International Congress on Environmental Modelling and Software

9th International Congress on Environmental Modelling and Software - Ft. Collins, Colorado, USA - June 2018

Jun 26th, 3:40 PM - 5:00 PM

Hydrologic Micro Services

Rajbir Parmar
U.S. Environmental Protection, parmar.rajbir@epa.gov

Chris Knightes
Environmental Protection Agency

Deron Smith
Independent Contractor, smith.deron@epa.gov

Kurt Wolfe
U.S. Environmental Protection Agency Office of Research and Development

Mike Galvin
U.S.Environmental Protection Agency

See next page for additional authors

Presenter/Author Information

Rajbir Parmar, Chris Knightes, Deron Smith, Kurt Wolfe, Mike Galvin, Josh Koblich, Jan Sitterson, John M. Johnston, and Tom Purucker

iEMSs International Congress on Environmental Modelling and Software

9th International Congress on Environmental Modelling and Software
Fort Collins, Colorado, USA, Mazdak Arabi, Olaf David, Jack Carlson, Daniel P. Ames (Eds.)
https://scholarsarchive.byu.edu/iemssconference/2018/

Hydrologic Micro Services

Rajbir Parmar[a], Christopher D. Knightes[a], Deron Smith[b], Kurt Wolfe[a], Mike Galvin[a], Joshua Koblich[c], Jan Sitterson[c], John M Johnston[a], Tom Purucker[a]

[a]United States Environmental Protection Agency
(parmar.rajbir@epa.gov, knightes.chris@epa.gov, wolfe.kurt@epa.gov, Galvin.mike@epa.gov, Johnston.johnm@epa.gov, Purucker.tom@epa.gov),
[b]Independent Contractor (smith.deron@epa.gov)
[c]Oak Ridge Associated Universities (Koblich.joshua@epa.gov, sitterson.jan@epa.gov),

Abstract: United States Environmental Protection Agency (EPA) has developed a collection of microservices called Hydrologic Micro Services (HMS) for building hydrologic and water quality modeling workflows. HMS components are available as RESTful web services as well as desktop libraries. An HMS component may have multiple implementations addressing varying levels of underlying physical process details and assumptions. HMS components can be used in desktop and web-based workflows. A workflow can call into a specific implementation of an HMS component depending upon the details suitable for the problem statement being addressed by the workflow. Building a workflow from HMS components enables modelers to address hydrologic and water quality problem statements more precisely, in contrast to the current state of modeling where using existing models forces modelers into a potentially sub-optimal workflow. Model selection to address a problem statement has several drawbacks: the selected model may not have the appropriate level of complexity, the model may not address all parts of the problem statement without making less desirable assumptions, or the model may have more features and requirements than necessary. HMS components include data provisioning and simulation algorithms for water quantity and quality modeling. Workflows built using HMS components can in turn be used as components in larger workflows. For example, precipitation data provisioning components can download data from various data sources such as NLDAS, GLDAS, DAYMET, NCDC, PRISM, and WGEN. A simple workflow was developed as an HMS component to compare precipitation data from different sources. Comparison is performed using multiple rainfall statistics.

Keywords: Environmental Modeling; Hydrological Microservices; Hydrologic Modeling; Web Service, Water Quality Modeling

1. INTRODUCTION

The current state of water quantity and water quality modeling has many drawbacks. Most environmental models are not inherently interoperable. The monolithic structure of environmental models is an obstacle in dealing with complex environmental problems and interoperability is inevitable for future generations of models (Hu & Bian, 2009). Another drawback is that most of the hydrologic and water quality models are long-term prediction models. Examples of long-term prediction models include SWAT (Arnold & Fohrer, 2005), HSPF (US EPA, 2014), AQUATOX (Park et. Al., 2008), and WASP (DiToro et al., 1983). Only a few generalized now-casting or forecasting hydrologic and water quality models exist; the most ambitious being the National Water Model (http://water.noaa.gov/about/nwm) which simulates streamflow for every NHDPlus stream segment on an hourly basis. In addition, many of the hydrologic and water quality models were developed before internet access to national data sources and sensor data were available, leaving the tasks of gathering and preprocessing model input data entirely up to the modelers.

