12-2019

Advancing Cyberinfrastructure for Collaborative Data Sharing and Modeling in Hydrology

Tian Gan
Utah State University

Recommended Citation

Gan, Tian, "Advancing Cyberinfrastructure for Collaborative Data Sharing and Modeling in Hydrology" (2019). *All Graduate Theses and Dissertations*. 7618.
https://digitalcommons.usu.edu/etd/7618

ADVANCING CYBERINFRASTRUCTURE FOR COLLABORATIVE DATA SHARING AND MODELING IN HYDROLOGY

by

Tian Gan

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Civil and Environmental Engineering

Approved:

David G. Tarboton, Ph.D. Major Professor	Curtis Dyreson, Ph.D. Committee Member
Jeffery S. Horsburgh, Ph.D. Committee Member	Bethany Neilson, Ph.D. Committee Member
David E. Rosenberg, Ph.D. Committee Member	Richard S. Inouye, Ph.D. Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2019

ABSTRACT

Advancing Cyberinfrastructure for Collaborative Data Sharing and Modeling in Hydrology

by

Tian Gan, Doctor of Philosophy

Utah State University, 2019

Major Professor: Dr. David G. Tarboton
Department: Civil and Environmental Engineering

Hydrologic research is increasingly data and computationally intensive, and often involves hydrologic model simulation and collaboration among researchers. With the development of cyberinfrastructure, researchers are able to improve the efficiency, impact, and effectiveness of their research by utilizing online data sharing and hydrologic modeling functionality. However, further efforts are still in need to improve the capability of cyberinfrastructure to serve the hydrologic science community. The goals of the research described in this dissertation were to use physically based snow modeling to improve operational water supply forecasts in the Colorado River Basin and to create new hydrologic information system functionality to address the challenges of utilizing cyberinfrastructure for hydrologic data sharing and modeling.

This dissertation first presents the evaluation of the Utah Energy Balance snowmelt model as an alternative to temperature index snowmelt modeling for water supply forecasts. Then it presents the design of the multidimensional space-time data sharing functionality of the HydroShare hydrologic information system. It also describes a web application developed to facilitate input preparation and model execution of the Utah Energy Balance snowmelt model and the storage of these results in HydroShare. The snowmelt model evaluation served as use cases to evaluate the cyberinfrastructure elements developed.

The comparison of snowmelt models showed that both physically based and temperature index models, when coupled to a runoff model and calibrated, provide reasonable basin snow and discharge simulations. However, the physically based model was able to better quantify evaporative water balance components and sensitivity to land cover change with fewer calibrated parameters, thus offering better transferability potential to remain valid for different climate and terrain conditions.

The contribution of the new hydrologic information system functionality presented is that it enables hydrologic researchers and water resources professionals to collaborate from initial data preparation to final data publication of multidimensional space-time data. Moreover, by integrating hydrologic modeling web services with the hydrologic information system we established web-based simulation functionality that improved hydrologic modeling research in terms of collaboration, computer platform independence, and reproducibility. In addition, the methods and technologies for cyberinfrastructure development in this research provide potential solutions for the challenges associated with the design and implementation of cyberinfrastructure for hydrologic data sharing and modeling.

(175 pages)

PUBLIC ABSTRACT

Advancing Cyberinfrastructure for Collaborative Data Sharing and Modeling in Hydrology

Tian Gan

Hydrologic research is increasingly data and computationally intensive, and often involves hydrologic model simulation and collaboration among researchers. With the development of cyberinfrastructure, researchers are able to improve the efficiency, impact, and effectiveness of their research by utilizing online data sharing and hydrologic modeling functionality. However, further efforts are still in need to improve the capability of cyberinfrastructure to serve the hydrologic science community. This dissertation first presents the evaluation of a physically based snowmelt model as an alternative to a temperature index model to improve operational water supply forecasts in the Colorado River Basin. Then it presents the design of the functionality to share multidimensional space-time data in the HydroShare hydrologic information system. It then describes a web application developed to facilitate input preparation and model execution of a snowmelt model and the storage of these results in HydroShare. The snowmelt model evaluation provided use cases to evaluate the cyberinfrastructure elements developed. This research explored a new approach to advance operational water supply forecasts and provided potential solutions for the challenges associated with the design and implementation of cyberinfrastructure for hydrologic data sharing and modeling.

ACKNOWLEDGMENTS

I quite appreciate my advisor, Dr. David Tarboton, for giving me the opportunity to study in the U.S. and work on great research projects. I am thankful for his patience, guidance, and generous funding support throughout the period of my study. His high academic standards and diligent work ethic made him a good role model for me to follow. I would also like to thank my committee members, Dr. Jeffery Horsburgh, Dr. Bethany Neilson, Dr. David Rosenberg, and Dr. Curtis Dyreson, for sharing their valuable research experiences and providing helpful suggestions.

I am grateful for those who were not part of my committee, but also contributed to help with the research in this dissertation. I want to acknowledge the kind support from Pabitra, Tseganeh, Nazmus, and all other lab friends, who were willing to help me with the problems in web app development and modeling research. I also want to thank the team members from other organizations, who collaborated in different projects to provide technical trainings or datasets to help accomplish my work.

Special thanks to my parents and my friends from the church, without their encouragements and sincere prayers, I would not have been able to reach the end of this long journey. It will be a precious and unforgettable experience in my life.

This work was supported by the National Science Foundation grants ACI-1148453 and OAC-1664061, and NASA award NNH15CM90C. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation and NASA.

Tian Gan

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Problem statement

In the western United States, snowmelt from mountainous areas is an important water source for regional streamflow, and snow models play an important role in predicting monthly to seasonal water supply for water resources management (Li et al., 2017). Current river forecasting system methods couple a temperature index snowmelt model with a rainfall runoff model to predict basin discharge conditions (Franz et al., 2008). The advantages of using a temperature index model are that it only requires climate forcing inputs of precipitation and temperature that are easy to obtain and process in real time for most places (Anderson, 2006), and that it has good model performance despite its simplicity (Hock, 2003). However, there are also limitations when applying a temperature index model for operational water supply forecasts. Model parameters are often not transferable among watersheds, and calibration for each watershed may require significant effort (Anderson, 2006). It is also questionable to use a temperature index model under the impact of climate change because of the high sensitivity of the model to temperature (Warscher et al., 2013) and reduced validity of calibrated parameters as the system changes from conditions used for calibration. Changes in seasonal water resources due to climate change have broad economic and ecologic impacts (Barnett et al., 2005; Sturm et al., 2017; Zierl and Bugmann, 2005), and it is necessary to advance the current method for water supply forecasting to address the challenges of future changing conditions and to provide reliable predictions to guide water resources management decision making.

Hydrologic models are essential tools to help provide reliable predictions for water supply forecasting. They are also applied in other research to address critical water issues to guide the formulation of water resources management strategies or as a tool of scientific inquiry (Dingman, 2008). However, modelers face a number of challenges. First, great effort is often required to discover and integrate heterogeneous and dispersed data from multiple sources to use as model inputs. Second, the need to install and configure advanced hydrologic models and their associated dependency code libraries may be difficult. Third, models may have a steep learning curve that needs to be overcome before they can start model simulations. Fourth, there is an increasing demand for modeling research to be curated and shared to enhance the reuse of data and models within the hydrologic science community and better enable reproducibility of the research (Archfield et al., 2015; Demir and Krajewski, 2013). Additionally, as collaboration among researchers from various disciplines and areas becomes a key factor to promote new research findings, an open platform for researchers to effectively communicate and collaborate becomes important.

Cyberinfrastructure development offers a new and promising approach to address these challenges in hydrologic research (Billah et al., 2016; Laniak et al., 2013; Wang, 2010). Generally, cyberinfrastructure consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks. Cyberinfrastructure is usually distributed beyond the scope of a single institution and is established to promote scientific research and education (Freeman et al., 2005; Stewart et al., 2010; Yang et al., 2010). Considerable effort has been put into cyberinfrastructure development and many people have benefitted from using or extending cyberinfrastructure for education or research in the field of hydrology (Conner et al., 2013; Hersh and Maidment, 2014; McEnery et al., 2013; Muste et al., 2012). For example, the CUAHSI Hydrologic Information System (HIS) is a cyberinfrastructure system for publishing environmental observations data (Horsburgh et al., 2009; Tarboton et al., 2009). This system is comprised of hydrologic databases, servers, and software for data publication, discovery, access, visualization, and analysis for time series data at stationary points. Researchers are able to easily discover and access the time series datasets for hydrologic modeling and data analysis. They could also adopt and adapt the technology to share their own time series datasets online. Another example is SWATShare (Rajib et al., 2016), which is a collaborative environment that provides the capability of publishing, sharing, discovering, and downloading of Soil and Water Assessment Tool (SWAT) models (Arnold et al., 1998). This cyberinfrastructure also supports model calibration with high performance computing (HPC) resources and visualization of model outputs.

In spite of these existing efforts, further improvements to cyberinfrastructure are still necessary to help address the challenges in hydrologic research. For example, there are many data sharing systems established to promote the community inputs from individual researchers or small research groups for data reuse and collaboration. However, because of the large diversity of hydrologic data types used in different models, one major issue is how to manage various datasets in different file types, formats, and semantics to facilitate data discovery, visualization, and analysis. Most existing systems support either a certain data type with advanced functionality (e.g., CUAHSI HIS) or multiple data types as generic file objects with basic functionality (e.g., Figshare). Therefore, the limitations of supported data types or advanced functionality inhibit the effectiveness of data sharing and reuse. For example, functions for metadata extraction, creation, and curation are still not available for some systems to enhance the reusability and discovery of shared datasets. Insufficient functionality to visualize or process different types of scientific datasets to help gain knowledge or insights from them is also a limitation for some data sharing platforms. For instance, while users can publish datasets with assigned digital object identifiers (DOI) in Figshare, they cannot enable subsetting or visualization to work with large datasets efficiently.

In terms of cyberinfrastructure to support hydrologic modeling, there are a number of model input preparation systems and web services. HydroTerre is a system developed to provide access to geospatial datasets for supporting physically-based numerical models (Leonard and Duffy, 2013). This system includes data workflows for web access to fundamental national datasets to run catchment models in the US. EcohydroLib (Miles, 2014) provides a series of Python scripts for ecohydrology data preparation workflows. The workflow scripts include tools for downloading and processing geospatial data from national data infrastructure or custom local datasets to prepare ecohydrology model inputs such as land cover data, digital elevation data, or vegetation leaf area index. RHESSys workflow (Miles, 2014) was an example built on EcohydroLib to support running the RHESSys model (Tague and Band, 2004) that simulates carbon, water, and nutrient fluxes. HydroDS is a system implemented to prepare model input for distributed hydrologic models (Gichamo, 2019). The HydroDS web services can process digital elevation data to delineate watersheds and create slope and aspect as terrain inputs. They can also process climate data such as precipitation, temperature, and wind speed as model inputs with required file format and content.

Despite that the existing model input preparation systems and web services have the potential to improve the work efficiency and support reproducible modeling research, a barrier still exists for those who may have limited programming skills to utilize them for modeling work. Moreover, these systems often do not provide a good data management mechanism, which makes it difficult for researchers to curate and share their model input/output with the hydrologic science community to improve research reproducibility and collaboration.

1.2 Objectives

The goals of this research were to use physically based snow modeling to improve operational water supply forecasts in the Colorado River Basin and to create new hydrologic information system functionality to address the challenges of utilizing cyberinfrastructure for hydrologic data sharing and modeling in the context of an advanced hydrologic information system, HydroShare. HydroShare (Tarboton et al., 2014, Horsburgh et al., 2015) was designed to expand the data types supported by CUAHSI HIS from time series to include types such as geographic raster data, geographic feature data, multidimensional space-time data, referenced time series (Sadler et al., 2015), model programs, and model instances (Morsy et al., 2017). It supports data discovery, access, publication, analysis, and visualization for different data types to facilitate the activities involved in the whole data life cycle. It also integrates social functionality to build up a collaborative environment for researchers to easily communicate and work around the shared datasets. Moreover, HydroShare provides a Representational State Transfer (REST) application programming interface (API) to help interact with other cyberinfrastructure systems, which makes information exchange possible among systems. For example, web apps hosted in other web servers and connected to HydroShare, such as web apps in HydroShare Tethys Apps portal (https://apps.hydroshare.org/), can use the API to retrieve shared datasets from HydroShare for visualization, analysis, or modeling, and can create new datasets from other sources and share them in HydroShare.

This research first presents the evaluation of the Utah Energy Balance model (Tarboton and Luce, 1996) as an alternative to temperature index snowmelt modeling for water supply forecasts. It then presents the functionality design and implementation in HydroShare to facilitate data management, sharing, and reuse of multidimensional space-time data, a widely used data type in hydrologic modeling. Finally, an approach to tackle the challenges associated with using web services as part of the modeling process is presented. Details related to each objective are presented below.

1.2.1 Objective 1: Evaluate temperature index and energy balance snow models to improve the operational water supply forecasts in the Colorado River Basin.

The current methodology for water supply forecasting at the Colorado Basin River Forecast Center (CBRFC) uses the SNOW-17 model (Anderson, 1973) to generate rain-plus-melt inputs to the Sacramento Soil Moisture Accounting (SAC-SMA) runoff model (Burnash et al., 1973). SNOW-17 is a temperature index model and, despite its simplicity, the use of its output as input to SAC-SMA produces generally good comparisons between simulated and observed discharges after calibration. However, the model parameters are often not transferable among watersheds, and much effort is needed to calibrate the model for new watersheds (Anderson, 2006). This is one limitation.

Additionally, the calibrated parameters such as the melting factor used to determine melting rate may have reduced validity if the system changes from conditions used for calibration, such as when watershed or climate conditions change.

In order to provide accurate predictions of seasonal water resources under the future changing conditions, energy balance modeling becomes a promising option to advance the current methodology. An energy balance model uses the energy and mass balance equations to simulate the physical process of snow accumulation and ablation. Because of its inherent physically based representation of processes, an energy balance model usually requires little model calibration and has potential to provide accurate forecasts under the impact of climate or land cover change.

Under this objective, we assessed and prototyped the application of an energy balance model for operational water supply forecasts in the Colorado River Basin. The SNOW-17 and the Utah Energy Balance (UEB) model were separately coupled with the SAC-SMA model for basin snowmelt and discharge simulation to evaluate the performance of the two snow models. Detailed research questions included:

1) Which snow model can provide better performance for snowmelt and discharge simulations in the watersheds?

2) What are the benefits or limitations of applying the energy balance model in the river forecasting system for the operational water supply forecasts?

1.2.2 Objective 2: Develop capability for multidimensional space-time data sharing in HydroShare to facilitate data management and reuse.

Hydrologic processes (e.g., snowmelt or rainfall-runoff) often involve physical phenomena, which are spatially and temporally variable (e.g., precipitation, temperature, and snow water equivalent). Modelers can utilize distributed hydrologic models to simulate detailed processes as a way to guide decisions for water resources management. These models often take multidimensional space-time data as input and/or output because they can represent the spatial and temporal variabilities of the physical phenomenon well. Many scientists and modelers create and use multidimensional space-time data in their work, and many wish to work collaboratively, share data with their colleagues, and publish the results of their work as research products. Therefore, it is essential to develop functionality in data sharing systems to help curate and share multidimensional space-time data to reuse these scientific results for hydrologic research.

Multidimensional space-time data is often stored and distributed in file formats such as Common Data Format (CDF http://cdf.gsfc.nasa.gov/), Hierarchical Data Form (HDF https://www.hdfgroup.org/), and Network Common Data Form (NetCDF http://www.unidata.ucar.edu/software/netcdf/). However, current data sharing systems have limitations that inhibit the management or reuse of these types of datasets. First, it is difficult to edit or extract the metadata within current file formats in many data sharing systems (e.g., Google Drive and Dropbox), so that they are often poorly described. Second, existing tools for visualization and analysis of this data type may require complicated software and/or server installation and configuration that is beyond the reach of many scientists who want to enable simple visualization and analysis for their shared data, but do not want to host a server or install software.

Thus, under this objective we developed new capabilities in HydroShare for storing and managing multidimensional space-time data to facilitate data sharing, visualization, and analysis. Detailed research questions included:

1) Which file format is most suitable to enable storage and management of multidimensional space-time data in a hydrologic information system?

2) What functionality can help enhance metadata capture to make metadata more visible and easier to edit and manage?

3) What functionality is needed to better facilitate data visualization or analysis?

The work under this objective also investigated available technology for implementation of this functionality.

1.2.3 Objective 3: Integrate hydrologic modeling web services with HydroShare to improve reproducibility of hydrologic modeling research.

With the development of hydrologic modeling web services, modelers can utilize them to simplify the process of model input preparation and/or model simulation, which saves time and energy and helps focus more on data analysis and interpreting results. Nonetheless, barriers still exist for people to utilize them, especially for those without advanced programming skills or knowledge of web services. In addition, systems that host modeling web services (Billah et al., 2016; Gichamo, 2019; Leonard and Duffy, 2013) often do not provide data curation and sharing functionality to help access the data, metadata, and the scripts to repeat or modify the modeling work for validation or deriving new results. This impedes the ability for the hydrologic science community to access and reproduce the work for collaboration.

Under this objective, we integrated hydrologic modeling web services with a data sharing system to resolve the limitations mentioned above, simplify the modeling process, and enhance the reproducibility of hydrologic research. As a case study, we integrated HydroShare with HydroDS, a system providing web services (https://github.com/CI-WATER/Hydro-DS) for input preparation and model simulation of the UEB snowmelt model. Specific research questions included:

1) How can a hydrologic information system provide easy access to hydrologic modeling web services?

2) How can hydrologic information system functionality be utilized to support data curation and to repeat or modify the modeling work created from the hydrologic modeling web services?

1.3 Chapter organization

Each of the above objectives is addressed within one chapter of this dissertation as follows.

Chapter 2 addresses the first objective and presents the evaluation of the model performance between the SNOW-17 and the UEB model. It first introduces the study sites located in the Colorado River Basin and the input datasets for model simulation as well as the observation data for model evaluation. Second, it describes the model calibration method and the multiple performance metrics to assess the snowmelt and discharge simulation results. Furthermore, simulated evaporative components of sublimation and evapotranspiration (ET) from snow and runoff models are also compared. Finally, advantages and challenges associated with the application of an energy balance model for operational water supply forecasts are discussed.

Chapter 3 addresses the second objective and presents the design and implementation of the functionality in HydroShare for multidimensional space-time data sharing. It first details the selection of the file format to store and organize multidimensional space-time data and introduces the design of metadata elements for describing this data type. Then, it presents development of a web application based on the file format and metadata design to manage the datasets in HydroShare. Thirdly, it describes the function implementations to facilitate metadata management, data subset, processing, and visualization. Finally, a case study is introduced to evaluate the data sharing functionality. The snow output datasets created from the first objective were organized and shared in HydroShare to support data curation and reuse.

Chapter 4 addresses the third objective and describes an approach to integrate a data sharing system with a system that hosts modeling web services to support hydrologic modeling research. This approach uses a three-layer architecture design to integrate the two systems. It describes a case study that uses HydroShare and HydroDS as an example for implementing this approach and tests the developed functionality for snowmelt modeling under the context of the first objective. Then, it provides the results for implementation details and functionality evaluation. Finally, it discusses the benefits of system integration to support hydrologic modeling research and summarizes the lessons learnt from the work.

1.4 Contribution

This research was driven by the need for advancing the methods used in operational water supply forecasts to adapt to changing future conditions (climate and watershed) and overcome common limitations identified from the application of cyberinfrastructure in hydrologic research.

The first objective evaluates the value of incorporating a more complex snow model within the river forecasting system used operationally by the CBRFC to facilitate water resources management in the Colorado River Basin. The analysis of retrospective model simulations in this research demonstrate the potential of applying the UEB model for operational water supply forecasts in the snow-dominated river basins in the western United States or other places with similar climate and terrain conditions.

The main contribution from the second objective is an approach to support the sharing and reuse of multidimensional space-time data in a system to help hydrologic researchers or water resources professionals collaborate from initial data preparation to final data publication. This approach organizes the datasets in a widely-used, standard file format to support data interoperability and enables users to extract and edit the metadata in the file to better support data annotation and discovery. It also automates the setup of standard data services to support data visualization and analysis without requiring data providers to establish and maintain the data services by themselves.

The third objective provides an approach to reuse different open source cyberinfrastructure to support web-based hydrologic simulation, which benefits hydrologic modeling in terms of enhancing opportunities for collaboration, promoting computer platform independence, and encouraging and facilitating greater reproducibility. It simplifies the use of hydrologic modeling web services to reach a broader community of users. It also expands the capabilities of the modeling web services by using the data sharing functionality from the HydroShare hydrologic information system. Through this work, users are enabled to create, curate, share, discover, access, repeat, or modify modeling work in an online-environment without using local storage and computing resources.

References

Anderson, E.A., 2006. Snow accumulation and ablation model – SNOW-17, NWSRFS user manual documentation. National Weather Service River Forecast System. https://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/part2/_pdf/22snow17.pdf

Anderson, E.A., 1973. National Weather Service river forecast system: snow accumulation and ablation model, NOAA technical memorandum: NWS HYDRO-17. US National Weather Service.

Archfield, S.A., Clark, M., Arheimer, B., Hay, L.E., McMillan, H., Kiang, J.E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W.H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., Over, T., 2015. Accelerating advances in continental domain hydrologic modeling. Water Resour. Res. 51, 10078–10091. https://doi.org/10.1002/2015WR017498

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development. J. Am. Water Resour. Assoc. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Barnett, T.P., Adam, J.C., Lettenmaier, D.P., 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. Nature 438, 303–309. https://doi.org/10.1038/nature04141

Billah, M.M., Goodall, J.L., Narayan, U., Essawy, B.T., Lakshmi, V., Rajasekar, A., Moore, R.W., 2016. Using a data grid to automate data preparation pipelines required for regional-scale hydrologic modeling. Environ. Model. Softw. 78, 31–39. https://doi.org/10.1016/j.envsoft.2015.12.010

Burnash, R.J.C., Ferral, R.L., McGuire, R.A., 1973. A generalized streamflow simulation system : conceptual modeling for digital computers. U.S. Department of Commerce National Weather Service and State of California Department of Water Resources.

Conner, L.G., Ames, D.P., Gill, R.A., 2013. HydroServer Lite as an open source solution for archiving and sharing environmental data for independent university labs. Ecol. Inform. 18, 171–177. http://dx.doi.org/10.1016/j.ecoinf.2013.08.006

Demir, I., Krajewski, W.F., 2013. Towards an integrated Flood Information System: Centralized data access, analysis, and visualization. Environ. Model. Softw. 50, 77–84. https://doi.org/10.1016/j.envsoft.2013.08.009

Dingman, S.L., 2008. Physcial Hydrology (second edition). Waveland Press, Inc, Long Grove, IL.

Franz, K.J., Hogue, T.S., Sorooshian, S., 2008. Operational snow modeling: Addressing the challenges of an energy balance model for National Weather Service forecasts. J. Hydrol. 360, 48–66. https://doi.org/10.1016/j.jhydrol.2008.07.013

Freeman, P.A., Crawford, D.L., Kim, S., Munoz, J.L., 2005. Cyberinfrastructure for science and engineering: Promises and challenges. Proc. IEEE 93, 682–690. https://doi.org/10.1109/JPROC.2004.842782

Gichamo, T.Z., 2019. Advancing streamflow forecasts through the application of a physically based energy balance snowmelt model with data assimilation and cyberinfrastructure resources (Doctoral dissertation). Utah State University, Logan, UT. https://digitalcommons.usu.edu/etd/7463

Hersh, E.S., Maidment, D.R., 2014. Extending hydrologic information systems to accommodate arctic marine observations data. Deep. Res. Part II Top. Stud. Oceanogr. 102, 9–17. https://doi.org/10.1016/j.dsr2.2013.07.014

Hock, R., 2003. Temperature index melt modelling in mountain areas. J. Hydrol. 282, 104–115. https://doi.org/10.1016/S0022-1694(03)00257-9

Horsburgh, J.S., Morsy, M.M., Castronova, A.M., Goodall, J.L., Gan, T., Yi, H., Stealey, M.J., Tarboton, D.G., 2015. HydroShare: Sharing diverse hydrologic data types and models as social objects within a hydrologic information system,. J. Am. Water Resour. Assoc. 27517. https://doi.org/10.1111/1752-1688.12363

Horsburgh, J.S., Tarboton, D.G., Piasecki, M., Maidment, D.R., Zaslavsky, I., Valentine, D., Whitenack, T., 2009. An integrated system for publishing environmental observations data. Environ. Model. Softw. 24, 879–888. https://doi.org/10.1016/j.envsoft.2009.01.002

Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M., Peckham, S., Reaney, S., Gaber, N., Kennedy, R., Hughes, A., 2013. Integrated environmental modeling: A vision and roadmap for the future. Environ. Model. Softw. 39, 3–23. https://doi.org/10.1016/j.envsoft.2012.09.006

Leonard, L., Duffy, C.J., 2013. Essential terrestrial variable data workflows for distributed water resources modeling. Environ. Model. Softw. 50, 85–96. https://doi.org/10.1016/j.envsoft.2013.09.003

Li, D., Wrzesien, M.L., Durand, M., Adam, J., Lettenmaier, D.P., 2017. How much runoff originates as snow in the western United States, and how will that change in the future? Geophys. Res. Lett. 44, 6163–6172. https://doi.org/10.1002/2017GL073551

McEnery, J.A., McKee, P.W., Shelton, G.P., Ramsey, R.W., 2013. Hydrologic information server for benchmark precipitation dataset. Comput. Geosci. 50, 145–153. https://doi.org/10.1016/j.cageo.2012.08.005

Miles, B.C., 2014. Small-scale residential stormwater management in urbanized watersheds: A geoinformatics-driven ecohydrology modeling approach (Doctoral dissertation). The University of North Carolina, Chappel Hill, NC. https://doi.org/10.17615/1m6y-vs10

Morsy, M.M., Goodall, J.L., Castronova, A.M., Dash, P., Merwade, V., Sadler, J.M., Rajib, M.A., Horsburgh, J.S., Tarboton, D.G., 2017. Design of a metadata framework for environmental models with an example hydrologic application in HydroShare. Environ. Model. Softw. 93, 13–28. https://doi.org/10.1016/j.envsoft.2017.02.028

Muste, M., Bennett, D., Secchi, S., Schnoor, J., Kusiak, A., Arnold, N., Mishra, S., Ding, D., Rapolu, U., 2012. End-to-End cyberinfrastructure for decision-making support in watershed management. J. Water Resour. Plan. Manag. 139, 565–573. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000289

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. Environ. Model. Softw. 75, 498–512. https://doi.org/10.1016/j.envsoft.2015.10.032

Sadler, J.M., Ames, D.P., Livingston, S.J., 2015. Extending HydroShare to enable hydrologic time series data as social media. J. Hydroinformatics 18, jh2015331. https://doi.org/10.2166/hydro.2015.331

Stewart, C.A., Simms, S., Plale, B., Link, M., Hancock, D.Y., Fox, G.C., 2010. What is cyberinfrastructure. Proceedings of SIGUCCS 2010, Norfolk, VA, pp. 37–44.

Sturm, M., Goldstein, M.A., Parr, C., 2017. Water and life from snow: A trillion dollar science question. Water Resour. Res. 53, 3534–3544. https://doi.org/10.1002/2017WR020840

Tague, C.L., Band, L.E., 2004. RHESSys: Regional hydro-ecologic simulation system—An object-oriented approach to spatially distributed modeling of carbon, water, and nutrient cycling. Earth Interact. 8, 1–42. https://doi.org/10.1175/1087-3562(2004)8<1:RRHSSO>2.0.CO;2

Tarboton, D.G., Horsburgh, J.S., Maidment, D.R., Whiteaker, T., Zaslavsky, I., Piasecki, M., 2009. Development of a community hydrologic information system. 18th World IMACS / MODSIM Congr. 988–994.

Tarboton, D.G., Idaszak, R., Horsburgh, J.S., Heard, J., Ames, D., Goodall, J.L., Band, L., Merwade, V., Couch, A., Arrigo, J., Hooper, R., Valentine, D., Maidment, D., 2014. HydroShare: advancing collaboration through hydrologic data and model sharing, in: 7th International Congress on Environmental Modelling and Software. International Environmental Modelling and Software Society (iEMSs).

Tarboton, D.G., Luce, C.H., 1996. "Utah Energy Balance Snow Accumulation and Melt Model (UEB)," Computer model technical description and users guide. Utah Water Research Laboratory and USDA Forest Service Intermountain Research Station.

Wang, S., 2010. A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. Ann. Assoc. Am. Geogr. 100, 535–557. https://doi.org/10.1080/00045601003791243

Warscher, M., Strasser, U., Kraller, G., Marke, T., Franz, H., Kunstmann, H., 2013. Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps. Water Resour. Res. 49, 2619–2637. https://doi.org/10.1002/wrcr.20219

Yang, C., Raskin, R., Goodchild, M., Gahegan, M., 2010. Geospatial Cyberinfrastructure: Past, present and future. Comput. Environ. Urban Syst. 34, 264–277. https://doi.org/10.1016/j.compenvurbsys.2010.04.001

Zierl, B., Bugmann, H., 2005. Global change impacts on hydrological processes in Alpine catchments. Water Resour. Res. 41, 1–13. https://doi.org/10.1029/2004WR003447

CHAPTER 2

EVALUATION OF TEMPERATURE INDEX AND ENERGY BALANCE SNOW MODELS FOR HYDROLOGICAL APPLICATIONS IN OPERATIONAL WATER SUPPLY FORECASTS [1]

Abstract

In the western United States, snow accumulation, storage, and ablation affect seasonal runoff. Thus, the prediction of snowmelt is essential to improve the reliability of water supply forecasts to guide water allocation and operational decisions. The current method used at the Colorado Basin River Forecast Center (CBRFC) couples the SNOW-17 temperature index snow model and the Sacramento Soil Moisture Accounting (SAC-SMA) runoff model in a lumped approach. Limitations in the transferability and calibration requirements for changing conditions with the temperature index model motivated this research where new avenues were investigated to assess and prototype the application of an energy balance snow model in a distributed modeling approach. The Utah Energy Balance (UEB) model was chosen to compare with the SNOW-17 model because it is simple and parsimonious making it suitable for distributed application with the potential to improve water supply forecasts. Each model was coupled with the SAC-SMA model and the Rutpix7 routing model to simulate basin snowmelt and discharge. All of the models were applied on grids over watersheds using the Research Distributed Hydrologic Model (RDHM) framework. Case studies were implemented for two study sites in the Colorado River Basin over a period of two decades. The model performance

---

[1] Coauthored by Tian Gan, David G. Tarboton, and Tseganeh Z. Gichamo.

was evaluated by comparing the model output with observed daily discharge and snow covered area data obtained from remote sensing sources. Simulated evaporative components of sublimation and evapotranspiration were also evaluated. Results showed similar model performance for both UEB and SNOW-17 after calibration, and both provided reasonable basin snow and discharge simulations in the two study sites. However, the UEB model has the advantage of being able to explicitly simulate sublimation for different land types and thus better quantify evaporative water balance components and their sensitivity to land cover change. It also has better transferability potential because it requires calibration of fewer parameters than SNOW-17. In UEB the majority of the parameters are physically based and regarded as constants characterizing spatially invariant properties of snow processes. Thus, the model remains valid for different climate and terrain conditions for multiple watersheds.

Keywords: snow modeling, operational water supply forecasts, SNOW-17 model, Utah Energy Balance model

2.1 Introduction

Snowmelt from mountainous areas is an important water source for regional streamflow in the western United States, and snow models play an important role in predicting monthly to seasonal water supply for water resources management (Li et al., 2017). The National Weather Service (NWS) Colorado Basin River Forecast Center (CBRFC) is responsible for basin wide seasonal water supply forecasts for several major watersheds in the western United States. Currently, the CBRFC produces water supply forecasts using the SNOW-17 snow model (Anderson, 1973) to generate inputs to the

Sacramento Soil Moisture Accounting (SAC-SMA) runoff model (Burnash et al., 1973). This forecasting method uses a lumped approach where the two models (SNOW-17+SAC-SMA) are applied over basins, with variability within basins represented through elevation zones.

SNOW-17 is a temperature index model that uses air temperature and precipitation as the model inputs to simulate snow accumulation and ablation. Using a temperature index model for operational water supply forecasts has the following advantages: (1) the climate forcing inputs are easy to obtain and process in real time for most places (Anderson, 2006), and (2) it has good model performance (e.g., good fit between observed and simulated discharge) despite its simplicity (Hock, 2003). However, the model parameters are often not transferable among watersheds, and calibration for each watershed may require significant effort (Anderson, 2006). It is also questionable to use a temperature index model under the impact of climate change because of the high sensitivity of the model to temperature and reduced validity of calibrated parameters as the system changes from conditions used for calibration.

Changes in seasonal water resources due to climate change have broad economic and ecologic impacts (Barnett et al., 2005; Sturm et al., 2017; Zierl and Bugmann, 2005), thus highlighting the importance of advancing the current method for forecasting water supply to address the challenges of future changing conditions and to guide water resources management decision making. The increased availability of meteorological data such as wind speed, vapor pressure, and solar radiation makes using an energy balance model a promising option for operational water supply forecasts. The advantages of an energy balance model are that it, in theory, requires less model calibration and has the potential to provide forecasts that account for climate or land cover change (Zeinivand and De Smedt, 2009). However, using an energy balance model may require more data and involve advanced computation that places high demand on computing resources, and model performance relies on the availability and quality of the additional needed climate input data.

Prior work comparing temperature index and energy balance models has not been conclusive as to whether one approach is better than the other (Essery et al., 2013; Magnusson et al., 2015; Shakoor et al., 2018). Franz et al. (2008) compared the Snow-Atmosphere-Soil Transfer (SAST) energy balance model with the SNOW-17 model to simulate basin streamflow by coupling them with the SAC-SMA model. They found that, although simulations of snowpack and streamflow from the two models were similar, the SNOW-17 model performed consistently well in general and in some years better than the SAST model. Debele et al. (2010) compared energy balance and temperature index models within the Soil and Water Assessment Tool (SWAT) model. They compared the runoff simulation results and found only insignificant differences between the two approaches, noting that, for practical application, the temperature index model can be utilized when net solar radiation rather than turbulent heat flux dominates the snowmelt process. Kumar et al. (2013) compared the Isnobal energy balance model with a temperature index model for snowmelt and streamflow simulation by linking them with the Penn State Integrated Hydrology Model (PIHM). Their results showed that both the Isnobal model and the calibrated temperature index model could provide reasonable streamflow results. Isnobal had the best accuracy, whereas the temperature index model without calibration had the poorest results. Thus, it is apparent that model complexity is not a determinant of the reliability of snow or runoff simulation results. Calibrated temperature index models may produce similar or better results, and the uncertainty of climate input data is a major factor affecting the performance of the energy balance models. Therefore, it is important to compare the model performance from different snow models before applying them in various contexts.

The purpose of this research was to assess and prototype the application of an energy balance model for operational water supply forecasts. The requirements for a snowmelt model to support operational water supply forecasts include not only model performance, but also computation time and input data availability. We separately coupled the SNOW-17 model and the Utah Energy Balance model (Tarboton and Luce, 1996) with the SAC-SMA runoff model and the Rutpix7 routing model (NWS, 2008a) and simulated basin snow and discharge to evaluate model performance (SNOW-17+SAC-SMA+Rutpix7, UEB+SAC-SMA+Rutpix7). We also adopted a distributed modeling approach that applied the models on grids over watersheds. This approach provides more accurate representation of the spatial distribution of the snowmelt process and leads to improved forecasts. We used the Research Distributed Hydrologic Model (RDHM) framework (NWS, 2008a) to support this approach. This framework consists of multiple modules to simulate hydrologic processes such as snowmelt, rainfall runoff, and routing. Individual modules are called from within the RDHM framework and new modules can also be developed and added into this framework.

