AWRA 2012 SPRING SPECIALTY CONFERENCE
New Orleans, Louisiana
March 26 – 28, 2012

USING HYDRODESKTOP TO SHARE HYDROLOGIC DATA AND ANALYSES

Tim Whiteaker, Yang Cao, Anthony M. Castronova*

ABSTRACT: To enhance accountability and collaboration, it is important to be able to share not only the results of a study, but also the analysis procedure and input data used to derive the results. This paper describes how HydroDesktop can be used to share data and analyses associated with water observations data. HydroDesktop is a free and open source desktop GIS program for discovering, accessing, and analyzing hydrologic observations data published via the CUAHSI Hydrologic Information System. While the core of HydroDesktop is focused on searching for and downloading data, several extensions have been written to add capabilities for analyzing data that have been retrieved. For example, the HydroR extension enables users to send downloaded data directly to the R software environment for processing. This paper highlights how geospatial and temporal data are shared using the HydroDesktop project file and database, while the sharing of analyses is discussed in the context of a use case involving the HydroR and HydroModeler extensions.
KEY TERMS: HydroDesktop; sharing; water data; hydrologic modeling; simulation evaluation

INTRODUCTION

To enhance accountability and collaboration, it is important to be able to share not only the results of a study, but also the analysis procedure and input data used to derive the results. This paper presents one mechanism for sharing from the perspective of HydroDesktop, a free and open source desktop program for accessing hydrologic data from the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrologic Information System. An example of sharing data and analyses is presented in the context of a use case involving one investigator performing a hydrologic simulation, delivering the output to another investigator who estimates errors in the simulation, and finally delivering the result to a third investigator.

SOFTWARE DESCRIPTION

The CUAHSI Hydrologic Information System (CUAHSI-HIS) provides standards such as WaterML for the communication of water data over the Internet, services such as HIS Central which maintains a catalog indexing the world's water data, and software such as HydroServer for publishing water data in CUAHSI-HIS (Tarboton et al., 2012). The system is focused primarily on the transmission of hydrologic time series data at point locations such as a series of stream discharge measurements from a gauge on a river, although there is limited support for other data types as well. Dozens of data publishers ranging from universities to local, state, and federal agencies have made their data available using CUAHSI-HIS.

HydroDesktop is open source software that enables users to discover, access, analyze, and manage water data available in CUAHSI-HIS (Ames, 2012). HydroDesktop is built upon an open source Geographic Information Systems (GIS) library called DotSpatial, allowing users to interact with CUAHSI-HIS data via maps and attribute tables.

When searching for data, users specify an area of interest, time period, keywords (such as "Streamflow" and "Temperature, air"), and optionally which data sources to use. HydroDesktop queries the HIS Central catalog for time series that match the search criteria and presents results as points on the map. The user can then further filter the search results to select only those time series which should be saved locally. HydroDesktop then downloads the time series data into a SQLite database. This database is specially designed to efficiently store water observations data. For spatial data such as watershed boundaries, HydroDesktop uses shapefiles or other popular GIS formats. Pointers to map layers and underlying data sources, the current SQLite database, and more are stored in the HydroDesktop project file.

*Respectively, Research Associate, The University of Texas at Austin, 1 University Station, R8000, Austin, TX 78712, Phone: (512) 471-0570, Fax: (512) 471-0072, E-Mail: twhit@mail.utexas.edu; Graduate Research Assistant, Idaho State University, 995 University Blvd., Idaho Falls, ID 83401-6290, Phone: (208) 533-8129, E-Mail: caoyang@isu.edu; Graduate Research Assistant, University of South Carolina, 300 Main St., Columbia, SC 29208, Phone: (585) 451-0667, E-Mail: castrona@engr.sc.edu

SHARING DATA AND PROJECT FILES

HydroDesktop is designed to be a consumer of CUAHSI-HIS services, but it currently has no capabilities for publishing data back into CUAHSI-HIS. Therefore, sharing HydroDesktop data typically involves zipping files of interest and either emailing them or placing them on a server accessible to potential collaborators.