To alleviate these shortcomings in the current state of hydrologic modeling EPA is in the process of developing a collection of hydrologic modeling microservices. In the context of modeling, microservices can be conceptualized as a new way to create modeling applications, where applications are broken down into smaller, interoperable, workflow agnostic, independent services that are not dependent on a specific computing platform or coding language. In other words, scientists, modelers, and software developers can contribute within their knowledge domain using the language tools they are most comfortable with and still achieve synergy in the application development process. Using the ideology of microservices, large complex modeling workflows can be broken up into smaller building blocks of executables, that when recomposed offer all the functionality of a large-scale, highly complex application. Foretta et. al. (2014) built a workflow using Object Modeling Software and found that encapsulation of code into components facilitates cooperation among researchers, the analysis of hydrological processes, the comparison among different modelling solutions, and the adoption of reproducible research strategies.

2. CURRENT STATE OF HYRDOLOGICAL MODELING

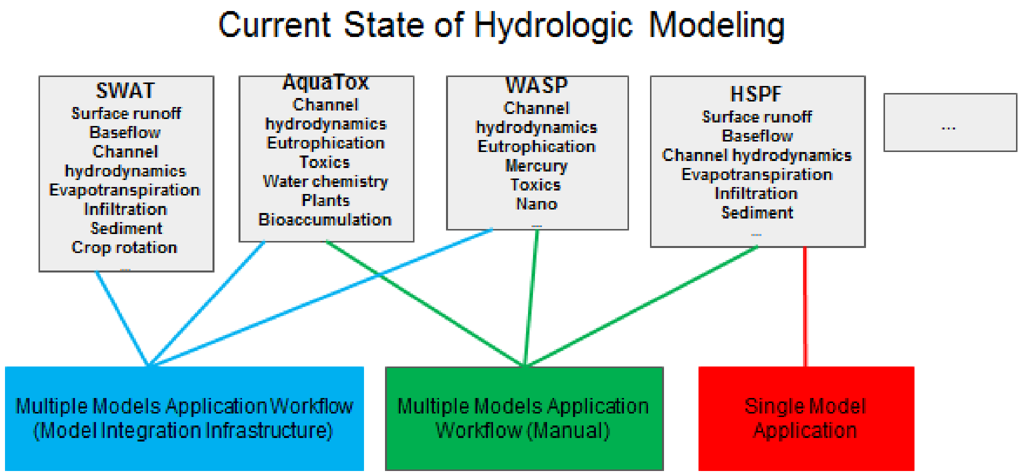
The current state of modeling often has modelers try fitting the problem statement to an existing model. In some cases, modelers integrate multiple models to address the problem statement. As shown in Figure 1, model integration can be performed either manually or using integration software such as BASINS (US EPA, 2015) and FRAMES (Whelan et. Al., 2014). In many cases the selected model or set of models either under- or over- fits the problem statement being addressed. An under-fit model lacks algorithms to address the processes of concern at the level of detail required by the problem statement. Only a small part of an over-fit model addresses the problem statement entirely. Modelers using an under-fit model must make assumptions simplifying the problem statement. On the other hand, the modeler using an over-fit model needs to provide extraneous inputs and handle extraneous outputs.

Currently, modelers must often gather and pre-process model input data even when programmatic access to data through web services is available. Many of the commonly used data in hydrologic modeling are available as national data sources through web services.

Another artifact of the current state of hydrologic modeling is that models are not inherently inter-operable. In other words, a model encapsulates and does not expose its implementation of a physical process for use by other models and applications. The implementation gets repeated in other models where the model developers choose to include the physical process. Multiple implementations of the same exact physical process can lead to confusion and discrepancy amongst the models.