To evaluate model performance, we applied the approach to study watersheds in the Colorado River Basin, USA. We evaluated the spatial distribution of the snowmelt simulation by comparing the simulated snow water equivalent (SWE) with the snow covered area (SCA) data from the MODIS Snow-Covered Area and Grain size (MODSCAG) product (Painter et al., 2009). We also evaluated the seasonal runoff simulation by selecting different evaluation metrics to compare the observed and simulated basin discharge. In addition, we compared the model outputs of sublimation and evapotranspiration from the snow and runoff models to discover the differences in simulating the evaporative components between the two model configurations (SNOW-17+SAC-SMA+Rutpix7, UEB+SAC-SMA+Rutpix7).

This research is an initial investigation into the feasibility of incorporating a more complex snow model within the CBRFC river forecasting system for use in water supply forecasts. The model simulation in the RDHM framework is also a first step exploration of a transition to operational distributed modeling at the CBRFC. Moreover, the approach used in this research shows the potential of applying the UEB model in other snow-dominated river basins for water supply forecasts in the western United States or other locations with similar conditions.

This paper is organized as follows. Section 2 describes the study area and research data. It then presents the model description, model calibration, and evaluation metrics. Section 3 provides the evaluation results and corresponding discussion. Finally, Section 4 summarizes the work and discusses the advantages and challenges associated with the application of an energy balance model for operational water supply forecasts.

2.2 Methods

2.2.1 Study sites and data

The study sites are within the Colorado River Basin and include watersheds of the Dolores River above McPhee reservoir and the Blue River above Dillon reservoir (referred to as the Dolores River watershed and the Blue River watershed in the following sections) (Figure 2.1 and Figure 2.2). Each site consists of head subwatersheds and local subwatersheds, with the details listed in (Table 2.1). The average elevation of the Blue River watershed (3347 m) is higher than that of the Dolores River watershed (2786.54m), whereas its total area (849.3 km^2) is much smaller than that of the Dolores River watershed (2080.1 km^2). We chose these two study sites because (1) they represented different terrain and climate conditions, and (2) they were high priority watersheds in the NASA applications project in collaboration with RTI International (https://www.rti.org). This work was part of that effort to improve water supply forecasts for the CBRFC watersheds.

We retrieved and processed data both from static datasets (e.g., topographic data and canopy cover data) and dynamic datasets (e.g., meteorological data) to prepare the model inputs. Precipitation and temperature are important model forcing inputs for both snow models. We utilized historical gridded precipitation and temperature datasets from the CBRFC. These 3-hour time step, 800-m resolution datasets were created using the Mountain Mapper algorithm based on quality controlled climate station data (Schaake et al., 2004). We also used the CBRFC temperature data to derive the daily maximum and minimum temperature as inputs to the UEB model for radiation flux calculation. Wind speed and vapor pressure for the UEB model were prepared using gridded data from the NLDAS-2 land surface forcing dataset with 1/8th degree (around 13 km) grid spacing and hourly time step (Xia et al., 2012). For this approach to be used operationally, methods to incorporate wind speed and vapor pressure forecasts from an operational weather model driving the predictions would need to be developed.

Static slope and aspect inputs for the UEB model were created using the 30-m National Elevation dataset (NED) (Gesch, 2007), and canopy coverage fraction, canopy height, and leaf area index inputs for the UEB model were prepared using the 30-m National Land Cover Dataset (NLCD) (Homer et al., 2015).

The RDHM framework uses the Hydrologic Rainfall Analysis Project (HRAP) grid system (Reed and Maidment, 1999). We defined the model resolution as 0.25 HRAP (around 1.2 km) and the model simulation time step as 6 hours for consistency with our RTI International collaborators. This choice was based on trading off computational considerations with explicit spatial detail. However, as UEB is a point model most meaningfully applied with a spatial footprint around 30 m (Tarboton et al., 2000), we applied UEB at grid cell centers within the 0.25 HRAP grid and with slope, aspect, and vegetation calculated from their respective 30-m scale datasets. This approach prevents the smoothing of the terrain that would occur if 1200-m grid cells were used, but does not represent the variability of slope and aspect within any one grid cell. Rather, the assumption is that, aggregated over the watershed, these center points are sufficiently representative. Dynamic forcing data for the 1988–2010 time span was resampled from its 800-m (temperature and precipitation) or 1/8-degree (humidity and wind) resolution by selecting the value for the grid cell as the value where the 0.25 HRAP grid cell centers falls.

The observed datasets used for performance evaluation included daily discharge and remotely sensed snow covered area (SCA) data. Daily historical natural discharge for 1988-2010 was obtained from the CBRFC values produced by adjusting the USGS streamflow using diversion and reservoir data to calculate historical natural flows without the impacts of regulation. The MODIS Snow Covered-Area and Grain size retrieval algorithm (MODSCAG) daily SCA data for 2000–2010 at 500-m resolution were used to evaluate the model snow outputs.

2.2.2 Models description

We compared two model configurations for simulating snow and basin discharge, each of which coupled a snowmelt model with the SAC-SMA runoff model and the Rutpix7 routing model. The first configuration used the SNOW-17 temperature index model, represented as SNOW-17+SAC-SMA+Rutpix7. The second used the UEB energy balance model, represented as UEB+SAC-SMA+Rutpix7. The RDHM framework was used to support each of these model configurations. SNOW-17, SAC-SMA, and Rutpix7 models were already part of RDHM, whereas the UEB model was added to the framework as a new module in this research. This took take advantage of the extensibility that RDHM provides for the addition of module code files configured following the developer's instructions (NWS, 2008b). Descriptions of the two model configurations are provided in the following subsections.

2.2.2.1 Utah Energy Balance model (UEB)

The UEB model is a physically based model for snow accumulation and melt developed to predict snowmelt rates that contribute to stream and river flows during the spring and summer. This model uses a single layer representation of the snowpack and a modified force-restore approach (Luce and Tarboton, 2001, 2010) that allows the snow surface temperature to be different from the snow average temperature. This design avoids modeling the complex processes within a snowpack and provides a parsimonious model with a small number of state variables that is applicable over a spatial grid with no or minimal calibration at different locations. In addition, the UEB model's vegetation component enhances its ability to model energy and mass balance processes in forested areas (Mahat et al., 2013; Mahat and Tarboton, 2012, 2014). The vegetation component estimates the transmission and attenuation of radiation through a forest canopy, precipitation interception and unloading, snowmelt and sublimation of intercepted snow, and turbulent energy exchanges between the ground surface, canopy, and atmosphere.

The UEB model inputs include air temperature, precipitation, wind speed, relative humidity, incoming solar radiation, and longwave radiation at a time step sufficient to resolve the diurnal cycle (e.g., hourly, three hourly, and six hourly). When the radiation inputs are not available, air temperature and the daily temperature range can be used to estimate them. Slope and aspect terrain conditions and canopy properties such as leaf area index, canopy height, and canopy cover are also required.

In the UEB model, the two major state variables of energy content, U, and water equivalence, W, are determined at each time step using the inputs mentioned above and the following energy and mass balance equations.

$$\frac{dU}{dt} = Q_{sn} + Q_{li} + Q_p + Q_g - Q_{le} + Q_h + Q_e - Q_m \qquad (1)$$

$$\frac{dW}{dt} = P_r + P_s - M_r - E \qquad (2)$$

In the energy balance equation, the state variable U is energy per unit of horizontal area (kJ m^{-2}). The flux terms are Q_{sn}, net shortwave radiation; Q_{li}, incoming longwave radiation; Q_p, advected heat from precipitation; Q_g, ground heat flux; Q_{le}, outgoing longwave radiation; Q_h, sensible heat flux; Q_e, latent heat flux due to sublimation/condensation; and Q_m, advected heat removed by meltwater, all of which are in units of energy per unit of horizontal area, per unit time (kJ $m^{-2}hr^{-1}$). In the mass balance equation, the state variable W is snow water equivalent (m). The flux terms are P_r, rainfall rate; P_s, snowfall rate; M_r, meltwater outflow from the snowpack; and E, sublimation from the snowpack, all in m/hr of water equivalent. Readers are referred to Mahat et al. (2013) and Mahat and Tarboton (2012, 2014) for details on how each process is modeled.

2.2.2.2 SNOW-17 model

The SNOW-17 model is a conceptual model that uses precipitation as the water input and air temperature as the index to determine the energy exchange across the snow-air interface. This model is mainly used for river forecasting and requires calibration of melt factors to generate reliable simulation results (Anderson, 2006).

The SNOW-17 model calculates snow surface melt differently depending on whether rain is present or not. For rain on snow, the model computes the surface melt based on the following equation (Anderson, 2006):

$$\mathrm{M} = \sigma \cdot \Delta\mathrm{t} \cdot [(\mathrm{T_a} + 273.15)^4 - 273.15^4] + 0.0125 \cdot \mathrm{P} \cdot \mathrm{f_r} \cdot \mathrm{T}_r + 8.5 \cdot \mathrm{UADJ} \cdot \Delta\mathrm{t} \cdot [(\mathrm{e_a} - 6.11) + 0.00057 \cdot \mathrm{P_a} \cdot \mathrm{T}_a] \qquad (3)$$

where M is the depth of melt (mm); σ is the Stefan-Boltzmann constant; Δt is the time interval (hours); T_a is air temperature (°C); P is the water equivalent of precipitation (mm); f_r is fraction of precipitation in the form of rain; T_r is rain temperature (°C); UADJ is the average wind function during rain-on-snow events ($mm \cdot mb^{-1} \cdot hr^{-1}$); e_a is vapor pressure of air (mb); P_a is atmospheric pressure (mb). This calculation is based on energy balance concepts but neglects solar radiation, assuming that the sky overcast. The first term represents longwave radiation, the second represents melt by rain, and the third represents melt by sensible and latent heat.

When there is no rain and the air temperature is above the base value, the SNOW-17 model uses a melt factor to calculate the snowmelt as follows:

$$\mathrm{M} = M_f \cdot (T_a - T_b) \qquad (4)$$

where M is the depth of melt (mm); M_f is a seasonally varying melt factor (mm/°C); T_a is air temperature; T_b is the base temperature above which melt starts (usually 0 °C). To represent the seasonal variation of the melt factor, M_f is calculated from a sinusoidal curve with maximum (MFMAX) and minimum (MFMIN) melt factor values as model parameters (Anderson, 2006).

The SNOW-17 model uses heat deficit to keep track of the net heat loss from the snow cover under conditions of no surface melt (Anderson, 2006). When the air temperature is below freezing, the snow cover can be losing or gaining heat depending on the thermal gradient in the upper layers of the snowpack. This gradient is estimated as the difference between the snow surface temperature T_{sur} and the temperature at some distance within the snowpack computed as the antecedent temperature index (ATI). When T_{sur} is less than ATI, the heat deficit is increasing; otherwise it is decreasing. When the heat deficit is zero and the amount of liquid water held in the pack equals the holding capacity, the snow cover is ripe and the excess liquid water will become the outflow. This is calculated using empirically derived equations to represent the lag and attenuation of water through the snow cover. Note that, unlike the UEB model, the SNOW-17 model does not have any representation of snow sublimation, and all snow water equivalent losses from SNOW-17 become snowmelt inputs to the SAC-SMA model of surface hydrology and runoff generation processes. For full details refer to Anderson (2006).

2.2.2.3 Sacramento Soil Moisture Accounting model (SAC-SMA)

The SAC-SMA model is a two-layer conceptual rainfall-runoff model (NWS, 2006). This model parameterizes the soil characteristics that are responsible for streamflow production and represents soil moisture storage, percolation, drainage, and evapotranspiration (ET) processes in a conceptual way. It uses rain-plus-melt data as its input, which can be obtained from the output of snow models such as the SNOW-17 or the UEB model, but it requires calibration of parameters quantifying processes such as soil water storage and percolation rate to produce runoff simulations.

The SAC-SMA model estimates evapotranspiration (ET) using the available tension water volume and potential evaporation (PE) demand. When ET occurs, the moisture is withdrawn from the upper and lower zone tension water. The PE demand is estimated using PE grids and PE adjustment factors. Twelve mean monthly PE grids are available for the model, and PE adjustment factors are used to account for the effects of vegetation.

2.2.2.4 Rutpix7

Rutpix7 is a hillslope and channel routing model (NWS, 2008a). Inputs to the Rutpix7 model include fast (surface) and slow (subsurface/ground) runoff from the SAC-SMA model. In each cell, fast runoff is routed over a conceptual hillslope to a channel. Then the channel inflow from the hillslopes, the slow runoff, and the upstream pixel outflows are routed through a cell conceptual channel, after which a topographically defined cell-to-cell connectivity sequence is used to move water from upstream to downstream. See Koren et al. (2004) for details.

2.2.3 Model calibration

We used code obtained from RTI International to automatically calibrate parameters for the snowmelt and rainfall runoff models to minimize the difference between simulated and observed discharge. The code that RTI International provided implemented the Nondominated Sorting-based Multiobjective Genetic Algorithm II (NSGA-II) (Deb et al., 2002). Three fitness functions were used in the algorithm: (1) Kling-Gupta efficiency (Gupta et al., 2009) based on the difference between simulated and observed discharge, (2) monthly volume difference between observed and simulated discharge, and (3) a penalty score to constrain model parameters within a prescribed valid range. To select a calibration parameter set on the pareto front defined by these metrics, root-mean-square error (RMSE), Nash-Sutcliffe efficiency (NSE), and bias were evaluated for the simulated discharge and used to rank parameter sets from which the best, in the judgment of the author, was chosen. This automatic-calibration was used to calibrate the SNOW-17 and the SAC-SMA model parameters given in Table 2.2. These parameters are either scalar, meaning that a single value applies to the whole domain, or gridded, meaning that they vary spatially. In the case of gridded parameters, RDHM provides procedures to compute a priori parameters based on topography, soils, and land cover information (NWS, 2008a). For our study watersheds, these a priori parameters were provided by RTI International. The spatial pattern from these geospatially derived a priori parameters was retained in the calibration algorithm by using a separate scalar multiplier for each grid parameter. Parameters (scalars and multipliers) were calibrated separately using the method described above for each subwatershed using all the available data from 1988–2010.

In the first model configuration (SNOW-17 + SAC-SMA + Rutpix7), the SNOW-17 and the SAC-SMA model parameters were automatically calibrated. As a result of separate calibration for each subwatershed, the scalar parameters such as snow correction factor (SCF) and PE adjustment factors differ between subwatersheds. In the second model configuration (UEB + SAC-SMA+Rutpix7), only SAC-SMA model parameters were automatically calibrated. Because the UEB model is physically based its parameters were held fixed at a priori published values and not calibrated. Initial results (not shown) revealed a low flow underestimation problem for some subwatersheds that was diagnosed to be due to bias in the precipitation inputs. This bias occurred in the Blue River watershed, and was indicated by SNOW-17 SCF being larger than 1.2, which means the calibration adjusted the precipitation input multiplier to increase the precipitation input. The UEB model parameter that accounts for bias in precipitation input is the drift factor. In cases where SCF was larger than 1.2, we increased precipitation input by setting drift factor to the SCF value. This was the only UEB parameter changed from a priori published values, and this change resolved the low flow underestimation problem.

Furthermore, Rutpix7 model parameters were kept constant using a set of pre-defined hillslope and channel parameters for both model configurations. RTI International completed the calibration for the first model configuration, and we calibrated the second model configuration.

2.2.4 Performance measures

The MODSCAG SCA data product was used to compare simulated snow water equivalent from the two snow models. MODSCAG SCA data at ~500 m resolution were resampled using a nearest neighbor approach to 0.25 HRAP resolution and then classified as snow (value 1) where SCA was larger than 5% and as no snow (value 0) elsewhere. Modeled SWE was classified into a binary snow/no snow dataset using a 1 mm SWE threshold. The binary snow cover maps were created only for the dates on which less than 10% of pixels were invalid (e.g., cloud cover or missing data) and at least one of the data sources (MODSCAG, UEB, SNOW-17) had snow in the watershed. Since there was insufficient valid observation data for the Blue River watershed, we focused comparison of the observed and simulated spatial distribution of the snowmelt process on the Dolores River watershed.

We used area and pixel-based methods to compare modeled and observed snow. The area-based comparison used fractional snow covered area (Equation 5) and calculated mean absolute error (MAE) as the difference between the modeled and observed SCA fractions (Equation 6). MAE calculations were made separately for each month to account for seasonality and then averaged over all of the years with data. We also used the daily fractional SCA to calculate both the annual and melting period (March-June) Nash-Sutcliffe efficiency (NSE) (Equation 7). The pixel-based evaluation compared observed and modeled binary snow cover maps using a fitness statistic (Equation 8) based on the number of pixels where snow was observed and modeled, observed and not modeled, not observed and modeled, and not observed and not modeled (Table 2.3) (Aronica et al., 2002; Bernhardt and Schulz, 2010). The fitness is the ratio between the number of pixels where both the simulation and observation have snow and the number of pixels where either the simulation or the observation has snow.

$$\text{Fractional SCA} = \frac{\mathrm{N_s}}{\mathrm{N_s} + \mathrm{N_d}} \qquad (5)$$

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum \mathrm{abs}(\mathrm{fO_i} - \mathrm{fM_i}) \tag{6}$$

$$\mathrm{NSE_{snow}} = 1 - \frac{\sum(\mathrm{fO_i} - \mathrm{fM_i})^2}{\sum\left(\mathrm{fO_i} - \overline{\mathrm{fO}}\right)^2} \tag{7}$$

$$\mathrm{Fitness} = \frac{\mathrm{A}}{\mathrm{A} + \mathrm{C} + \mathrm{B}} \tag{8}$$

In these equations, N_s is the total number of pixels with snow in the binary snow cover map; N_d is the total number of pixels without snow in the binary snow cover map; fO_i and fM_i are the fractional SCA from observation and simulation; A, B, and C in the fitness function are the number of pixels in each group as defined in Table 2.3.

Basin discharge was simulated at a 6-hour time step and averaged as a daily time step for evaluation. Moreover, before using this result, we removed the first water year (water year 1989) as the system spin up period. Observed daily discharge was compared with the simulation results using metrics of RMSE, NSE, bias, and percent April to July volume error:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum(\mathrm{O_i} - \mathrm{M_i})^2} \tag{9}$$

$$\mathrm{NSE_{discharge}} = 1 - \frac{\sum(\mathrm{O_i} - \mathrm{M_i})^2}{\sum(\mathrm{O_i} - \overline{\mathrm{O}})^2} \tag{10}$$

$$\mathrm{Bias} = \frac{1}{\mathrm{N}}\sum(\mathrm{O_i} - \mathrm{M_i}) \tag{11}$$

$$\mathrm{Volume\ error} = \left[\frac{\sum(\mathrm{VO_i} - \mathrm{VM_i})}{\sum \mathrm{VO_i}}\right] \cdot 100\% \tag{12}$$

where O_i and M_i are the daily discharge (m^3s^{-1}) from observation and simulation and VO_i and VM_i are the daily discharge volume (m^3) of observation and simulation from April to July.

Aside from the snow and discharge analysis, we also compared the model outputs of sublimation and ET from the snow and runoff models to discover the differences between the two model configurations in simulating the evaporative components. We compared the water mass balance by calculating the simulated interannual domain average of precipitation, sublimation, and ET. We also examined the sublimation results of the UEB model in different land types to evaluate the model performance.

2.3 Results and discussion

2.3.1 Snow process simulation

Both observation and simulation datasets for the Dolores River watershed were converted into binary snow cover maps, and the evaluation metrics were calculated using results from 2000–2010. Table 2.4 shows the annual and melting period NSE results. Table 2.5 shows the monthly MAE and fitness (except for July–Sept). The UEB model produced higher NSE and lower MAE in most of the months compared to the SNOW-17 model, indicating that the UEB model performed better for the area-based evaluation. As for fitness results, the SNOW-17 model had a higher fitness value most of the time, and hence a better pixel-based performance than the UEB model. Additionally, both models have higher fitness during the snow accumulation period (Dec–Mar) than during the melting (Apr–Jun) and early snowfall (Oct–Nov) periods. This is because both observation and simulation have high SCA over the watershed during the snow accumulation period, which increases the possibility of matching pixels between the simulation and observation binary snow cover maps.

In order to gain a better understanding of the spatial and temporal dynamics of the SCA in the watershed, we further examined the results in water year 2006, which has the largest number of satellite observation images with sufficient valid data. A time series plot of the modeled and observed fractional SCA during water year 2006 (Figure 2.3), shows that both models generally follow the observed SCA pattern. During the snow accumulation period, the SNOW-17 model tended to have higher peaks (e.g., during October and November) and overestimate SCA more than the UEB model does. During the melting period, both models simulated the snowmelt process with reasonable timing and amount, compared to the observational data. The binary snow cover maps (Figure 2.4) show the spatial distribution of snow cover from the two snow models and the MODSCAG observations for various dates in water year 2006. The four days correspond to the accumulation (October 13, December 6) and snowmelt (April 9, May 1) periods. The maps show that the UEB model better captures the reduction in area during melt-out, whereas the SNOW-17 model overestimates SCA. This is also a problem during the snow accumulation period (October 13th). Examining the UEB SCA simulations, a scattered or pixelated pattern is present (e.g., October 13th). This is due to the UEB model using terrain parameters (slope and aspect) at the center point of each 0.25 HRAP grid cell. These center point values do not represent the larger grid cell as a whole and may have slope and aspect disassociated with the slope and aspect of adjacent large grid cell centers, leading to the pixelated SCA pattern.

The UEB model's better performance in the area-based evaluation can be explained as follows. First, the automatic-calibration adjusted SCF (SCF > 1) from the SNOW-17 model leads to greater snow accumulation, which may delay snow disappearance. Second, the UEB model does simulate sublimation, which may lead to more rapid snow depletion and disappearance than SNOW-17. Third, the SNOW-17 model uses a melt factor to implicitly represent the energy input and corresponding topographic effect for snowmelt, whereas the UEB model directly calculates the radiation fluxes using slope, aspect, and canopy data as inputs. This makes the UEB model more sensitive to the variability in melting caused by different terrain conditions.

For pixel-based evaluation, the UEB model uses the slope and aspect at the center point of each pixel to represent the terrain features of the corresponding grid area. However, terrain features at the center points are different from the grid cell as a whole, especially when the grid spacing is large, and this may lead to the mismatch with observed SCA and thus lower fitness. However, the similarity of aggregate observed and UEB SCA suggests that over the basin these points may be sufficient to represent basin terrain variability, something that is not done by SNOW-17 that does not account for slope and aspect.

2.3.2 Basin discharge simulation

We evaluated the basin discharge performance by comparing the observed and simulated daily discharge with different evaluation metrics for water years 1990–2010. According to the values of the different performance metrics, the overall model performance for the basin discharge simulation indicated a satisfactory calibration for each model configuration in the two watersheds (Table 2.6 and Table 2.7). In the Dolores River watershed, the UEB model had somewhat better performance than the SNOW-17 model, with higher NSE and lower values for the other metrics in most of the subwatersheds, whereas the SNOW-17 model outperforms the UEB model somewhat in the Blue River watershed when comparing these metrics. In addition, the head watershed LCCC2 had much lower NSE indicating the model performance was not as good as for the other subwatersheds. This is because LCCC2 has much less precipitation input than other subwatersheds and generates intermittent streamflow that mainly happens during the spring melt season, with almost no streamflow during July to September. Since neither snow model can simulate the streamflow during dry periods well (results not shown here), the model performance for this subwatershed is not as good as for the others.

Figure 2.5 and Figure 2.6 present the simulated domain average SWE and observed and simulated discharge in different water years (1994, 1997, 2001, and 2008) for the Dolores River and Blue River watersheds, respectively. These years were chosen because they were typical of and spanned the range of model performance over the two decades (22 years), except for one year that was exceptionally dry and where there was poor model performance from both models (2002). These results show that the two snow models coupled to SAC-SMA and Rutpix7 provide reasonable discharge simulations for the two watersheds, each of which has different snowmelt and discharge patterns. In the Dolores River watershed, the snowpack ripens to melt fast around late March or the beginning of April, with total melt out around late June or the beginning of July, while a similar process happens approximately one month later in the Blue River watershed. This difference in snowmelt patterns also affects the corresponding discharge patterns. For instance, the timing of spring pulse is often influenced by the temperature increase that ripens the snow pack to melt and trigger the surge in discharge. In the Dolores River watershed, the spring discharge increase starts around April, which corresponds to the early snowmelt that is about one month earlier than in the Blue River watershed (around May). Also, snowpack size is the major controlling factor for the discharge decline process. Therefore, because of its later melt out process, the flow recession process in the Blue River watershed lasts longer than in the Dolores River watershed.

Aside from the discharge results, both the SNOW-17 and the UEB model were found to have similar timing for snow accumulation and snowmelt. The SNOW-17 model has a higher SWE during the accumulation period mainly because the UEB model simulates water loss from sublimation, leading to less snow accumulation than the SNOW-17 model. As a result, the SNOW-17 model actually provides more rain-plus-melt input to feed the SAC-SMA model that simulates the runoff and ET processes. This leads to differences in the simulated quantity of ET from the two model configurations. This will be discussed in the next subsection.

2.3.3 Evaporative components simulation

The UEB model coupled to SAC-SMA simulates both sublimation and ET. However, since SNOW-17 does not simulate sublimation, the only evaporative component in the SNOW-17 model coupled to SAC-SMA is ET. To better understand the consequences of this difference, we compared the water mass balance from the two model configurations. We calculated the watershed average of annual mean precipitation, sublimation, and ET for the two watersheds over the simulation period (Figure 2.7). Precipitation adjustments made to SNOW-17 through the SCF parameter, and to UEB through the drift factor parameter are shown. Precipitation inputs to both snowmelt models were adjusted by a similar amount in the Blue River watershed, whereas only the SNOW-17 model was adjusted in the Dolores River watershed, noting from the calibration section above that drift factor was not adjusted when SCF was less than 1.2. Since the simulated precipitation inputs are similar and the models were calibrated against the same observed discharge, both model configurations have similar total evaporative components for each of the watersheds. The UEB model, however, explicitly simulated the portion due to sublimation. The results show that the water loss due to sublimation is a considerable amount (12%–13% of annual mean precipitation), and should not be neglected in the snow mass balance for these watersheds.

We further examined the canopy and ground sublimation simulated by the UEB model for different land types (Figure 2.8). This figure shows the watershed average of the annual mean precipitation, as well as the canopy and ground sublimation for forest and open areas. The canopy sublimation in the forest area dominates the process, and the total water loss from sublimation in the forest areas is about twice as much as in the open area. In addition, the annual mean precipitation and canopy sublimation were compared for different forest types and at different elevations using the simulated results at each grid cell over the watershed domain. Figure 2.9 shows that annual precipitation increases with increased elevation, and the canopy sublimation increases with increased elevation, precipitation, and forest density, determined by LAI, canopy cover, and canopy height of different forest types (Table 2.8). These results are similar to findings from other work that evaluates sublimation variability in semi-arid mountainous regions (Montesi et al., 2004; Sexstone et al., 2018).

Since a large fraction of both watersheds consists of forest area (87% in the Dolores River watershed and 53% in the Blue river watershed), land type changes may affect sublimation and thus impact the water mass balance in the watersheds (Biederman et al., 2014; Harpold et al., 2014; Penn et al., 2016). This analysis highlights the advantages of using the UEB model, including it better quantifying the proportions of the different evaporative components and providing the means to evaluate the impact of land cover change on the sublimation process and the corresponding influence on the water availability in the watershed. Using SNOW-17 to accomplish these same tasks would be difficult or impossible because the model doesn't directly account for the sublimation process.

2.4 Summary and conclusions

The objective of this study was to assess whether applying an energy balance model in the river forecasting system used by CBRFC would improve water supply forecasts in the Colorado River Basin. This research used analysis of historical, or retrospective, model simulations to evaluate model performance in comparison to snow covered area, daily discharge, and water mass balance. The UEB and SNOW-17 models were evaluated by coupling them with the SAC-SMA model and the Rutpix7 model within the RDHM framework for distributed modeling of basin snow and discharge in the Dolores and the Blue River watersheds. Parameters for the SNOW-17 and the SAC-SMA models were calibrated using an automated multi-objective procedure. In the UEB model, the drift factor parameter was adjusted to account for the precipitation input bias, but other parameters were held fixed at their literature values.

Comparison of the simulated and observed SCA data showed that both snow models were able to simulate the spatial and temporal change of the SCA in the Dolores River watershed with reasonable timing and amount (e.g., annual NSE of SCA is larger than 0.7). Results indicated that both model configurations were also able to provide good discharge simulation results for the study sites (e.g., NSE of discharge is between 0.85 and 0.94 for most subwatersheds). Although both models have similar performance, the UEB model showed its potential for application in the river forecasting system to advance water supply forecasts for future changing conditions. First, the UEB model was able to simulate the sublimation process for different land cover types, whereas sublimation is not represented in the SNOW-17 model. Sublimation is an important evaporative component during the snow season in the Colorado River Basin, and the UEB model demonstrated its capability to evaluate sublimation water loss and its impact on the water mass balance when the land type alters. Second, the UEB model held parameters (except for drift factor) constant and achieved fit metrics comparable to the SNOW-17 model, where parameters were calibrated for each subwatershed. This suggests that the UEB model parameters are more transferable and provide the ability to simulate the snowmelt process under different terrain or climate conditions, thus reducing the intensive model calibration work required within the temperature index model to provide a reliable simulation. Moreover, the maximum/minimum melt factors in the SNOW-17 model were calibrated against historical data, which may not well represent the melt rate under potential future conditions given a changing climate. In contrast, the UEB model accounts for the physical process of snowmelt based on energy and water mass balance, which means it is more capable of providing reliable predictions when climate patterns change. However, the performance of the UEB model was found to be affected by biases in the input precipitation. It was necessary to adjust the UEB model's drift factor based on the SNOW-17 model's SCF values to resolve low flow underestimation caused by the precipitation input bias in the Blue River watershed. Without the reference SCF value, it may be challenging to estimate the data bias and calibrate the UEB model parameters, demanding more simulation time and computing resources than the SNOW-17 model for automatic-calibration.

In the future, additional work is needed to further understand the UEB model performance for operational water supply forecasts. One direction is to evaluate the two snow models under forecasting conditions, which involves data assimilation and ensemble forecasting techniques to compare the UEB and the SNOW-17 models using historical forcing over decades as representative of future possible weather conditions. Another direction is to evaluate model performance when running the UEB model at higher spatial resolutions. It is assumed that the energy balance model will provide better performance at finer resolution because of the better representation of the spatial variation in topographic and vegetation features. However, higher model resolution will require more computing resources and longer simulation time. Balancing the trade-offs between model performance and computational demand of model operation is an important issue for operational water supply forecasts.

Acknowledgements

This work was supported by NASA award NNH15CM90C. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA.

References

Anderson, E.A., 2006. Snow accumulation and ablation model – SNOW-17, NWSRFS user manual documentation. National Weather Service River Forecast System. https://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/part2/_pdf/22snow17.pdf

Anderson, E.A., 1973. National Weather Service river forecast system: snow accumulation and ablation model, NOAA technical memorandum: NWS HYDRO-17. US National Weather Service.

Aronica, G., Bates, P.D., Horritt, M.S., 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. Hydrol. Process. 16, 2001–2016. https://doi.org/10.1002/hyp.398

Barnett, T.P., Adam, J.C., Lettenmaier, D.P., 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. Nature 438, 303–309. https://doi.org/10.1038/nature04141

Bernhardt, M., Schulz, K., 2010. SnowSlide: A simple routine for calculating gravitational snow transport. Geophys. Res. Lett. 37. https://doi.org/10.1029/2010GL043086

Biederman, J.A., Brooks, P.D., Harpold, A.A., Gochis, D.J., Gutmann, E., Reed, D.E., Pendall, E., Ewers, B.E., 2014. Multiscale observations of snow accumulation and peak snowpack following widespread, insect-induced lodgepole pine mortality. Ecohydrology 7, 150–162. https://doi.org/10.1002/eco.1342

Burnash, R.J.C., Ferral, R.L., McGuire, R.A., 1973. A generalized streamflow simulation system : conceptual modeling for digital computers. U.S. Department of Commerce National Weather Service and State of California Department of Water Resources.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6, 182–197. https://doi.org/10.1109/4235.996017

Debele, B., Srinivasan, R., Gosain, A.K., 2010. Comparison of process-based and temperature-index snowmelt modeling in SWAT. Water Resour. Manag. 24, 1065–1088. https://doi.org/10.1007/s11269-009-9486-2

Essery, R., Morin, S., Lejeune, Y., B Ménard, C., 2013. A comparison of 1701 snow models using observations from an alpine site. Adv. Water Resour. 55, 131–148. https://doi.org/10.1016/j.advwatres.2012.07.013

Franz, K.J., Hogue, T.S., Sorooshian, S., 2008. Operational snow modeling: Addressing the challenges of an energy balance model for National Weather Service forecasts. J. Hydrol. 360, 48–66. https://doi.org/10.1016/j.jhydrol.2008.07.013

Gesch, D.B., 2007. The National Elevation Dataset, in: Maune, D. (Ed.), Digital Elevation Model Technologies and Applications—the DEM Users Manual. American Society for Photogrammetry and Remote Sensing, Bethesda, MD, pp. 99–118.

Gupta, H. V, Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Harpold, A.A., Biederman, J.A., Condon, K., Merino, M., Korgaonkar, Y., Nan, T., Sloat, L.L., Ross, M., Brooks, P.D., 2014. Changes in snow accumulation and ablation following the Las Conchas Forest Fire, New Mexico, USA. Ecohydrology 7, 440–452. https://doi.org/10.1002/eco.1363

Hock, R., 2003. Temperature index melt modelling in mountain areas. J. Hydrol. 282, 104–115. https://doi.org/10.1016/S0022-1694(03)00257-9

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K., 2015. Completion of the 2011 national land cover database for the conterminous United States – Representing a decade of land cover change information. Photogramm. Eng. Remote Sens, 81(0):345-354.

Koren, V., Reed, S., Smith, M., Zhang, Z., Seo, D.J., 2004. Hydrology laboratory research modeling system (HL-RMS) of the US national weather service. J. Hydrol. 291, 297–318. https://doi.org/10.1016/j.jhydrol.2003.12.039

Kumar, M., Marks, D., Dozier, J., Reba, M., Winstral, A., 2013. Evaluation of distributed hydrologic impacts of temperature-index and energy-based snow models. Adv. Water Resour. 56, 77–89. https://doi.org/10.1016/j.advwatres.2013.03.006

Li, D., Wrzesien, M.L., Durand, M., Adam, J., Lettenmaier, D.P., 2017. How much runoff originates as snow in the western United States, and how will that change in the future? Geophys. Res. Lett. 44, 6163–6172. https://doi.org/10.1002/2017GL073551

Luce, C., Tarboton, D.G., 2001. A modified force-restore approach to modeling snow-surface heat fluxes, in: Proceedings of the 69th Annual Western Snow Conference. Sun Valley, Idaho.

Luce, C., Tarboton, D.G., 2010. Evaluation of alternative formulae for calculation of surface temperature in snowmelt models using frequency analysis of temperature observations. Hydrol. Earth Syst. Sci. 14, 535–543. https://doi.org/10.5194/hess-14-535-2010

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A., Jonas, T., 2015. Evaluating snow models with varying process representations for hydrological applications. Water Resour. Res. 51, 2707–2723. https://doi.org/10.1002/2014WR016498

Mahat, V., Tarboton, D.G., 2012. Canopy radiation transmission for an energy balance snowmelt model. Water Resour. Res. 48, 1–16. https://doi.org/10.1029/2011WR010438

Mahat, V., Tarboton, D.G., 2014. Representation of canopy snow interception, unloading and melt in a parsimonious snowmelt model. Hydrol. Process. 28, 6320–6336. https://doi.org/10.1002/hyp.10116

Mahat, V., Tarboton, D.G., Molotch, N.P., 2013. Testing above- and below-canopy representations of turbulent fluxes in an energy balance snowmelt model. Water Resour. Res. 49, 1107–1122. https://doi.org/10.1002/wrcr.20073

Montesi, J., Elder, K., Schmidt, R.A., Davis, R.E., 2004. Sublimation of intercepted snow within a subalpine forest canopy at two elevations. J. Hydrometeorol. 5, 763–773. https://doi.org/10.1175/1525-7541(2004)005<0763:SOISWA>2.0.CO;2

NWS, 2006. Conceptualization of the sacramento soil moisture accouting model. National Weather Service. https://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/part2/_pdf/23sacsma.pdf

NWS, 2008a. Hydrology laboratory-research distributed hydrologic model (HL-RDHM) user manual V. 2.4.2. National Weather Service. https://www.cbrfc.noaa.gov/present/rdhm/RDHM_User_Manual.pdf

NWS, 2008b. The HL research distributed Hydrologic Model (HL-RDHM) developer's manual v. 2.0. National Weather Service. https://www.cbrfc.noaa.gov/present/rdhm/rdhm_developers_manual.pdf

Painter, T.H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R.E., Dozier, J., 2009. Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. Remote Sens. Environ. 113, 868–879. https://doi.org/10.1016/j.rse.2009.01.001

Penn, C.A., Bearup, L.A., Maxwell, R.M., Clow, D.W., 2016. Numerical experiments to explain multiscale hydrological responses to mountain pine beetle tree mortality in a headwater watershed. Water Resour. Res. 52, 3143–3161. https://doi.org/10.1002/2015WR018300

Reed, S.M., Maidment, D.R., 1999. Coordinate transformations for using NEXRAD data in GIS-based hydrologic modeling. J. Hydrol. Eng. 4, 174–182. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(174)

Schaake, J., Henkel, A., Cong, S., 2004. Application of prism climatologies for hydrologic modeling and forcasting in the Western U.S. Proc. 18th Conf. on Hydrology, Seattle, WA.

