2018-12-01

A Flexible Service-Oriented Approach to Address Hydroinformatic Challenges in Large-Scale Hydrologic Predictions

Michael Antonio Souffront Alcantara
Brigham Young University

A Flexible Service-Oriented Approach to

Address Hydroinformatic Challenges in

Large-Scale Hydrologic Predictions

Michael Antonio Souffront Alcantara

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

E. James Nelson, Chair
Rollin H. Hotchkiss
Daniel P. Ames
Norman L. Jones
Gus P. Williams

Department of Civil and Environmental Engineering

Brigham Young University

ABSTRACT

A Flexible Service-Oriented Approach to
Address Hydroinformatic Challenges in
Large-Scale Hydrologic Predictions

Michael Antonio Souffront Alcantara
Department of Civil and Environmental Engineering, BYU
Doctor of Philosophy

Water security is defined as a combination of water for achieving our goals as a society, and an acceptable level of water-related risks. Hydrologic modeling can be used to predict streamflow and aid in the decision-making process with the goal of attaining water security.

Developed countries usually have their own hydrologic models; however, developing countries often lack hydrologic models due to factors such as the maintenance, computational costs, and technical capacity needed to run models. A global streamflow prediction system (GSPS) would help decrease vulnerabilities in developing countries and fill gaps in areas where no local models exist by providing extensive results that can be filtered for specific locations.

The development of a GSPS has been deemed a grand challenge of the hydrologic community. To this end, many scientists and engineers have started to develop large-scale systems to an acceptable degree of success. Renowned models like the Global Flood Awareness System (GloFAS), the US National Water Model (NWM), and NASA's Land Assimilation System (LDAS) are proof that our ability to model large areas has improved remarkably. Even so, during this evolution the hydrologic community has started to realize that having a large-scale forecasting system does not make it immediately useful. New hydroinformatic challenges have surfaced that prevent these models from reaching their full potential. I have divided these challenges in four main categories: big data, data communication, adoption, and validation.

I present a description of the background leading to the development of a GSPS including existing models, and the components needed to create an operational system. A case study with the NWM is also presented where I address the big data and data communication challenges by developing cyberinfrastructure and accessibility tools such as web applications and services.

Finally, I used the GloFAS-RAPID model to create a forecasting system covering Africa, North America, South America, and South Asia using a service-oriented approach that includes the development of web applications, and services for providing improved data accessibility, and helping address adoption and validation challenges. I have developed customized services in collaboration with countries that include Argentina, Bangladesh, Colombia, Peru, Nepal, and the Dominican Republic. I also conducted validation tests to ensure that results are acceptable. Overall, a model-agnostic approach to operationalize a GSPS and provide meaningful results at the local level is provided with the potential to allow decision makers to focus on solving some of the most pressing water-related issues we face as a society.

Keywords: hydrologic modeling, cyberinfrastructure, data visualization, hydroinformatics

ACKNOWLEDGEMENTS

I want to first thank God for the opportunity to come to an academic institution like Brigham Young University where I have been able to grow both intellectually and spiritually. Second, I want to thank all the people that collaborated and contributed to my research. I am very grateful to Jim Nelson for giving me the opportunity to be part of his research group. Thanks to present and past members of the Hydroinformatics Lab. Especially to Alan Snow for providing a starting point for this project; and to Corey Krewson, Spencer McDonald, Elise Jackson, Wade Roberts, Chris Edwards, and Jackson Flood for helping develop tests, web apps, and other relevant parts of this research. Also, thanks to Zhiyu Li for continuing the development of the US National Water Model app, and to Fidel Perez for first introducing me to Dr. Nelson.

Thanks to my graduate committee Norm Jones, Dan Ames, Rollin Hotchkiss, and Gus Williams for their teachings, and advice throughout the process of writing this dissertation. I am grateful to Steve Kopp, Angelica Gutierrez, and Christel Prudhomme for their key role in developing this global streamflow prediction system. Thanks to NASA SERVIR for providing the funds and connections necessary to make this work relevant, and to Microsoft's AI for Earth team for providing the cloud services necessary to host such a project. I would also like to thank Kim Glade and Jolene Johnson for helping with all kinds of paperwork these last three years.

Lastly, thanks to my family for their love and support. To Jolyn Souffront for being such a great wife. To my children Alexa Souffront, and Marcus Souffront, and to my family back in the Dominican Republic: Tamara Alcantara, Rafael Souffront, and Yosayris Souffront.

¡Muchas Gracias!

