platforms to enhance teaching hinged on the formalized pedagogical structure provided by HydroLearn. The emphasis on constructive alignment between teaching tools, learning activities and objectives facilitates development of activities that integrate data and tools from multiple sources while explicitly targeting mindfully crafted learning objectives across multiple levels (e.g., understand, apply, analyze). Effective framing of questions can encourage the students to think critically about the underlying processes while learning the basics of the data analysis tools rather than getting lost in the mechanics of the calculations. Following completion of each section, instructors are encouraged to facilitate additional discussion regarding which settings the equations and models worked well in and which settings gave strange results and why that might be. This would provide an opportunity for students to critically evaluate model assumptions and requirements based on the shared and varied experiences of the students themselves, rather than a learning activity that relies on pre-canned data from a well-behaved system.

Data-driven learning was facilitated through the flexible data and programming language integration capacity of the open web-distributed platforms. In Section 1, students were taught how to access streamflow data for any USGS stream gage of interest using a computational notebook written in *R* and work through the learning activity with no prior programming experience. In general, the computational notebooks allowed the students to generate transparent and reproducible outcomes and compare their results to those for other locations. The notebooks were designed to accommodate different physical settings, parameter values, or datasets with minimal effort to explore how results varied across systems. For instance, students could simply input a different digital elevation model or change the storm depth value and re-run the notebook to update results and figures. Furthermore, using short and modular lines of code, students could accomplish all necessary tasks without the need for familiarity with a particular software interface, including the location of specific tools and features and any extra tools that may need to be installed.

Guidance is provided in the module for obtaining input datasets for other locations in learning activities where students are provided with example data to simplify the exercise. This guidance is intended to promote reproducibility and empower students to use the data analysis techniques and concepts covered in the module at other locations. Data retrieval guidance also supports module implementation by other instructors. Instructors can adapt learning activities associated with the Logan River watershed to other watersheds that are more relevant for their students. The collaborative nature of the HydroLearn platform allows instructors to create a new instance of the learning module that is customized with data for a local watershed where students may better appreciate the context. Instructors may adopt the entire or part of the module and modify the content to better reflect the specific goals defined in their syllabus. The Physical Hydrology module is currently being implemented in this manner by two instructors at different universities. In the case of modules sections linked to specific datasets and dynamic notebooks stored in HydroShare, instructors may need to develop and link separate HydroShare resources referencing their preferred watershed. Alternatively, in the case of user-selected datasets such as in Section 1, there is no need to modify the resources at all, and each student will still have an authentic, personalized learning experience due to the modular data-driven nature of these tools.

Opportunities to enhance data-driven learning

Instructors using this or similar learning modules are encouraged to focus students' efforts on how to apply the JH computational notebooks in different contexts rather than to fully understand or be able to generate every line of code and function themselves. The learning module was not intended to be an introduction to programming, although basic programming literacy is necessary and the module did provide some level of context-based learning of coding that serves as a motivator and entry point for non coders to approach coding in a limited and practical way. An instructor may go so far as to clearly articulate that the students do *not* need to fully understand the code to effectively apply the notebooks - particularly if they do not have a strong programming background - just as they do not need to understand all the code that supports calculations in ArcGIS or other software programs they may have worked with. Students should be helped to understand in general what code does, not how it does it. We posit that an instructor's main role in these modules should be to guide students regarding key concepts and how to implement, reproduce, and adapt an analysis in various settings - including the ability to identify key inputs, outputs, and parameters in the code.

Students may need encouragement to trust the parts of the code that they do not understand until they develop confidence through repeated evaluation of results. Instructors should, of course, still promote critical evaluation of results. Interestingly, to some extent, having the black box opened up - including all the programming scripts that are not visible in user-friendly GUIs - led to more rather than less skepticism by the students. Incorporating some videos narrated by the instructor and embedded with the module (e.g., illustrating different steps) has been shown to help support student buy-in and reduce their "shock" to the use of open platforms (Habib et al., 2018).

The data-driven learning activities had a steep learning curve, as evident from the number of functions and packages required as well as student feedback. The technical and technological challenges students encountered as a result required substantial technical support. Both times the learning module was implemented, the instructor had a graduate teaching assistant who provided detailed walk-throughs of the notebooks when they were first assigned and offered technical sessions and troubleshooting support for students. These additional requirements for effective implementation were a large time sink for both the instructor and teaching assistants. For subsequent applications of this or other learning modules that use open web-based platforms and programming, we encourage the use of code that is easy to understand, troubleshoot, and requires limited prior programming or operating system knowledge of students or instructors, particularly if the students have a range of backgrounds and programming experience.