Sexstone, G.A., Clow, D.W., Fassnacht, S.R., Liston, G.E., Hiemstra, C.A., Knowles, J.F., Penn, C.A., 2018. Snow sublimation in mountain environments and its sensitivity to forest disturbance and climate warming. Water Resour. Res. 54, 1191–1211. https://doi.org/10.1002/2017WR021172

Shakoor, A., Burri, A., Bavay, M., Ejaz, N., Razzaq Ghumman, A., Comola, F., Lehning, M., 2018. Hydrological response of two high altitude Swiss catchments to energy balance and temperature index melt schemes. Polar Sci. 17, 1–12. https://doi.org/10.1016/j.polar.2018.06.007

Sturm, M., Goldstein, M.A., Parr, C., 2017. Water and life from snow: A trillion dollar science question. Water Resour. Res. 53, 3534–3544. https://doi.org/10.1002/2017WR020840

Tarboton, D.G., Blöschl, G., Cooley, K., Kirnbauer, R., Luce, C., 2000. Spatial Snow Cover Processes at Kühtai and Reynolds Creek, in: Grayson, R., Blöschl, G. (Eds.), Spatial Patterns in Catchment Hydrology: Observations and Modelling. Cambridge University Press, Cambridge, pp. 158–186.

Tarboton, D.G., Luce, C.H., 1996. “Utah Energy Balance Snow Accumulation and Melt Model (UEB),” Computer model technical description and users guide. Utah Water Research Laboratory and USDA Forest Service Intermountain Research Station.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., Mocko, D., 2012. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. J. Geophys. Res. Atmos. 117. https://doi.org/10.1029/2011JD016048

Zeinivand, H., De Smedt, F., 2009. Hydrological modeling of snow accumulation and melting on river basin scale. Water Resour. Manag. 23, 2271–2287. https://doi.org/10.1007/s11269-008-9381-2

Zierl, B., Bugmann, H., 2005. Global change impacts on hydrological processes in Alpine catchments. Water Resour. Res. 41, 1–13. https://doi.org/10.1029/2004WR003447

Table 2.1 Details of subwatersheds in the two study sites.

Index	Area (km^2)	Elevation range (m)	Type
Dolores River watershed			
DRRC2	275.11	2569.89 - 4323.56	Headwater
LCCC2	172.42	2114.00 - 3393.22	Headwater
DOLC2	1026.32	2111.56 - 4297.89	Local
MPHC2	606.24	2093.00 - 2964.40	Local
Blue River watershed			
TCFC2	228.83	2776.98 - 4242.29	Headwater
SKEC2	148.84	2838.89 - 4349.08	Headwater
BUEC2	110.60	2999.36 - 4344.82	Headwater
BSWC2	204.03	2750.39 - 4166.63	Local
DIRC2	156.97	2687.40 - 3923.83	Local

Table 2.2 Parameters of the SNOW-17 and the SAC-SMA models used in calibration.

Parameter	**Description**	**Type**
snow_SCF	Snow correction factor	Scalar
snow_PXTMP	Temperature that separates rain from snow [°C]	Scalar
snow_MFMAX	Maximum melt factor [mm $(6hr)^{-1}$ $°C^{-1}$]	Grid
snow_MFMIN	Minimum melt factor [mm $(6hr)^{-1}$ $°C^{-1}$]	Grid
snow_UADJ	Wind function factor during rain-on-snow periods [mm mb^{-1}]	Grid
sac_peadj	Potential evaporation adjustment factor (12 factors in total)	Scalar
sac_UZTWM	Upper zone tension water maximum storage [mm]	Grid
sac_UZFWM	Upper zone free water maximum storage [mm]	Grid
sac_LZTWM	Lower zone tension water maximum storage [mm]	Grid
sac_LZFPM	Lower zone free water primary storage [mm]	Grid
sac_LZFSM	Lower zone free water supplementary storage [mm]	Grid
sac_UZK	Upper zone free water storage depletion coefficient [day^{-1}]	Grid
sac_LZPK	Lower zone primary storage depletion coefficient [day^{-1}]	Grid
sac_LZSK	Lower zone supplementary storage depletion coefficient [day^{-1}]	Grid
sac_ZPERC	Maximum percolation capacity coefficient [dimensionless]	Grid
sac_REXP	Exponent for the percolation equation	Grid
sac_PFREE	Percent of percolated water which always goes directly to lower zone free water storages (decimal fraction)	Grid

* Prefix “snow” denotes the SNOW-17 model and “sac” denotes the SAC-SMA model.

Table 2.3 Four pixel types used in the fitness evaluation.

Number of pixels	**Observed snow**	**Observed no snow**
Modeled snow	A	B
Modeled no snow	C	D

Table 2.4 Annual and melting period (Mar-June) NSE of SCA in the Dolores River watershed evaluated over 2000-2010.

Models	Annual NSE	Melting period NSE
SNOW-17	0.739	0.822
UEB	0.886	0.891

Table 2.5 Monthly MAE and fitness of the SCA in the Dolores River watershed evaluated over 2000-2010.

	MAE (%)		**Fitness**	
Month	**UEB**	**SNOW-17**	**UEB**	**SNOW17**
Jan	7.5	10.2	0.878	0.898
Feb	7.0	15.6	0.801	0.837
Mar	6.7	10.3	0.819	0.860
Apr	12.4	18.3	0.557	0.582
May	7.2	6.3	0.375	0.372
Jun	2.3	1.6	0.185	0.184
Oct	4.5	8.0	0.083	0.177
Nov	13.2	22.3	0.196	0.237
Dec	14.6	24.7	0.747	0.741

Table 2.6 Results of evaluation metrics for basin discharge in the Dolores River watershed.

Watershed index	**Model**	**NSE**	**RMSE (cms)**	**BIAS (cms)**	**Vol Err (%)**
DRRC2	SNOW-17	0.851	2.201	0.112	1.285
	UEB	0.897	1.827	0.023	-0.138
LCCC2	SNOW-17	0.654	0.871	0.027	1.796
	UEB	0.684	0.832	0.013	0.151
DOLC2	SNOW-17	0.905	5.494	0.159	0.826
	UEB	0.915	5.231	0.132	0.184
MPHC2	SNOW-17	0.900	6.834	0.425	3.343
	UEB	0.913	6.411	0.572	2.777

Table 2.7 Results of evaluation metrics for basin discharge in the Blue River watershed.

Watershed index	Model	NSE	RMSE (cms)	BIAS (cms)	Vol Err (%)
TCFC2	SNOW-17	0.937	1.174	0.068	-1.922
	UEB	0.928	1.257	0.045	0.09
SKEC2	SNOW-17	0.921	0.776	-0.02	-2.404
	UEB	0.931	0.727	-0.024	-0.834
BUEC2	SNOW-17	0.912	0.576	-0.008	-2.381
	UEB	0.896	0.629	-0.013	-3.373
BSWC2	SNOW-17	0.924	1.125	0.0	-1.248
	UEB	0.915	1.191	-0.011	-1.665
DIRC2	SNOW-17	0.947	2.794	-0.056	-2.462
	UEB	0.937	3.063	-0.186	-2.819

Table 2.8 LAI, canopy cover, and canopy height for each forest type used in the simulation.

Land type	LAI	Canopy cover	Canopy height
evergreen forest	4.5	0.7	15
deciduous forest	1	0.5	8
shrub	1	0.5	3

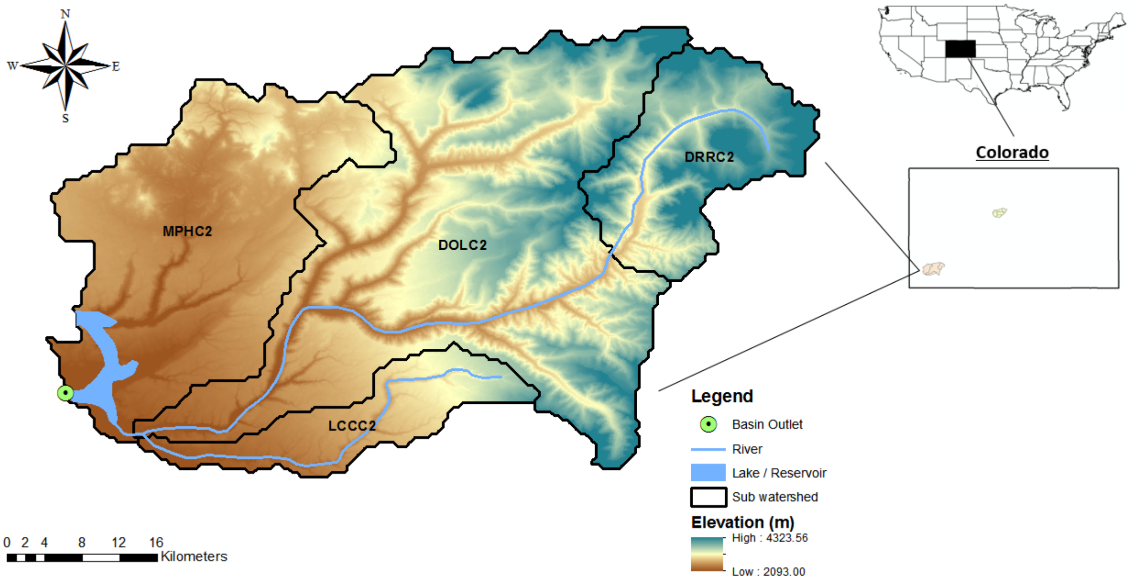


Figure 2.1 Dolores River above McPhee Reservoir study site.

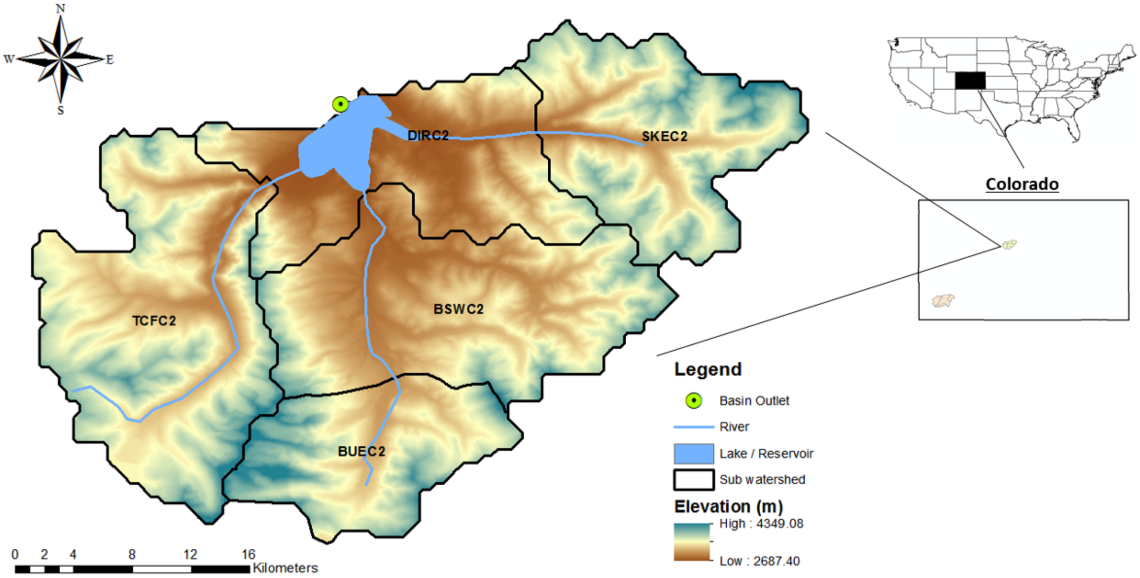


Figure 2.2 Blue River above Dillon Reservoir study site.

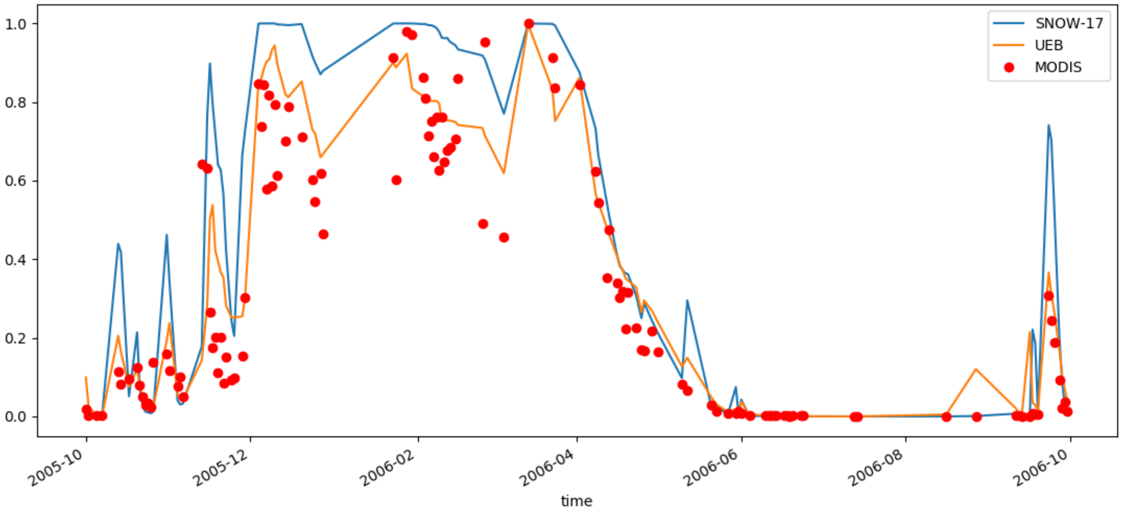


Figure 2.3 Modeled and observed MODIS fractional SCA in water year 2006 in the Dolores River watershed.

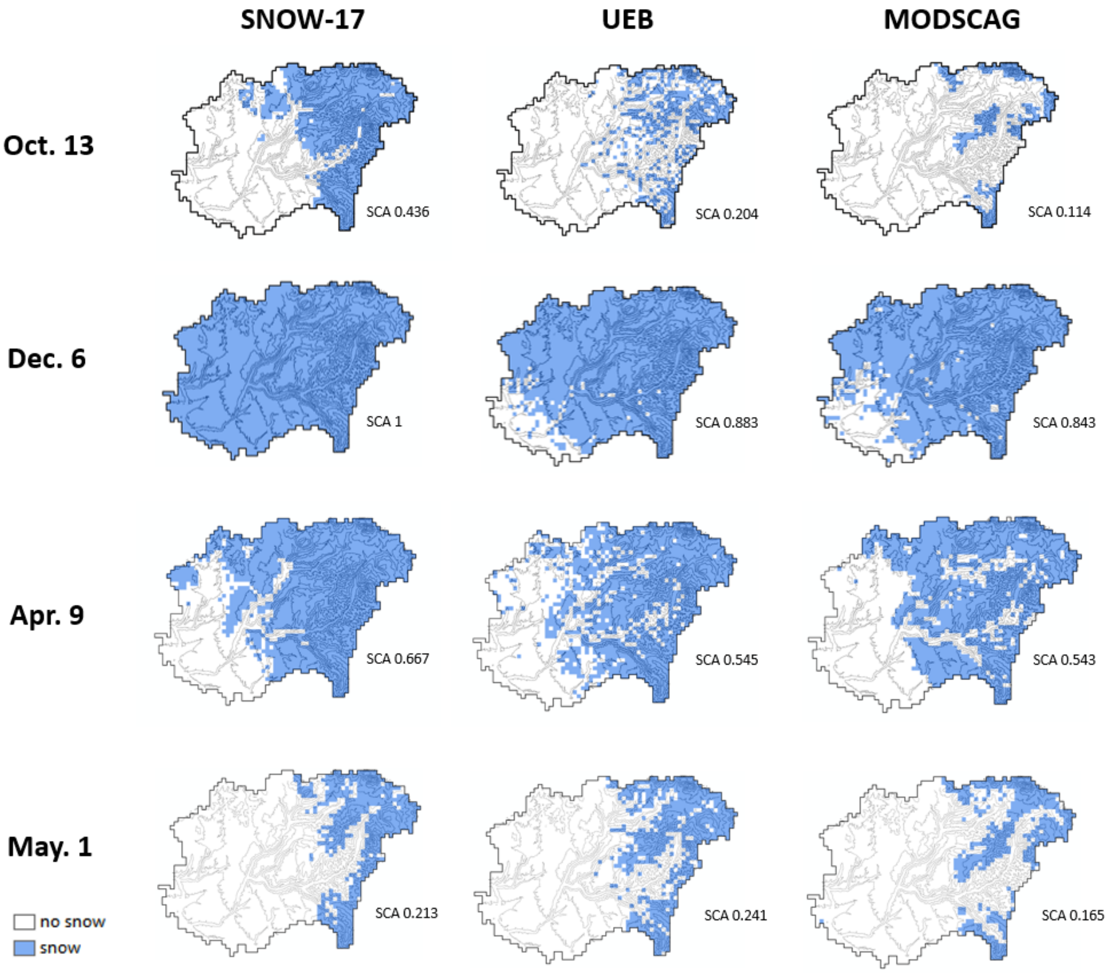


Figure 2.4 Snow cover maps for model simulations and MODIS observation at different dates in water year 2006 in the Dolores River watershed.

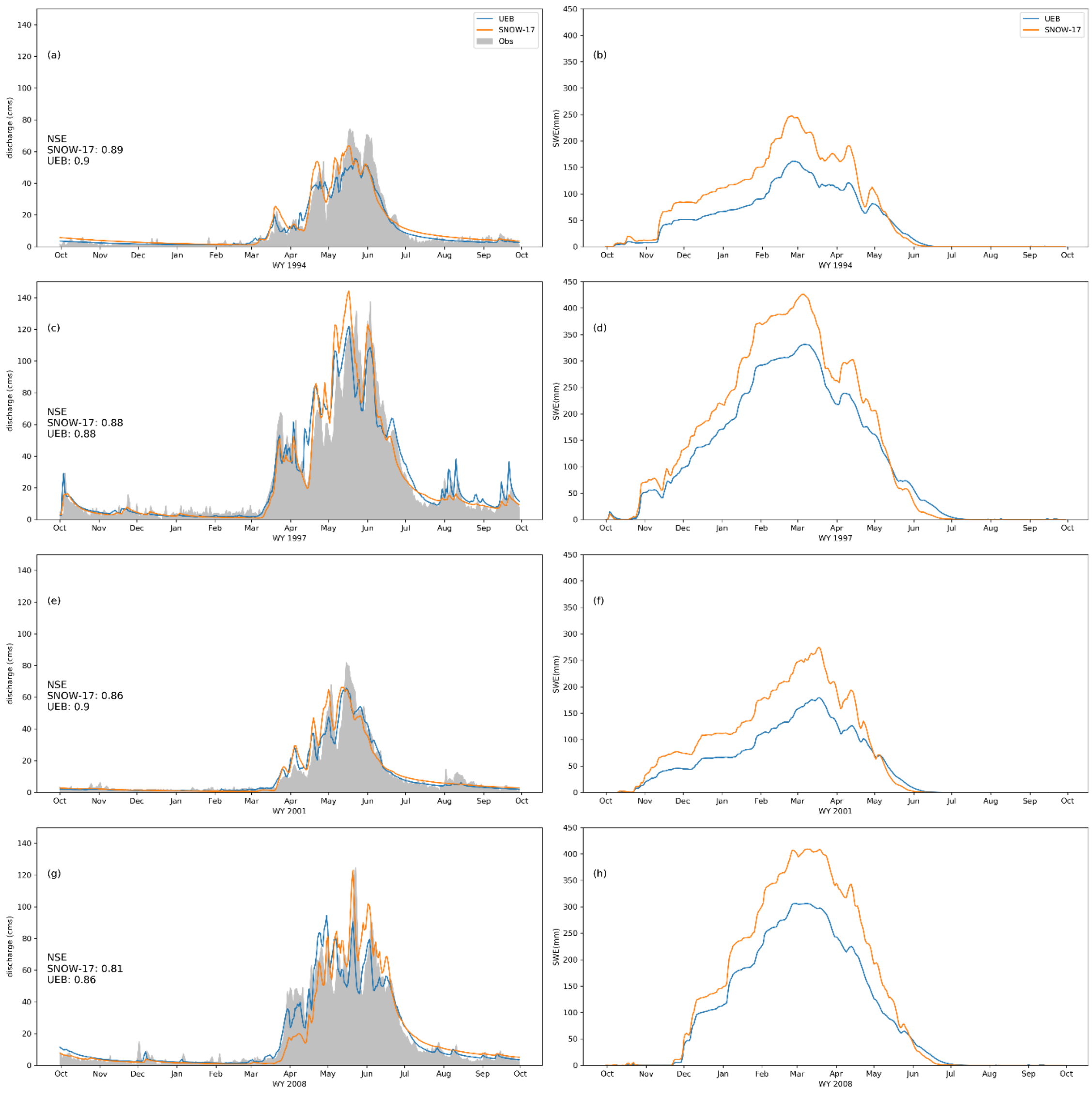


Figure 2.5 Simulated domain average SWE, and simulated and observed daily discharge in the Dolores River watershed for water years (WY) 1994, 1997, 2001, and 2008.

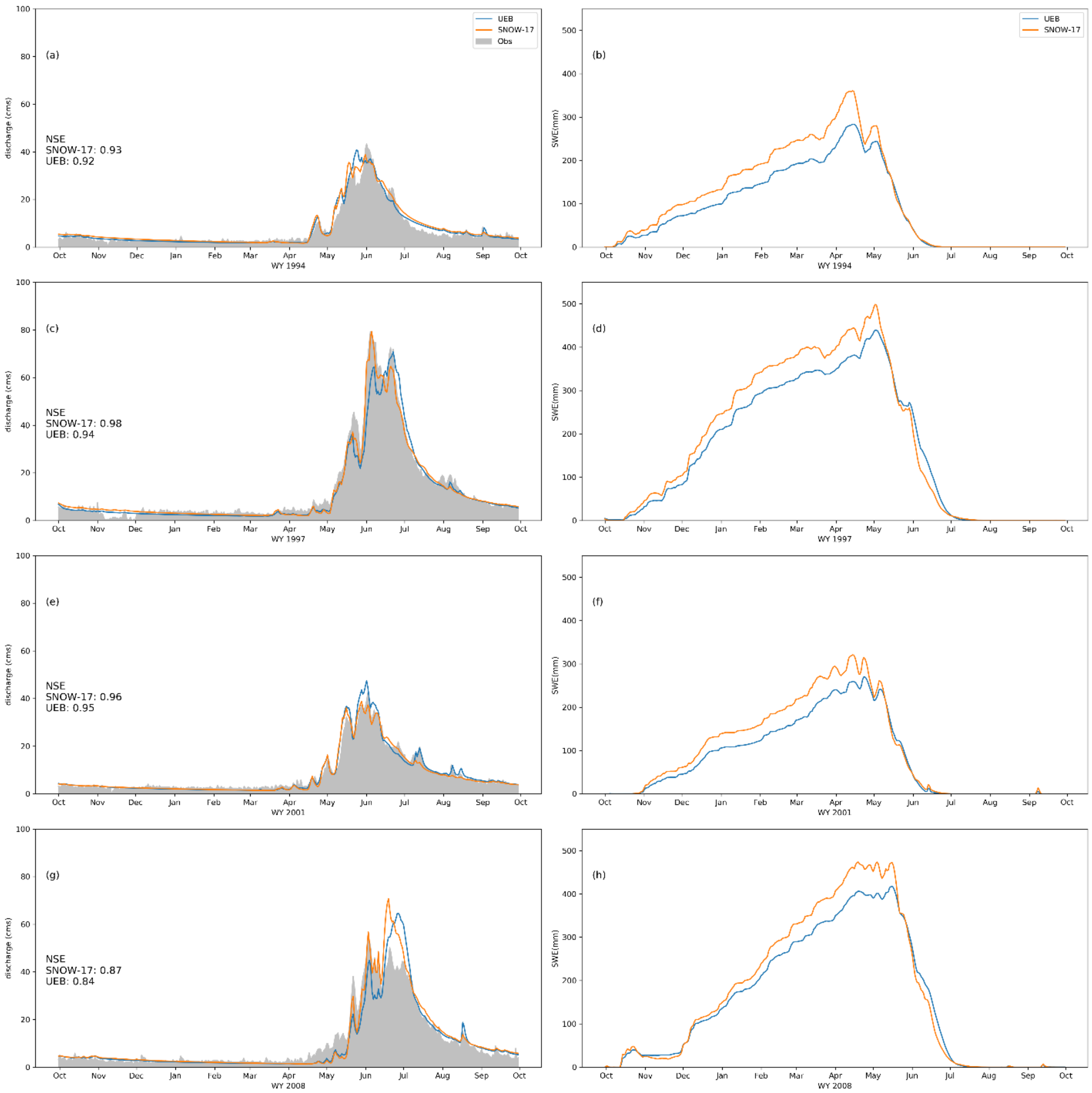
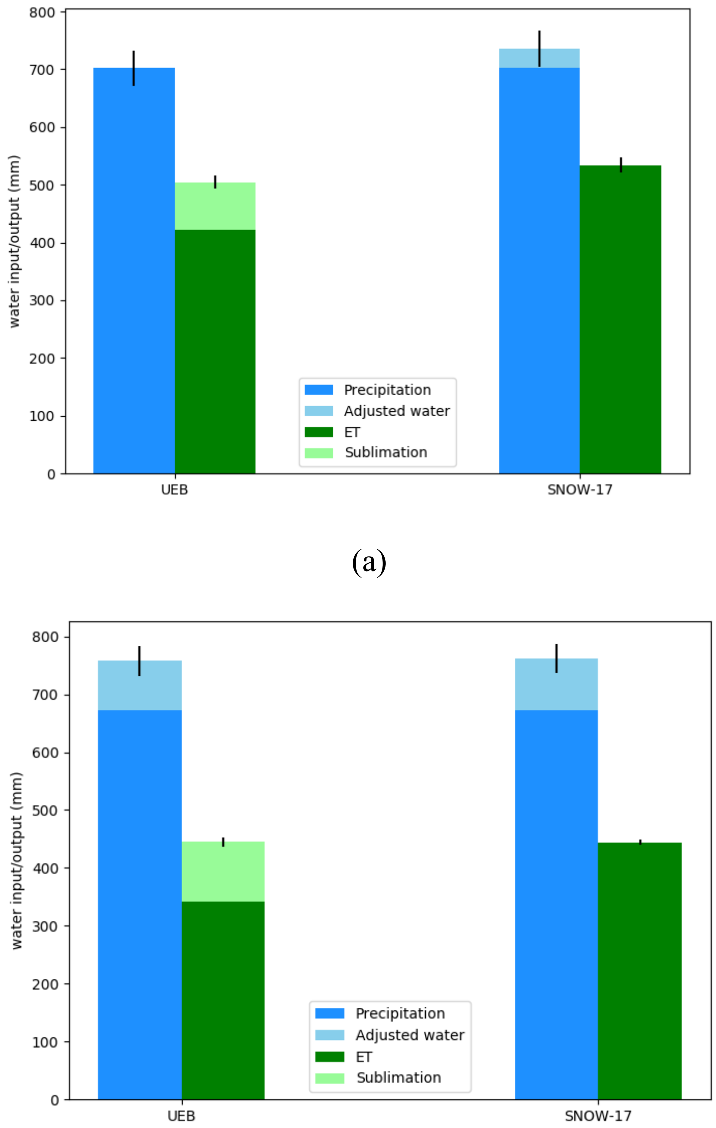


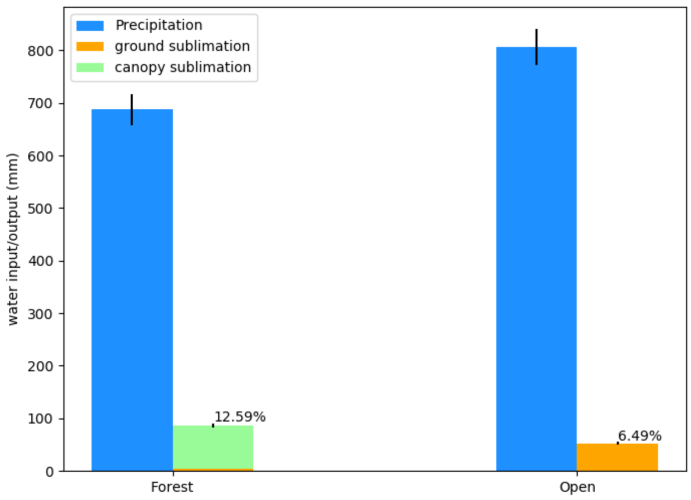
Figure 2.6 Simulated domain average SWE, and simulated and observed daily discharge in the Blue River watershed for water years (WY) 1994, 1997, 2001, and 2008.



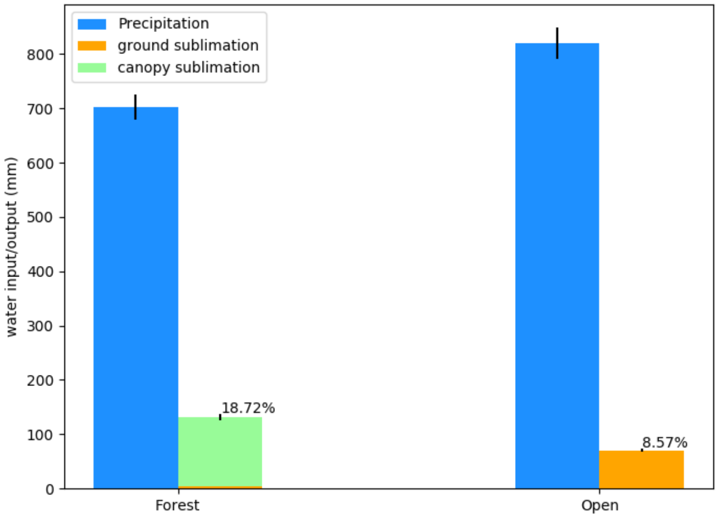
(a)

(b)

Figure 2.7 Domain average of annual mean precipitation, sublimation, and ET fluxes simulated from the two model configurations. Error bars denote the standard error of the mean. Panel (a) is for the Dolores River watershed; Panel (b) is for the Blue River watershed.

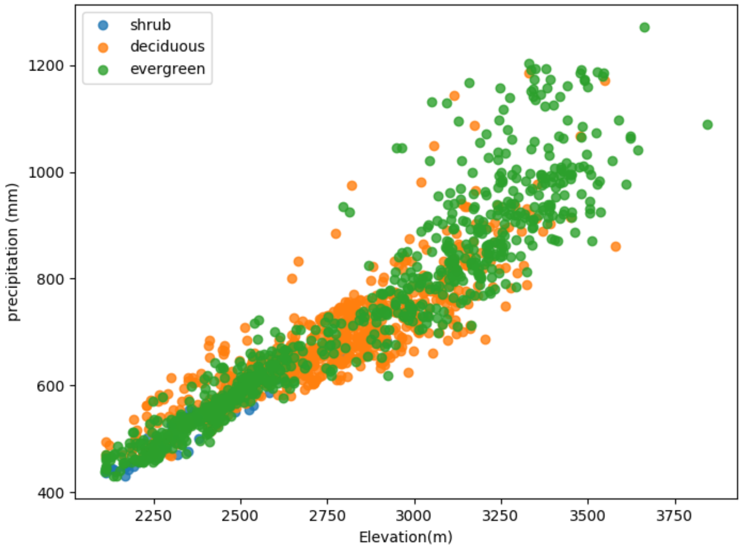


(a)

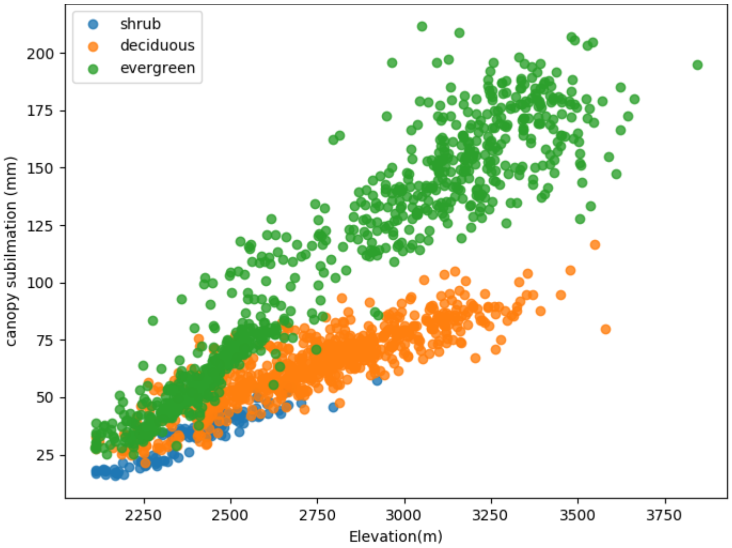


(b)

Figure 2.8 Simulated domain average annual mean sublimation fluxes compared to annual mean precipitation in forest and open areas. The percentage listed above the sublimation column represents the percentage of annual mean precipitation that was sublimated in each land cover type. Error bars denote the standard error of the mean. Panel (a) is for the Dolores River watershed; Panel (b) is for the Blue River watershed.



(a)



(b)

Figure 2.9 Annual mean precipitation (a) and canopy sublimation (b) versus elevation simulated at each grid cell over the Dolores River watershed for different vegetation types.

CHAPTER 3

COLLABORATIVE SHARING OF MULTIDIMENSIONAL SPACE-TIME DATA IN A NEXT GENERATION HYDROLOGIC INFORMATION SYSTEM [1]

Abstract

In hydrologic research, there is a need to manage, archive, and publish data in a discoverable way to increase data reuse, transparency, and reproducibility. Multidimensional space-time data are commonly used in hydrologic research, and systems are needed for sharing and exchanging such data. Simply exchanging files is not always convenient given file sizes, may result in loss of metadata and provenance information, and does not take advantage of server-based functionality available for serving these types of data. We developed an approach to manage, share, and publish multidimensional space-time data in HydroShare, a next generation hydrologic information system and domain specific repository. This paper presents the design, development, and testing of this approach. We selected the Network Common Data Form (NetCDF) as the file format and defined metadata elements to store and manage multidimensional space-time data. We adopted and adapted existing software to automatically harvest, support entry of metadata, and establish standardized data services.

Keywords: multidimensional space-time data, NetCDF, collaborative data sharing, HydroShare, cyberinfrastructure

---

[1] Coauthored by Tian Gan, David G. Tarboton, Jeffery S. Horsburgh, Pabitra Dash, Ray Idaszak, and Hong Yi.

3.1 Introduction

With advances in hydrologic monitoring and model simulation technologies, hydrologic research has become data and computationally intensive, resulting in large volumes of scientific data generated or collected by individual researchers and organizations. However, advances in hydrologic understanding now tends to require discovery, access, and integration of heterogeneous and dispersed data from multiple sources. Moreover, large-scale hydrologic problems often need to be solved by collaboration among researchers, thus working as a team to collaborate around data has become indispensable. These emerging trends in hydrologic research are key drivers that demand new tools to support the entire research cycle of data creation, discovery, access, curation, publication, and analysis to help achieve new scientific breakthroughs (Hey et al., 2009; Hey and Trefethen, 2005).

The Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI) has devoted great effort to the development of cyberinfrastructure (CI) to satisfy this need, including HydroShare (http://www.hydroshare.org), a next generation Internet-based Hydrologic Information System (HIS) (Tarboton et al., 2014). HydroShare was developed to extend the capability of the earlier, server based CUAHSI HIS, which focused on the sharing of point observation time series data (Horsburgh et al., 2008, 2009; Tarboton et al., 2009). Given that the needs of hydrology researchers go well beyond time series data, HydroShare was established to add support for sharing a broader range of hydrologic datasets and models that are widely used in the hydrologic science community. These include time series, geographic raster, geographic feature, multidimensional space-time data, model instances, and model programs (Morsy et al., 2017). As a first step in HydroShare development, a data model was designed that enabled storing, transmitting, and cataloging of resources comprised of these diverse hydrologic data types and models to facilitate discovery (Horsburgh et al., 2015). Details for each of HydroShare's supported data types were also specified, including required file format and content, metadata elements, and functions for data processing, analysis, or visualization. HydroShare's resource data model was designed to generalize the way datasets and models were managed and shared, while at the same time supporting specific metadata elements and functions required for enhancing the hydrologic analysis capability and interoperability for different data types and models.

HydroShare enables users to upload datasets stored in a recognized file format and annotate them with metadata. For known content types, available functions can be used to visualize or analyze the datasets for further insights. Functions exist for data discovery, access, versioning, and formal publication, as well as social functions for commenting on, rating, and managing access to the shared datasets. Thus, the shared datasets and models in HydroShare are "social objects" that can be published, collaborated around, annotated, discovered, and accessed (Horsburgh et al., 2015).

One very important and ubiquitous data type used in hydrologic modeling research is multidimensional space-time data. This type of data is usually derived either from computational hydrologic models or from observations to represent the values for a physical phenomenon over a geospatial region within a time period. Examples include time slices of spatially distributed precipitation, temperature, wind speed, humidity, or snow water equivalent. While commonly used, there are several challenges associated with multidimensional space-time data that can make data sharing more difficult. For example, there is no single, accepted file format for storing this type of data to support the interoperability of data sharing and analysis. They are often stored and distributed in file formats such as Common Data Format (CDF http://cdf.gsfc.nasa.gov/), Hierarchical Data Form (HDF https://www.hdfgroup.org/), and Network Common Data Form (NetCDF http://www.unidata.ucar.edu/software/netcdf/). File size can also be a challenge. Multidimensional space-time datasets can be quite large, making it inconvenient to download large files when potential users may need only a subset or slice of them. Recognizing these challenges, this paper describes our efforts to establish functionality to support the sharing of multidimensional space-time data in HydroShare.

Currently, several websites and software tools can be used for sharing multidimensional space-time data, and each has its own strengths and limitations. For example, general-purpose file sharing systems such as Google Drive (https://www.google.com/drive/) and Dropbox (https://www.dropbox.com/) may be used to exchange multidimensional space-time data files. Users can easily upload and privately share preliminary and intermediate data products with these systems. However, they do not support permanent data publication, and little or no metadata is captured for the shared datasets. In addition, anyone other than the dataset creator would have difficulty discovering or accessing the datasets because no public metadata cataloging or search services are provided.

Figshare (http://figshare.com/) is a website that enables users to manage their research output in the cloud to be stored, shared, published, and discovered. It supports permanent data publication and provides citation information for shared datasets to give the data provider credit and make their datasets citable. Figshare also supports social functions such as commenting and access control to facilitate collaboration around the datasets. However, although Figshare has functions to capture simple metadata and preview file contents for commonly used file formats such as Microsoft Word, PDF, and Microsoft Excel, no functions are provided to preview or edit the metadata or contents of the more advanced, scientific data formats used for multidimensional space-time data (e.g., NetCDF and CDF). These limitations hinder users' ability to describe, preview, access, and interpret file contents through the website, which can be a barrier to data sharing, inhibiting data reuse and the reproducibility of scientific analyses.

The Thematic Real-time Environmental Distributed Data Services (THREDDS) and Hyrax data servers can provide cataloging functionality and support access to metadata and data for scientific datasets through various data access protocols (OPeNDAP, 2017; Unidata, 2017a). Multidimensional space-time data stored in file formats such as CDF or NetCDF can be stored and served with a single installation of these data servers. The THREDDS catalog and the Open-source Project for a Network Data Access Protocol (OPeNDAP) service are common services supported by these two data servers. THREDDS catalogs are logical directories of available online datasets that help discover data. The OPeNDAP services enable to subset or preview the contents of remote datasets and metadata. Moreover, OPeNDAP client software programs exist that can help retrieve remote datasets for analysis and visualization. These include NetCDF Operators (NCO) (Zender, 2008), Integrated Data Viewer (IDV) (Unidata, 2017b), and Panoply (NASA, 2018), etc. While this software stack can provide powerful and performant access to large volumes of multidimensional space-time data, one limitation is that sharing data requires server hardware and software to be set up and maintained. In many cases, adopting this approach would not be practical, especially with small amounts of data or small research groups with limited information technology expertise or server resources. Additionally, the web server provides limited search capabilities, which may prevent or impede scientists from discovering datasets using search terms or the geolocation of the dataset, etc.

The Repository for Archiving, Managing and Accessing Diverse Data (RAMADA) (http://ramadda.org/) is another web-based application framework that provides a broad suite of services for content and data management, publishing and collaboration. With RAMADA, users can search, access, upload, or comment on datasets. The system incorporates the OPeNDAP service and data analysis tools to provide functions for file content preview, metadata capture and curation, and data subsetting and analysis for multidimensional space-time data. However, as with the THREDDS and Hyrax data servers, sharing multidimensional space-time data with RAMADA requires setting up and maintaining the services, which may make sharing research datasets impractical for individual researchers or small research groups.

The HydroShare multidimensional space-time data representation design and implementation reported in this paper was developed to address and overcome some of these limitations of existing methods. It provides functionality to help share multidimensional space-time data to promote data curation, publication, and reuse. The approach is different from that of other data sharing systems that take scientific datasets in various file formats as generic file objects with basic data sharing functionality. Rather, for HydroShare we developed functionality to support data management and reuse based on the features of each specific data type supported. This includes selection of a file format and design of type specific metadata elements. We also designed type specific functions to facilitate data visualization, processing, and metadata management.

This approach enables users to share multidimensional space-time data in the NetCDF file format and annotate them with descriptive metadata in HydroShare. Users can access shared datasets in HydroShare for file content preview, visualization, and processing. When made public, the data are automatically registered in a HydroShare-connected THREDDS server that enables access using the OPeNDAP service without data providers being required to provision any server hardware or install or configure any software. In addition, with HydroShare's inherent data discovery, versioning, publication, and social functions, users can collaborate around datasets from initial data preparation to final data publication, and the sharing, discovery, and reuse of multidimensional space-time data is simplified.

In this paper, we describe the design, development, and testing of this approach. We first introduce the HydroShare system and briefly describe its functionality, system architecture, and resource data model design. Next, we discuss our selection of the NetCDF file format to represent multidimensional space-time data in HydroShare and the metadata elements used to describe this data type. Then, we describe the functions developed based on this file format to enable remote data access for content preview, subsetting, visualization, and processing, as well as metadata capture and editing. Next, we provide an experimental case study in which we tested this approach to demonstrate the functions in HydroShare that can facilitate collaboration among users for data preparation, publication, and reuse. Finally, we summarize the work and describe future directions for development.

3.2 Background

The goal for this work was to develop functionality for HydroShare to share multidimensional space-time data and provide a collaborative environment for curation, publication, discovery, and reuse of this data. The design and system architecture of HydroShare are extensible, which allows developers to add new elements for different hydrologic data types and incorporate them with existing system functions as was done here for multidimensional space-time data. We took advantage of the data sharing and social functions already developed within the HydroShare system and built additional functionality to support multidimensional space-time data sharing. Thus, a brief overview of HydroShare is given to provide context for the work described in this paper.

HydroShare is a web based hydrologic information system, operated by CUAHSI, in which anyone can create an account to share datasets for research. A goal of HydroShare is to advance hydrologic science by enabling the scientific community to more easily and freely share products resulting from their research — not just the scientific publication summarizing a study, but also the data and models used to create the scientific publication. HydroShare provides functionality for metadata capture and curation, data manipulation, data publication, data discovery, and collaboration, including access control functionality for sharing with individuals, groups, or the public, and for enhancing the social value of resources through commenting and rating. Together, these functions represent a new paradigm in data sharing systems, supporting discovery through the integration of information from multiple sources, team work, collaboration, reuse of data, and transparency to enhance trust in research findings.

HydroShare's system architecture is centered on several open source components

(Heard et al., 2014). Figure 3.1 demonstrates how these components interact with each other. The major components include Django (https://www.djangoproject.com) and iRODS (http://iRODS.org/). Django is an open source Python web framework that provides the functionality used to build the web user interface to help users interact with HydroShare and manage their shared datasets and models. iRODS is a data system that supports data storage shared between distributed servers. Additionally, HydroShare's Representational State Transfer (REST) application programming interface (API) and iRODS interface (e.g., iRODS python API) enable web applications, or Apps, deployed by the HydroShare development team or third party organizations (Rajib et al., 2016) to interact with the Django and iRODS components. This design enables HydroShare, in addition to providing core collaboration and data sharing functionality via its primary website, to support additional web services that interact via HydroShare's RESTful web services to enhance the capability for data analysis and visualization as well as model simulation.

The HydroShare resource data model was designed and implemented to manage various types of hydrologic datasets and models (Horsburgh et al., 2015). A HydroShare resource is the granular unit of shared content for access control, serialization for transport over the Internet, and cataloging for discovery within the system. A resource consists of a resource-level metadata file and resource content files. The resource-level metadata file is encoded using extensible markup language (XML) and is generated by HydroShare from user-created metadata that documents the resource. The resource content files may include single or multiple files that make up different hydrologic datasets or models uploaded into HydroShare by users. Additional informational or readme files for the datasets or models may also be included as resource content files. In addition, HydroShare's resource data model allows for definition of specific content types for widely used and well-known hydrologic data types and models, which extends the core resource data model through specification of a content type data model. The resource content files associated with a supported content type are grouped together into an OAI-ORE aggregation (Lagoze et al., 2008). The content type data model defines the file format and contents of the aggregated files along with aggregation-level metadata that provides the provenance information and describes the file contents.

Figure 3.2 shows the organization of an example HydroShare resource. This example includes a physical file (a Microsoft Word document) and different content types representing time series, geographic features, and multidimensional space-time data. Each content type includes a single or multiple file(s) aggregated to represent one logical object (or dataset) and the aggregation-level metadata created as an XML file by HydroShare. The advantage of the resource data model design is that HydroShare can manage (e.g., storage on disk, packaging for delivery over the Internet, access control, and cataloging for discovery) multiple types of datasets and models in the same way, regardless of the file formats and contents. Meanwhile, a content type data model enables users to standardize file formats and syntax and to add additional metadata to describe the hydrologic dataset or model. Developers can then use the standardized file formats and metadata to create advanced functions to facilitate metadata management, data analysis, or visualization. This paper specifically reports the design of the content type data model for multidimensional space-time data and serves as an example demonstrating how to extend the HydroShare resource data model with a new content type.

3.3 Methods

We designed the multidimensional content type in HydroShare to support the sharing of multidimensional space-time data. We first specified the format and syntax of the files that comprise a multidimensional aggregation and the metadata elements associated with the aggregation. Then the multidimensional content type was implemented within the HydroShare system. In addition, we developed automated functions to extract metadata from the uploaded files (e.g., extracting metadata automatically from uploaded files rather than having users enter it manually) and to set up the OPeNDAP web services to facilitate remote data access and subsetting for data preview, analysis, and visualization. Finally, we validated the design with an experimental use case to demonstrate how sharing multidimensional space-time data in HydroShare can help users collaborate around datasets to facilitate the activities involved in the data management life cycle. The detailed methods are described in the following sections.

3.3.1 Multidimensional content type data model

3.3.1.1 Content type files

Since there are many scientific file formats capable of storing multidimensional space-time data, we evaluated the benefits and tradeoffs of these file formats and chose the one that we felt most suitable for data storage and management in HydroShare. We established the following criteria to decide the file format used to represent multidimensional space-time data in HydroShare. First, the data stored in the file needed to be organized in a way that helps understand the data structure and retrieve a subset for analysis. Second, the file format needed to be widely recognized and used in hydrologic research, with available open source software or libraries to help analyze or visualize the data. Third, widely accepted standards needed to be available to guide users in how to organize the data contents and metadata in the file to promote interoperability for data processing and sharing.

Based on these criteria, we compared CDF, HDF5, and NetCDF file formats and adopted the NetCDF file format to represent multidimensional space-time data in HydroShare. These three file formats use similar data models to hold multidimensional space-time data. Open source software programs for these file formats are also available to support data analysis or visualization. The reasons for selecting NetCDF were its wide use in modeling research in hydrology and aligned fields such as atmospheric science, its adoption as a standard (OGC, 2011) and support for standards for its metadata (Eaton et al., 2017; ESIP, 2017).

A NetCDF file usually includes three components: dimensions, variables, and attributes. Dimensions may be used to represent one or more physical dimensions such as time, latitude, longitude, or height. They may also be used to index other quantities like station (e.g., the location of a monitoring site) or model run number. Variables are used to store an array of values to represent a physical phenomenon such as precipitation, temperature, or snow water equivalent. The array shape of each variable is defined by the dimensions, and different variables can be defined with different array shapes. Attributes are used to store metadata information. Global attributes provide information about the NetCDF dataset as a whole. These may include information about the data creator, a descriptive abstract, key words, and spatial and temporal coverage, etc. Variable attributes are used to provide information for a specific variable, such as the variable unit and variable descriptive name.

The NetCDF file format supports the creation of array-oriented datasets and can contain metadata that makes them self-describing and helps researchers understand the structure and the properties of the data. This file format is widely used to represent multidimensional space-time datasets as the input or output for hydrologic models (David et al., 2011; Sen Gupta et al., 2015; Thornton et al., 1997). It has also been used for data management and curation of data converted from other file formats (Guo et al., 2015). Moreover, many software programs and libraries available for NetCDF data processing, analysis, or visualization are widely applied among the research community (Unidata, 2017c). Thus, researchers with these tools can easily manipulate NetCDF files. This capability was also an advantage that enabled us to develop the functions in HydroShare without starting from scratch. Furthermore, several conventions are available to promote the processing and sharing of data in NetCDF format. The Climate and Forecast (CF) (Eaton et al., 2017) convention specifies how to define the dimensions, variables, and attributes to represent multidimensional space-time data as regular grid data or point time series. Additionally, the Attribute Conventions for Data Discovery (ACDD) (ESIP, 2017) were designed to define the metadata attributes to describe the whole NetCDF dataset to a discovery system. These attributes can be extracted from the file and stored in a data sharing system to support data discovery or data processing.

In HydroShare, aggregated files of a specific content type may consist of one or multiple files used to represent one logical object (or dataset). Thus, we specified that a multidimensional aggregation should include only one NetCDF file uploaded by the user and one metadata header information text file automatically generated by the system from the uploaded file to provide a brief summary of the contents in the NetCDF file.

For the NetCDF file, it is recommended that users define the dimensions, variables, and attributes by following the CF and the ACDD conventions. HydroShare does not prevent users from sharing multidimensional space-time data in NetCDF files that do not follow these conventions. However, the functions developed to harvest metadata were based on these conventions and when they are not followed metadata will not be automatically extracted and users will need to enter it manually.

The metadata header information text file is generated using the NetCDF “ncdump -h” command, and is in a format that is widely recognizable to researchers who are familiar with NetCDF. This text file includes information about the defined attributes and data structures extracted from the NetCDF file and can provide users a brief summary of the file contents without needing to download the full data file that may be large.

3.3.1.2 Content type metadata

A HydroShare resource holding a multidimensional aggregation has two sets of metadata elements. One is the resource-level metadata that are the standard Dublin Core metadata elements common to all HydroShare resources describing the general attributes of a resource (title, creator, abstract, etc.). The other consists of content type metadata (or aggregation-level metadata) that are designed to describe the multidimensional aggregation. The content type metadata includes general elements to capture the basic information of any content type (e.g., keywords and coverage), and extended elements to capture the data features in the NetCDF file (e.g., spatial reference and variable information). Figure 3.3 shows resource-level and content type metadata elements and example metadata information for a resource containing a multidimensional aggregation. Some of the metadata elements also contain sub-elements. For example, the “netcdfVariable” metadata element includes sub-elements to describe the name, data type, units, etc., for a given variable.

In designing the multidimensional content type, we chose to extract metadata elements held within the NetCDF file and explicitly list them as the resource-level or content type metadata for two main reasons. First, this made it easier to present the full metadata description on a resource’s landing page in HydroShare, which is the web page for the user to view and manage the resource, making it more accessible to potential users of the data (e.g., potential users are not required to download or open the NetCDF file to learn about its contents). Second, explicit metadata helps HydroShare (and potentially other web services) catalogue the information to enable data discovery or facilitate interoperability for data processing and analysis functions.

3.3.1.3 Content type implementation

In HydroShare, a general pattern can be followed to add a new content type. A new content type will inherit from the abstract content type, and the new content type metadata will inherit from the abstract content type metadata. Since the abstract content type metadata includes general elements that apply to all content types, the extended metadata elements are added by inheriting from the abstract metadata element class. A new content type will also include specific data or metadata functions to, for example, provide functionality such as automatically harvesting metadata from data files or updating data files with user edits to metadata in HydroShare. Moreover, since a content type can’t exist independently outside of a resource, a “composite” resource type was implemented in HydroShare as a container for different content types with in a resource.

Given this general extensibility pattern, we implemented a new content type to manage multidimensional space-time data in HydroShare. A UML diagram of the logical database design for the multidimensional content type within a composite resource in HydroShare is shown in Figure 3.4. This presents only major classes, attributes, and methods and demonstrates the organization of this content type in HydroShare.

In this diagram, the green frame contains the classes that define the composite resource type (CompositeResource class) and its corresponding resource-level metadata elements (ResourceMetadata class). The red frame contains the classes that define the new content type, which include four main categories: (1) the abstract classes, including the AbstractContentType class, the AbstractContentTypeMetadata class, and the AbstractMetadataElement class that are inherited by any new content type; (2) the class to define the new content type (MDContentType); (3) the class to manage the content type metadata (MDMetadata); and (4) the classes to define the extended metadata elements of the content type (e.g. SpatialReference class and NetcdfVariable class).

The CompositeResource class defines the composite resource type, which manages all the resource content files and provides data access control, data publication, and social functions for the resource. This class also contains the ResourceMetadata class that manages the resource-level metadata and creates the XML metadata file. Additionally, the CompositeResource class can include different content type classes (multidimensional as developed here, time series, geographic raster, etc.) to manage different types of hydrologic datasets in the resource.

The MDContentType class inherits from the AbstractContentType class, which is the abstract class that provides the interface to represent a content type in HydroShare. The AbstractContentType class includes the properties and methods for the system to manage a content type and provides a common interface to enable the content type related functions. Functionality specific to the multidimensional content type had to be developed by overriding some methods of AbstractContentType class. For example, the AbstractContentType class has a set_file_types() method that was overridden in the MDContentType class, which is used to check the uploaded multidimensional space-time data in NetCDF file format to create a multidimensional aggregation.

The MDMetadata class inherits from the AbstractContentTypeMetadata class, which is the base class used by all content types to manage the content type metadata elements in the system. By inheriting from the abstract class, the MDMetadata class only needs to contain classes that represent the extended metadata elements, which are the SpatialReference and the NetcdfVariable classes. These classes all inherit from the AbstractMetadataElement, which is the base class used to represent a metadata element and defines its sub-elements and methods. In addition, the get_xml() method and has_all_required_elements() methods in the MDMetadata class override the corresponding methods from the abstract class. The get_xml() method is used to generate the content type metadata XML file for multidimensional content type. The has_all_required_elements() method is used to check if the required multidimensional content type metadata elements are provided by the user before the resource is shared to the public.

3.3.2 Additional content type functions

As described above, HydroShare provides a base set of functionality for each resource that includes access control, publication, social functions, etc. However, one of the advantages of the design and implementation we describe here is that additional functionality can be developed for a specific content type to support specialized metadata management and sharing of the data via content type specific web services. In the following sections, we describe how this functionality was created for the multidimensional content type.

3.3.2.1 Metadata management functionality

In order to simplify the work required to record the metadata for sharing multidimensional space-time data in HydroShare, we designed two functions to (1) extract information (where it exists) from the NetCDF file to populate the resource and content type metadata elements, and (2) generate the metadata header information text file. When a user uploads a file with the ".nc" extension, HydroShare will test whether the file holds valid NetCDF content, and if successful, execute these functions to create a multidimensional aggregation from the file.

We used the NetCDF utility "ncdump" and NetCDF4 Python library to implement these metadata extraction functions. Furthermore, we established a mapping between HydroShare's metadata elements and the ACDD and CF conventions (Table 3.1). Thus, for files that follow either of these conventions, the automated metadata extraction function retrieves and populates matched HydroShare metadata elements. Additionally, for files without ACDD metadata elements, but with spatial or temporal coordinate variables given following the CF conventions, spatial and temporal coverage metadata elements determined by reading these data variables are populated in the content type and resource coverage metadata.

We also implemented functionality for editing the metadata in the NetCDF file through HydroShare. When a user edits the metadata in HydroShare, the system utilizes the metadata mapping (Table 3.1) to check for consistency between the NetCDF file and the HydroShare metadata. If there is a need to update the metadata in the NetCDF file, the system will notify the user, and the user can have the system update the file based on the new metadata edits. This functionality helps a user easily update the NetCDF file without having to download and manually edit it. When the initial file includes little metadata, this functionality makes it easy to create metadata in the file that follows NetCDF conventions.

The metadata editing functionality described above was implemented using the NetCDF4 Python library and HydroShare iRODS client interface (Figure 3.1). When metadata needs to be updated, the system first copies the original NetCDF file from iRODS to a temporary folder. Second, the system writes HydroShare's metadata into the copied file using the NetCDF4 Python library. Then, the system generates a new metadata header information text file from the updated copied file. Finally, the system replaces the original NetCDF file and the metadata header information text file in iRODS with these newly created files.

3.3.2.2 OPeNDAP service

OPeNDAP services add value to the web sharing of NetCDF files by enabling users to access the data for previewing, visualization, and processing from programs that consume these services such as NCO and Panoply. These services help users learn about and work with the contents of the datasets without being required to download them first. They also enable users to retrieve a subset of the data for use cases that require smaller spatial or temporal extent. To provide this capability for HydroShare users, we automated the process of creating an OPeNDAP web service for all publicly shared multidimensional space-time data in HydroShare. Users can access and subset the dataset stored in HydroShare through an OPeNDAP data access form in a web browser or through existing OPeNDAP client software for data visualization or analysis.

In HydroShare, support for OPeNDAP services was created by setting up a THREDDS data server to interact with HydroShare's iRODS file storage system. In the system architecture shown in (Figure 3.1), the data server plays the role of a web service provided as part of HydroShare's "Actions on Resources" functionality. The data server requires direct file system access to the NetCDF files for its OPeNDAP services. Thus, we used existing iRODS client software to interface to the iRODS Network file system (yellow arrow connecting the orange and purple frames in Figure 3.1). We developed a script that copies HydroShare public resources containing multidimensional aggregations using the iRODS "iget" command to a directory on the data server. This copying occurs: 1) when access control for a private resource is changed to public; and 2) when the time stamp of a public resource on the data server is older than that in HydroShare and a data update is needed. This use of "iget" takes advantage of iRODS' high performance data transfers, but in the present implementation does require duplicate storage of NetCDF files. Moreover, since the data server does not support file level user access control as would be required for access to private files in HydroShare, the OPeNDAP service is limited to NetCDF files stored in public or formally published resources in HydroShare. This functionality saves users the work that would be required to set up a server to host OPeNDAP services for their datasets and gives them the freedom to control when to make their datasets accessible via OPeNDAP services by using HydroShare's access control settings.

3.3.3 Case study design

Figure 3.5 depicts the cycle of activities involved in collaborative research and the HydroShare functions that support this. HydroShare enables users to incrementally add metadata to the initial dataset to prepare and describe it for permanent publication. Sharing data in HydroShare also enables users to easily discover and access datasets for reuse. The availability of detailed metadata can assist potential data users in determining whether the data are appropriate for reuse, and the availability of the OPeNDAP service means that potential data users can access and retrieve a subset of the data for visualization or analysis to derive new results. With social functions that include access control, commenting on, and rating of resources, HydroShare provides a collaborative environment in which users can work together to edit or describe datasets for formal publication or to communicate, evaluate, and iterate on datasets to improve data quality and potential for reuse.

As a method for evaluating HydroShare's capability to enable collaborative research around the workflow shown in Figure 3.5 that focuses on multidimensional space-time data, we considered the case where a researcher simulates the snowmelt process for the Dolores River watershed in the Colorado River Basin from 1988 to 2010. This was part of a study that the authors were involved in on snowmelt modeling and operational water supply forecasting within the Colorado River Basin. The model used in this study initially stored snow water equivalent output as separate two-dimensional geospatial data files for each 6-hour time step. This results in thousands of model output files for a 22-year simulation. Sharing of these original model output files has limitations that make data management and reuse difficult. First, information may be lost if any file is missed during the file transfer process. Second, when the original model output files are in a format not widely used by the research community, it is inconvenient to extract subsets that involve thousands of files and difficult to find available software for data analysis or visualization.

Thus, we developed a Python script to reorganize and convert the multiple original model output files into one NetCDF file, which includes simulated results for snow water equivalent with CF conventions used to define the dimensions and variables. Additionally, metadata elements from ACDD convention were added to the NetCDF file as global attributes that describe the whole dataset (e.g., “title” and “keywords”). Figure 3.6 shows the data structure and the attributes defined in the NetCDF file. This file includes three dimensions to represent the spatial and temporal dimensions (“time,”“x,” and “y” dimensions). It also includes five variables, one of which stores the snow water equivalent data (“ueb_swe” variable). There are three variables that store the spatial and temporal coordinate data for the three dimensions (“x,”“y,” and “time” variables) and the last variable holds spatial reference information (“polar_stereographic” variable).

Once this data was organized in a single NetCDF file, it was shared via HydroShare to enable others to discover and access the data, and add additional metadata, or use it in further analysis. The results for this case study presented below are used to validate the functions for sharing multidimensional space-time data in HydroShare to support data reuse. This case study involving multiple hypothetical users, was implemented by the first author acting as these users from separate HydroShare accounts.

3.4 Results

3.4.1 HydroShare resource landing page and basic functions

Upon uploading the case study dataset into an empty HydroShare composite resource, the type of data file was automatically recognized and a multidimensional aggregation was automatically created in the resource. HydroShare generated a resource landing page for this composite resource, which shows the resource content files and metadata as well as different functions for the user to interact with the system to manage the resource.

In this resource landing page (Figure 3.7), there are buttons to trigger the functions for editing, managing access, deleting the resource, creating a new version, copying the resource as a new resource, and formal publication. The functions for managing access help the user control whether a dataset is private and shared only with trusted colleagues to prepare and annotate the dataset, or whether a dataset is exposed to the public for anyone to discover and access. The data versioning functionality can help the user manage the shared datasets with multiple versions when the original dataset evolves. The publication function is used to formally publish the final data product with an assigned digital object identifier (DOI) in HydroShare. The suggested citation information is also provided to encourage proper citation of this dataset. Additionally, this resource landing page provides the commenting and rating functions for users to communicate with each other and evaluate the shared datasets. The metadata panel at the right of the contents area shows the content type metadata for the multidimensional aggregation, in which the title, keywords, spatial/temporal coverage, spatial reference, and variable metadata were automatically extracted from the NetCDF file.

After the multidimensional aggregation was created in the resource, the user can manage access to share the resource only with trusted collaborators. One example could be a collaborator who edits the aggregation's metadata to correct information in the NetCDF file or add information that is not in the NetCDF file. For instance, after a collaborator added information in the metadata panel, HydroShare's consistency check identified the presence of newly added metadata and showed an "Update NetCDF File" button (Figure 3.7) to inform the user that the NetCDF file could be updated with the new information. Then, the collaborator clicked the button to have HydroShare update the metadata in the NetCDF file. This is an example of how, using HydroShare, multiple users can collaborate to annotate the resource with metadata. This metadata editing function also enhances NetCDF files to have more attributes that follow NetCDF conventions.

After editing the metadata elements in HydroShare, the resource was made public, and another user discovered it using HydroShare's search and filter functions (Figure 3.8). This user provided a search term, and HydroShare listed matching resources by querying the HydroShare metadata elements such as title, abstract, and keywords. This user also used HydroShare's map search function to determine the geographic location associated with this dataset. HydroShare can also query the spatial coverage metadata to identify resources that match coordinates input by the user. In addition, search results can be filtered based on different metadata facets, such as content type, author, and subject.

3.4.2 OPeNDAP service

When the resource was made public, the OPeNDAP service was enabled for the case study dataset. Any user can use the OPeNDAP service to access and subset this dataset for analysis or visualization. Generally, there are two ways to use it: 1) through the OPeNDAP data access form in a web browser, or 2) through OPeNDAP client software.

In HydroShare, the user can right click on the multidimensional aggregation folder or the NetCDF file within the folder to directly open the OPeNDAP data access form for the dataset (Figure 3.9 (a)), which allows users to preview or download a data subset from the NetCDF file through the website. In this data access form (Figure 3.9 (b)), users can select the variable names and specify the spatial and temporal dimension indexes to subset the dataset. Moreover, the data access form also provides the "Data URL" that can be used in OPeNDAP client software to access and subset the dataset for visualization or analysis.

For example, consider a user who discovered this resource in HydroShare and wanted to reuse a subset of the model results for water year 2009. This user could use the OPeNDAP service and different client software for data visualization and analysis without downloading the whole NetCDF file to a local computer. As a demonstration of this, Figure 3.10 (a) shows a two-dimensional graph of the distribution of snow water equivalent in the test watershed at a single time step generated by entering the OPeNDAP "Data URL" into the Panoply visualization tool. As the case study dataset was saved in the NetCDF file format following the CF convention, Panoply can easily interpret the data contents and retrieve the subset via the OPeNDAP service for visualization. The Panoply user only specified the dimension index information in the software for data subsetting, and the software then automatically retrieved the data from HydroShare via the OPeNDAP service to generate the plots.

There are also some free tools and libraries for data processing and analysis for NetCDF files such as NCO and the NetCDF4 Python library. For instance, NCO can be used to programmatically access and subset the discovered dataset for analysis. Figure 3.10 (b) shows the NCO commands used to access, subset, and process the case study dataset using the OPeNDAP service. The code first subsets the data from January 1st to May 31st, 2009, identifies the maximum snow water equivalent for each grid cell within this period, and writes the result to a new NetCDF file (max.nc). This provides the model result for maximum snow accumulation (assumed to occur within this period) for that year. The code then retrieves the data for April 1st and April 15th (april_1.nc and april_15.nc) and evaluates the snow water equivalent difference between the two dates to create a new NetCDF file (diff.nc). This provides the analysis result for accumulation (increase) or ablation (decrease) during this period. Water managers often track such snow water equivalent changes in water supply forecasts.

This use case demonstrated the activities shown in the collaboration cycle depicted in Figure 3.5, and how to use OPeNDAP and client software for data analysis in HydroShare for collaborative research. After the original user organized the model output files into one NetCDF file and shared it in HydroShare (Gan, 2019a), other users were able to directly subset the data for visualization and analysis without downloading the whole dataset to local computers. This way to share the data makes it more convenient for data analysis when compared with sharing thousands of model output files in a not widely used file format. Additionally, the data analysis code and the derived NetCDF files can be uploaded into HydroShare as a new resource to support data reuse and improve research reproducibility (Gan, 2019b).

3.5 Discussion

The case study illustrated how organizing multidimensional space-time data using the NetCDF file format and sharing it in HydroShare provided added value in terms of functionality for metadata management, data analysis, and visualization. When compared with other data sharing methods for multidimensional space-time data, this approach has several advantages.

First, this approach provides functionality to capture, expose, and edit the metadata stored in the NetCDF file. The creator or manager of the dataset can add metadata through forms in a web browser and have it encoded following widely used conventions. Metadata is more accessible to potential data consumers on the resource landing page, header text file, and via the machine readable metadata XML files. Other data sharing methods such as Dropbox, Google Drive, or Figshare do not automatically expose the metadata from a NetCDF file for viewing or editing, making it harder to read, edit, and understand the file contents and determine appropriate uses for the data. Although THREDDS or RAMADA can expose the metadata, it is difficult to edit the metadata in the file directly. Moreover, the manage access function in HydroShare enables users to collaborate on metadata editing and thus improve its description of the data.

Second, this approach provides OPeNDAP services for shared datasets, which support data analysis, visualization, and reuse that enhance opportunities for collaboration around the data in the derivation of new results or data products. In HydroShare, users have the freedom to decide when to expose shared datasets through an OPeNDAP service by simply changing the resource sharing status. They also do not need to setup and maintain the data server themselves. Other available methods either do not provide an OPeNDAP service or require effort to set up and maintain a server and service.

Third, this approach provides better data discovery functionality for the shared datasets. For example, Hyrax or THREDDS servers require the users to know the naming and directory hierarchy of files. Our approach improves data discovery functionality by supporting keyword and geolocation searches based on a catalog of metadata extracted from the NetCDF file or input by the data provider.

In addition, there is other HydroShare functionality useful but not available for some other exiting methods for sharing multidimensional space-time data. For example, HydroShare’s data publication functionality helps users formally publish their datasets and obtain a citable DOI, which can formally link published datasets with published research manuscripts to enhance reproducibility and help others cite published datasets. This supports users in receiving citation credit for their data.

In our approach, two key factors make this advantageous functionality available. First, we adopted a standard file format (NetCDF) to organize multidimensional space-time data. This file format has conventions that standardize how data and metadata are organized in the file to improve the interoperability of datasets. Based on this file format, we utilized existing tools and standard data services to develop additional functions for metadata management, data analysis, and visualization to promote data reuse. Second, we created automated functionality (e.g., metadata extraction, OPeNDAP service creation) that makes sharing of multidimensional space-time data easier. HydroShare's data sharing functionality applies to all data types, while at the same time allowing value added functionality for specific data types. This system design helps improve consistent data discovery, access, and publishing across the broad range of data types used by scientists in the hydrology domain.

However, there are limitations that need further improvements for sharing multidimensional space-time data in HydroShare. One limitation is that the NetCDF file format may be a bit obscure to some users. As with any file format, users need to learn how to organize multidimensional space-time data in this file format for data sharing. One way to facilitate this would be to make the system support automatic file format conversion to transfer data from other formats into NetCDF format. Another limitation is web-based visualization. There is a need for additional web applications developed to provide researchers with greater capacity to process and visualize datasets directly without transferring the data or subsets of the data between the data sharing system and their local computers.

3.6 Conclusions

HydroShare is a web based hydrologic information system that provides researchers with a platform to share their hydrologic data and models. As multidimensional space-time data is one of the widely used data types in hydrologic research, we developed an approach to support sharing of this data type within HydroShare.

This work has demonstrated sharing multidimensional space-time data in a standard file format (NetCDF) and with value added functions, which are supported in the framework of HydroShare’s resource data model and web based collaboration platform to enhance analysis, visualization, and reuse of this data. In concert with existing HydroShare functionality, the work described here enables relatively straightforward sharing and formal publication of multidimensional space-time data. This increases transparency and reproducibility of the associated research. This also enables and promotes reuse of data, and the derivation of additional value from research data investments.

We demonstrated how the new functionality developed solves issues faced by researchers who are using alternative or more traditional methods of sharing this type of data, including difficulty in previewing or processing datasets without downloading them and the lack of advanced metadata editing, sharing, and social functions that encourage collaboration around shared datasets. In HydroShare, researchers can preview and edit the metadata for datasets in a NetCDF file and access or subset them with the automatically configured OPeNDAP service and existing tools. They can also discover datasets using HydroShare's flexible metadata-based data discovery capabilities. Along with other functions such as data versioning and social functions, researchers can manage their multidimensional space-time datasets and collaborate with colleagues for data preparation, description, publication, discovery, and analysis.