Michael Souffront

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Background

On September 25th 2015, the United Nations (UN) put forth a collection of 17 goals aimed towards ending poverty, protecting our planet, and ensuring prosperity for all (Nam, 2015). This set of goals is known as the Sustainable Development Goals (SDGs), which has been adopted by 193 countries. The goal categories are: no poverty, zero hunger, good health and well-being, quality education, gender equality, clean water and sanitation, affordable and clean energy, decent work and economic growth, industry innovation and infrastructure, reduced inequalities, sustainable cities and communities, responsible consumption and production, climate action, life below water, live on land, peace justice and strong institutions, and partnerships for the goals. The direct effect of water in more than half of these categories can be easily recognized, and it can be argued that water indirectly affects some of the other categories. Complementary to the UN's SDGs, the SENDAI Framework for Disaster Risk Reduction constitutes an agreement endorsed by the UN to reduce risk due to natural disasters, and subsequently the losses of lives, livelihoods, and environmental assets at the individual, community, and country scale. Water management in the form of prediction and preparedness has the potential to significantly improve risk reduction, especially in developing countries lacking the resources to develop their own prediction system.

Grey and Sadoff (2007) defined the relatively new term water security as the availability of an acceptable quantity and quality of water for health, livelihoods, ecosystems and production, coupled with an acceptable level of water-related risks to people, environments and economics. In other words, water security is determined by a combination of the main water-related issues addressed by the SDGs and the SENDAI framework; namely water for achieving our goals as a society, and water risk management. Water security can only be attained by having an understanding of current water resources available and future availability with enough lead-time to prepare and respond to changes that may affect production or risk factors. Moreover, this type of understanding can only be reached by having easily accessible and accurate water data.

Water is the most important natural resource on earth. As such, it needs to be managed properly from all angles. Water data needs to be readily accessible and easily understood by all involved groups. These groups usually include scientists, decision-makers, emergency responders, and the general public. The common use level and key functionality needed by each group are shown in Table 1-1.

Table 1-1. Water data use levels.

Group	Use Level	Key Functionality
Scientists	Research	Comprehensive
Decision Makers	Planning	Relevant
Engineers/Emergency Responders	Action	Accessible
General Public	Awareness	Intuitive

Fluvial flooding is one of the most recurrent and costly natural disasters around the world in both human lives and property damage. In the United States (US), the last 50 years have seen over 100 such floods affecting about 12 million people and costing about $50 billion (Guha-

Sapir, Below, & Hoyois, 2017). In Europe, the European Environmental Agency estimated that floods caused economic losses of over €60 Billion and over 1000 fatalities between 1998 and 2009 (Wehrli, Herkendell, & Jol, 2010). Over the years, engineers and scientists have developed a variety of tools to timely respond to the problems posed by floods. For a long time, human efforts were concentrated in the prevention and mitigation of damaging floods through the construction of physical barriers to protect specific areas. This approach has been successful to an extent; however, there is a limit to the protection that these structures can offer. In the last few decades, flood preparedness has been incorporated into flood risk management as a way to minimize the impact of inevitable floods. A comprehensive flood risk management strategy should be a combination of prevention, preparation, response, and recovery (van Alphen, Martini, Loat, Slomp, & Passchier, 2009). This is evident in the main goals of organizations like the Global Flood Partnership (GFP), whose goal is to establish a partnership for global flood forecasting, monitoring and impact assessment to strengthen preparedness and response and to reduce global disaster losses; and NOAA's Office of Water Prediction (OWP), whose goal is to develop and deliver state-of-the-science national hydrologic analyses, forecast information, data, decision-support services and guidance to support and inform essential emergency services and water management decisions.

Hydro-meteorological models play a critical role in flood management systems. Hydrologic modeling is an important tool that helps us understand how to better respond to extreme events, and plan according to predicted expectations. Forecast predictions are used in the development of preparedness and mitigation strategies. More specifically, they are used as one of the main triggers in warning systems along with earth observations. Furthermore, the scarcity, and sometimes unavailability of the latter together with insufficient lead times to respond usually make hydrologic models the main input to such warning systems with streamflow observations being assimilated when available.

During the past few decades, engineers and scientists have increased our ability to predict floods. Hydrology and hydrologic modeling are continually evolving. There has been an internal expansion thanks to technological advances. We now have better physically-based distributed models with better grid resolution, larger coverage, and faster computation times. Second, there has been a vertical expansion, where Meteorology is now incorporating Global Circulation Models (GCMs), Land Surface Models (LSMs), and routing models so they can hydrologically route their runoff estimates. At the same time, Hydrology is looking at weather parameters not simply as inputs but as part of the overall cycle. As a result, new hydro-meteorological models that take advantage of the respective strengths of Meteorology and Hydrology are being developed. Finally, hydrology is also experiencing a multidisciplinary expansion, where other earth sciences are not only consuming model data but are also actively seeking to better understand hydrologic principles. Furthermore, not only are the earth sciences taking advantage of these models, but because this new integrated approach lowers the barrier to produce more accessible water data, other disciplines are also taking advantage of these hydro-meteorological models (Figure 1-1).