The HydroDesktop project file uses relative path names, meaning that it keeps track of where project data are located relative to the project file's current location, and it does not necessarily keep track of the absolute path to the data on the computer's file system. This makes sharing easy if the project data are located in the same folder as the project file or a subfolder of the project file's parent folder. The user would zip the project file, SQLite database of time series data, and shapefiles of spatial data into a single file and deliver that file to a collaborator who would unzip the file on his own computer. The collaborator can then open the unzipped project file in HydroDesktop and have immediate access to the temporal and spatial data, with the HydroDesktop map already zoomed in to the study area that the initial user defined.

If a collaborator is not interested in the spatial data or doesn't want to use HydroDesktop to view the data, a couple of other options exist for sharing data. One option is to only deliver the SQLite database of time series data. Another option is to use HydroDesktop's export functionality to create a delimited text file of the data and deliver the text file to a collaborator. In either of these approaches, the means of sharing data still involve "zip and ship."

SHARING ANALYSES

HydroDesktop extensions enable analysis capabilities beyond simply viewing the data in table or graph formats. These extensions typically have their own software requirements and related data formats, but in many cases the data and details of analyses used can be zipped and shipped along with the HydroDesktop SQLite database, shapefiles, and project data. In the case study that follows, data and analyses are shared between three collaborators. The analyses involve using data related to the Blanco River near Wimberly, TX, to simulate streamflow and evaluate the results by calculating volume error. The analyses are performed using the HydroModeler and HydroR extensions of HydroDesktop, respectively.

CASE STUDY: BLANCO RIVER HYDROLOGIC SIMULATION

The Blanco River in Central Texas flows in a generally eastward direction before joining with the San Marcos River. The USGS operates gauge 08171000 on the Blanco River near Wimberley, whose watershed is shown in Figure 1.

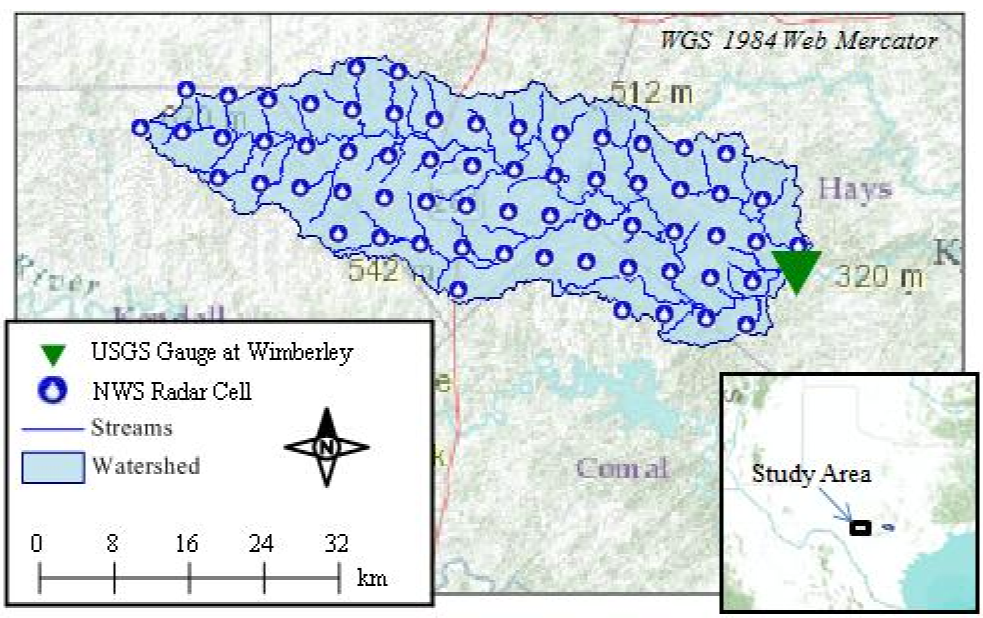

Figure 1. Watershed for USGS Gauge 08171000 with Precipitation and Streamflow Time Series Locations.