Currently, modelers spend significant time gathering and pre-processing input data for modeling applications because most of the existing models lack built-in data provisioning services. In addition, since most existing models do not implement standard interfaces with standard data formats, each model requires its own nuances to be considered when preparing input data.

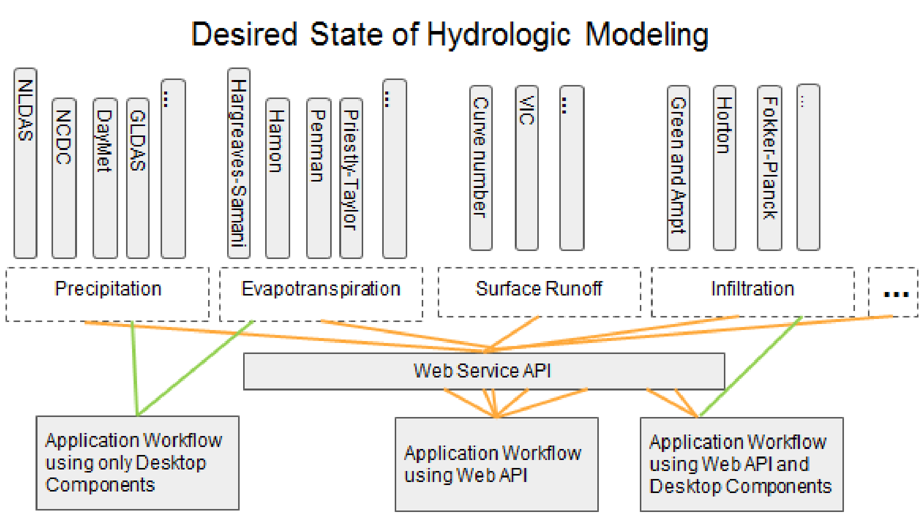
Figure 1. Current State of Hydrologic Modeling



3. DESIRED STATE OF HYDROLOGIC MODELING

A desired state of hydrologic modeling is where an appropriate workflow could be rapidly composed to fit the problem statement. Such a state requires developing smaller, independent services or building blocks that are inherently inter-operable and are agnostic of the workflows where they would be used. The services must include data provisioning, modeling physical processes, and other services in support of composing workflows. It would be desirable to have multiple implementations of each service to address varying levels of details of the modeled processes. It is important for the services to implement standard interfaces and data exchange formats to be inter-operable across a wide range of applications. Transparency represented as detailed metadata and thorough documentation are other desired attributes of the services. Although the components wrapped in microservices can be made available as desktop components, their inter-operability may be limited in terms of computer language and platform. Implementing microservices using industry standard communication protocols and data exchange formats, the modelers and developers would be able to build workflows by selecting services from other sources such as the Object Modeling Software (David et al., 2013) in addition to the services provided by HMS.