Support mechanisms to guide learners through the data-driven procedures and provide just-in-time assistance are critical to the success of online learning activities (Kolodner, Owensby, & Guzdial, 2004; Habib et al., 2018). This is particularly true when multiple new tools are being presented at once, and it may be difficult to foresee where students might make mistakes or need assistance. For these reasons, the material should be presented with appropriate curricular expectations and include embedded interactive tools to support students’ progression through the lessons and activities (Habib et al., 2018). The issues described above could likely be addressed in large part by incorporating additional technical support within the HydroLearn module. These user-support tools might include narrated video tutorials, additional CYU questions or check-in points, and formative feedback quizzes.

There are inevitable costs to the emphasis on new tools and software, and student feedback indicated some difficulty focusing on the key concepts and higher-level learning objectives with so much emphasis on using tools and performing calculations. Furthermore, with any technology and particularly open and open-source, there is always the challenge of future-proofing learning activities to limit the need to re-write or adjust scripts. Already, in the year since the module was first developed, several scripts had to be revised to accommodate a major transition in CUAHSI JH platform structuring. The open nature of HydroLearn allows for easy updates of the resources and content, as opposed to it being more difficult to update static, closed material (e.g., textbook, slides, pdfs, etc.). There are also numerous and growing options for platforms (e.g., Google Colab, GitHub) that may work as well or better than those applied in this module and have long-term support and cyberinfrastructure at a much larger scale.

Conclusion

As an applied and interdisciplinary science, hydrology relies on direct experience with many different data sets and analyzing many systems. Teaching students to explore different datasets and how well models or equations capture hydrological processes in particular settings or to solve a particular problem is essential. The learning module described in this paper is a proof-of-concept for harnessing state-of-the-science open web-based technology that is increasingly utilized by the hydrology professional community to enhance physical hydrology education and prepare students to apply open and reproducible tools and practices. The data-driven learning activities allowed students to work with datasets for systems that they were particularly interested in, and enabled critical evaluation of results and assumptions. Challenges included some resistance to open software and programming, time consuming technical and technological difficulties, and difficulty meeting higher-level learning objectives with the heavy emphasis on new tools. Opportunities to further enhance data-driven learning include better articulating the benefits of using open web-based platforms upfront, incorporating additional user-support tools, and focusing methods and study questions on implementing and adapting codes to explore fundamental processes rather than tools and syntax. The profound shift in the field of hydrology toward using open data and data analysis platforms requires hydrology instructors to rethink traditional content delivery and focus instruction on using these data and data analysis tools in the preparation of future hydrologists and engineers.

Data Availability Statement

The Physical Hydrology HydroLearn module described here is available at https://edx.hydrolearn.org/courses/course-v1:Utah_State_University+CEE6400+2019_Fall/about (Lane and Garousi-Nejad, 2020). Data and Jupyter Notebooks are in HydroShare (Garousi-Nejad, I. and Lane, B. 2020, 2021a, 2021b).

Citations

Battersby, S. E., & Remington, K. C. (2013). Story maps in the classroom. *Arcuser Magazine Spring 2013*. https://www.esri.com/about/newsroom/wp-content/uploads/2018/11/storymaps.pdf

Beven, K. (1989). Changing ideas in hydrology—The case of physically-based models. *Journal of Hydrology*, *105*(1–2), 157–172. https://doi.org/10.1016/0022-1694(89)90101-7

Biggs, J., & Tang, C. (2011). Train−the−trainers: Implementing outcomes−based teaching and learning in Malaysian Higher Education. *Malaysian Journal of Learning and Instruction*, *8*, 1–19. https://files.eric.ed.gov/fulltext/EJ1137298.pdf

Bourget, P. G. (2006). Integrated water resources management curriculum in the United States: Results of a recent survey. *Journal of Contemporary Water Research & Education*, *135*, 107–114. https://doi.org/10.1111/j.1936-704X.2006.mp135001013.x