The expansion of hydrology has made the possibility of larger-scale, more-accurate models that are useful at local scales a reality. Likewise, probabilistic forecasts now offer an alternative to incorporate the uncertainty introduced by the inputs used to run a model and deliver to decision-makers the reality of hydrologic forecasts. Nevertheless, major challenges remain. For example, the inherent uncertainty introduced by the model itself is usually neglected, but can be significant (Butts, Payne, Kristensen, & Madsen, 2004). In addition, integrating and communicating model results has historically been a major challenge due to the evolving nature of hydrologic models (Beran & Piasecki, 2009). This challenge has begun to be answered with the adoption of standards like the Open Geospatial Consortium (OGC), a push to create Earth Observation Systems (EOS) and cloud computing that allows model results to be accessed as a service via the Internet, and the creation of derived tools that facilitate the visualization and interpretation of "big" water data through light-weight web applications.

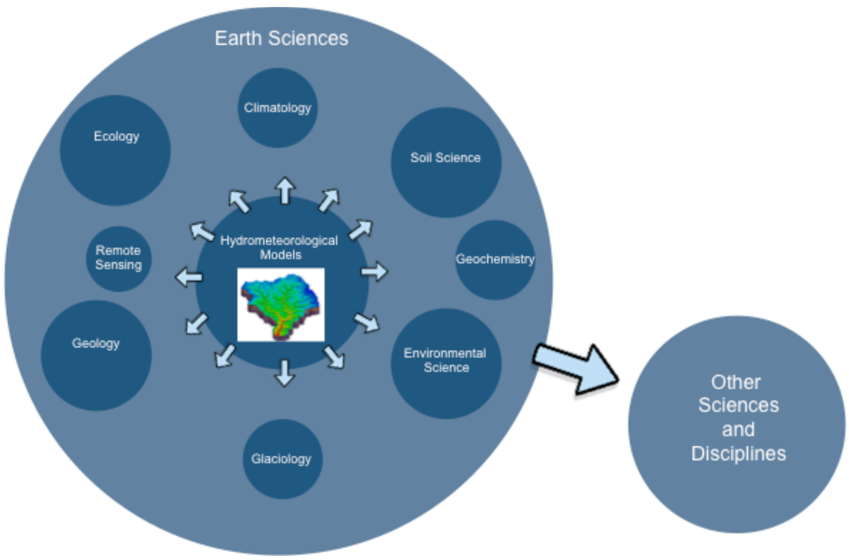


Figure 1-1. The expansion of Hydrology.

Flood maps are an example of a tool that can help us communicate hydrologic model results by identifying areas that are susceptible to flooding. Such maps allow decision makers to properly respond to flood events by providing intelligence regarding flood extent, depth, velocity, affected communities, and access routes for specific areas. In the US, the Federal Emergency Management Agency (FEMA) is responsible for conducting flood studies and producing flood maps. However, these 'static' flood maps have a relatively short lifespan due to the dynamic character of the inputs involved in their production. As a result, these maps are often used to identify flood hazard zones instead of incorporating into a dynamic awareness system. Furthermore, flood maps are created using a number of different methods, which usually complicate the development of other flood management tools derived from them, often making the tools unique to a specific flood map.

Hydrologic modeling can help make flood maps dynamic. Streamflow or water depth is an essential input to generate a flood map. These values can come from observations (i.e. historical, real-time), or from a hydrologic model (prediction). A workflow that continually links the inputs of a monitoring system or a hydrologic model to generate a flood map can be used as early warning system (EWS).

In Europe, the European Flood Risk Management Directive, in force since 2007, requires all member countries of the European Union (EU) to assess all water courses and determine if they are at risk of flooding; to map the flood extent, assets, and humans at risk in these areas; and to take adequate and coordinated measures to reduce flood risk. Attempts to standardize the approaches used to create flood maps have been made in EU under this directive (van Alphen et al., 2009). Similarly, the Open Water Data Initiative (OWDI) in the US was started in 2014 and seeks to standardize and facilitate water data integration and sharing. The conceptual model developed for the OWDI includes four key functionalities that are essential for engaging the broader community of water data providers and users: Water Data Catalog, Water Data as a Service, Enriching Water Data, and Community for Water Data (Blodgett, Read, Lucido, Slawecki, & Young, 2016). Nevertheless, the standardization of water data remains an ongoing challenge around the world due in part to the wide variety of methods, water data outputs, and aimed audiences.