For this case study, three HydroDesktop collaborators share analyses and data involving the Blanco River. Collaborator 1 uses the HydroModeler extension to simulate the streamflow response on the Blanco River at Wimberley. This calculation is delivered to Collaborator 2 who uses the HydroR extension to estimate the accuracy of simulation output. All results are then sent to Collaborator 3 who examines the data and is able to see not only analytical results, but also details about how the analyses were performed and the input data that drove the analyses.

PREDICTING STREAMFLOW USING HYDROMODELER

HydroModeler is an OpenMI-based modeling environment that has been developed as an extension for HydroDesktop. It enables users to create unique model compositions by linking together OpenMI-compliant models (Moore and Tindall, 2005). Integration with HydroDesktop provides the capability to read CUAHSI-HIS observation data via the HydroDesktop database repository and save results to the same repository. This study demonstrates how a HydroModeler configuration can utilize such data to predict streamflow at the Blanco River at Wimberley. The model simulation can then be "zipped " and "shipped" to another colleague for further analysis.

Collaborator 1 constructs a watershed model from OpenMI-compliant components that are driven by readily available observation data. The method used for predicting watershed runoff, outlined by Beven, 1997, is known as the TOPography based hydrologic MODEL (TOPMODEL). The TOPMODEL was originally developed for simulating small humid watersheds; however, it has also been extensively applied to drier catchments (Beven, 2004). The full details of TOPMODEL are beyond the scope of this paper. The key points related to CUAHSI-HIS and HydroDesktop are that TOPMODEL requires (among other things) measured precipitation and evapotraspiration as input, and it computes streamflow as its output. Measured precipitation can be searched for and downloaded using HydroDesktop. However, the latter input, daily potential evapotransipration (PET), must be calculated by another component.

PET is estimated by a separate model component that uses the approach outlined by Hargreaves and Samani, 1982. This is an approximation of the typically complicated relationships that govern evapotranspiration (i.e. Monteith, 1965; Priestley and Taylor, 1972) and requires only temperature observations as input (Equation 1)

$$PET = 0.0075 * Ra * Kt * \Delta T^{0.5} * Tave \qquad 1$$

where ΔT is the difference between maximum and minimum temperatures, T_{ave} is the daily average temperature, and K_t is an empirically derived temperature coefficient (Hargreaves and Samani, 1982). Furthermore, total extraterrestrial solar radiation, R_a, can be estimated as a function of the time of year and geographic location (Duffie and Beckman, 1980).

Using HydroDesktop, observed temperature measurements are found at the SW Medina meteorological station near D'Hanis TX. Similarly, multi-sensor precipitation estimates at centroids of NEXRAD radar cells from the National Weather Service are found covering the entire watershed (shown in Figure 1). HydroDesktop is used to download these data into the local database repository where they can then be used as inputs to the HydroModeler composition.

Observation input data are extracted at runtime from the HydroDesktop database repository and supplied to HydroModeler components. This task is accomplished using a specialized model component, called the DbReader, which has been designed to read observation data series from the underlying HydroDesktop database and construct OpenMI objects during component initialization. This is possible through a series of queries made to the HydroDesktop database that occur when components are linked to the DbReader. Once extracted, the data are stored in memory and retrieved upon request from other model components.

Simulation results are written from the HydroModeler composition back into the HydroDesktop database with a similar model component called the DbWriter. Using information discovered across model links, the DbWriter is able to construct a HydroDesktop data model object. During simulation, it retrieves computations from the desired model components and saves them in the data model object. Once simulation is complete, the data model objects are written to the HydroDesktop database.