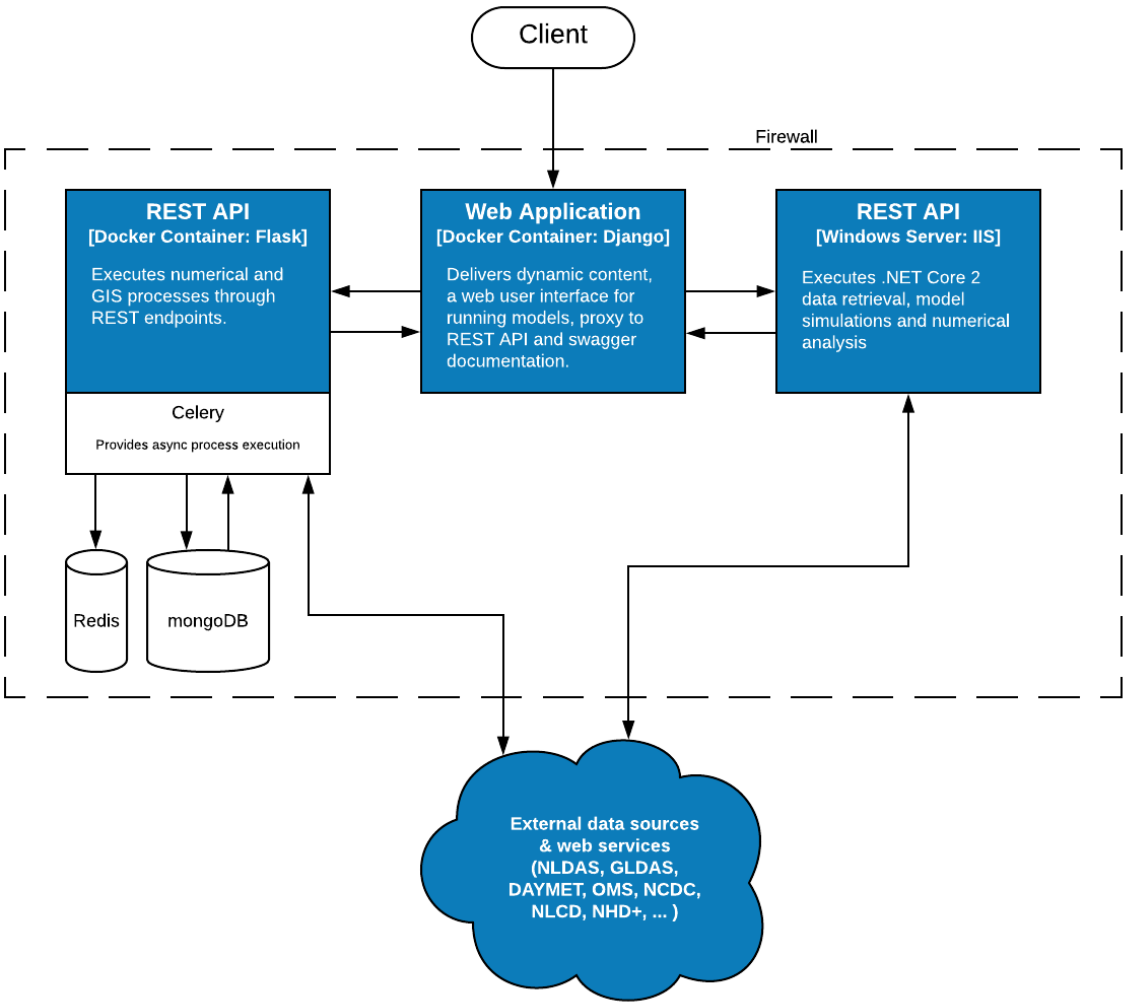
Figure 2. Desired State of Hydrologic Modeling



4. HMS ARCHITECTURE

Figure 3 shows EPA's HMS architecture. All the science logic of HMS components is written in open source .NET Core C# using a Model View Controller architectural pattern. Currently, HMS science logic web services are hosted on a Microsoft Windows server running IIS web server. Encapsulating HMS science logic web services in a dockerized container is planned. Containerized HMS web services would provide the flexibility of being hosted on a Windows or Linux server as well as facilitate future scalability. Currently the Windows server hosting HMS services is not exposed to the public, but the services are exposed through a public-facing Linux server, as explained later. HMS web services depend on a collection of geoprocessing services. The geospatial processing functionality is written in Python using an open source Geospatial Data Abstraction Library (GDAL, http://www.gdal.org) and runs in its own dockerized container as a Flask application (http://flask.pocoo.org) on a Linux server. Although HMS is a library of hydrologic data provisioning and modeling components, EPA has developed a HMS front-end to showcase HMS components and derived workflows. HMS front-end consists of web pages, API endpoints, and API endpoint documentation. The front-end is written as a Django (https://www.djangoproject.com) application running in its own dockerized container on a Linux server. API end-point documentation is implemented using the Django REST Swagger (https://django-rest-swagger.readthedocs.io/en/latest) framework. HMS web pages have been developed as a combination of HTML, CSS, JavaScript, and D3.js library (https://d3js.org). A Celery (http://www.celeryproject.org) task queue using redis (https://redis.io/) and mongoDB (https://github.com/mongodb) databases for managing longer-running distributed tasks has been implemented. All APIs have been implemented as RESTful web services and the data exchange is implemented as JSON.