Beyond the context of the new functionality we have demonstrated within the HydroShare system, another contribution of this work is that the methods we developed for improving sharing of multidimensional space-time data can be used as examples for supporting other data types in HydroShare or for better supporting multidimensional space-time data in other systems. CI developers who are going to build or have built a data sharing system to support multidimensional space-time data sharing can use the recommendations of this work to organize data in a standard file format and document the datasets using the standards-based metadata. They may also be able to establish standard data services or develop new functionality to facilitate metadata management, data analysis, or visualization. Adopting standard formats and techniques across data repositories could lead to a level of interoperability that is worth considering in the future.

Acknowledgements

This work was supported by the National Science Foundation under collaborative grants OCI-1148453 and OCI-1148090. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We are also thankful to others on the HydroShare development team for providing suggestions on function design and implementation for HydroShare system.

References

David, C.H., Maidment, D.R., Niu, G.Y., Yang, Z.L., Habets, F., Eijkhout, V., 2011. River network routing on the NHDPlus dataset. J. Hydrometeorol. 12, 913–934. https://doi.org/10.1175/2011JHM1345.1

Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Blower, J., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., Pamment, A., Juckes, M., Raspaud, M., 2017. NetCDF climate and forecast (CF) metadata conventions. http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.pdf (accessed 8.18.19).

ESIP, 2017. Attribute convention for data discovery 1-3. http://wiki.esipfed.org/index.php/Attribute_Convention_for_Data_Discovery (accessed 8.18.19).

Gan, T., 2019a. UEB model simulation of snow water equivalent for test watershed in the Colorado River Basin from 1988 to 2010. HydroShare. https://doi.org/10.4211/hs.deb0250d446441fd886a2e288b5a8a90

Gan, T., 2019b. Data analysis of snow water equivalent for test watershed in the Colorado River Basin in 2009. HydroShare. https://doi.org/10.4211/hs.5f685bb517eb4343b1b293a921f14639

Guo, Q., Zhang, Y., He, Z., Min, Y., 2015. Web-based data integration and interoperability for a massive spatial-temporal dataset of the Heihe River Basin eScience framework. Adv. Meteorol. Volume 201. http://dx.doi.org/10.1155/2015/982062

Heard, J., Tarboton, D.G., Idaszak, R., Horsburgh, J.S., Bedig, A., Castronova, A.M., Couch, A., Frisby, C., Gan, T., Goodall, J., Jackson, S., Livingston, S., Maidment, D., Martin, N., Miles, B.C., Mills, S., Sadler, J., Valentine, D., Zhao, L., 2014. An architectural overview of HydroShare, a next-generation hydrologic information system. Proceedings of the 11th International Conference on Hydroinformatics, HIC 2014, New York City, USA.

Hey, T., Tansley, S., Tolle, K., Stweart, T., Tolle, K., 2009. Jim Gray on eScience: A transformed scientific method, in: Hey, T., Tansley, S., Tolle, K (Eds.), The Fourth Paradigm Data-Intensive Scientific Discovery. Microsoft, pp. 18–31.

Hey, T., Trefethen, A.E., 2005. Cyberinfrastructure for e-Science. Science 308, 817–21. https://doi.org/10.1126/science.1110410

Horsburgh, J.S., Morsy, M.M., Castronova, A.M., Goodall, J.L., Gan, T., Yi, H., Stealey, M.J., Tarboton, D.G., 2015. HydroShare: Sharing diverse hydrologic data types and models as social objects within a Hydrologic Information System,. J. Am. Water Resour. Assoc. 27517. https://doi.org/10.1111/1752-1688.12363

Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I., 2008. A relational model for environmental and water resources data. Water Resour. Res. 44. https://doi.org/10.1029/2007WR006392

Horsburgh, J.S., Tarboton, D.G., Piasecki, M., Maidment, D.R., Zaslavsky, I., Valentine, D., Whitenack, T., 2009. An integrated system for publishing environmental observations data. Environ. Model. Softw. 24, 879–888. https://doi.org/10.1016/j.envsoft.2009.01.002

Lagoze, C., Sompel, H. Van de, Johnston, P., Nelson, M., Sanderson, R., Warner, S., 2008. Open archives initiative object reuse and exchange: ORE user guide – primer. http://www.openarchives.org/ore/1.0/primer (accessed 8.18.19).

Morsy, M.M., Goodall, J.L., Castronova, A.M., Dash, P., Merwade, V., Sadler, J.M., Rajib, M.A., Horsburgh, J.S., Tarboton, D.G., 2017. Design of a metadata framework for environmental models with an example hydrologic application in HydroShare. Environ. Model. Softw. 93, 13–28. https://doi.org/10.1016/j.envsoft.2017.02.028

NASA, 2018. Panoply netCDF, HDF and GRIB data viewer. http://www.giss.nasa.gov/tools/panoply/ (accessed 8.18.19).

OGC, 2011. OGC Network Common Data Form (NetCDF) core encoding standard version 1.0. http://www.opengis.net/doc/IS/netcdf/1.0 (accessed 8.18.19).

OPeNDAP, 2017. The Hyrax data server installation and configuration guide. https://opendap.github.io/hyrax_guide/Master_Hyrax_Guide.html (accessed 8.18.19).

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. Environ. Model. Softw. 75, 498–512. https://doi.org/10.1016/j.envsoft.2015.10.032

Sen Gupta, A., Tarboton, D.G., Hummel, P., Brown, M.E., Habib, S., 2015. Integration of an energy balance snowmelt model into an open source modeling framework. Environ. Model. Softw. 68, 205–218. https://doi.org/10.1016/j.envsoft.2015.02.017

Tarboton, D.G., Horsburgh, J.S., Maidment, D.R., Whiteaker, T., Zaslavsky, I., Piasecki, M., 2009. Development of a community hydrologic information system, in: 18th World IMACS / MODSIM Congr. 988–994.

Tarboton, D.G., Idaszak, R., Horsburgh, J.S., Heard, J., Ames, D., Goodall, J.L., Band, L., Merwade, V., Couch, A., Arrigo, J., Hooper, R., Valentine, D., Maidment, D., 2014. HydroShare: Advancing collaboration through hydrologic data and model sharing, in: 7th International Congress on Environmental Modelling and Software. International Environmental Modelling and Software Society (iEMSs).

Thornton, P.E., Running, S.W., White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. J. Hydrol. 190, 214–251. https://doi.org/10.1016/S0022-1694(96)03128-9

Unidata, 2017a. THREDDS data server (TDS). http://www.unidata.ucar.edu/software/thredds/current/tds/ (accessed 8.18.19).

Unidata, 2017b. Integrated data viewer (IDV). http://www.unidata.ucar.edu/software/idv/docs/userguide/toc.html (accessed 8.18.19).

Unidata, 2017c. Software for manipulating or displaying NetCDF Data. http://www.unidata.ucar.edu/software/netcdf/software.html (accessed 8.18.19).

Zender, C.S., 2008. Analysis of self-describing gridded geoscience data with netCDF Operators (NCO). Environ. Model. Softw. 23, 1338–1342. https://doi.org/10.1016/j.envsoft.2008.03.004

Table 3.1 Mapping between HydroShare metadata terms and the NetCDF conventions metadata terms.

HydroShare metadata terms	**NetCDF conventions metadata terms**
creator: name	creator_name (ACDD)
creator: url	creator_url (ACDD)
creator: email	creator_email (ACDD)
contributor: name	contributor_name (ACDD)
coverage (temporal): start	time_coverage_start (ACDD)
coverage (temporal): end	time_coverage_end (ACDD)
coverage (spatial): northlimit	geospatial_lat_max (ACDD)
coverage (spatial): southlimit	geospatial_lat_min (ACDD)
coverage (spatial): eastlimit	geospatial_lon_max (ACDD)
coverage (spatial): westlimit	geospatial_lon_min (ACDD)
description	summary (ACDD)
relation: cites	references (ACDD)
rights	license (ACDD)
source	source (CF)
subject	keywords (ACDD)
title	title (ACDD)
identifier	id (ACDD)
netcdfVariable: unit	unit (CF)
netcdfVariable: descriptiveName	long_name (CF)
netcdfVariable: missingValue	missing_value (CF)
netcdfVariable: comment	comment (CF)
spatialReference: box	attributes for grid mapping variable (CF)

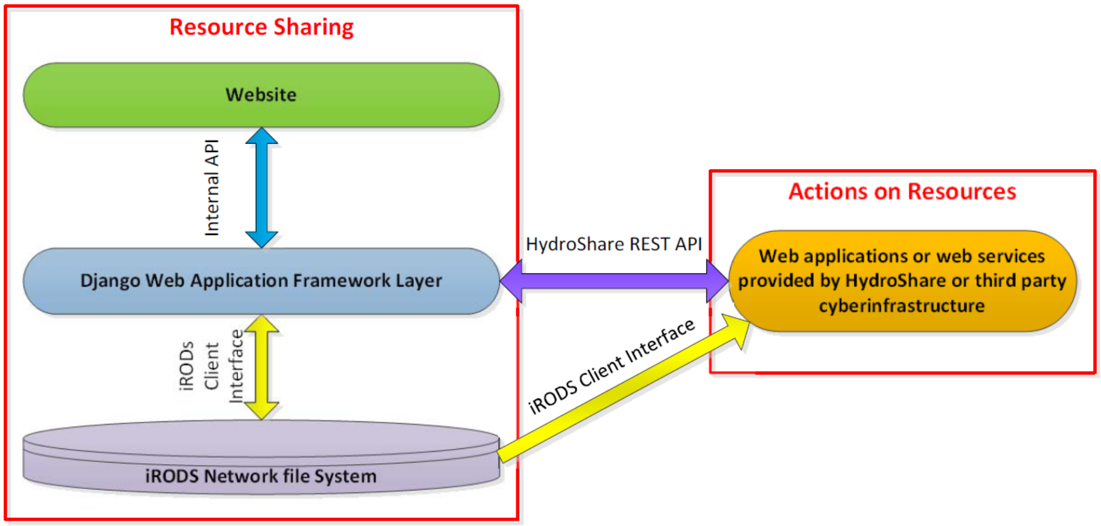


Figure 3.1 High level system architecture of HydroShare.

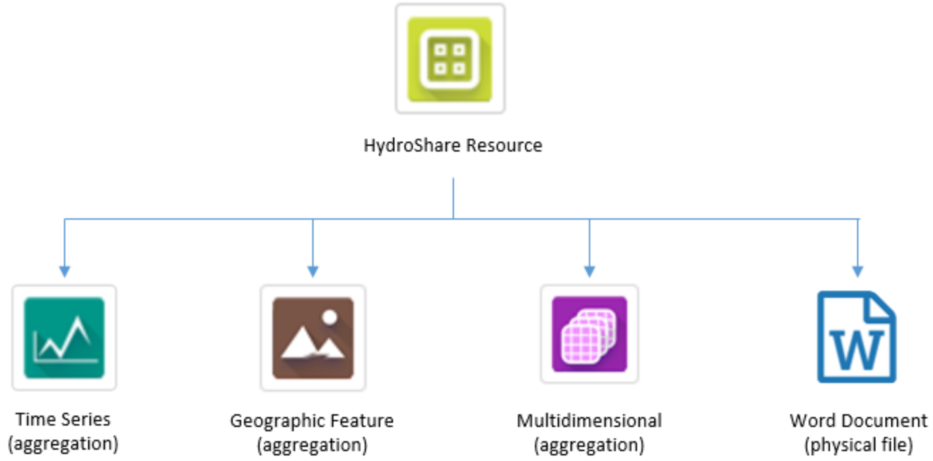
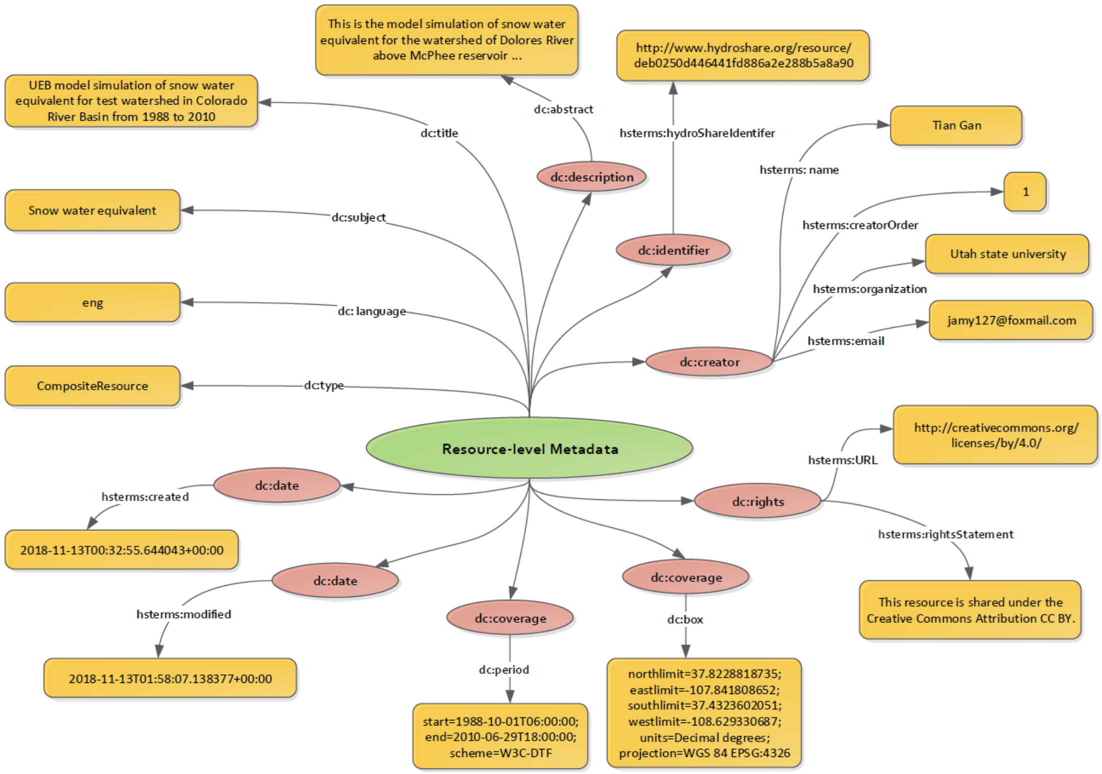
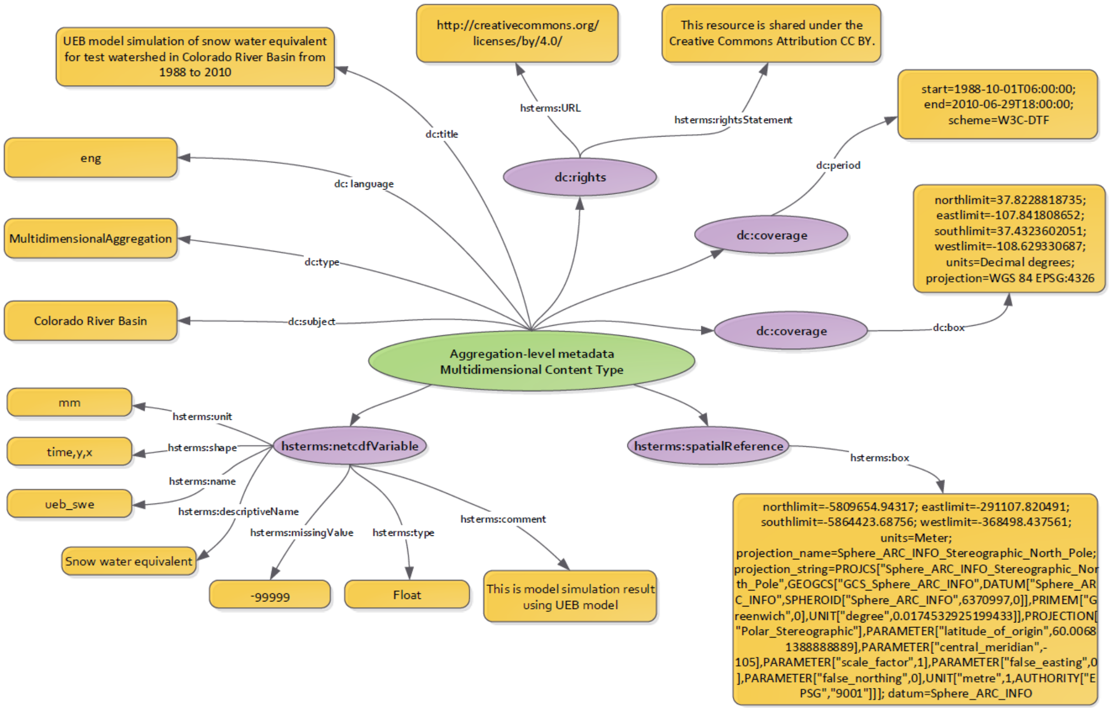


Figure 3.2 Example HydroShare resource including a physical file and different OAI-ORE aggregations.


This is the model simulation of snow water equivalent for the watershed of Dolores River above McPhee reservoir …
http://www.hydroshare.org/resource/deb0250d446441fd886a2e288b5a8a90
UEB model simulation of snow water equivalent for test watershed in Colorado River Basin from 1988 to 2010
dc:abstract
dc:title
hsterms:hydroShareIdentifer
Tian Gan
dc:description
hsterms: name
1
Snow water equivalent
dc:subject
hsterms:creatorOrder
dc:identifier
Utah state university
hsterms:organization
eng
dc: language
jamy127@foxmail.com
hsterms:email
dc:creator
CompositeResource
dc:type
Resource-level Metadata
http://creativecommons.org/licenses/by/4.0/
hsterms:URL
dc:date
hsterms:created
dc:rights
2018-11-13T00:32:55.644043+00:00
hsterms:rightsStatement
dc:date
dc:coverage
hsterms:modified
dc:coverage
This resource is shared under the Creative Commons Attribution CC BY.
dc:box
2018-11-13T01:58:07.138377+00:00
dc:period
northlimit=37.8228818735; eastlimit=-107.841808652; southlimit=37.4323602051; westlimit=-108.629330687; units=Decimal degrees; projection=WGS 84 EPSG:4326
start=1988-10-01T06:00:00; end=2010-06-29T18:00:00; scheme=W3C-DTF

(a)



(b)

Figure 3.3 Metadata elements for a HydroShare resource holding a multidimensional aggregation. Panel (a) shows Dublin Core metadata elements held at the resource level. Panel (b) shows metadata elements specific to the multidimensional content type. Each Dublin Core metadata element is prefixed with "dc"; each metadata element defined by HydroShare is prefixed with "hsterms." Individual metadata element names are labeled on the arrows, and examples of their values are shown in the rectangle boxes.

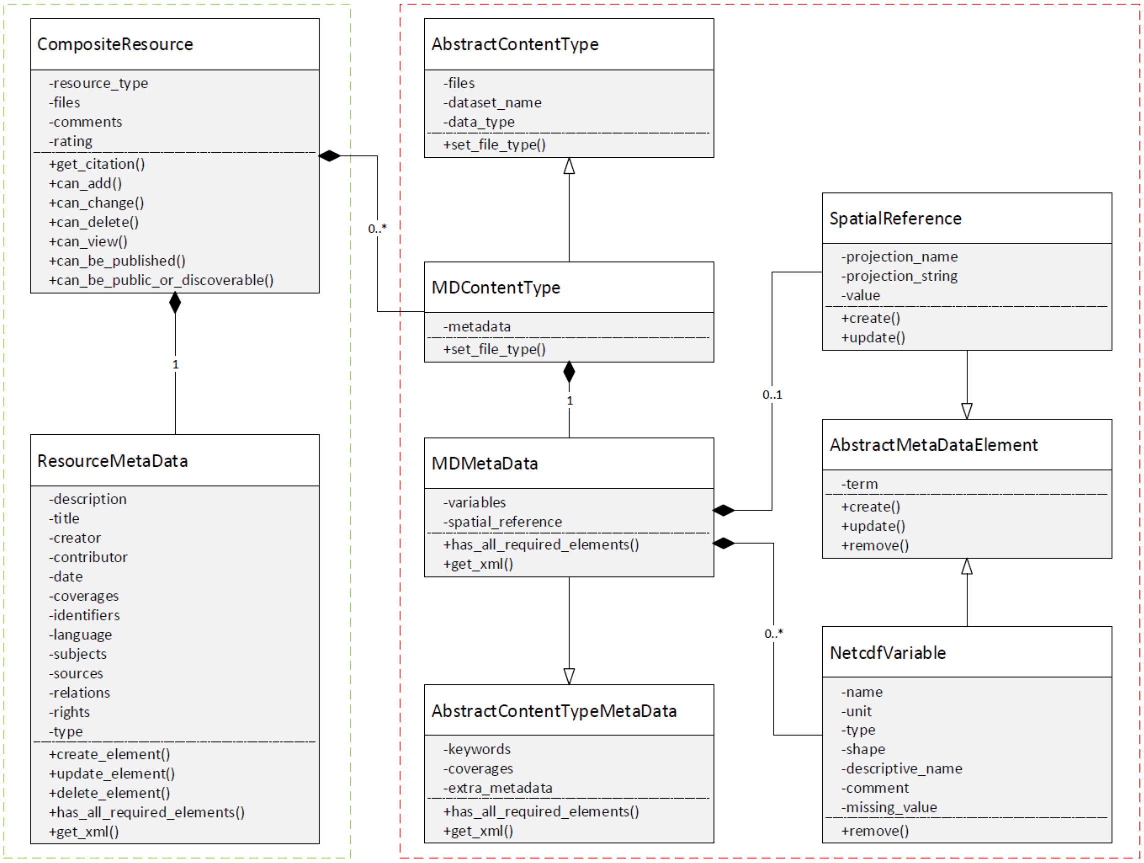


Figure 3.4 UML class diagram for the multidimensional content type data model in HydroShare.

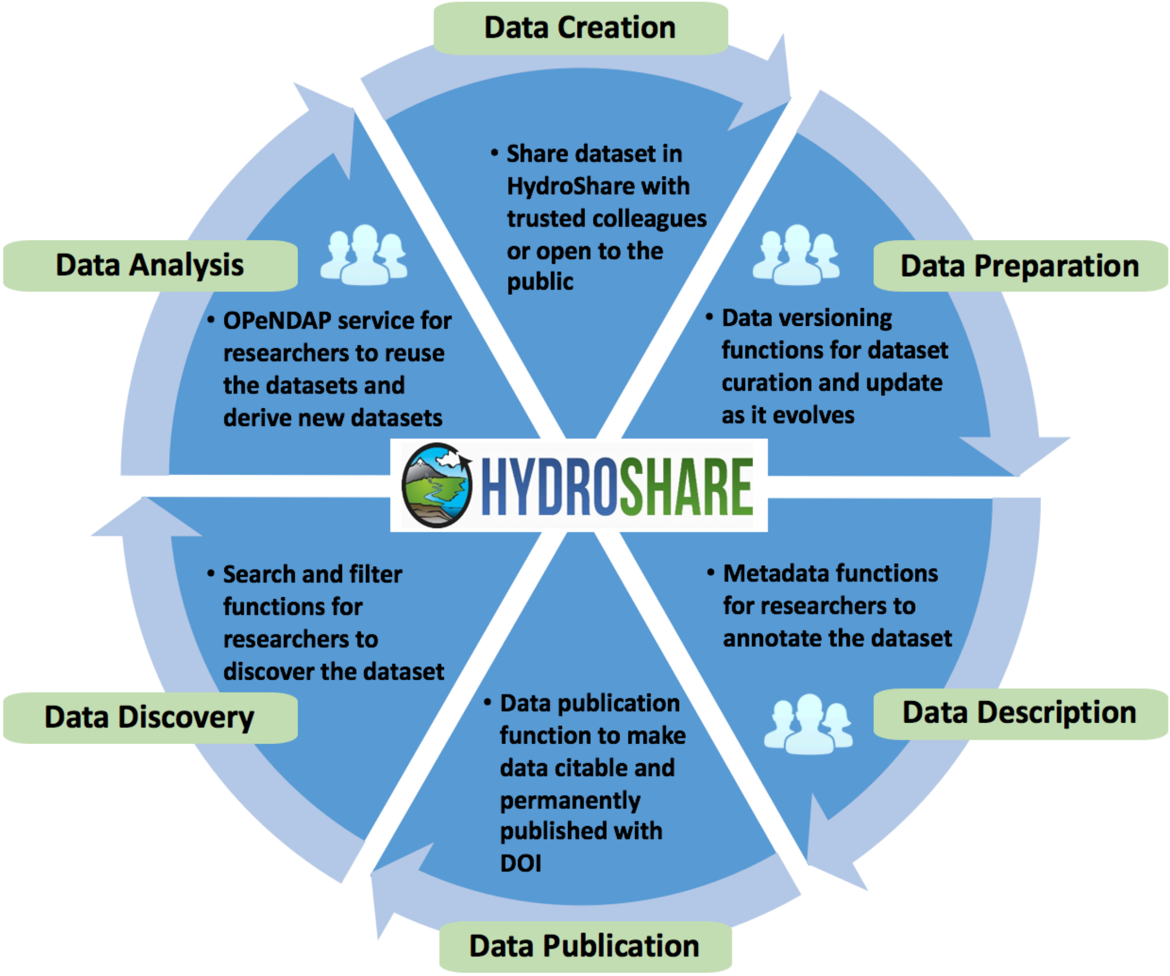
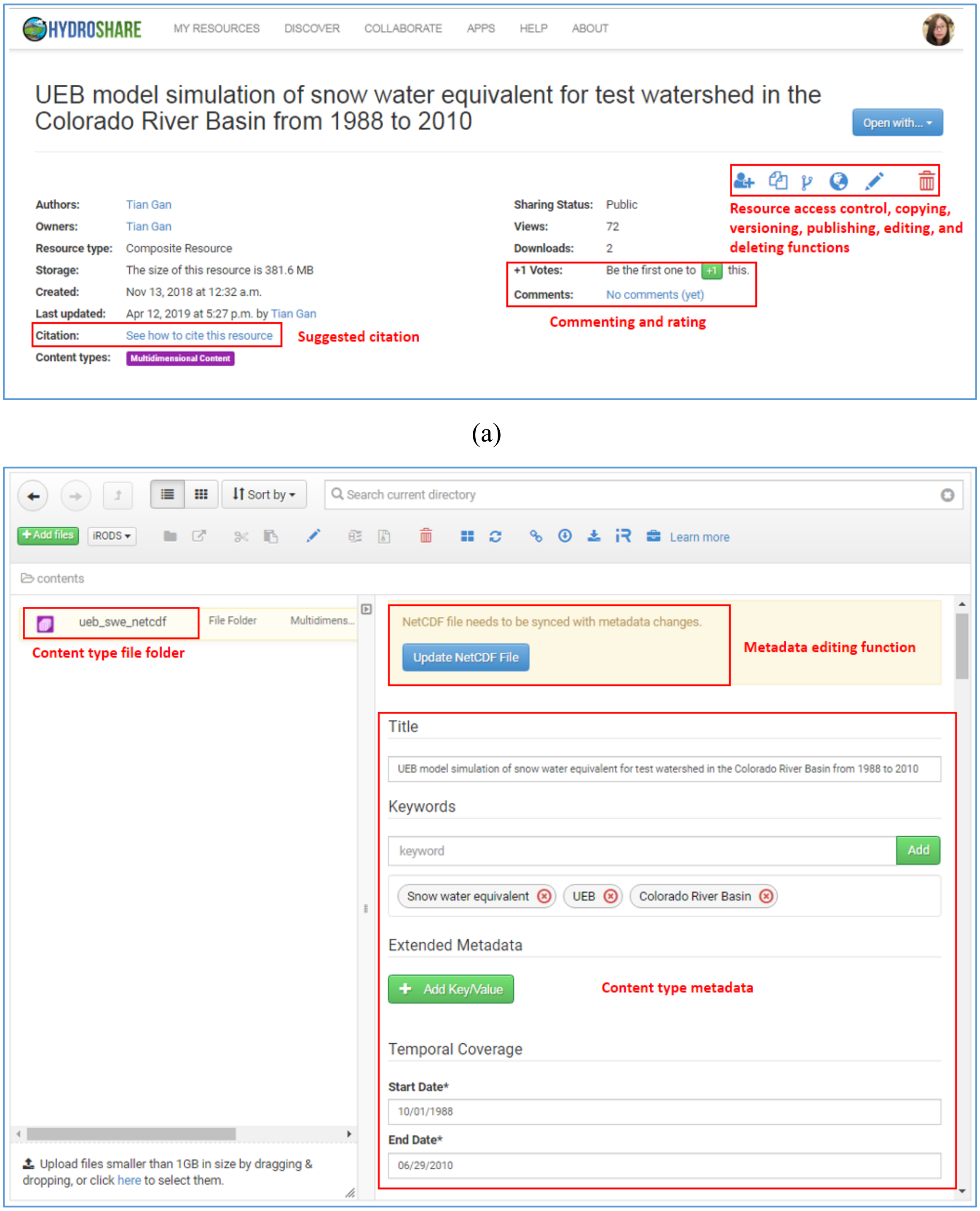


Figure 3.5 HydroShare supports the collaborative sharing of multidimensional space-time data with multiple functions that facilitate the cycle of data sharing activities involved in collaborative research.

```
netcdf ueb_swe_netcdf {
dimensions:
    time = UNLIMITED ; // (31767 currently)
    y = 47 ;
    x = 66 ;
variables:
    char polar_stereographic ;
        polar_stereographic:grid_mapping_name = "polar_stereographic" ;
        polar_stereographic:straight_vertical_longitude_from_pole = -105. ;
        polar_stereographic:false_easting = 0. ;
        polar_stereographic:false_northing = 0. ;
        polar_stereographic:latitude_of_projection_origin = 90. ;
        polar_stereographic:standard_parallel = 60.0068138888889 ;
        polar_stereographic:long_name = "CRS definition" ;
        polar_stereographic:longitude_of_prime_meridian = 0. ;
        polar_stereographic:semi_major_axis = 6370997. ;
        polar_stereographic:inverse_flattening = 0. ;
        polar_stereographic:spatial_ref =
"PROJCS[\"Sphere_ARC_INFO_Stereographic_North_Pole\",GEOGCS[\"GCS_Sphere_ARC_INFO\",DATUM[\"Sphere_ARC_INFO\",SPHEROID[\"Sphere_ARC_INFO\",6370997
,0]],PRIMEM[\"Greenwich\",0],UNIT[\"degree\",0.0174532925199433]],PROJECTION[\"Polar_Stereographic\"],PARAMETER[\"latitude_of_origin\",60.00681388888889],PA
RAMETER[\"central_meridian\",-105],PARAMETER[\"scale_factor\",1],PARAMETER[\"false_easting\",0],PARAMETER[\"false_northing\",0],UNIT[\"metre\",1,AUTHORITY[\"EP
SG\",\"9001\"]]]" ;
        polar_stereographic:GeoTransform = "-369093.75 1190.624877999998 0 -5809059.630734 0 -1190.624877999998 " ;
    double time(time) ;
        time:units = "hours since 1988-10-01 00:00:00.0 UTC" ;
        time:calendar = "standard" ;
    float ueb_swe(time, y, x) ;
        ueb_swe:_FillValue = -99999.f ;
        string ueb_swe:grid_mapping = "polar_stereographic" ;
    double x(x) ;
        x:standard_name = "projection_x_coordinate" ;
        x:long_name = "x coordinate of projection" ;
        x:units = "m" ;
    double y(y) ;
        y:standard_name = "projection_y_coordinate" ;
        y:long_name = "y coordinate of projection" ;
        y:units = "m" ;

// global attributes:
        :GDAL_AREA_OR_POINT = "Area" ;
        :Conventions = "CF-1.5" ;
        :GDAL = "GDAL 2.1.3, released 2017/20/01" ;
        :history = "Mon Oct 29 13:41:52 2018: ncks -x -v Band1
/Projects/Tian_workspace/rdhm_ueb_modeling/McPhee_MPHC2/Xmrg_to_netCDF/McPhee_xmrg_to_netCDF/ueb_swe_temp.nc
/Projects/Tian_workspace/rdhm_ueb_modeling/McPhee_MPHC2/Xmrg_to_netCDF/McPhee_xmrg_to_netCDF/ueb_swe_netcdf.nc\nMon Oct 29 13:25:57 2018: GDAL
CreateCopy( /Projects/Tian_workspace/rdhm_ueb_modeling/McPhee_MPHC2/Xmrg_to_netCDF/McPhee_xmrg_to_netCDF/ueb_swe_temp.nc, ... )" ;
        :NCO = "4.4.4" ;
        :title = "UEB model simulation of snow water equivalent for test watershed in Colorado River Basin from 1988 to 2010" ;
        :keywords = "Snow water equivalent, UEB, Colorado River Basin" ;
        :creator_name = "Tian Gan" ;
        :summary = "This is the model simulation of snow water equivalent for the watershed of Dolores River above McPhee reservoir in Colorado River Basin from 1988 to
2010. The model used is the Utah Energy Balance model which is a physically based snow melt model." ;
}
```

Figure 3.6 Data structure and attributes of the case study dataset expressed in NetCDF common data language (CDL) and derived from the ncdump command.



UEB model simulation of snow water equivalent for test watershed in the Colorado River Basin from 1988 to 2010

Authors: Tian Gan
Owners: Tian Gan
Resource type: Composite Resource
Storage: The size of this resource is 381.6 MB
Created: Nov 13, 2018 at 12:32 a.m.
Last updated: Apr 12, 2019 at 5:27 p.m. by Tian Gan
Citation: See how to cite this resource
Content types: Multidimensional Content

Sharing Status: Public
Views: 72
Downloads: 2
+1 Votes: Be the first one to +1 this.
Comments: No comments (yet)

(a)

NetCDF file needs to be synced with metadata changes.

Title
UEB model simulation of snow water equivalent for test watershed in the Colorado River Basin from 1988 to 2010

Keywords
keyword
Snow water equivalent ⊗ UEB ⊗ Colorado River Basin ⊗

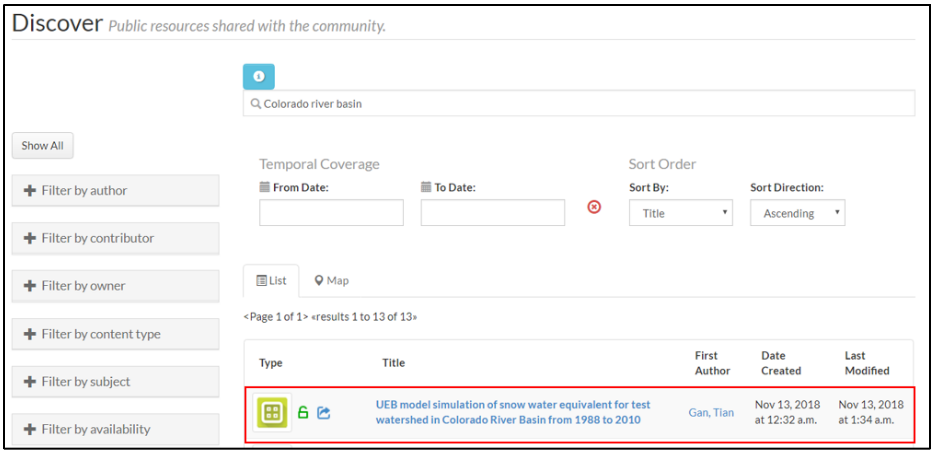
Extended Metadata

Temporal Coverage
Start Date*
10/01/1988
End Date*
06/29/2010

Upload files smaller than 1GB in size by dragging & dropping, or click here to select them.

(b)

Figure 3.7 Resource landing page for the case study dataset. Panel (a) shows the basic data sharing functionality for a resource. Panel (b) shows the content type files, content type metadata, and the “Update NetCDF file” button that appears after the user edits the metadata.