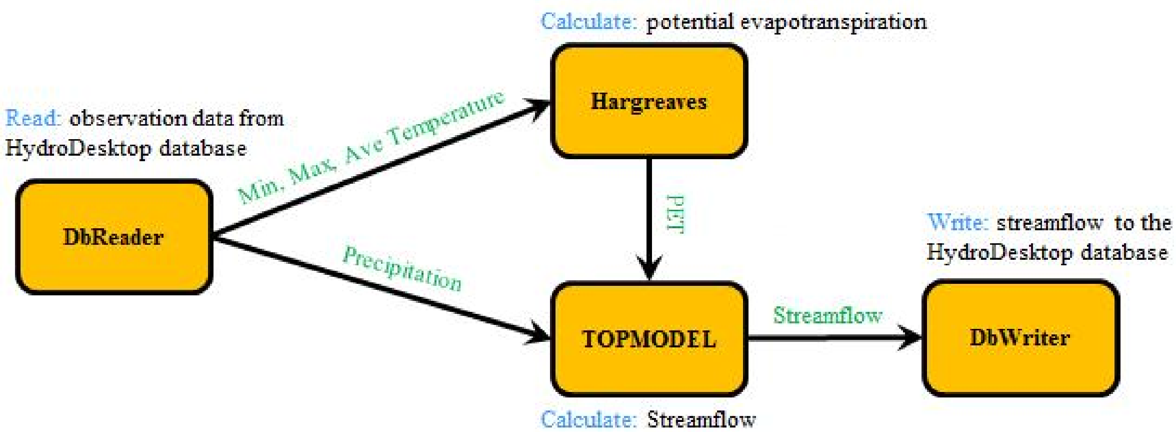

Figure 2. The flow of data during a single time step of a HydroModeler simulation.

The communication pattern amongst components within a HydroModeler configuration follows the same "Request" and "Reply" paradigm outline by Gregersen et al., 2007. Furthermore, model components walk through simulation on a time step basis, i.e. the simulation does not advance in time until all model components have done so. Prior to simulation, the DbReader queries the HydroDesktop database for the data required by its "downstream" components and stores these values in memory. Similarly, the DbWriter prepares data objects to store the calculations that will be retrieved during simulation. Both the DbReader and DbWriter components rely on information defined on the links to perform these tasks. Figure 2 illustrates the flow of data on each time step of a HydroModeler simulation. On the first time step, the DbReader retrieves precipitation and temperature data at the current simulation time and supplies them to the TOPMODEL and Hargreaves components, respectively. Next, the Hargreaves component computes potential evapotranspiration and supplies this as input to TOPMODEL. TOPMODEL then uses both of these inputs to calculate streamflow at the desired location. Finally, the DbWriter requests the calculated streamflow from TOPMODEL and saves it locally. This interaction is repeated every time step of simulation. Once the simulation is complete, the DbWriter saves the calculated streamflow to the HydroDesktop database. The predicted streamflow calculations (Figure 3) are now readily available for viewing and/or statistical analysis using other extensions within HydroDesktop.