Figure 3. Hydrologic Micro Services Architecture.



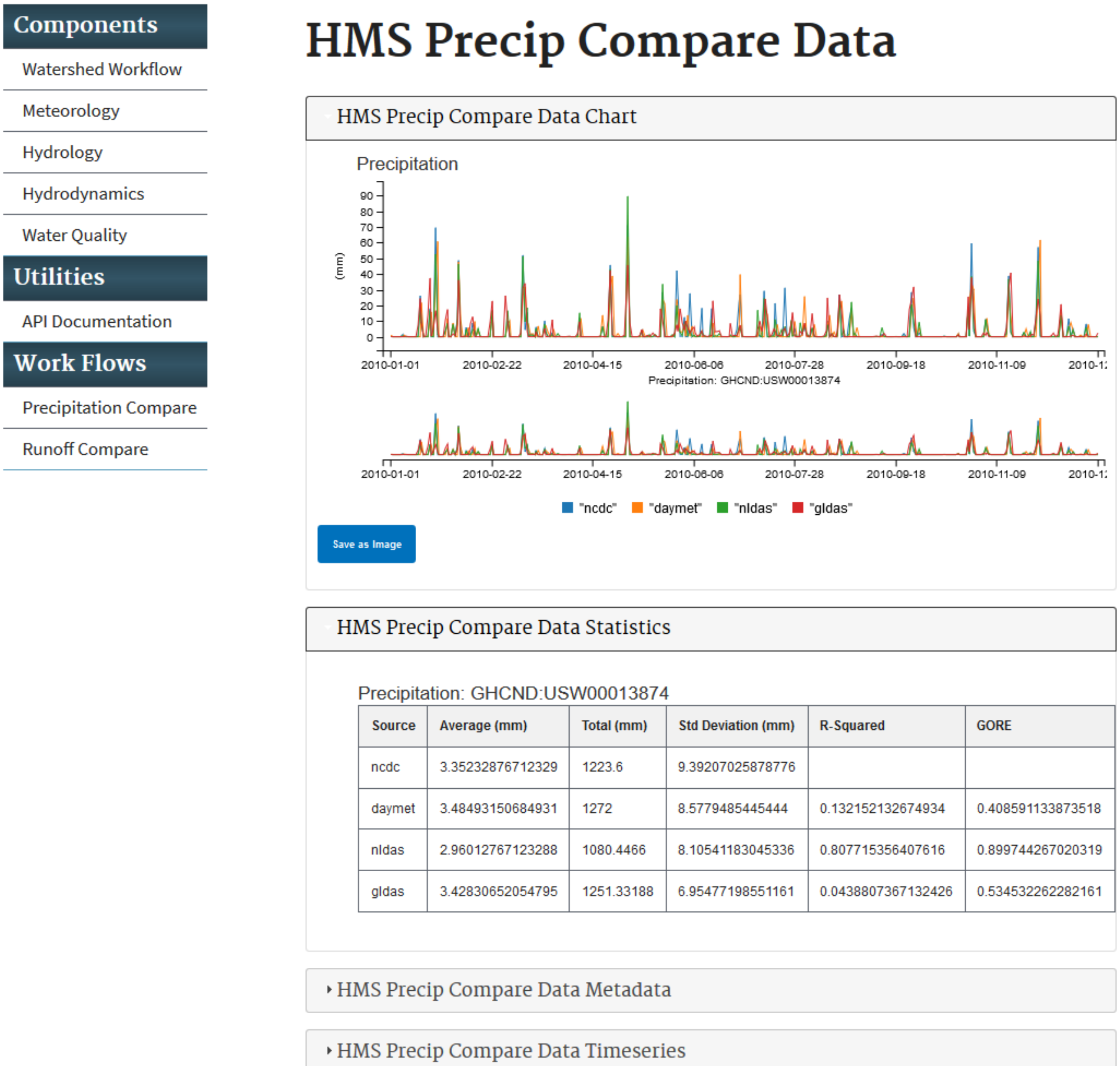
5. CURRENT STATE OF HMS

Table 1 shows a list of completed and in-progress RESTful HMS web services. The table does not include future components and services. Components behind microservices are also available as a desktop library. The components and services are currently available within the EPA firewall and EPA plans to release them publicly. In addition to the web services and components, EPA is in the process of developing workflows to demonstrate the utility of the HMS services and components. For example, Precipitation Compare is a workflow that demonstrates the use of the precipitation-related components. The workflow returns time series from all precipitation data provisioning components (NLDAS, GLDAS, DAYMET, NCDC, and PRISM). In addition, the workflow provides statistics for each data source as compared to data from NCDC. A collection of web pages has also been developed to demonstrate the utility of HMS services. Each HMS component has its own web page where the user can select an implementation of the component and provide inputs pertinent to the selected implementation. A few web pages have also been developed to demonstrate work flows composed using HMS services. Figure 4 shows a workflow web page for comparing precipitation time series data from multiple sources.

Table 1. HMS Components/Web Services completed or in progress

Component/Web Service	Functionality	Component/Web Service	Functionality
NLDAS Precipitation Data	Downloads & pre-processes NLDAS Precipitation Data	NLDAS Surface Runoff Data	Downloads & pre-processes NLDAS Surface Runoff Data for an NHDPlus HUC8 or HUC12
GLDAS Precipitation Data	Downloads & pre-processes GLDAS Precipitation Data	NLDAS Subsurface Runoff Data	Downloads & pre-processes NLDAS Subsurface Runoff Data for an NLDAS cell