Colorado river basin

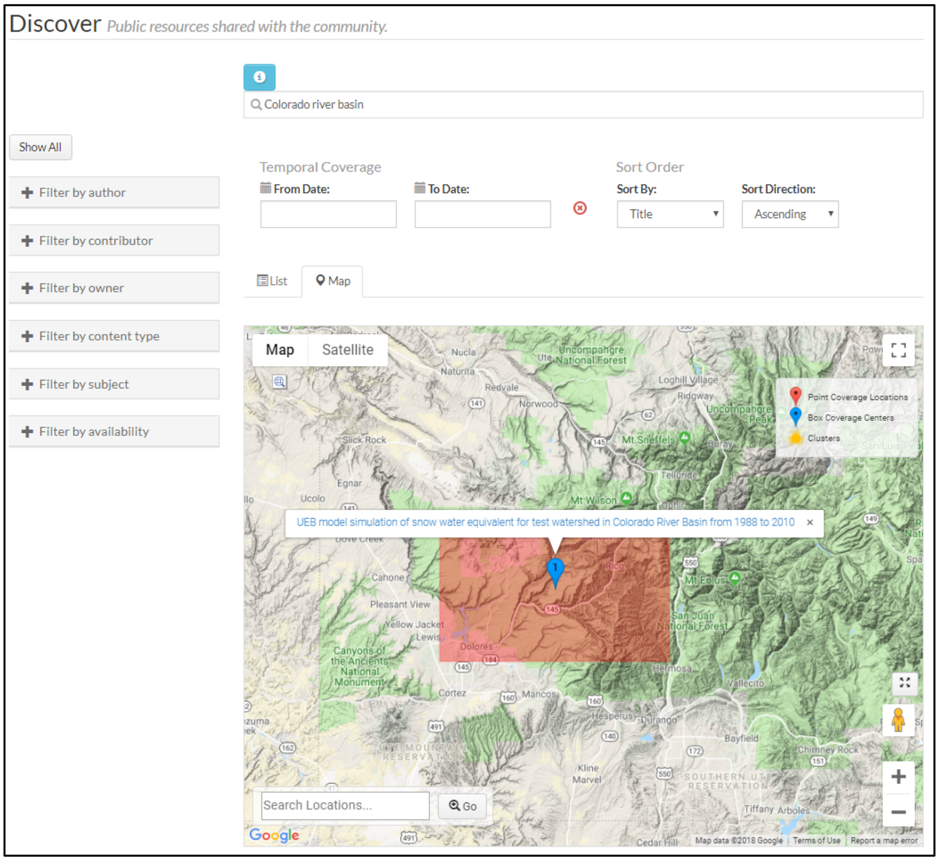
Sort By: Title

Sort Direction: Ascending

<Page 1 of 1> «results 1 to 13 of 13»

Type	Title	First Author	Date Created	Last Modified
	UEB model simulation of snow water equivalent for test watershed in Colorado River Basin from 1988 to 2010	Gan, Tian	Nov 13, 2018 at 12:32 a.m.	Nov 13, 2018 at 1:34 a.m.

(a)



Colorado river basin

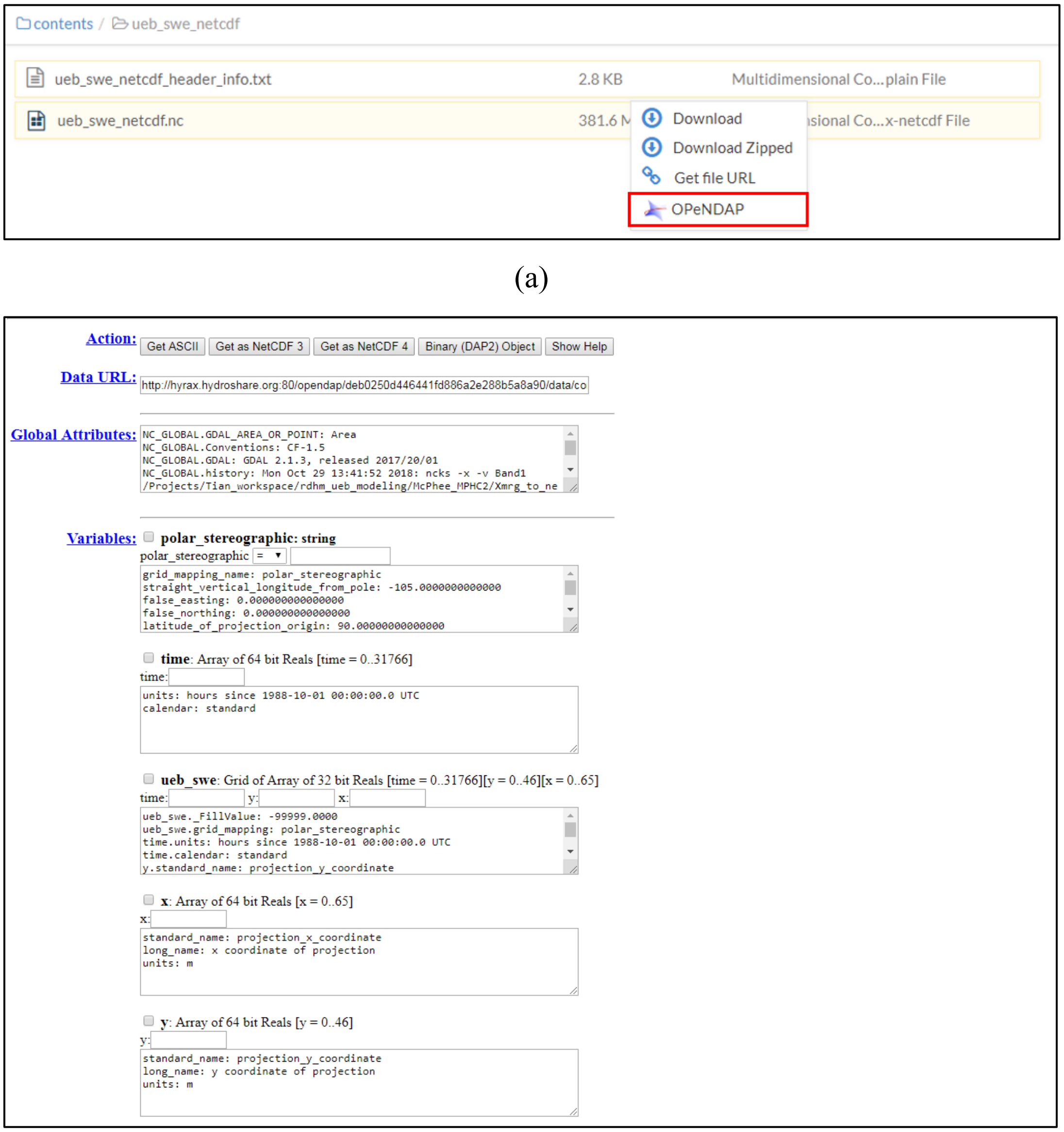
Sort By: Title

Sort Direction: Ascending

UEB model simulation of snow water equivalent for test watershed in Colorado River Basin from 1988 to 2010

(b)

Figure 3.8 Data discovery of the case study dataset with the search and filter functions in HydroShare. Panel (a) shows the data discovery with a search term. Panel (b) shows the data discovery with geolocation of the dataset.



ueb_swe_netcdf_header_info.txt	2.8 KB	Multidimensional Co...plain File
ueb_swe_netcdf.nc	381.6 M	sional Co...x-netcdf File

(a)

Data URL: http://hyrax.hydroshare.org:80/opendap/deb0250d446441fd886a2e288b5a8a90/data/co

Global Attributes:

```
NC_GLOBAL.GDAL_AREA_OR_POINT: Area
NC_GLOBAL.Conventions: CF-1.5
NC_GLOBAL.GDAL: GDAL 2.1.3, released 2017/20/01
NC_GLOBAL.history: Mon Oct 29 13:41:52 2018: ncks -x -v Band1
/Projects/Tian_workspace/rdhm_ueb_modeling/McPhee_MPHC2/Xmrg_to_ne
```

Variables: ☐ **polar_stereographic: string**

polar_stereographic = ▾

```
grid_mapping_name: polar_stereographic
straight_vertical_longitude_from_pole: -105.0000000000000
false_easting: 0.000000000000000
false_northing: 0.000000000000000
latitude_of_projection_origin: 90.00000000000000
```

☐ **time**: Array of 64 bit Reals [time = 0..31766]

time:

```
units: hours since 1988-10-01 00:00:00.0 UTC
calendar: standard
```

☐ **ueb_swe**: Grid of Array of 32 bit Reals [time = 0..31766][y = 0..46][x = 0..65]

time: y: x:

```
ueb_swe._FillValue: -99999.0000
ueb_swe.grid_mapping: polar_stereographic
time.units: hours since 1988-10-01 00:00:00.0 UTC
time.calendar: standard
y.standard_name: projection_y_coordinate
```

☐ **x**: Array of 64 bit Reals [x = 0..65]

x:

```
standard_name: projection_x_coordinate
long_name: x coordinate of projection
units: m
```

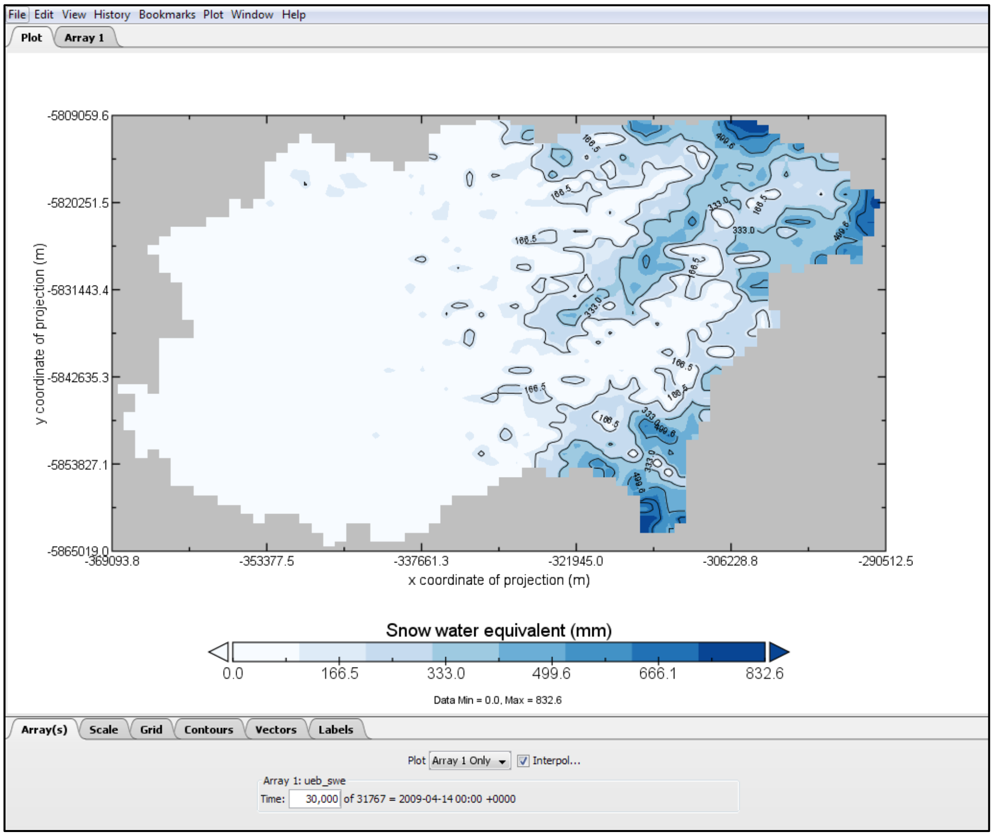
☐ **y**: Array of 64 bit Reals [y = 0..46]

y:

```
standard_name: projection_y_coordinate
long_name: y coordinate of projection
units: m
```

(b)

Figure 3.9 OPeNDAP data access form for the case study multidimensional space-time data. Panel (a) shows the way to open the OPeNDAP dataset access form through the resource landing page. Panel (b) shows the OPeNDAP data access form to help researchers directly subset the NetCDF file shared in HydroShare.

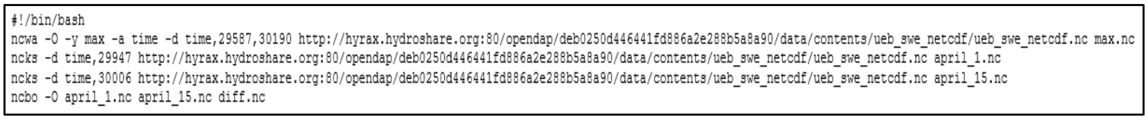


Plot Array 1 Only ☑ Interpol...

Array 1: ueb_swe

Time: 30,000 of 31767 = 2009-04-14 00:00 +0000

(a)



```
#!/bin/bash
ncwa -O -y max -a time -d time,29587,30190 http://hyrax.hydroshare.org:80/opendap/deb0250d446441fd886a2e288b5a8a90/data/contents/ueb_swe_netcdf/ueb_swe_netcdf.nc max.nc
ncks -d time,29947 http://hyrax.hydroshare.org:80/opendap/deb0250d446441fd886a2e288b5a8a90/data/contents/ueb_swe_netcdf/ueb_swe_netcdf.nc april_1.nc
ncks -d time,30006 http://hyrax.hydroshare.org:80/opendap/deb0250d446441fd886a2e288b5a8a90/data/contents/ueb_swe_netcdf/ueb_swe_netcdf.nc april_15.nc
ncbo -O april_1.nc april_15.nc diff.nc
```

(b)

Figure 3.10 Data visualization and analysis for the case study multidimensional space-time data by using the OPeNDAP service and its client software programs. Panel (a) shows a 2D graph of snow distribution using Panoply. Panel (b) shows data processing code to derive new results using NCO.

CHAPTER 4

INTEGRATING HYDROLOGIC MODELING WEB SERVICES WITH ONLINE DATA SHARING TO PREPARE, STORE, AND EXECUTE HYDROLOGIC MODELS [1]

Abstract

Web based applications, web services, and online data and model sharing technology are becoming increasingly available to support hydrologic research. This promises benefits in terms of collaboration, computer platform independence, and reproducibility of modeling workflows and results. In this research, we designed an approach that integrates hydrologic modeling web services with an online data sharing system to support web-based simulation for hydrologic models. We used this approach to integrate example systems as a case study to support reproducible snowmelt modeling for a test watershed in the Colorado River Basin, USA. We demonstrated that this approach enabled users to work within an online environment to create, describe, share, discover, repeat, modify, and analyze the modeling work. This approach encourages collaboration and improves research reproducibility. It can also be adopted or adapted to integrate other hydrologic modeling web services with data sharing systems for different hydrologic models.

Keywords: hydrologic modeling, data sharing, reproducibility, web services, HydroShare

---

[1] Coauthored by Tian Gan, David G. Tarboton, Tseganeh T. Gichamo, Pabitra Dash, and Jeffery S. Horsburgh.

4.1 Introduction

Hydrologic modeling is essential as a guide to formulating strategies for water resources management or as a tool of scientific inquiry (Dingman, 2008). However, hydrologic modeling research presents a number of challenges. Modelers need to discover and collect data from various sources (Archfield et al., 2015) and use it to prepare model inputs. Model input preparation can be time consuming and may require a substantial learning curve, especially where programming is needed (Miles, 2014). Furthermore, modelers may need to access high performance computing (HPC) resources to effectively handle large scale or complicated hydrologic model simulations (Kumar et al., 2008; Laloy and Vrugt, 2012). Curating and sharing modeling datasets and metadata publicly is also important to improving reproducibility (Demir and Krajewski, 2013; Archfield et al., 2015; Hutton et al., 2016). Collaboration among people from various disciplines and areas is one of the key factors in catalyzing new research findings (Silliman et al., 2008). For instance, the water-food-energy nexus research requires expertise from various fields and cross-sector collaboration to enhance the water, energy, and food security. A system that provides a way to effectively communicate and collaborate in their research projects would be of significant value.

With the development of web technologies and standards, one promising direction is to provide web services or web applications to help people overcome these challenges and improve the efficiency of hydrologic modeling work. Some systems help acquire or preprocess datasets as model input files for hydrologic models (Leonard and Duffy, 2013). For instance, Billah et al. (2016) developed web services that help to automate the grid data pre-processing workflow for preparation of model inputs for the Variable

Infiltration Capacity (VIC) model (Liang et al., 1996). The workflow includes the information that allows others to independently reproduce the model results and acts as a means for documenting the steps used to create model input files. Some systems focus on simulation using a specific hydrologic model while others couple different hydrologic models to simulate integrated hydrologic processes. For example, SWATShare (Rajib et al., 2016) established a collaborative environment to publish, share, discover, and download Soil and Water Assessment Tool (SWAT) models. This cyberinfrastructure also supports SWAT model calibration running on HPC resources and visualization of model outputs. The Community Surface Dynamics Modeling System (CSDMS) (Peckham et al., 2013) created an environment that promotes the sharing, reuse, and integration of open-source modeling software. Many models in CSDMS are installed and maintained on its high-performance cluster. CSDMS members can access these resources and integrate them for complex model simulation. Other systems support both model input preparation and simulation to facilitate modeling work. The framework of AWARE, which is described as “A tool for monitoring and forecasting Available WAter REsource in mountain environments,” was developed to offer online geospatial processing services and other tools to help users monitor and forecast water resources in Alpine regions (Granell et al., 2010).

Although these web services or web applications improve the efficiency of hydrologic modeling work, they do have limitations. One limitation is that it may require programming to use the web services and thus be difficult for those without the required programming skills or knowledge to use them. Another limitation is related to the reproducibility of the modeling work, an essential principle in scientific research (Hutton et al., 2016). The model input/output files and the programming code for data processing and analysis are often not well curated and shared with the public (Stagge et al., 2019). This hinders the ability for the modeling community to reproduce and verify the modeling work and reuse the results.

In this research, our goal was to integrate hydrologic modeling web services with a data sharing system to provide web-based simulation that improves the reproducibility of the modeling work and the usability of these web services. We define web-based simulation as the use of web technologies to develop, execute, and analyze simulation models with the web browser playing an active role in the modeling process, either as a graphical user interface or as a container for the simulation engine (Byrne et al., 2010; Walker and Chapra, 2014). We sought to provide an online environment within which users can prepare model input, execute the model, share and analyze the results, and repeat or modify the modeling work for collaboration.

To achieve this goal, we designed an approach for system integration. The general idea was to add a browser-based graphical user interface for the modeling web services to simplify the way of using them without programing and to take advantage of a data sharing system to provide advanced data curation and management capability beyond existing modeling web services and add value to them. As a case study, we used this approach to integrate two example systems, HydroDS and HydroShare, to support web-based simulation for a snowmelt model. The functionality implemented was evaluated using use cases based on snowmelt modeling in a test watershed of the Colorado River Basin, USA. HydroDS is a system that provides hydrologic modeling web services to process terrain and climate datasets as model inputs for distributed hydrologic models such as the Utah Energy Balance (UEB) snow model (Tarboton and Luce, 1996). This system also provides a Python client for the Representational State Transfer (REST) web service application programming interface (API) for users to write Python code to automate data processing workflows. Model input and output files can be temporarily saved in the HydroDS system and are then downloadable for further analysis. HydroShare is a hydrologic information system and repository for sharing hydrologic data, models, and analysis tools (Tarboton et al., 2014). In HydroShare, the hydrologic datasets or models can be shared as resources (Horsburgh et al., 2015) which can be published, collaborated around, annotated, discovered, and accessed. Aside from the data sharing functions, HydroShare also provides a REST API and corresponding Python client library that enables other systems including web applications (or apps) to interact with HydroShare.

This approach makes the hydrologic modeling web services available to a broader community of users who have limited programming skills. By sharing the datasets, scripts, and metadata of modeling work in a data sharing system, the research community will be better able to discover and access them for reuse and collaboration. This approach also facilitates research validation and experimentation within an online environment without model configuration and data transfer or using the storage and computing resources of the user’s local machine. Additionally, this approach reuses and extends open source software to promote reproducible research. It can be adopted or adapted to integrate other hydrologic modeling web services with data sharing systems for different hydrologic models.

In Section 2, we introduce the general architecture design and the case study design that uses this approach to integrate two example systems (HydroDS and HydroShare). In Section 3, we present the case study results, which describes the integration of the functionality implemented and snow modeling use cases for functionality test. Sections 4 is discussion and Section 5 summary and conclusions.

4.2 Methods

4.2.1 General approach

The purpose of the system integration is to support web-based simulation that : 1) provides easy access through a web browser to use the modeling web services, 2) provides online data curation and sharing to support management and reuse of the modeling work, and 3) avoids the complexity of changing existing systems to achieve system integration.

Based on these criteria, we designed a three-layer architecture to integrate hydrologic modeling web services with a data sharing system. This architecture includes a user interface layer, a data service layer, and a data storage layer (Figure 4.1). The user interface layer can be a web app that provides browser based user interface for modelers to use the hydrologic modeling web services without programming. This user interface layer web app can be hosted on other web servers separate from the data service layer or the data storage layer and interact with them through REST APIs. This design decouples the user interface web app from the other two layers and avoids significant changes in the existing systems. The data service layer is a system that hosts hydrologic modeling web services. This layer can receive web requests from the user interface layer to prepare model input datasets or execute hydrologic models. The data storage layer is a data sharing system to store and share modeling work created and transferred from the data service layer. This design uses the emerging functionality of data sharing systems to avoid additional software development work and provide the storage and data curation needs for systems that host hydrologic modeling web services.

4.2.2 Case study design

Our case study was designed to use this general approach and integrate example systems to test if the system integration can support web-based simulation to improve research reproducibility and reduce the need for coding to use the modeling web services. Thus, we used the three-layer architecture to integrate HydroShare and HydroDS, and designed use cases to evaluate the application of implemented functionality for snowmelt modeling in a test watershed of the Colorado River Basin.

We chose these systems because: 1) they represent the general functionality of hydrologic modeling web services (HydroDS) and data sharing systems (HydroShare); and 2) the author has access to both systems and is thus able to work on them for integration. In the following, we first provide background on these systems and then present the case study design.

HydroDS is a system that provides web services to simplify model input preparation for distributed hydrologic models (Gichamo, 2019). Modelers can use these web services to create model input files and save the time and energy often spent collecting datasets from multiple sources and developing code to preprocess the data into required file formats. For example, Table 4.1 shows the UEB model input variables and the major HydroDS Python client functions used to prepare them. The UEB model requires climate, terrain, and canopy datasets as model input and uses Network Common Data Form (NetCDF http://www.unidata.ucar.edu/software/netcdf/) as its input/output file format. Modelers can use HydroDS Python client functions to write data processing code for input preparation. These web services store generated datasets in HydroDS and preprocess them to NetCDF format for a given area.

The HydroDS system was built using Django, an open-source Python web framework for web development (https://www.djangoproject.com/) (Figure 4.2). Several open-source libraries and software programs for processing NetCDF and raster datasets were installed in HydroDS, such as NetCDF4 Python module, NCO (Zender, 2008), GDAL (http://www.gdal.org/), and TauDEM (Tarboton, 1997). Additionally, datasets from multiple sources for input preparation were also stored in this system, including the National Elevation Dataset (NED) (Gesch, 2007), National Land Cover Datasets (Homer et al., 2015), and Daymet climate data (Thornton et al., 2016). A Python client program for the HydroDS web services called “Hydrogate” (https://github.com/CI-WATER/hydrogate_Python_client) is available for users to write Python code and make web requests to HydroDS.

HydroShare’s system architecture (Figure 4.3) is centered on several open source components (Heard et al., 2014). The major components include Django and iRODS (http://iRODS.org/). Django provides the functionality that was used to build the web user interface to help users manage their shared datasets or models. iRODS is open source data management software that is used for data storage and access control. Aside from data sharing functionality, web apps hosted on other web servers can also connect to HydroShare. For instance, the CUAHSI JupyterHub web app (http://jupyter.cuahsi.org) is an example that is developed by others (Castronova, 2016) and connected to HydroShare. This web app was built with the JupyterHub software stack (https://jupyter.org/hub) and configured with many scientific Python libraries and tools and provides an online programming environment where researchers can load data from HydroShare and develop Python code for data analysis and visualization. Another example is the HydroShare Tethys Apps portal (https://apps.hydroshare.org/apps/), a system established by the HydroShare team to host multiple web apps and interact with HydroShare resources. This web portal was built using the Tethys platform (Swain et al., 2016) that includes various software suites and development kits to alleviate the difficulties non-professional programmers may have in developing web apps for environmental data visualization, analysis, and modeling applications. In order to enable information exchange between HydroShare and the HydroShare Tethys Apps portal, Oauth (https://oauth.net/) is used to support user authentication and authorization, and the HydroShare REST API Python client "hs_restclient" (https://github.com/hydroshare/hs_restclient) is used to transfer the datasets between the two systems.

In our case study design, we applied the three-layer architecture based on the features of HydroDS and HydroShare to support modeling work of the UEB model (Figure 4.1). A Tethys web app (the UEB web app) was developed and hosted in the HydroShare Tethys Apps portal and serves as the user interface layer to provide easy access to the HydroDS web services. HydroDS is the data service layer used to prepare the model input files and execute the model. HydroShare acts as the data storage layer to store and share the results created from HydroDS. The main activity between the UEB web app and HydroDS is the transfer of user input information to make web requests for model input preparation or model simulation. Between HydroDS and HydroShare, the activity is mainly the transfer of model input/output files and associated metadata for modeling work. The UEB web app also interacts with HydroShare to retrieve the metadata of shared model input files to facilitate model simulation.

We evaluated the system integration for two snowmelt modeling use cases. These use cases were designed to use the web-based simulation functionality to test the sensitivity of the UEB model to grid cell resolutions and find out if different grid cell resolutions for the model input files can lead to different snow outputs. This finding can help modelers to evaluate the tradeoffs between model performance and computational as well as data storage requirements. In the first use case, a user prepares model input, executes the model, and curates the results in HydroShare. In the second use case, another user discovers the shared modeling work in HydroShare and modifies the work to derive new results with higher grid cell resolution and compares the snow outputs from the two use cases.

4.3 Results

4.3.1 System integration

4.3.1.1 User interface layer

The UEB web app was developed as a Tethys web app and hosted in the HydroShare Tethys Apps portal to provide a graphical user interface for HydroDS web services. We chose this web portal to host the UEB web app for several reasons. First, this decouples the user interface application from the systems that hosts data or hydrologic modeling web services. Loosely coupled systems allow changes in one system without big changes in the other systems making them easier to maintain. Second, this web portal is built on Tethys platform that provides software development kits to simplify the web app development and lower the requirement of learning multiple languages, which reduces the coding required (Swain et al., 2016).

The UEB web app was designed to provide three functions: model input preparation, model execution, and job status checking. Users can interact with this web app to perform modeling work without writing program code to simplify access to HydroDS. Figure 4.4 (a) shows the user interface for model input preparation. This has two main sections: the user input form section on the left and the map view section in the center. The user input form section allows the user to enter settings to create a complete model input package for model simulation. The map view section helps the user draw a bounding box and/or an outlet point to specify the modeling domain. After the user fills out the form and clicks on the "Input Data Preparation" button, the web request is sent to HydroDS and a corresponding job ID is returned so that the UEB web app can monitor the status of the submitted job. Figure 4.4 (b) shows the user interface for model execution. It also has two main sections: the model input information section on the left and the map view section. The model input information section allows the user to select a model input package stored in HydroShare. When the user selects a model input package, its corresponding metadata is retrieved from HydroShare and shown in this section. Furthermore, if the metadata includes the bounding box and outlet point information for the modeled domain, it will be automatically shown on the map to orient the user geographically. After the user clicks on the "Submit Model Execution" button, the web request is sent to HydroDS, and the corresponding job ID is returned so that the UEB web app can monitor the job status. Figure 4.5 shows the job status checking user interface where the status of submitted model input preparation or model simulation jobs is shown. When the job is completed successfully, the user is provided with a link to the resource in HydroShare that stores the model input package (in the green frame) or model output files (in the red frame). If the job fails, the user will be provided with detailed error information (in the yellow frame).

The UEB web app was built based on Tethys, which by default includes a narrow left panel and a wide right panel in the main app section. We designed the app to display a map in the main app section and parameter entry form with control buttons on the left. Menu bars at the top were used to switch between steps in the designed use of the app, which can provide the user with guidance on the functionality of each page. Implementing this design required Hypertext Markup Language (HTML) and cascading style sheets (CSS) to customize the default layout provided by Tethys. The user input forms in the left panel were implemented using Bootstrap, an open-source front-end web framework (http://getbootstrap.com/) and the Template Gizmos API (http://docs.tethysplatform.org/en/latest/tethys_sdk/gizmos.html) from the software development kit of Tethys platform. The map view in the right panel was implemented using the Google Maps JavaScript API (https://developers.google.com/maps/). Additionally, the HydroShare REST API Python client was used to manage all the interactions between the user interface layer and the data storage layer. For example, the metadata for existing model input packages from HydroShare can be retrieved using the Python client and displayed on the model execution interface.

4.3.1.2 Data service layer

In the HydroDS system, we implemented new web services and job submission capability, which were used by the UEB web app for model input preparation, model simulation, and job status checking.

This was a departure from the original design for the HydroDS web services, which required users to make multiple web requests to process various datasets for input preparation (Table 4.1). It is inefficient for the UEB web app to send multiple web requests to HydroDS and periodically check for completion. Thus, we used the existing data processing functionality in HydroDS and implemented a new web service for model input preparation, which enables the UEB web app to submit a single web request to HydroDS to accomplish the work. Figure 4.6 (a) shows the detailed tasks done by this new web service. It first creates a complete UEB model input package that includes both the input data files and the model parameter files. Then, it generates a Python file to document the details of how the model input package can be created using the HydroDS Python client. Finally, it transfers all of the files and associated metadata to HydroShare. In this web service, the Python script created was designed to provide input preparation details instead of hiding the processing work behind the scenes as a black box to users. This Python script can be reused to reproduce or derive new model input for the UEB model. It can also be used as an example to learn how to use HydroDS web services and create input preparation workflows for other hydrologic models.

We also implemented a new web service that helps the UEB web app to make a single web request to HydroDS for model simulation. Figure 4.6 (b) presents the specific tasks accomplished by this web service. It first downloads the model input package from HydroShare into HydroDS. Then, it validates the model input package to check if there are missing files required for executing the model. If the validation is successful, HydroDS executes the UEB model and then transfers the model output files and stores them with the model input package in HydroShare. To support data transfer between the data service and data storage layers, the HydroShare REST API Python client “hs_restclient” was used for reading and writing files and metadata to and from HydroShare.

In order to improve the user experience, we also added job submission capability for the two new web services. When users use the UEB web app to make web requests to HydroDS, the system responds with a job ID, and the model input preparation or model execution process can be accomplished asynchronously so that users are able to check the job status any time after the job submission. Web services for querying the job status from HydroDS were also implemented, and were used by the UEB web app to get the job details and present them on the user interface.

4.3.1.3 Data storage layer

In HydroShare, we chose the “model instance” resource type (Morsy et al., 2017) to support curation and sharing of the data files and metadata generated by HydroDS. This resource type was specifically designed to support the collaborative sharing of model input/output files and their associated metadata, which best suits our requirement to improve reproducibility of hydrologic modeling research (Figure 4.7). For example, users can store model input/output files in a HydroShare model instance resource and describe them with predefined resource-level metadata as well as user-defined key-value pair metadata. This can help others discover and access the model instance with enough information for reuse. Users can also manage the resource access control, so that it can be kept as private and accessed only by trusted users to prepare and edit the contents, or it can be shared to the public so that anyone can discover and reuse it for validation or deriving new results. In addition, users can formally publish their modeling work in HydroShare to get a digital object identifier (DOI) and suggested citation information. This encourages proper citation of the shared work.

When the UEB web app is used for model input preparation, a new model instance resource is created in HydroShare to store the model input package. The information entered in the user input form of the UEB web app is stored as user-defined resource metadata in HydroShare, which saves users from manual metadata editing work to provide detailed information about the input package. When the UEB web app is used for model simulation, the model instance resource is downloaded from HydroShare into HydroDS for execution, and the resulting model output files are sent back to the corresponding model instance resource in HydroShare. In the case where a user submits a model simulation job but deletes the model instance resource before the job completes, a new model instance resource is created that includes model input package and output files after the model simulation. The user can run the simulation to generate model output multiple times with all the results stored in the same resource. Additionally, other users can use the resource copy function in HydroShare to duplicate the model instance resource as their own new resource to repeat or build on the modeling work.

In addition to using the model instance resource for data curation and sharing, we also used the CUAHSI JupyterHub web app, an online programming environment that supports the development and execution of program code from a Jupyter Notebook file. The benefit of using this web app is that users do not need to download the modeling work and install software on their local computers for post-modeling analysis or to reproduce or reuse a shared model instance. Instead, the model instance resource can be directly retrieved from HydroShare into this web app for reuse. They can develop and execute Python code in a Jupyter Notebook file to visualize or analyze the model input/output datasets (Figure 4.8). Other users can also use this web app and the Python script from the model instance resource to repeat or modify the model input preparation workflow to validate the existing model input package or generate a new model input package (Figure 4.9). This provides another option for model input preparation, that is more scripted, but less graphical user interface friendly than the UEB web app.

4.3.2 Snowmelt modeling

We used the Animas watershed in the Colorado River Basin (Figure 4.10) as the study area to implement our two use cases for model input preparation, then simulation of snowmelt for water year 2010. This served to validate the implemented functionality and test if the system integration can provide web-based simulation to support hydrologic modeling.

In the first use case the UEB web app was used to prepare the model input package, execute the model, and then have all the results automatically copied into a HydroShare resource. Figure 4.4 and Table 4.2 show the interfaces and detailed settings information that were used in the UEB web app for model input preparation and model simulation for the Animas watershed. Figure 4.5 shows the job status of the corresponding results. The green frame is the status for model input preparation, and the red frame for model simulation. Figure 4.7 is the resource landing page for the model instance resource (Gan, 2019a), which was created to store the model input/output files, the associated metadata, and the Python script of the input preparation workflow for the first use case.

The second use case demonstrated collaboration and showed how the modeling work created in the first use case could be discovered, modified, and reused to derive new findings. Assume that the user who prepared the model in the first use case was user 1, and the user who collaborated and reused the model was user 2. The first author of this paper actually acted as both users with separate HydroShare accounts to prepare this illustration. The second use case included the following steps. First, user 2 discovered and got access to the model instance resource created by user 1. Second, user 2 retrieved the resource into the CUAHSI JupyterHub web app, which was used to modify the Python script from the model input package of the first use case to create a new model input package and store it in a new model instance resource in HydroShare. Third, the UEB web app was used to execute the model with the new model instance resource. Finally, the CUAHSI JupyterHub web app was used to develop Python code in a Jupyter Notebook to compare the model outputs from the two use cases.

Figure 4.11 shows the discovery page in HydroShare where the model instance resource created in the first use case can be discovered. In HydroShare, users can search for resources with text or geolocation information and filter the listed results with different facets (e.g., authors or keywords) to find the needed content.

Figure 4.9 shows the Python script loaded into a cell in a Jupyter Notebook within the CUAHSI JupyterHub web app. This Python script is from the model instance resource of the first use case and documents the workflow of model input preparation for creating the climate forcing datasets and parameter files. Figure 4.9 highlights where the user modified the Python script and changed the model resolution from 1200 meters to 600 meters, a model configuration change being tested by user 2 in the second use case (reuse of a model previously established). This modification was designed to test the sensitivity of the model to grid cell resolution and determine whether different resolutions lead to different snow outputs. After the modification, the Jupyter Notebook file was used to execute the script and to create a new model instance resource in HydroShare to store the results, which includes the modified Python script and the new model input package (Gan, 2019b). After the new model instance resource was created, the UEB web app was used to execute the model to create the model output files, which were automatically stored in the same resource.

Finally, the CUAHSI JupyterHub web app was used to retrieve the two resources from HydroShare and to develop data visualization code (Figure 4.8) to compare the snow output from the two use cases. It was found that in the Animas watershed, the comparison of 600 meters versus 1200 meters grid cell resolutions resulted in only very small differences in the model output for snow water equivalent and total surface water input (Figure 4.12 and Figure 4.13). This is mainly because the spatial variability of the terrain and canopy input for the UEB model at the two grid cell resolutions only has small differences, which leads to similar performance for the snowmelt results. The user can also test with higher grid cell resolutions (e.g., 100m or 300m) and compare the model outputs.