Chen, M., Voinov, A., Ames, D. P., Kettner, A. J., Goodall, J. L., Jakeman, A. J., ... Lü, G. (2020). Position paper: Open web-distributed integrated geographic modelling and simulation to enable broader participation and applications. *Earth-Science Reviews*, *207*, 103223. https://doi.org/10.1016/j.earscirev.2020.103223

Choi, Y.-D., Goodall, J. L., Sadler, J. M., Castronova, A. M., Bennett, A., Li, Z., … Tarboton, D. G. (2021). Toward Open and Reproducible Environmental Modeling by Integrating Online Data Repositories, Computational Environments, and Model Application Programming Interfaces. *Environmental Modelling & Software*, *135*: 104888, https://doi.org/10.1016/j.envsoft.2020.104888

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., ... Ceola, S. (2016). Improving the theoretical underpinnings of process-based hydrologic models. *Water Resources Research*, *52*(3), 2350–2365. https://doi.org/10.1002/2015WR017910

Daly, C., Taylor, G. H., Gibson, W. P., Parzybok, T. W., Johnson, G. L., & Pasteris, P. A. (2000). High-quality spatial climate data sets for the United States and beyond. *Transactions of the ASAE*, *43*(6), 1957. https://pubag.nal.usda.gov/catalog/27078

Eagleson, P. S. (1991). Hydrologic science: A distinct geoscience. *Reviews of Geophysics*, *29*(2): 237-248. https://doi.org/10.1029/90RG02615

Essawy, B. T., Goodall, J. L., Voce, D., Morsy, M. M., Sadler, J. M., Choi, Y. D., ... Malik, T. (2020). A taxonomy for reproducible and replicable research in environmental modeling. *Environmental Modelling & Software, 134***,** 104753. https://doi.org/10.1016/j.envsoft.2020.104753

Garousi-Nejad, I., B. Lane (2020). Hydrologic Statistics and Data Analysis (M1), HydroShare, http://dio.org/10.4211/hs.873b5490f986428bbd23e37c81906402.

Garousi-Nejad, I., B. Lane (2021). Digital Elevation Models and GIS in Hydrology (M2), HydroShare, https://doi.org/10.4211/hs.9c4a6e2090924d97955a197fea67fd72.

Garousi-Nejad, I., B. Lane (2021). Calculating Runoff using TOPMODEL (M6), HydroShare, https://doi.org/10.4211/hs.ea30176c717d4b7baeb85c16427f1d6f.

Goodall, J. L., Merwade, V., Zaslavsky, I., Tarboton, D. G., Horsburgh, J. S., Ames. D. P., & Couch, A. (2017). Hydrologic Information Systems, Chapter 7 in *Handbook of Applied Hydrology*, Ed. V. P. Singh, McGraw Hill, pp.7-1 to 7-9.

Habib, E., Ma, Y., Williams, D., Sharif, H. O., & Hossain, F. (2012). HydroViz: Design and evaluation of a Web-based tool for improving hydrology education. *Hydrology and Earth System. Sciences*., *16*(10), 3767–3781. https://doi.org/10.5194/hess-16-3767-2012

Habib, E., Deshotel, D., & Williams, D. (2018). Unlocking the educational value of large-scale, coastal-ecosystem restoration projects: Development of student-centered, multidisciplinary learning modules. *Journal of Coastal Research*, *34*(3), 738–751. https://doi.org/10.2112/JCOASTRES-D-17-00064.1

Habib, E., Deshotel, M., Lai, G. L., & Miller, R. (2019). Student perceptions of an active learning module to enhance data and modeling skills in undergraduate water resources engineering education. *International Journal of Engineering Education*, *35*(5), 1353–1365. https://par.nsf.gov/biblio/10189082

Horsburgh, J. S., Tarboton, D. G., Maidment, D. R., & Zaslavsky, I. (2008). A relational model for environmental and water resources data. *Water Resources Research*, *44*(5). https://doi.org/10.1029/2007WR006392

Horsburgh, J. S., Aufdenkampe, A. K., Mayorga, E., Lehnert, K. A., Hsu, L., Song, L., ... Whitenack, T. (2016). Observations Data Model 2: A community information model for spatially discrete Earth observations. *Environmental Modelling & Software*, *79*: 55-74. http://dx.doi.org/10.1016/j.envsoft.2016.01.010