The accuracy of a flood map largely depends on the hydrologic model used. A hydrologic model is a simplification of a water system that helps us understand, predict, and manage water resources. We can better understand patterns, trends, and changes; we can predict where, when, and how much water we can expect; and most importantly, we can identify, plan, and respond to water-related issues (Figure 1-2). Hydrologic models have allowed scientists and engineers to expand both the temporal and spatial scale at which flood maps can be created. Flood Warning Systems (FWS) derived from hydrologic models require high-resolution results that can capture the effects of forecast predictions on the areas of interest. “Hyper-resolution” or “street-level” hydrologic and hydraulic models have become an essential part in the development of a FWS. Furthermore, these models can be used to help us understand how to better respond to hydrologic events without the need of a flood map. Flood maps are only one of the different tools available to help us communicate the results of hydrologic models. On the other end of the spectrum, hydrologic models can help in the prediction and monitoring of extended droughts.

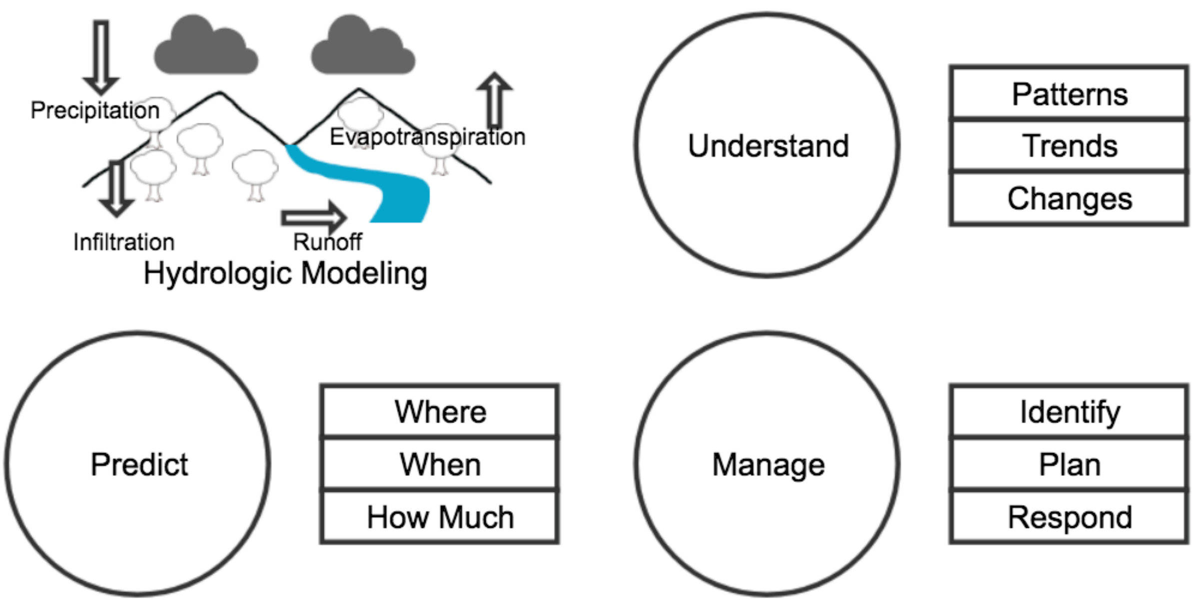


Figure 1-2. Hydrologic modeling and its uses.

1.2 Global Streamflow Prediction New Challenges

Having a hyper-resolution global model does not automatically make it useful. New challenges, which are considered an extension to the initial grand challenge, have surfaced. I have subdivided these challenges in the following categories: big data issues like storage, maintenance, and metrics tracking; communication issues like accessibility, relevancy, and clarity, particularly when working with different groups; adoption issues like ownership, partnering, branding, and overall implementation at the local level (country, region, etc.); and validation issues. This section discusses some of the recent advances regarding the core challenge of creating a hyper-resolution global hydrological model.

The benefits of a large-scale hyper-resolution prediction system are of critical importance in many water-related areas. Some of these areas include food production, climate change, and disaster risk reduction. Specifically, the main concept behind any disaster risk reduction or mitigation is to lower the costs of such inevitable events. The effectiveness of flood preparedness has been proven by various general and localized estimates that compare the initial cost of the initiative with the potential cost of a given flood event or a number of them (Godschalk, Rose, Mittler, Porter, & West, 2009; Kelman, 2013; Kull, Mechler, & Hochrainer - Stigler, 2013). Developed countries usually possess their own flood warning system. The US National Water Model and the European Flood Awareness System are examples of this. While most of the developed world has good data, tools, and experience, developing countries often lack the capacity to produce and maintain their own models, resulting on an increase in vulnerability. Entities like the World Bank and the Inter-American Development Bank have recognized that international assistance is essential for developing countries to overcome vulnerability.