DAYMET Precipitation Data	Downloads & pre-processes DAYMET Precipitation Data	GLDAS Precipitation Data	Downloads & pre-processes GLDAS Subsurface Runoff Data for an GLDAS cell
NCDC Precipitation Data	Downloads & pre-processes NCDC Precipitation Data	Curve Number Based Surface Runoff	Calculates Surface Runoff using Curve Number method
PRISM Precipitation Data	Downloads & pre-processes PRISM Precipitation Data	NLDAS Temperature Data	Downloads & pre-processes NLDAS Air Temperature Data at different heights above ground for a given location (latitude and longitude)
WGEN Precipitation Data	Generates synthetic precipitation Data using Weather Generator (WGEN)	GLDAS Temperature Data	Downloads & pre-processes GLDAS Air Temperature Data at different heights above ground for a given location (latitude and longitude).
NLDAS Evapotranspiration Data	Downloads & pre-processes NLDAS Evapotranspiration Data	Stream Hydrology/Hydrodynamics – Constant Volume	Simulates stream hydrology and hydrodynamics using constant volume algorithm (volume, velocity, and depth remain constant)
GLDAS Evapotranspiration Data	Downloads & pre-processes GLDAS Evapotranspiration Data	Stream Hydrology/Hydrodynamics – Varying Volume	Simulates stream hydrology and hydrodynamics using varying volume algorithm (volume, velocity, depth, and surface area vary)
NLDAS Soil Moisture Data	Downloads & pre-processes NLDAS Soil Moisture Data at different soil depths	Stream Hydrology/Hydrodynamics – Kinematic Wave	Simulates stream hydrology and hydrodynamics using kinematic wave algorithm
GLDAS Soil Moisture Data	Downloads & pre-processes GLDAS Soil Moisture Data at different soil depths	Stream Hydrology/Hydrodynamics – Constant Volume	Simulates stream hydrology and hydrodynamics using constant volume algorithm
NLDAS Surface Runoff Data	Downloads & pre-processes NLDAS Surface Runoff Data for an NHDPlus catchment	Solar radiation absorption	Calculates direct photolysis rates and half-lives of pollutants in aquatic environments

Figure 4. Precipitation comparison workflow web page.