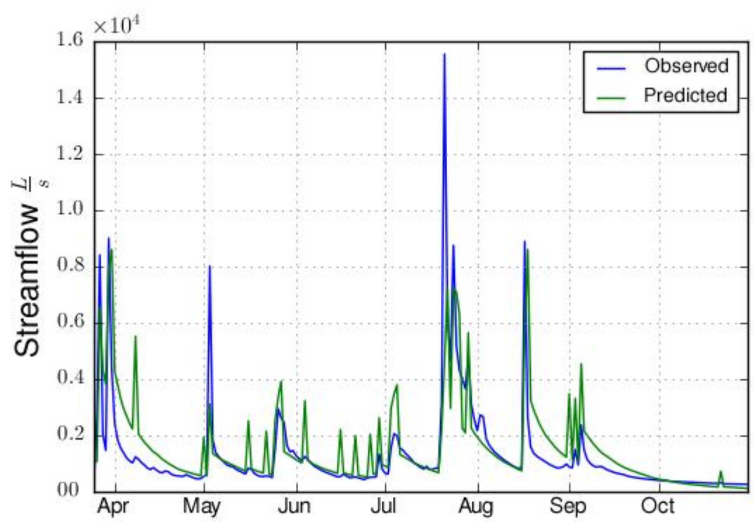

Figure 3. Predicted and observed (USGS gauge 08171000) streamflow at the Blanco River near Wimberley from April-November, 2007.

HydroModeler simulations can be "zipped" and "shipped" to other hydrologists for further analysis. This is made possible by the independent nature of the HydroModeler components. Since model components are developed completely independent of one another, individual models can be shared by simply archiving the model DLL (Dynamic-Link Library) and any model specific input files, along with the HydroDesktop database repository housing the input observation data. By giving another researcher this archive, they are able to instantly evaluate the model calculations as well as re-run or modify the simulation. This feature encourages collaboration between multiple institutions, and also provides a mechanism for documenting model runs.

EVALUATING STREAMFLOW SIMULATION OUTPUT WITH HYDROR

The HydroR extension integrates R software in HydroDesktop, where R is an open source statistical computing language for data analysis and graphing (Ihaka and Gentleman, 1996). The HydroR extension includes an R package (also named HydroR) with functions such as getDataSeries and saveDataSeries for reading from and writing to the HydroDesktop database. Users can interactively select time series and automatically generate code to send the data to R for analysis.

Following the use case, Collaborator 1 sends the HydroModeler components, HydroDesktop database, and HydroDesktop project file to Collaborator 2. Collaborator 2 decides to evaluate the accuracy of the simulation results using HydroR. The volume error (Equation 2) is used to estimate the streamflow simulation accuracy (Singh et al., 2012). The workflow for calculating the volume error by using HydroR is shown in Figure 4.

$$Volume\ error = \frac{\sum Modeled\ streamflow - \sum Observation\ streamflow}{\sum Observation\ streamflow} \quad (2)$$

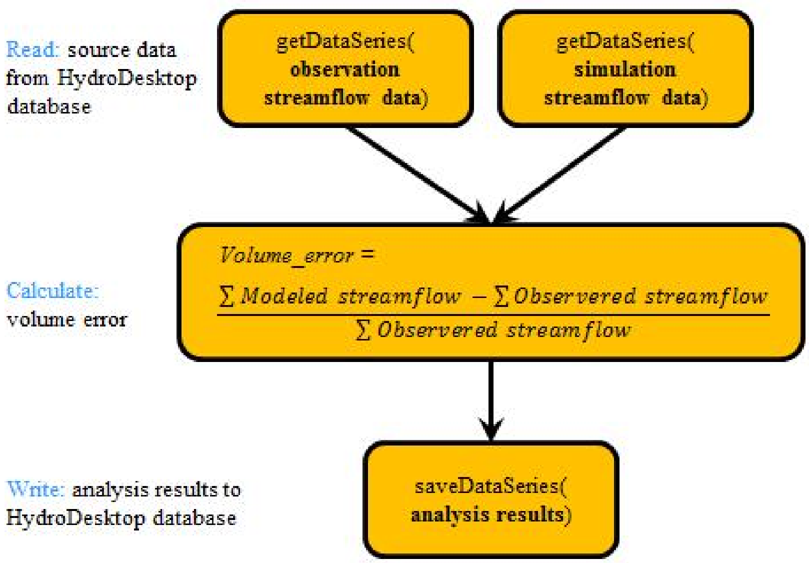

Figure 4. The workflow of evaluating simulation accuracy.