This sensitivity test is useful because UEB modelers may choose a coarser cell resolution for model simulation to decrease the simulation time and the size of input and output datasets if there is no significant difference for the snowmelt output. In addition, users may also reuse the first use case to conduct model experiments for parameter sensitivity analysis and find out the relationship between different parameter settings and model performance. The modeling and analysis process can be conducted using the web-based simulation without using the local computing and storage resources. The corresponding results for model experiments can be directly curated and shared with others for validation or reuse.

4.4 Discussion

This case study demonstrated that after using the three-layer architecture to integrate example systems, users were able to develop, share, and reuse modeling work in an online environment by interacting with HydroShare and HydroShare Apps (Figure 4.14). The UEB web app helped to prepare the model input and execute the model through a graphical web user interface. The model instance resource in HydroShare was used to curate and share the modeling results as well as the associated metadata, which enabled others to discover and access them. The CUAHSI JupyterHub web app also provided a web-based tool with which users can modify the work and analyze the results without using data storage or computing resources on their own local computers.

We also compared three ways to accomplish the same tasks involved in the snow modeling use cases: (1) conducting research without HydroDS web services, (2) conducting research with HydroDS before system integration, and (3) conducting research with HydroDS after system integration (Table 4.3). The first option represents how modelers are doing modeling research now. The second option represents the use of modeling web services to simplify the work involved in the first option, which might still be difficult in a real application because of the requirement for learning and writing program code. The third option represents a new way of using the modeling web services, which provides a graphical user interface to lower the requirement of programming and the functionality to support data curation and sharing.

This comparison allowed us to evaluate whether the system integration could accomplish the modeling work with less need for coding, and fewer manual operations or data transitions among different environments. We found that the system integration provided benefits in several aspects. First, the system integration lowered the requirement for writing code to interact with HydroDS web services. The UEB web app only requires knowledge of the UEB model, which allows users to overcome the programming barrier, saving the time required to write Python code. Additionally, the Python script created by HydroDS to document the input preparation workflow also helps to learn and use the web services from example code.

Second, the system integration simplifies data curation and management efforts. The data files, metadata, and script are automatically curated in the data sharing system without manually moving the files among different environments (HydroDS, local computer, and data sharing systems), a process that can be error prone with potential for information loss. This automatic data transfer capability can encourage the preparation and sharing of modeling work rather that retaining it only on local computers. This also supports collaboration and makes it easier to comply with open data mandates and document reproducibility.

Third, the system integration can simplify the way for others to validate reproducibility of the modeling work, and reuse or extend it for their own work. Users can use the UEB web app and the CUAHSI JupyterHub web app to repeat or modify the modeling work without downloading the files to their local computers or configuring their local environments for model execution or data analysis.

4.5 Conclusions

In hydrologic modeling research, we are starting to see the availability of more and more hydrologic modeling web services that enable users to write code and make their work more efficient. However, limitations still exist in real application of such services in terms of their usability and the reproducibility of the modeling work. Users need to learn and write code to utilize these web services, which may be a barrier for those with limited programming skills. In addition, a good mechanism is needed for curation and sharing of not only the data and metadata, but also the script of the modeling work, which can improve the research reproducibility and encourage collaborations around them.

In this paper, we presented an approach that uses a three-layer architecture to integrate open source software to enable web-based simulation to support hydrologic modeling research. As an example, we integrated the HydroDS hydrologic modeling web services with a data sharing system, HydroShare, and tested the implemented functionality with use cases of snowmelt modeling for the Animas watershed in the Colorado River Basin. The results demonstrated that the system integration enabled users to work within an online environment to create, describe, share, discover, modify and analyze the modeling work, which encourages collaboration around the hydrologic modeling research and significantly reduces the need for coding and manual operation for data transfer and model configuration. This approach has the advantage of reusing open source software to support hydrologic modeling research in terms of collaboration, computer platform independence, and reproducibility of modeling workflows and results.

In addition, the general design of the three-layer architecture can be adopted or adapted to other open source data sharing and modeling software. Furthermore, other modeling web services can be integrated with a data sharing system such as HydroShare using the methods we described to support automated data curation and post-modeling analysis without repeating development of similar functionality. While we used HydroShare for our work, other data sharing systems could also be used. We found that the following data sharing system features were needed to ease integration with other cyberinfrastructure and add value to them. First, the system should have well-developed data sharing functionality and corresponding web service API for interoperating with other systems over the Internet. For example, HydroShare has the REST API Python client, which helped us to develop new web services in HydroDS that enable automatic data transfer between the two systems to support data curation and sharing. Secondly, the data sharing system needs to be a platform where new functionality for interacting with the shared datasets can be added as loosely coupled components (e.g., as web apps) without requiring significant changes to the existing system. For instance, the HydroShare Tethys Apps portal established by HydroShare team was used to host the UEB web app, which provided a user interface layer to interact with HydroDS and HydroShare with minimal changes in both systems.

In the future, possible development could include a new web app that provides graphical user interface for multiple data processing web services from HydroDS. This would benefit researchers by making it easier for them to reuse and combine different web services based on their need and to prepare inputs for other hydrologic models without writing code while having the results directly curated in HydroShare.

Acknowledgements

This work was supported by the National Science Foundation under collaborative grants OCI-1148453 and ICER-1639655. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Archfield, S.A., Clark, M., Arheimer, B., Hay, L.E., McMillan, H., Kiang, J.E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W.H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., Over, T., 2015. Accelerating advances in continental domain hydrologic modeling. Water Resour. Res. 51, 10078–10091. https://doi.org/10.1002/2015WR017498

Billah, M.M., Goodall, J.L., Narayan, U., Essawy, B.T., Lakshmi, V., Rajasekar, A., Moore, R.W., 2016. Using a data grid to automate data preparation pipelines required for regional-scale hydrologic modeling. Environ. Model. Softw. 78, 31–39. https://doi.org/10.1016/j.envsoft.2015.12.010

Byrne, J., Heavey, C., Byrne, P.J., 2010. A review of Web-based simulation and supporting tools. Simul. Model. Pract. Theory 18, 253–276. https://doi.org/10.1016/j.simpat.2009.09.013

Castronova, T., 2016. HydroShare Jupyterhub Notebook Server. Github repository. https://github.com/hydroshare/hydroshare-jupyterhub

Demir, I., Krajewski, W.F., 2013. Towards an integrated Flood Information System: Centralized data access, analysis, and visualization. Environ. Model. Softw. 50, 77–84. https://doi.org/10.1016/j.envsoft.2013.08.009

Dingman, S.L., 2008. Physcial Hydrology (second edition). Waveland Press, Inc, Long Grove, IL.

Gan, T., 2019a. Animas watershed snowmelt modeling in the 2010 water year (use case 1). HydroShare. https://doi.org/10.4211/hs.1be4d7902c87481d85b93daad99cf471

Gan, T., 2019b. Animas watershed snowmelt modeling in the 2010 water year (use case 2). HydroShare. https://doi.org/10.4211/hs.a2b87a2f25d046958ac604e522f449c0

Gesch, D.B., 2007. The National Elevation Dataset, in: Maune, D. (Ed.), Digital Elevation Model Technologies and Applications—the DEM Users Manual. American Society for Photogrammetry and Remote Sensing, Bethesda, MD, pp. 99–118.

Gichamo, T.Z., 2019. Advancing streamflow forecasts through the application of a physically based energy balance snowmelt model with data assimilation and cyberinfrastructure resources (Doctoral dissertation). Utah State University, Logan, UT. https://digitalcommons.usu.edu/etd/7463

Granell, C., Dıaz, L., Gould, M., Díaz, L., Gould, M., 2010. Service-oriented applications for environmental models: Reusable geospatial services. Environ. Model. Softw. 25, 182–198. https://doi.org/10.1016/j.envsoft.2009.08.005

Heard, J., Tarboton, D.G., Idaszak, R., Horsburgh, J.S., Bedig, A., Castronova, A.M., Couch, A., Frisby, C., Gan, T., Goodall, J., Jackson, S., Livingston, S., Maidment, D., Martin, N., Miles, B.C., Mills, S., Sadler, J., Valentine, D., Zhao, L., 2014. An architectural overview of HydroShare, a next-generation hydrologic information system. Proceedings of the 11th International Conference on Hydroinformatics, HIC 2014, New York City, USA.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K., 2015. Completion of the 2011 national land cover database for the conterminous United States – Representing a decade of land cover change information. Photogramm. Eng. Remote Sens, 81(0):345-354.

Horsburgh, J.S., Morsy, M.M., Castronova, A.M., Goodall, J.L., Gan, T., Yi, H., Stealey, M.J., Tarboton, D.G., 2015. HydroShare: Sharing diverse hydrologic data types and models as social objects within a Hydrologic Information System,. J. Am. Water Resour. Assoc. 27517. https://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 Resour. Res. 52, 7548–7555. https://doi.org/10.1002/2016WR019285

Kumar, S.V., Peters-Lidard, C.D., Eastman, J.L., Tao, W.K., 2008. An integrated high-resolution hydrometeorological modeling testbed using LIS and WRF. Environ. Model. Softw. 23, 169–181. https://doi.org/10.1016/j.envsoft.2007.05.012

Laloy, E., Vrugt, J.A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing. Water Resour. Res. 48, 1–18. https://doi.org/10.1029/2011WR010608

Leonard, L., Duffy, C.J., 2013. Essential terrestrial variable data workflows for distributed water resources modeling. Environ. Model. Softw. 50, 85–96. https://doi.org/10.1016/j.envsoft.2013.09.003

Liang, X., Lettenmaier, D.P., Wood, E.F., 1996. One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model. J. Geophys. Res. Atmos. 101, 21403–21422. https://doi.org/10.1029/96JD01448

Miles, B.C., 2014. Small-scale residential stormwater management in urbanized watersheds: A geoinformatics-driven ecohydrology modeling approach (Doctoral dissertation). The University of North Carolina, Chapel Hill, NC, https://doi.org/10.17615/1m6y-vs10

Morsy, M.M., Goodall, J.L., Castronova, A.M., Dash, P., Merwade, V., Sadler, J.M., Rajib, M.A., Horsburgh, J.S., Tarboton, D.G., 2017. Design of a metadata framework for environmental models with an example hydrologic application in HydroShare. Environ. Model. Softw. 93, 13–28. https://doi.org/10.1016/j.envsoft.2017.02.028

Peckham, S.D., Hutton, E.W.H., Norris, B., 2013. A component-based approach to integrated modeling in the geosciences: The design of CSDMS. Comput. Geosci. 53, 3–12. https://doi.org/10.1016/j.cageo.2012.04.002

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. Environ. Model. Softw. 75, 498–512. https://doi.org/10.1016/j.envsoft.2015.10.032

Silliman, S.E., Hamlin, C., Crane, P.E., 2008. International collaborations and incorporating the social sciences in research in hydrology and hydrologic Engineering. J. Hydrol. Eng. 13, 13–19. https://doi.org/10.1061/(ASCE)1084-0699(2008)13:1(13)

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, 190030, https://doi.org/10.1038/sdata.2019.30

Swain, N.R., Christensen, S.D., Snow, A.D., Dolder, H., Goharian, E., Jones, N.L., Nelson, E.J., Espinoza-d, G., Ames, D.P., Burian, S.J., 2016. A new open source platform for lowering the barrier for environmental web app development. Environ. Model. Softw. 85, 11-26. https://doi.org/10.1016/j.envsoft.2016.08.003

Tarboton, D.G., 1997. A new method for the determination of flow directions and contributing areas in grid digital elevation models. Water Resour. Res. *33*, 309-319, http://dx.doi.org/10.1029/96WR03137

Tarboton, D.G., Idaszak, R., Horsburgh, J.S., Heard, J., Ames, D., Goodall, J.L., Band, L., Merwade, V., Couch, A., Arrigo, J., Hooper, R., Valentine, D., Maidment, D., 2014. HydroShare: Advancing collaboration through hydrologic data and model sharing, in: 7th International Congress on Environmental Modelling and Software. International Environmental Modelling and Software Society (iEMSs).

Tarboton, D.G., Luce, C.H., 1996. “Utah Energy Balance Snow Accumulation and Melt Model (UEB),” Computer model technical description and users guide. Utah

Water Research Laboratory and USDA Forest Service Intermountain Research Station.

Thornton, P.E., Thornton, M.M., Mayer, B.W., Wei, Y., Devarakonda, R., Vose, R. S., Cook, R.B., 2016. Daymet: Daily surface weather data on a 1-km grid for north america. ORNL Distributed Active Archive Center, v3. https://doi.org/10.3334/ORNLDAAC/1328

Walker, J.D., Chapra, S.C., 2014. A client-side web application for interactive environmental simulation modeling. Environ. Model. Softw. 55, 49–60. https://doi.org/10.1016/j.envsoft.2014.01.023

Zender, C.S., 2008. Analysis of self-describing gridded geoscience data with netCDF Operators (NCO). Environ. Model. Softw. 23, 1338–1342. https://doi.org/10.1016/j.envsoft.2008.03.004

Table 4.1 UEB model input variables and HydroDS Python client functions for input preparation.

Input type	Specific variables	Major Python client functions for preparation
Model domain	Watershed grid	subset_raster() delineate_watershed() raster_to_netcdf()
Terrain	Slope Aspect	create_raster_aspect() create_raster_slope() raster_to_netcdf()
Canopy	Canopy cover Canopy height Leaf area index	project_clip_raster() get_canopy_variable()
Climate	Incoming shortwave radiation Minimum air temperature Maximum air temperature Air vapor pressure Precipitation	subset_netcdf() concatenate_netcdf() subset_netcdf_by_time() project_subset_resample_netcdf()

Table 4.2 Inputs set for model input preparation in the first case study.

Item	Value	Required? (Yes/No)
Bounding box [north, south, west, east]	[37.9695, 37.2626, -108.0505, -107.5150] in degrees	Yes
Energy content initial condition	0	Yes
Snow water equivalent initial condition	0	Yes
Snow surface dimensionless age initial condition	0	Yes
Snow water equivalent of canopy condition	0	Yes
Snow surface temperature one day prior to the model starting time	0	Yes
Spatial coordinate system	NAD83/UTM zone 13N	Yes
Time period [start date, end date]	[2009/10/01, 2010/10/01]	Yes
Cell size for model simulation [dx, dy]	[1200, 1200] in meter	Yes
Watershed outlet [longitude, latitude]	[-107.8797, 37.27917] in degree	No
HydroShare resource title	Animas watershed snowmelt modeling in 2010 water year (case study1)	No
HydroShare resource keywords	snow melt, UEB Utah Energy Balance Model	No

Table 4.3 Comparison of three ways to accomplish tasks for the snowmelt modeling use cases.

Modeling task	**Option1: Traditional method**	**Option2: Use HydroDS before integration**	**Option3: Use HydroDS after integration**
Prepare input and execute model	**Local PC:** • Collect data from multiple sources • Learn and write code • Install software to run script • Install and configure model	**Local PC:** • Learn HydroDS and write code • Install software to run script	**Data sharing system:** • Enter required information in the UEB web app
Curate and share results	**Local PC:** • Manually upload data and script to a data sharing system • Manually add metadata	**Local PC:** • Download model input/output from HydroDS • Manually upload data and script to a data sharing system • Manually add metadata	**Data sharing system:** • Data, script, and metadata directly stored in HydroShare
Repeat or modify modeling work	**Data Sharing system** • Download script and data **Local PC** • Learn and modify script • Install software to run script • Install and configure model	**Data Sharing system** • Download script and data **Local PC** • Learn and modify script • Install software to run script	**Data sharing system:** • Enter required information in the UEB web app or • Use HydroShare's Jupyter Notebook web app to modify and run script (if familiar with HydroDS)

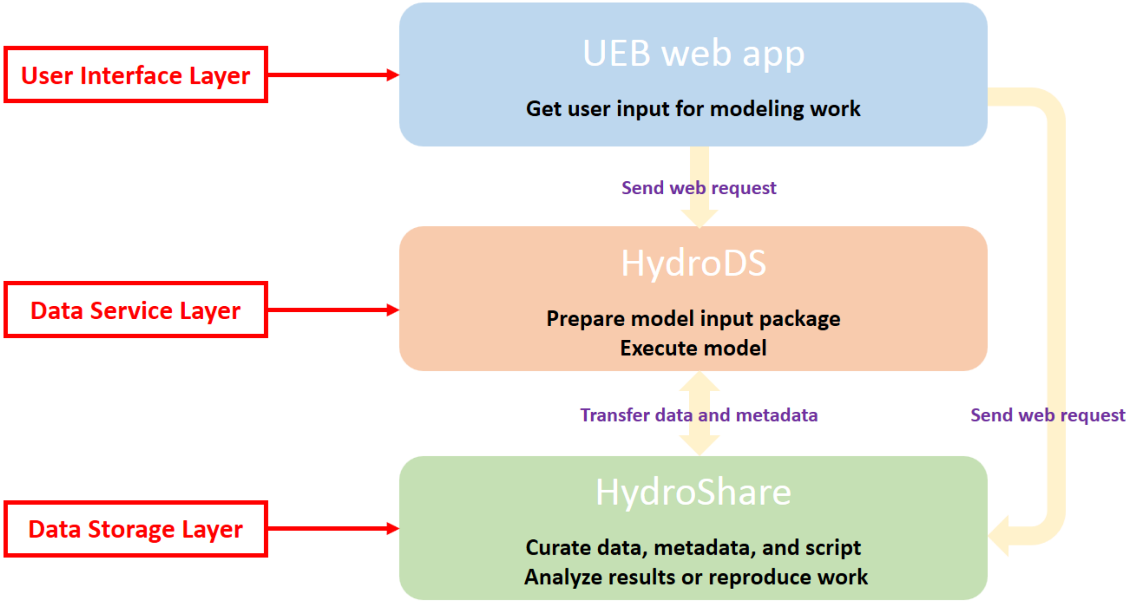


Figure 4.1 A three-layer architecture to integrate hydrologic modeling web services (e.g., HydroDS) with a data sharing system (e.g., HydroShare).

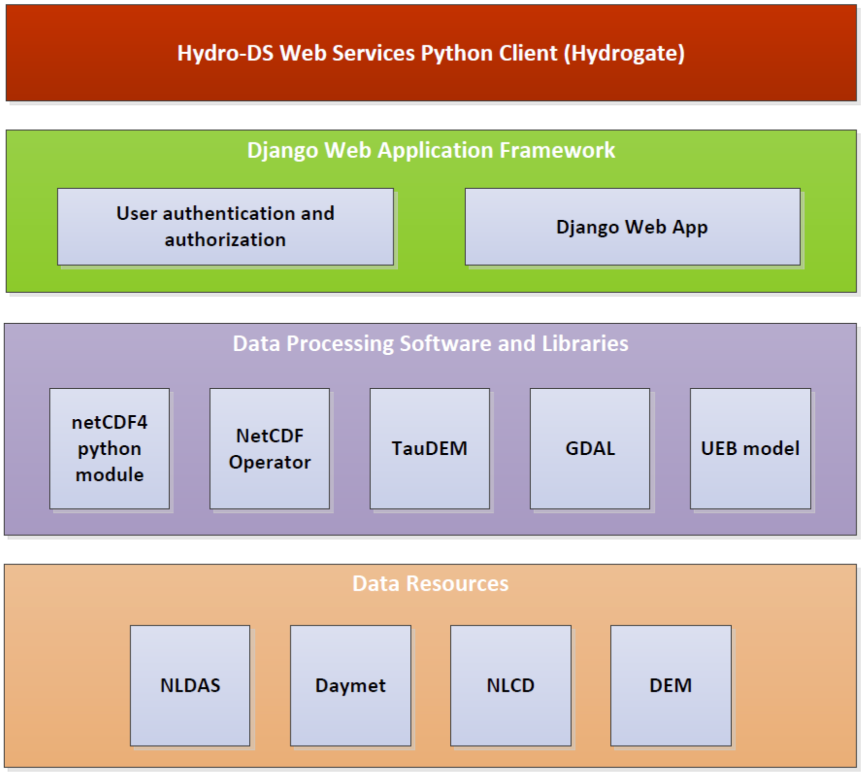


Figure 4.2 The HydroDS system architecture.

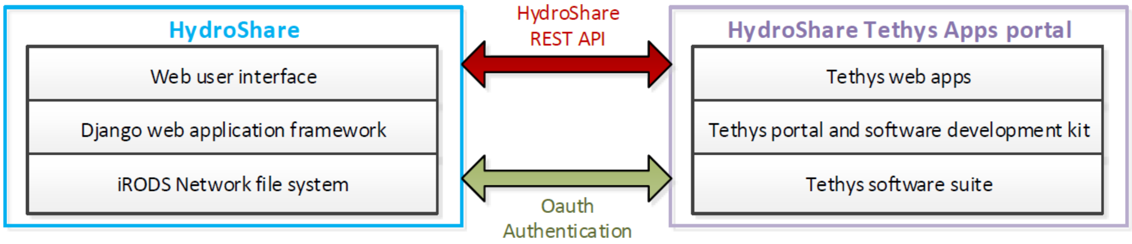
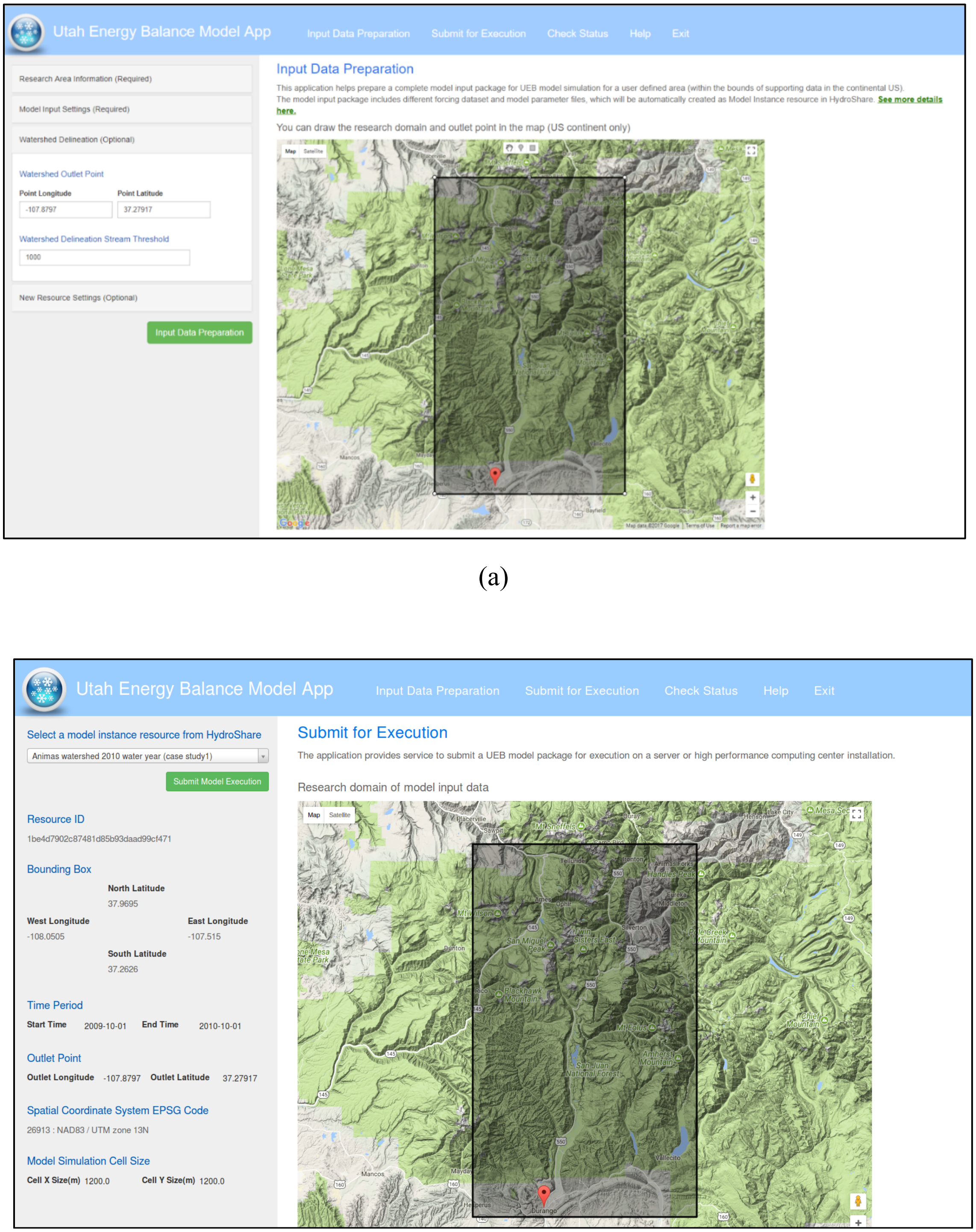


Figure 4.3 System architecture of HydroShare and HydroShare Tethys Apps portal



Input Data Preparation

This application helps prepare a complete model input package for UEB model simulation for a user defined area (within the bounds of supporting data in the continental US). The model input package includes different forcing dataset and model parameter files, which will be automatically created as Model Instance resource in HydroShare. See more details here.

You can draw the research domain and outlet point in the map (US continent only)

Watershed Outlet Point

Point Longitude: -107.8797
Point Latitude: 37.27917

Watershed Delineation Stream Threshold: 1000

(a)

Submit for Execution

The application provides service to submit a UEB model package for execution on a server or high performance computing center installation.

Research domain of model input data

Select a model instance resource from HydroShare: Animas watershed 2010 water year (case study1)

Resource ID: 1be4d7902c87481d85b93daad99cf471

Bounding Box

North Latitude: 37.9695
West Longitude: -108.0505
East Longitude: -107.515
South Latitude: 37.2626

Time Period

Start Time 2009-10-01 End Time 2010-10-01

Outlet Point

Outlet Longitude -107.8797 Outlet Latitude 37.27917

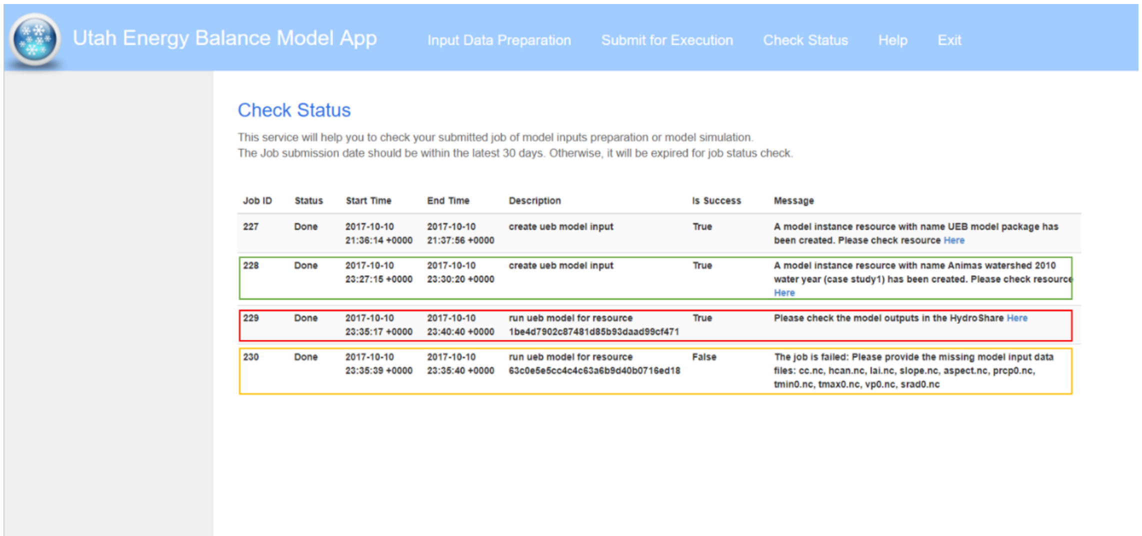
Spatial Coordinate System EPSG Code: 26913 : NAD83 / UTM zone 13N

Model Simulation Cell Size

Cell X Size(m) 1200.0 Cell Y Size(m) 1200.0

(b)

Figure 4.4 User interface of the UEB web app for input preparation (a) and model execution (b).



Check Status

This service will help you to check your submitted job of model inputs preparation or model simulation.
The Job submission date should be within the latest 30 days. Otherwise, it will be expired for job status check.

Job ID	Status	Start Time	End Time	Description	Is Success	Message
227	Done	2017-10-10 21:36:14 +0000	2017-10-10 21:37:56 +0000	create ueb model input	True	A model instance resource with name UEB model package has been created. Please check resource Here
228	Done	2017-10-10 23:27:15 +0000	2017-10-10 23:30:20 +0000	create ueb model input	True	A model instance resource with name Animas watershed 2010 water year (case study1) has been created. Please check resource Here
229	Done	2017-10-10 23:35:17 +0000	2017-10-10 23:40:40 +0000	run ueb model for resource 1be4d7902c87481d85b93daad99cf471	True	Please check the model outputs in the HydroShare Here
230	Done	2017-10-10 23:35:39 +0000	2017-10-10 23:35:40 +0000	run ueb model for resource 63c0e5e5cc4c4c63a6b9d40b0716ed18	False	The job is failed: Please provide the missing model input data files: cc.nc, hcan.nc, lai.nc, slope.nc, aspect.nc, prcp0.nc, tmin0.nc, tmax0.nc, vp0.nc, srad0.nc

Figure 4.5 User interface of the UEB web app for job status checking.

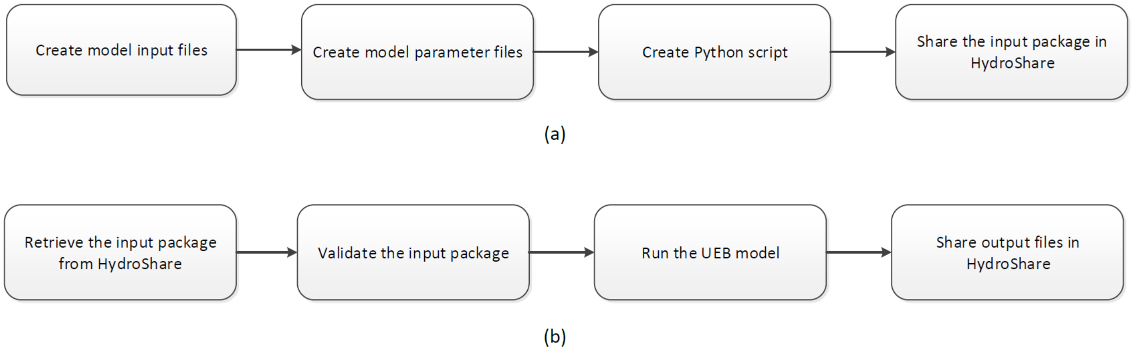
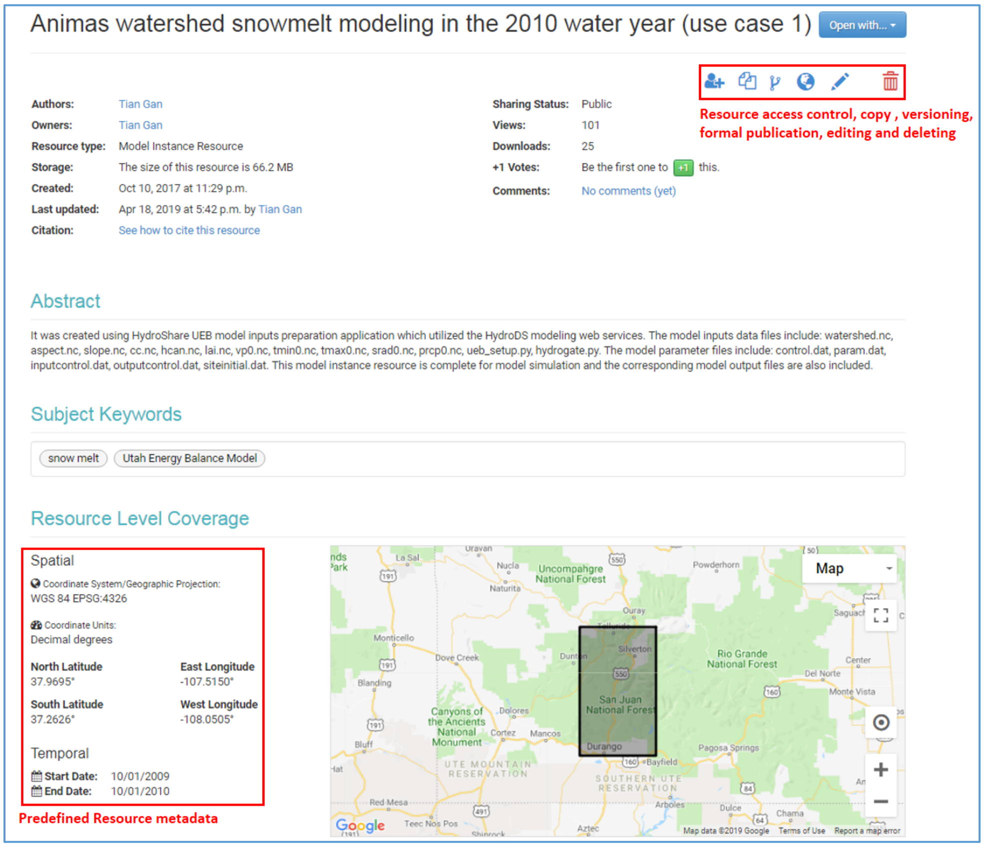


Figure 4.6 The functionality of the added web services in HydroDS. Panel (a) for model input package preparation; Panel (b) for model simulation.



Animas watershed snowmelt modeling in the 2010 water year (use case 1)

Authors:	Tian Gan
Owners:	Tian Gan
Resource type:	Model Instance Resource
Storage:	The size of this resource is 66.2 MB
Created:	Oct 10, 2017 at 11:29 p.m.
Last updated:	Apr 18, 2019 at 5:42 p.m. by Tian Gan
Citation:	See how to cite this resource

Sharing Status:	Public
Views:	101
Downloads:	25
+1 Votes:	Be the first one to +1 this.
Comments:	No comments (yet)

Resource access control, copy , versioning, formal publication, editing and deleting

Abstract

It was created using HydroShare UEB model inputs preparation application which utilized the HydroDS modeling web services. The model inputs data files include: watershed.nc, aspect.nc, slope.nc, cc.nc, hcan.nc, lai.nc, vp0.nc, tmin0.nc, tmax0.nc, srad0.nc, prcp0.nc, ueb_setup.py, hydrogate.py. The model parameter files include: control.dat, param.dat, inputcontrol.dat, outputcontrol.dat, siteinitial.dat. This model instance resource is complete for model simulation and the corresponding model output files are also included.