Horsburgh, J. S., Morsy, M. M., Castronova, A. M., Goodall, J. L., Gan, T., Yi, H., Stealey M. J. & Tarboton, D. G. (2016), HydroShare: Sharing Diverse Environmental Data Types and Models as Social Objects with Application to the Hydrology Domain, *JAWRA Journal of the American Water Resources Association*, 52(4): 873-889, http://dx.doi.org/10.1111/1752-1688.12363

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., & Arheimer, B. (2016). Most computational hydrology is not reproducible, so is it really science? *Water Resources Research*, *52*(10), 7548-7555. https://doi.org/10.1002/2016WR019285

Kalelioğlu, F. (2015). A new way of teaching programming skills to K-12 students: Code. org. *Computers in Human Behavior*, 52, 200-210. https://doi.org/10.1016/j.chb.2015.05.047

Kandlbinder, P. (2014). Constructive alignment in university teaching. *HERDSA Review of Higher Education*, *36*(3), 5-22. https://www.herdsa.org.au/herdsa-review-higher-education-vol-1/5-22

Kerski, J. (2019). A geomorphology field trip in Northwest Indiana story map. *ArcGIS Story Map*. Accessed at https://storymaps.arcgis.com/stories/79348b0fd4474d8c87819dc95ef36c1e

Kolodner, J. L., Owensby, J. N., & Guzdial, M. (2004). Case-based learning aids. *Handbook of research on educational communications and technology (2nd ed.),* 829-861.

Lane, B. A. & Garousi Nejad, I. (2019), Physical Hydrology, HydroLearn, https://edx.hydrolearn.org/courses/course-v1:Utah_State_University+CEE6400+2019_Fall/about

Litzinger, T., Lattuca, L. R., Hadgraft, R., & Newstetter, W. (2011). Engineering Education and the Development of Expertise. *Journal of Engineering Education*, *100*(1), 123–150. https://doi.org/10.1002/j.2168-9830.2011.tb00006.x

Lyon, S. W., & Teutschbein, C. (2011). Problem−based learning and assessment in hydrology courses: Can non−traditional assessment better reflect intended learning outcomes?. *Journal of Natural Resources and Life Sciences Education*, *40*(1), 199-205. https://doi.org/10.4195/jnrlse.2011.0016g

Marshall, J. A., Castillo, A. J., & Cardenas, M. B. (2013). Assessing student understanding of physical hydrology. *Hydrology & Earth System Sciences*, *17*(1), 829-836. https://doi.org/10.5194/hess-17-829-2013

Merck, M. F., Gallagher, M. A., Habib E., and Tarboton, D. (2021). Engineering Students' Perceptions of Mathematical Modeling in a Learning Module Centered on a Hydrologic Design Case Study. International Journal of Research in Undergraduate Mathematics Education, http://doi.org/10.1007/s40753-020-00131-8.

Merwade, V., & Ruddell, B. L. (2012). Moving university hydrology education forward with community-based geoinformatics, data and modeling resources. *Hydrology and Earth System Sciences*, *16*(8), 2393-2404. https://doi.org/10.5194/hess-16-2393-2012

Merwade, V., & Rajib, M. A. (2014). RWater–A Cyber Enabled Tool For Hydrologic Modeling And Analysis. Proceedings of the 11th International Conference on Hydroinformatics.

National Research Council. (1991). Opportunities in the Hydrologic Sciences. *The National Academies Press*. https://doi.org/10.17226/1543

Ngambeki, I., Thompson, S. E., Troch, P. A., Sivapalan, M., & Evangelou, D. (2012). Engaging the students of today and preparing the catchment hydrologists of tomorrow: student-centered approaches in hydrology education. *Hydrology & Earth System Sciences Discussions*, *9*(1), 707-740. https://doi.org/10.5194/hessd-9-707-2012

Rajib, M. A., Merwade, V., Kim, I. L., Zhao, L., Song, C., & Zhe, S. (2016). SWATShare – A web platform for collaborative research and education through online sharing, simulation and visualization of SWAT models. *Environmental Modelling & Software*, *75*, 498–512. https://doi.org/10.1016/j.envsoft.2015.10.032