6. CONCLUSION

HMS is a collection of data provisioning and hydrologic modeling components along with RESTful web services that can be used in rapid development of workflows to more precisely address hydrologic and water quality modeling problems. HMS is being developed to address drawbacks in the current state of water quantity/quality modeling. Existing monolithic models tend be under- or over-fit for many environmental modeling problem statements. In many instances monolithic models may not address the process of interest to the level needed by the problem statement, resulting in the modeler making simplifying assumptions. In other instances, modelers need to handle extraneous input and output data when the model is over-fit. Integrating models to solve complex problems is difficulty because existing models lack inherent interoperability. HMS divides the environmental modeling universe into micro-level interoperable building blocks which can be used to compose complex modeling workflows. Since HMS implements industry standard communication and data exchange standards, workflow developers are able to combine web services from other sources such as the Object Modeling System (http://oms.colostate.edu/). In addition to physical process modeling services, HMS also includes data provisioning services, making it easier for modelers to rapidly characterize the geospatial feature of interest and parameterize workflows. Since transparency is paramount, especially in a regulatory environment, detailed metadata and documentation of each service is an integral part of HMS. HMS is an on-going effort, adding more services as scientists, modelers, and software engineers pursue contributions. Many HMS microservices are currently available, and EPA plans to develop more services in the future. EPA has constructed a few workflows to demonstrate the utility of HMS services, with plans to build additional flows. Currently, HMS services and components are only available within the EPA firewall, but will be publicly available in the future.

REFERENCES

Arnold, J. G., & N. Fohrer, 2005. SWAT2000: Current capabilities and research opportunities in applied watershed modeling. Hydrol. Process, 19(3): 563-572.

David, O., Ascough II, J. C., Lloyd, W., Green, T. R., Rojas, K. W., Leavesley, G. H., & Ahuja, L. R., 2013. A software engineering perspective on modeling framework design: The Object Modeling System. Environmental Modeling and Software, 39, 201-213.

DiToro, D., Fitzpatrick, J., & Thomann, R., 1983. Documentation for Water Quality Analysis Simulation Program (WASP) and Model Verification Program (MVP). U.S. Environmental Protection Agency, Washington, D.C., EPA/600/3-81/044.

Formetta, G., Antonello, A., Franceschi, S., David, & O., Rigon, R., 2014. Hydrological modelling with components: a GIS-based open source framework. Environ. Modell. Softw. 55,190-200.

Hu, Shixiong & Bian Ling (2009). Interoperability of functions in environmental models – a case study in hydrological modeling, International Journal of Geographical Information Science, 23:5, 657-681, DOI: 10.1080/13658810902733674.

Park, Richard A., Clough, Jonathan S., & Wellman, Marjorie Coombs, 2008. AQUATOX: Modeling Environmental Fate and Ecological effects in Aquatic Ecosystems. Ecological Modeling, 213(1), 1-15.

US EPA, 2015. BASINS 4.1 (Better Assessment Science Integrating point & Non-point Sources) Modeling Framework. National Exposure Research Laboratory, RTP, North Carolina. https://www.epa.gov/exposure-assessment-models/basins. Accessed April 3, 2018.

US EPA, 2014. Hydrological Simulation Program – FORTRAN (HSPF). National Exposure Research Laboratory, RTP, North Carolina. https://www.epa.gov/ceam/hydrological-simulation-program-fortran-hspf. Accessed April 3, 2018.

Whelan, G., Kim, K., Pelton, M. A., Castleton, K. J., Laniak, G. F., Wolfe, K., Parmar, R., Babendreier, J., & Galvin, M., 2014. Design of a component-based integrated environmental modeling framework. Environmental Modeling and Software, 55, 1-24.