The simulation streamflow results and observation data are read from the Hydrodesktop database into the HydroR computing environment by using the getDataSeries() function. The monthly volume errors from April, 2007, to November, 2007, are calculated according to Equation 2. Additional intrinsic functions, such as sum, subtract, divide, and abs, are also used in this analysis. The volume errors are written back to HydroDesktop database by using the saveDataSeries() function.

All of the R code used in this workflow is saved as an R script file via the HydroR controls. These scripts can be opened by other users who have HydroR installed, or from the standalone R user interface independently of HydroDesktop.

SHARING RESULTS WITH A THIRD COLLABORATOR

Collaborator 2 sends a zip file with the following items to Collaborator 3: the HydroDesktop database, the HydroDesktop project file, the exported R script file for computing volume error, and HydroModeler components. By this point, the HydroDesktop database not only includes observations data for temperature, precipitation, and streamflow, but also the simulated streamflow and calculated volume errors. Collaborator 3 unzips and opens the project file to view the study area, input data, and analysis results in maps, tables, and graphs. Perhaps Collaborator 3 evaluates the acceptability of the volume errors (the values of which are beyond the focus of this paper). With R and OpenMI installed, Collaborator 3 can also load the scripts used in HydroR and the model used by HydroModeler to see the exact steps performed by the previous collaborators in their analyses. Collaborator 3 could even run these scripts and models to reproduce results created by the other collaborators if desired.

CONCLUSIONS

Integration of analysis environments such as R and OpenMI into HydroDesktop enable the management of hydrologic data and analyses from within a single user interface. Collaborators can share data and reproduce analyses by transferring HydroDesktop databases, project files, and related scripts and models. Users can easily open the files in HydroDesktop to view what collaborators have done. However, the transfer of these files currently relies on arcane methods such as "zip and ship." The authors recommend that future work should include more elegant means of indexing and sharing HydroDesktop projects.

REFERENCES

Ames, D., 2012. What is HydroDesktop? Why Should You Care? and When Will it Work on a Mac? Proceedings for AWRA Spring Specialty Conference in New Orleans, Louisiana.

Beven, K. J., 1997. TOPMODEL: a critique. *Hydrological Processes* 11(9): 1069-1086.

Beven, K. J., 2004. Rainfall-runoff Modelling: The Primer. John Wiley & Sons Ltd, Chichester, UK.

Duffie, J. A., and W. A. Beckman, 1980. Solar Engineering of Thermal Processes. Wiley, N.Y., pp. 1-109

Gregersen, J. B., P. J. A. Gijsbers, and S. J. P. Westen, 2007. OpenMI: Open modelling interface. *Journal of Hydroinformatics* 9(3): 175-191.

Hargreaves, G. H., and Z. A. Samani, 1982. Estimating potential evapotranspiration. *Journal of Irrigation and Drainage Engineering* 108(3): 225-230.

Ihaka, R. and R. Gentleman, 1996. R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics 5(3), 299–314.

Monteith, J. L., 1965. Evaporation and environment. In: *The State and Movement of Water in Living Organsims*. Fogg G.R. (Editor), Symposium of the Society for Experimental Biology, Academic Press Inc., NY, pp 205-234.

Moore, R. V., and C. Tindall, 2005. An overview of the open modelling interface and environment (the OpenMI). *Environmental Science & Policy* 8 (3): 279-286.

Priestley, C. H. B., and Taylor R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Rev. 100: 81-92.*

Singh, J., H. V. Knapp, and M. Demissie. 2004. Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT. ISWS CR 2004-08. Champaign, Ill.: Illinois State Water Survey. Available at: www.sws.uiuc.edu/pubdoc/CR/ISWSCR2004-08.pdf. *Accessed on January 1, 2012.*

Tarboton, D., D. R. Maidment, I. Zaslavsky, D. P. Ames, J. L. Goodall, R. P. Hooper, and J. S. Horsburgh, 2012. The CUAHSI Community Hydrologic Information System. Proceedings for AWRA Spring Specialty Conference in New Orleans, Louisiana.