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rosenberg, D. E., Filion, Y., Teasley, R., Sandoval-Solis, S., Hecht, J. S., Van Zyl, J. E., ... Tarboton, D. G. (2020). The next frontier: Making research more reproducible. *Journal of Water Resources Planning and Management*, *146*(6). https://doi.org/10.1061/(ASCE)WR.1943-5452.0001215

Ruddell, B. L., & Wagener, T. (2015). Grand challenges for hydrology education in the 21st century. *Journal of Hydrologic Engineering*, *20*(1). https://doi.org/10.1061/(ASCE)HE.1943-5584.0000956

Ryberg, K. R., & Vecchia, A. V. (2012). An R package for retrieval, analysis, and anomaly calculation of daily hydrologic time series data, version 1.0 (No. 2012–1168; Open-File Report). US Department of the Interior, US Geological Survey. https://doi.org/10.3133/ofr2012116

Sanchez, C. A., Ruddell, B. L., Schiesser, R., & Merwade, V. (2016). Enhancing the T-shaped learning profile when teaching hydrology using data, modeling, and visualization activities. *Hydrology and Earth System Sciences*, *20*(3), 1289-1299. https://doi.org/10.5194/hess-20-1289-2016

Slater, L. J., Thirel, G., Harrigan, S., Delaigue, O., Hurley, A., Khouakhi, A., ... Smith, K. (2019). Using R in hydrology: a review of recent developments and future directions. *Hydrology and Earth System Sciences*, *23*(7), 2939-2963. https://doi.org/10.5194/hess-23-2939-2019

Stagge, J. H., Rosenberg, D. E., Abdallah, A. M., Akbar, H., Attallah, N. A., & James, R. (2019). Assessing data availability and research reproducibility in hydrology and water resources. *Scientific data*, *6*(1). https://doi.org/10.1038/sdata.2019.30 .

Tarboton, D. G. (2003). Rainfall-runoff processes. Utah State University, 1(2).

Tarboton, D. G., R. Idaszak, J. S. Horsburgh, J. Heard, D. Ames, J. L. Goodall, L. Band, V. Merwade, A. Couch, J. Arrigo, R. Hooper, D. Valentine and D. Maidment, (2014), HydroShare: Advancing Collaboration through Hydrologic Data and Model Sharing, in D. P. Ames, N. W. T. Quinn and A. E. Rizzoli (eds), *Proceedings of the 7th International Congress on Environmental Modelling and Software*, San Diego, California, USA, International Environmental Modelling and Software Society (iEMSs), https://scholarsarchive.byu.edu/iemssconference/2014/Stream-A/7/.

Tarboton, D. G. (2018). Terrain Analysis Using Digital Elevation Models (TauDEM), Utah Water Research Laboratory, *Utah State University*, http://hydrology.usu.edu/taudem, accessed 10/5/2019.

Tesfa, T. K., Tarboton, D. G., Watson, D. W. , Schreuders, K. A. T. , Baker M. E., & Wallace, R. M. (2011). Extraction of hydrological proximity measures from DEMs using parallel processing. *Environmental Modelling & Software*, *26*(12): 1696-1709. http://dx.doi.org/10.1016/j.envsoft.2011.07.018.

Vogel, R. M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P. K., Hooper, R. P., & Matalas, N. C. (2015). Hydrology: The interdisciplinary science of water. *Water Resources Research*, *51*(6), 4409-4430. https://doi.org/10.1002/2015WR017049

Wagener, T., Weiler, M., McGlynn, B., Gooseff, M., Meixner, T., Marshall, L., ... McHale, M. (2007). Taking the pulse of hydrology education. *Hydrological Processes*, *21*(13), 1789-1792. https://doi.org/10.1002/hyp.6766

Wagener, T., Kelleher, C., Weiler, M., McGlynn, B., Gooseff, M., Marshall, L., ... & Zappe, S. (2012). It takes a community to raise a hydrologist: the Modular Curriculum for Hydrologic Advancement (MOCHA). *Hydrology and Earth System Sciences*, *16*(9), 3405-3418. https://doi.org/10.5194/hess-16-3405-2012

Ward, A., Herzog, S., Bales, J., Barnes, R., Ross, M., Jefferson, A., ... Maertens, J. (2020). Educational Resources for Hydrology & Water Resources. *HydroShare*, http://www.hydroshare.org/resource/148b1ce4e308427ebf58379d48a17b91

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1). https://doi.org/10.1038/sdata.2016.18