Subject Keywords

snow melt
Utah Energy Balance Model

Resource Level Coverage

Spatial

Coordinate System/Geographic Projection:
WGS 84 EPSG:4326

Coordinate Units:
Decimal degrees

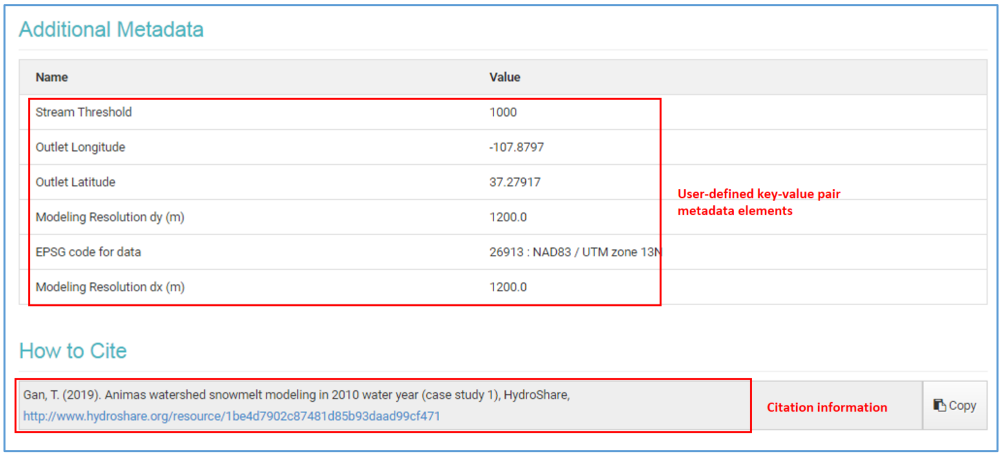
North Latitude 37.9695°	**East Longitude** -107.5150°
South Latitude 37.2626°	**West Longitude** -108.0505°

Temporal

Start Date: 10/01/2009
End Date: 10/01/2010

Predefined Resource metadata

(a)



Additional Metadata

Name	Value
Stream Threshold	1000
Outlet Longitude	-107.8797
Outlet Latitude	37.27917
Modeling Resolution dy (m)	1200.0
EPSG code for data	26913 : NAD83 / UTM zone 13N
Modeling Resolution dx (m)	1200.0

User-defined key-value pair metadata elements

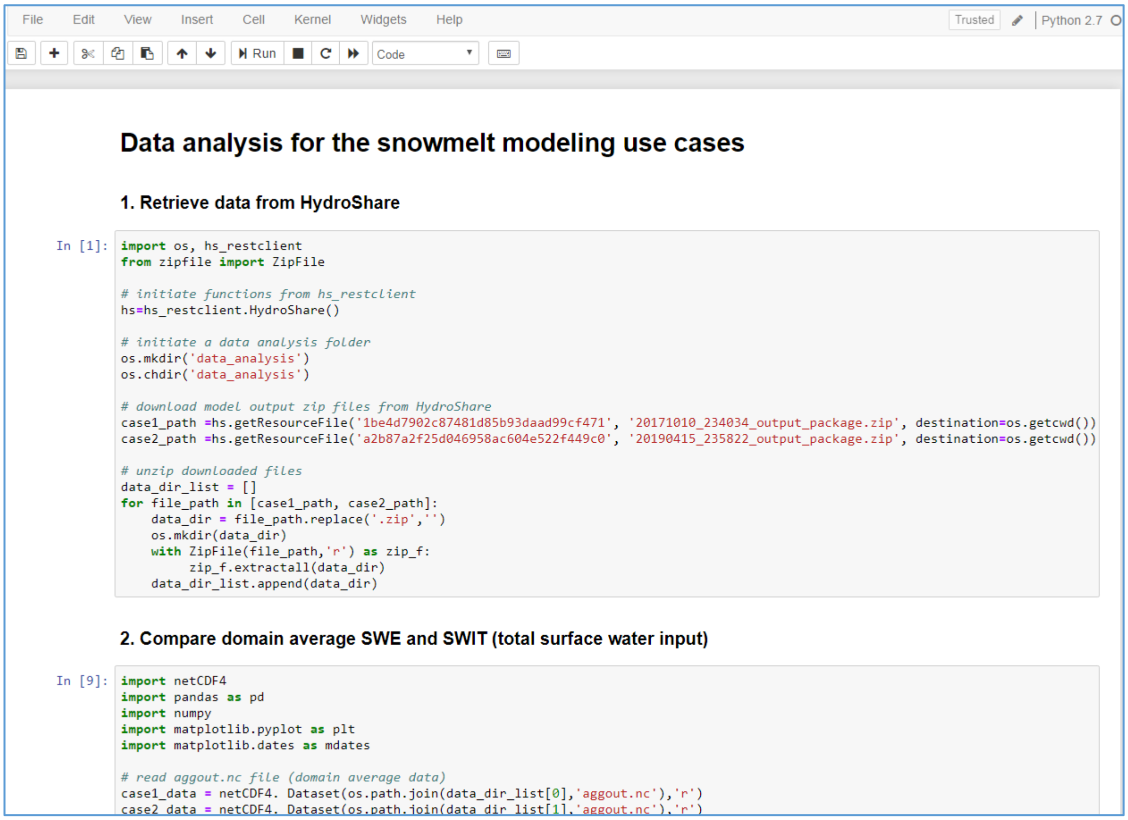
How to Cite

Gan, T. (2019). Animas watershed snowmelt modeling in 2010 water year (case study 1), HydroShare, http://www.hydroshare.org/resource/1be4d7902c87481d85b93daad99cf471

Citation information

(b)

Figure 4.7 Example model instance resource in HydroShare. Panel (a) shows different resource functions and predefined metadata; Panel (b) shows the user-defined metadata and suggested citation information.



Data analysis for the snowmelt modeling use cases

1. Retrieve data from HydroShare

In [1]:

```python
import os, hs_restclient
from zipfile import ZipFile

# initiate functions from hs_restclient
hs=hs_restclient.HydroShare()

# initiate a data analysis folder
os.mkdir('data_analysis')
os.chdir('data_analysis')

# download model output zip files from HydroShare
case1_path =hs.getResourceFile('1be4d7902c87481d85b93daad99cf471', '20171010_234034_output_package.zip', destination=os.getcwd())
case2_path =hs.getResourceFile('a2b87a2f25d046958ac604e522f449c0', '20190415_235822_output_package.zip', destination=os.getcwd())

# unzip downloaded files
data_dir_list = []
for file_path in [case1_path, case2_path]:
    data_dir = file_path.replace('.zip','')
    os.mkdir(data_dir)
    with ZipFile(file_path,'r') as zip_f:
         zip_f.extractall(data_dir)
    data_dir_list.append(data_dir)
```

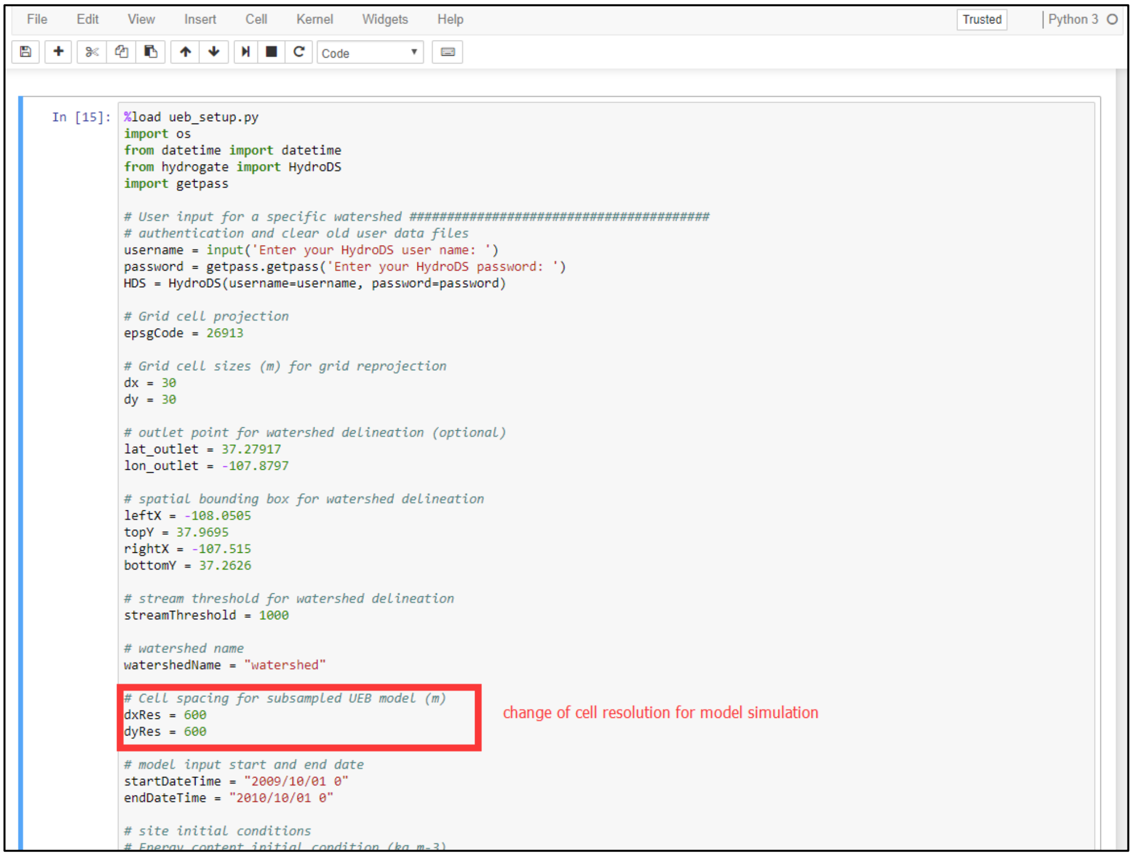
2. Compare domain average SWE and SWIT (total surface water input)

In [9]:

```python
import netCDF4
import pandas as pd
import numpy
import matplotlib.pyplot as plt
import matplotlib.dates as mdates

# read aggout.nc file (domain average data)
case1_data = netCDF4. Dataset(os.path.join(data_dir_list[0],'aggout.nc'),'r')
case2_data = netCDF4. Dataset(os.path.join(data_dir_list[1],'aggout.nc'),'r')
```

Figure 4.8 Python code for post-modeling analysis comparison plots in the CUAHSI JupyterHub web app.



```
In [15]: %load ueb_setup.py
         import os
         from datetime import datetime
         from hydrogate import HydroDS
         import getpass

         # User input for a specific watershed ########################################
         # authentication and clear old user data files
         username = input('Enter your HydroDS user name: ')
         password = getpass.getpass('Enter your HydroDS password: ')
         HDS = HydroDS(username=username, password=password)

         # Grid cell projection
         epsgCode = 26913

         # Grid cell sizes (m) for grid reprojection
         dx = 30
         dy = 30

         # outlet point for watershed delineation (optional)
         lat_outlet = 37.27917
         lon_outlet = -107.8797

         # spatial bounding box for watershed delineation
         leftX = -108.0505
         topY = 37.9695
         rightX = -107.515
         bottomY = 37.2626

         # stream threshold for watershed delineation
         streamThreshold = 1000

         # watershed name
         watershedName = "watershed"

         # Cell spacing for subsampled UEB model (m)
         dxRes = 600
         dyRes = 600

         # model input start and end date
         startDateTime = "2009/10/01 0"
         endDateTime = "2010/10/01 0"

         # site initial conditions
         # Energy content initial condition (kg m-3)
```

change of cell resolution for model simulation

Figure 4.9 Python script for model input preparation loaded into a Jupyter Notebook file in the CUAHSI JupyterHub web app.

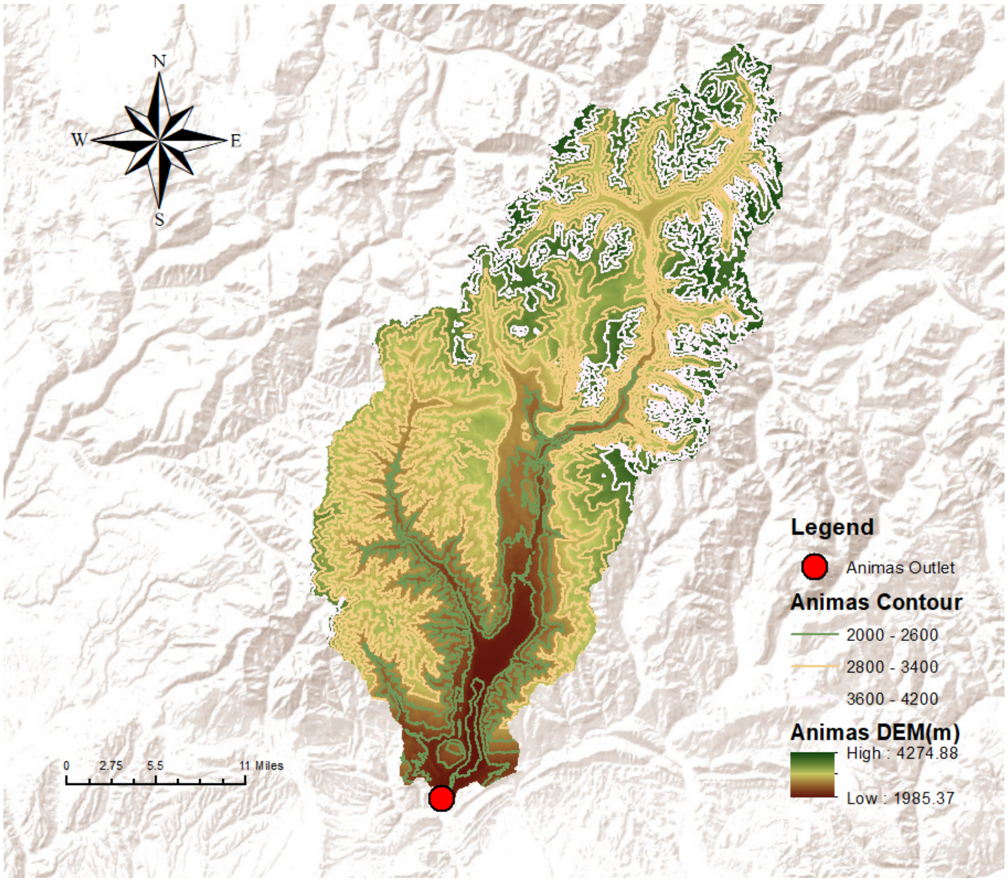
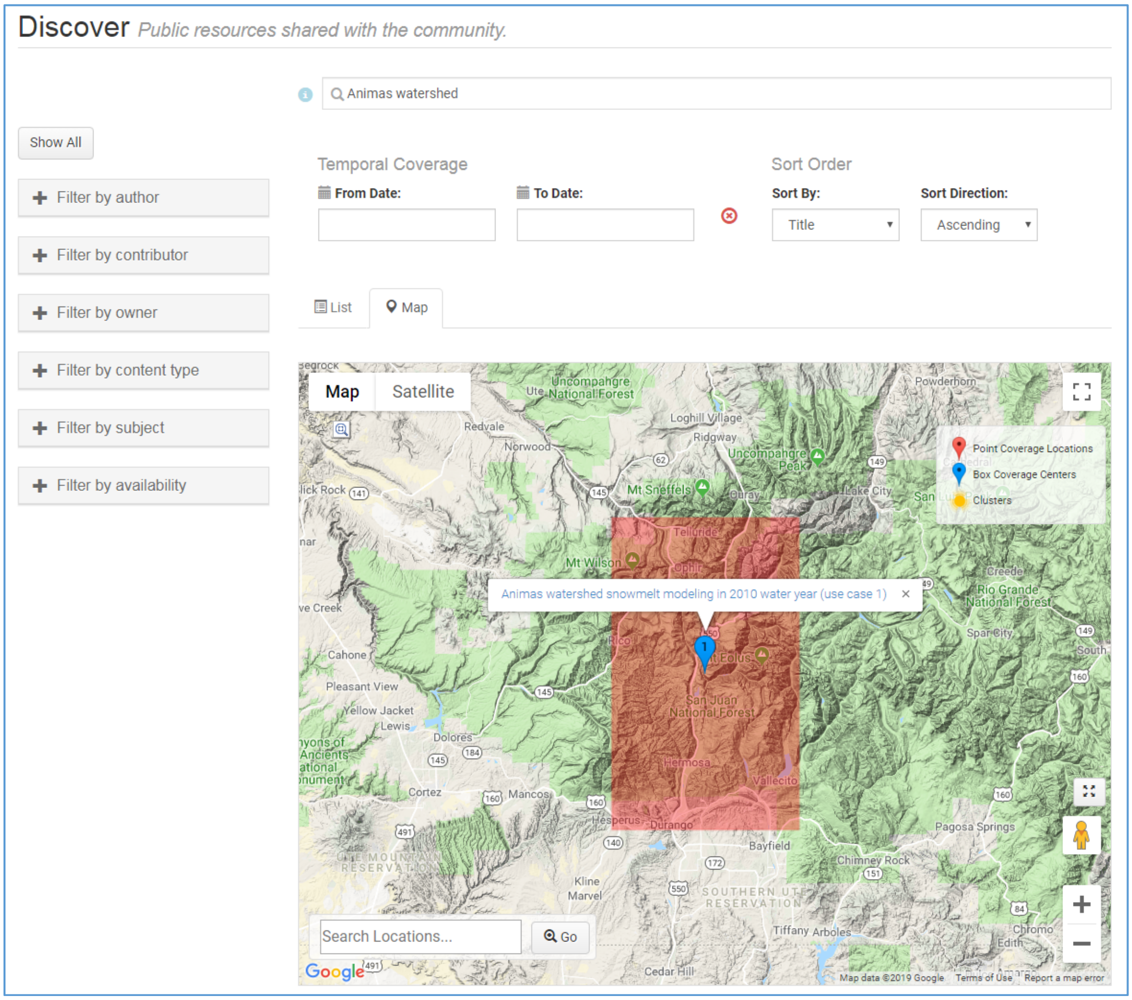


Figure 4.10 The Animas watershed in the Colorado River Basin.



Discover *Public resources shared with the community.*

Animas watershed

Sort By: Title

Sort Direction: Ascending

Animas watershed snowmelt modeling in 2010 water year (use case 1)

Figure 4.11 The HydroShare discovery page used to search for the model instance resource created in the first case study.

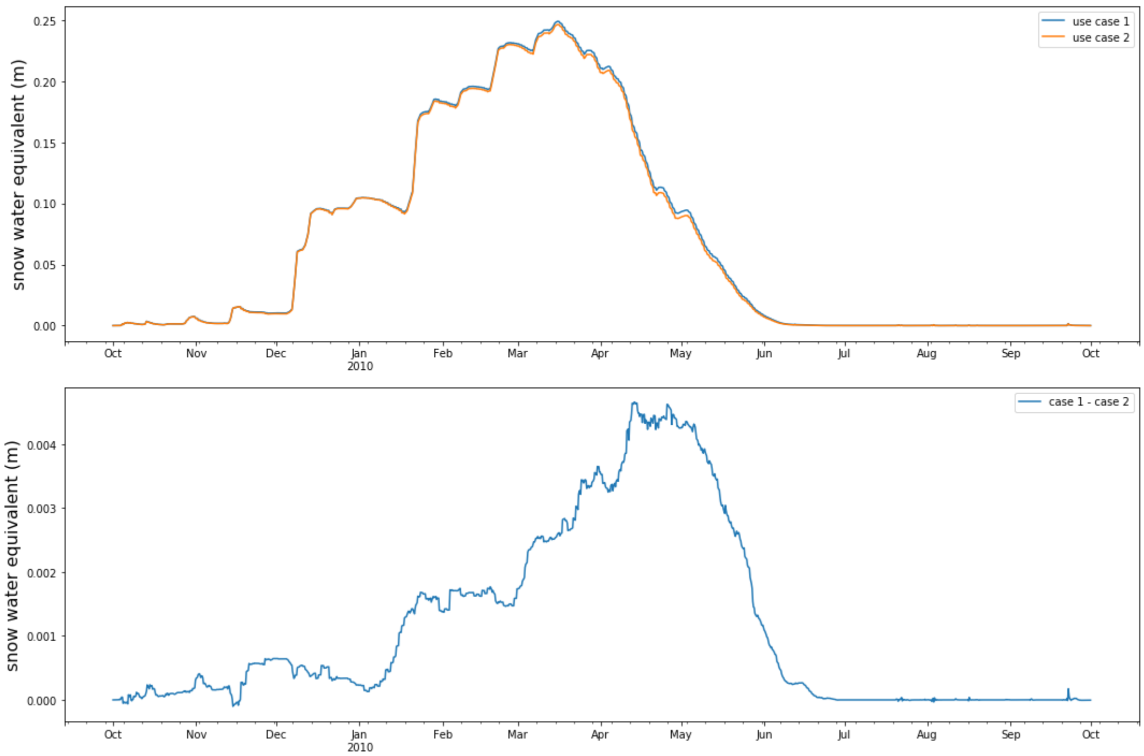


Figure 4.12 Comparison of snow water equivalent created by the two uses cases.

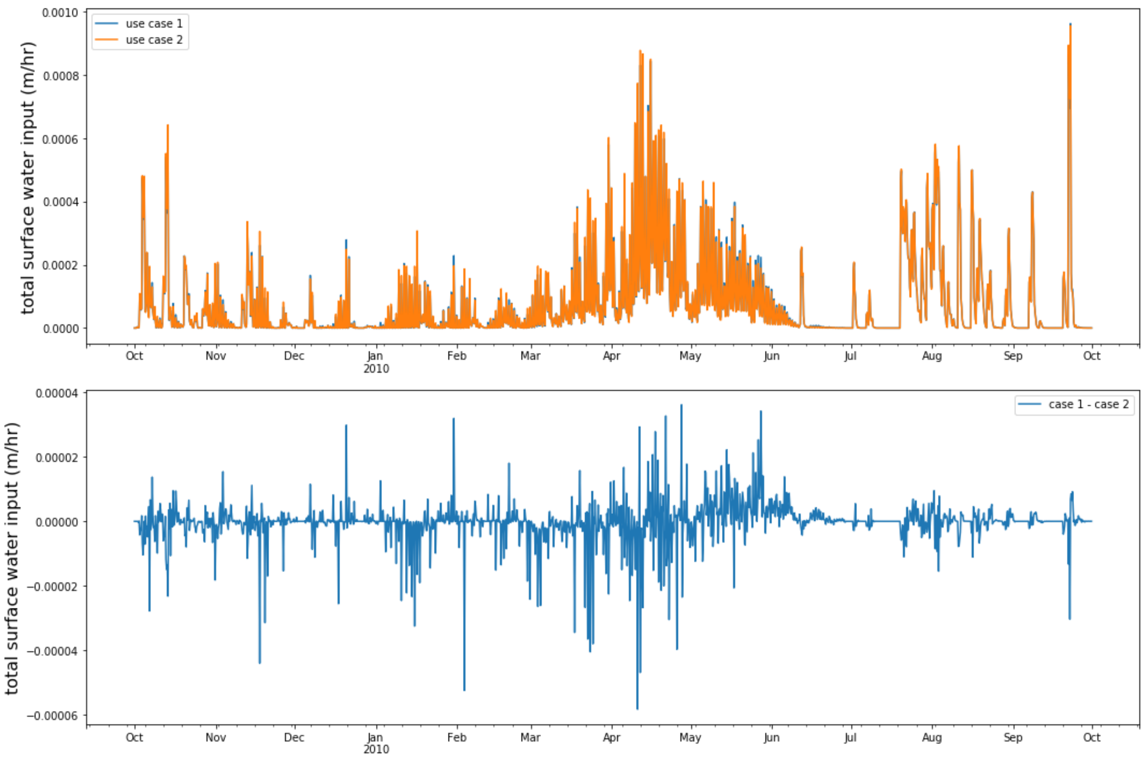


Figure 4.13 Comparison of total surface water input created by the two uses cases.

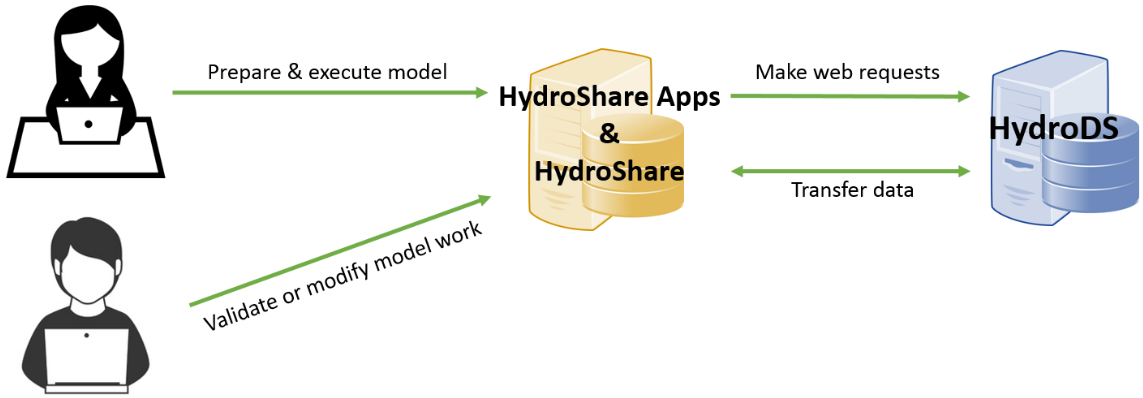


Figure 4.14 The system integration enables users to interact with HydroShare and HydroShare Apps for multiple modeling tasks.

CHAPTER 5

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

This dissertation presents research that advances cyberinfrastructure for collaborative data sharing and modeling in hydrology. Modeling to improve methods for operational water supply forecasts in the Colorado River Basin was a driving use case. There were three objectives:

1. Evaluate temperature index and energy balance snow models to improve operational water supply forecasts in the Colorado River Basin.
2. Develop capability for multidimensional space-time data sharing in HydroShare to facilitate data management and reuse.
3. Integrate hydrologic modeling web services with HydroShare to improve reproducibility of hydrologic modeling research.

Chapter 2 describes the work addressing the first objective. The SNOW-17 temperature index model and the Utah Energy Balance (UEB) physically based snowmelt model were coupled with the SAC-SMA runoff model and the Ruptix7 routing model to simulate basin snowmelt and discharge in two test watersheds in the Colorado River Basin over a time span of two decades. The modeling and analysis results demonstrate that both the SNOW-17 and the UEB models can provide reliable simulation of the basin snowmelt and discharge with reasonable timing and amount. While the performance of each model was similar, the UEB model has the following advantages. First, the UEB model is able to simulate sublimation, which is an important evaporative component that may affect seasonal water resources availability. Since the SNOW-17 model does not explicitly represent the sublimation process, its evaporative losses were calibrated into the SAC-SMA land surface model. On the other hand, the UEB model's direct representation of sublimation can be utilized to help modelers get a better understanding of the different evaporative components and evaluate their sensitivities under conditions of changing land cover or land use. Second, the UEB model requires less calibration than that for the SNOW-17 model, which can reduce some intensive calibration work required in model application. In the two test watersheds, most of the UEB model parameters were held constant (except for drift factor), while the SNOW-17 model requires calibration in each of the subwatersheds. This shows that the UEB model parameters are more transferable under different terrain and climate conditions than the SNOW-17 model parameters, and also indicates that the UEB model is potentially better for simulating snowmelt processes in ungauged catchments. Furthermore, the SNOW-17 model is sensitive to the maximum and minimum melting factors, which are calibrated against historical data that may not well represent the future melting conditions under global warming. On the contrary, the UEB model takes advantage of the physical mechanisms that account for energy and water mass balance to simulate the process of snowmelt without relying on the conceptual parameters calibrated based on the historical data. However, there are also limitations that were identified when utilizing the UEB model in this research. For example, manual adjustment of the drift factor was required to account for precipitation input bias. Its value was set based on the calibrated snow correction factor (SCF) of the SNOW-17 model in this research. Without a reference SCF value, it may be challenging to estimate the model input error and adjust the UEB model's drift factor parameter in an efficient and timely manner.

Chapter 3 describes the work addressing the second objective, which developed capability for sharing multidimensional space-time data in HydroShare. In this work, we defined a content type data model for managing multidimensional space-time data and justified the selection of NetCDF as the file format for data storage in HydroShare. The case study showed that by storing multidimensional space-time datasets in a standard NetCDF file, rather than multiple two-dimensional files, it was easier to curate and manage the data. Moreover, the metadata functionality implemented in HydroShare can help simplify the work involved in preparing the metadata from metadata attributes in the NetCDF file. In addition, since the standard OPeNDAP data service can be easily enabled through HydroShare, data reuse is simplified. The case study showed how OPeNDAP enabled subsetting for data reuse without downloading the whole dataset from HydroShare to local PC for visualization and analysis. The HydroShare OPeNDAP service also saves users the work needed to host and maintain the corresponding data servers. Along with the other data sharing and social functions, HydroShare provides a collaborative environment in which users can work together to edit or describe datasets for formal publication or to communicate, evaluate, and iterate on datasets to improve data quality and potential for reuse.

Chapter 4 presents the research addressing the third objective, which designed a three-layer architecture to integrate hydrologic modeling web services with a data sharing system. As a case study, two example systems, HydroDS and HydroShare, were used to implement this approach. This case study demonstrated that users were able to accomplish multiple tasks involved in the modeling process in HydroShare without intensive programing or manually transferring the data among different environments for data curation and sharing. Moreover, other users can discover, access, repeat, or build on the shared work in HydroShare. This provides a way to reproduce the modeling work for collaboration without downloading and configuring the model for use on local computing or storage resources.

The work described in this dissertation contributes to hydrologic research in several aspects. First, the historical analysis of the snowmelt model simulations is an initial step in investigation of incorporating a physically based model within a river forecasting system such as the one used at the CBRFC. This research shows the advantages of potentially applying the UEB model for operational water supply forecasts. This encourages further investigation to evaluate the model performance under forecasting conditions, and to establish the workflow required for practically applying the UEB model in a river forecasting system. Moreover, the proposed approach using the UEB model for basin snow and discharge simulation also holds promise for application in other snow-dominated river basins that have similar climate and topographic conditions. Second, the data sharing functionality developed in HydroShare provides a new way of sharing multidimensional space-time datasets, which enables relatively straightforward data sharing and formal publication to promote collaboration and data reuse to advance hydrologic understanding. Third, the system integration functionality developed provides a web-based simulation environment for hydrologic modeling, which not only simplifies the way of using the modeling web services, but also improves research reproducibility by providing easy access to the modeling work and corresponding web apps to repeat or modify the work for validation and collaboration.

The software results from this research also benefit future cyberinfrastructure development. The method developed to share multidimensional space-time data within

HydroShare can be transferred to other data sharing systems. For the design and implementation of such functionality, one key point is to select a widely used standard file format and metadata elements to organize the datasets to improve the data interoperability for processing and sharing. Another key point is to evaluate the trade-offs between building functionality from scratch and adopting existing standardized tools or services to support the advanced functionality. Moreover, the method to integrate hydrologic modeling web services with HydroShare can be applied to other similar systems for different hydrologic models. A benefit of this method is that it reuses multiple cyberinfrastructure elements to exploit their respective advantages without needing to re-develop similar functionality. In addition, a well-developed API and good mechanism to add new functionality as loosely coupled components are the essential features for a data sharing system to support successful integration with other systems.

While this work has made important advances to address challenges in hydrologic research and cyberinfrastructure development, further work is still needed. For the research to advance the methods used for operational water supply forecasts, additional work is needed to validate calibrated models using recent years of climate forcing input, to have a better understanding of the model performance. It is also important to evaluate the time and computing resources required for both snowmelt models when applied in the river forecasting system under practical forecasting conditions. Moreover, there is a need to figure out an effective way of estimating the climate forcing errors and adjusting the inputs or the model parameters when the UEB model is applied in the river forecasting system.

In terms of the sharing of multidimensional space-time data, additional functionality is needed to guide researchers to convert the datasets stored in other formats into NetCDF format and help those who are not familiar with NetCDF to easily share their datasets. Moreover, although there is client software to support data visualization and processing, it is important to develop functionality to help researchers directly visualize or process the data in HydroShare without installing software in the local PC and transferring the datasets between the online and local environment. One way is to utilize the CUAHSI JupyterHub web app in HydroShare to provide well-developed code to support data processing and visualization. Another way is to implement new web apps that can be connected to HydroShare to provide graphical user interface to support advanced visualization and data analysis functionality.

As for the work of the system integration to support hydrologic modeling research, there are two directions for future work. First, the Tethys web app in current research only supports the complete workflow of model input preparation and model simulation for the UEB model. Future research could use this web app as a template to make the implementation of web apps for other models more efficient. Further, in the general process of hydrologic model input preparation and analysis, it may be possible to identify general and common methods that apply to many models and develop web services and applications that support to minimize the model specific work needed.

Second, when there are no available hydrologic modeling web services for models such as the SNOW-17 and the SAC-SMA models, how to provide an effective way to reproduce the modeling process conducted in the local environment is also a challenge because of the difficulties in setting up exactly the same modeling environment. Sciunit (https://sciunit.run/), which is software for creating self-contained and annotated containers that describe and package computational experiments, provides a possible solution. A future opportunity is to take the SNOW-17 and the SAC-SMA modeling process as a case study to explore ways for using HydroShare and Sciunit to help improve the reproducibility of the modeling work conducted in the local environment.

APPENDICES

CURRICULUM VITAE

TIAN GAN

Department of Civil and Environmental Engineering
Utah State University
Logan, UT 84322, United States
jamy127@foxmail.com

EDUCATION

Ph.D. Hydrology and Water Resources expected 8/ 2019
Utah State University, Logan, Utah

M.Sc. Hydrology and Water Resources 7/ 2012
North China Institute of Water Conservation and Hydroelectricity, Zhengzhou, China

B.Sc. Management of Environmental Resources 7/ 2009
North China Institute of Water Conservation and Hydroelectricity, Zhengzhou, China

RESEARCH EXPERIENCE

Graduate Research Assistant
Utah State University, Logan, Utah

Advancing Water Supply Forecasts in the Colorado River Basin (NASA funded)

(1) Used two snowmelt models (SNOW-17 model and Utah Energy Balance model) to couple with the runoff model (SAC-SMA model) to simulate the basin discharge for study watersheds. This evaluated empirical and physically based snow models to help advance the water supply forecasts methodology used at the Colorado Basin River Forecast Center (CBRFC).

(2) Used high performance computing resources to accomplish automatic-calibration of snowmelt and runoff model parameters.

(3) Used ArcGIS, GDAL and Python to analyze the model results and evaluated the model performance. The results show the potential of applying physically based snowmelt model to improve the reliability of water supply forecasts under future changing conditions.

HydroShare: an Interactive Software Infrastructure for Sustaining Collaborative Community Innovation in the Hydrologic Sciences (NSF funded)

(1) Collaborated with a team of hydrologists and software engineers on the design and development of HydroShare. HydroShare (https://www.hydroshare.org/) is a web based hydrologic information system developed using the Django web framework that supports scientific data management, visualization and analysis.

(2) Designed and developed functions in HydroShare for sharing multidimensional space-time data using NetCDF file format. The functions can facilitate data management and reuse.

(3) Designed and developed functions in HydroShare to simplify model input preparation and model simulation for a snowmelt model. The functions can improve the reproducibility of hydrologic modeling research.

SKILLS

(1) **Modeling**: Utah Energy Balance snow model, SNOW-17 model, SAC-SMA runoff model
(2) **Programming**: Python, R, Git, Django
(3) **Geospatial analysis**: ArcGIS, GDAL, Arcpy
(4) **Presentation**: Numerous presentations of research to project teams, at AGU and CUAHSI Hydroinformatics Conferences.

PUBLICATIONS

(5) **Gan, T.**, Tarboton, D. G., Gichamo, T. Z. (2019), Evaluation of temperature index and energy balance snow models for hydrological applications in operational water supply forecast, in preparation for submission to *Journal of Hydrology.*

(4) **Gan, T.**, Tarboton, D. G., Gichamo, T. Z., Dash, P., Horsburgh, J. S. (2019), Integrating hydrologic modeling web services with online data sharing to prepare, store, and execute models in hydrology, in preparation for submission to *Environmental Modeling and Software*.

(3) **Gan, T.**, Tarboton, D. G., Horsburgh, J. S., Dash, P., Idaszak, R., Yi, H. (2019), Collaborative sharing of multidimensional space-time data in a next generation hydrologic information system, in preparation for submission to *Environmental Modeling and Software.*

(2) Horsburgh, J. S., Morsy, M. M., Castronova, A., Goodall, J. L., **Gan, T.**, Yi, H., Stealey, M. J., and Tarboton, D. G. (2015), HydroShare: Sharing diverse hydrologic data types and models as social objects to the hydrology domain, *Journal of the American Water Resources Association*, 52(4), 873-889, doi:10.1111/1752-1688.12363.

(1) Chen, N. X., **Gan, T.**, Du, Q. H. (2011), Hydrologic frequency analysis using SCEM-UA algorithm, *Journal of Northwest A&F University*, 39(8), 210-214.

CONFERENCE PRESENTATIONS (selected list)

(1)	2018 American Geophysical Union Fall Meeting, Washington, D.C. (poster)	2018
(2)	2017 American Geophysical Union Fall Meeting, New Orleans, LA (poster)	2017
(3)	2015 American Geophysical Union Fall Meeting, San Francisco, CA (poster)	2015
(4)	3rd CUAHSI Conference on Hydroinformatics, Tuscaloosa, AL (poster)	2015
(5)	2nd CUAHSI Conference on Hydroinformatics, Logan, UT (poster)	2013

PROFESSIONAL SERVICES

(1) Affiliations:
American Geophysical Union (2015 - present)
(2) Journal reviewer:
Hydrological Processes (2019)

TEACHING EXPERIENCE

(1) CEE6740: Environmental Quality Modeling (course assistant)
(2) CEE6440: GIS in Water Resources (course assistant)