The benefits of a global streamflow forecast are numerous. Streamflow prediction is critical for assessing/predicting extreme events such as floods and droughts, but it also has the ability to positively impact areas like agriculture, biodiversity, and climate change by providing essential data for the development and advancement of problem-solving tools in these areas.

A global streamflow prediction system would not only benefit developing countries by offering predictions in areas that would have no coverage otherwise. Developed countries can also benefit in a similar manner, especially in areas where there are no local models or enough observed data. Other advantages of having a large-scale modeling system include the ability to predict events at a larger scale (e.g. multi-watershed level), and to provide a secondary prediction to serve as comparison to or support an already standing forecast prediction system.

Organizations like the Global Flood Partnership (GFP) seek to establish an international partnership to help strengthen preparedness and response to global disasters, while at the same time incentivizing cooperation, and the development of such a global system that can benefit everyone.

The development of a global hyper-resolution hydrologic model has been deemed a grand challenge mainly because of the significant computational cost, and the amount of data needed to run such a model. However, other scientists have also started to incorporate the communication and usefulness of such a model as part of the challenge. For example, Emerton et al. (2016) included quantifying, understanding, and communicating the values and benefits derived from a global forecast as part of the many challenges of a global forecast prediction system.

A description of some of the main accompanying challenges that come with a global streamflow prediction system is discussed below.

1.2.1 Big Data

A global Streamflow Prediction System requires a solid cyberinfrastructure where results can be stored, and retrieved. Moreover, a continuous operational forecast system requires a workflow that can be run automatically. This would include the download and organization of model inputs, which would add to the already large amount of data produced by the model. Therefore the infrastructure for a global model is bound to include organization tasks to download, archive, and delete data. Traditionally, hydrologic models have been run on local servers, however with the latest advances in Information and Communication Technologies cloud storage and computing has become a common alternative. Souffront Alcantara, Crawley, et al. (2017) provides an example of a large-scale forecast prediction system that is based on an on-premise architecture for storing the US National Water Model. However, it is important to note that this server is separately located at the Renaissance Computing Institute (RENCI), which is a leader in data science innovations, and it is managed and maintained by them (see chapter 2). Therefore, in practical terms, this cyberinfrastructure was able to combine some of the benefits provided by both cloud and on-premise storage and computing.

Cloud computing offers a number of advantages for the development of an operational global forecast prediction system, especially in developing countries. Some of the most obvious advantages include: the removal of expensive computing hardware and space, easy to scale, machines are maintained centrally by the provider, model improvements and other updates only need to be done on the server rather than numerous desktops/servers, and the entire system can be managed from one place (usually a dashboard).

A global forecast prediction system is meant to be accessed by a large number of users. Therefore, an easy to scale system would be ideal. Furthermore, cloud computing and big data, such as the output of a large-scale streamflow forecast system, are conjoined in the sense that big data provides users the ability to use commodity computing to consume subsets of the data that are of interest, while cloud computing provides the underlying engine that makes the data available in the first place (Hashem et al., 2015).

1.2.2 Communication

Communicating water data has been a common challenge for the hydrologic community (Beran & Piasecki, 2009). This is due in part to the evolving nature of hydrologic models. In the last few decades, the emergence of standards for the sharing and distribution of hydrologic data has made communicating and sharing water data much easier. Some of these standards include WaterML, which offers a simple structure for working with time series data; netCDF, which offers a more solid structure for working with multi-dimensional data; and GIS open web service standards like the Web Mapping Service (WMS), Web Feature Service (WFS), and Web Processing Service (WPS) standards, which offer a common denominator for exposing water data in a dynamic way that is compatible with the most used visualization tools available.

The adoption of the standards mentioned above has helped reduce the existing gap between data producers and data users in the hydrologic and decision-making communities. However, most of the focus on data communication is usually placed on scientific/research users. Furthermore, water data needs to be effectively communicated not only to the scientific community, but also to decision-makers, emergency responders, and the general public. Water data needs to be presented as actionable information that is accessible and understandable for all user levels (Souffront Alcantara, Crawley, et al., 2017).

A solution for communicating results to other groups is to develop web applications that allow users to interact with the data through thin-clients depending on their specific needs. Water data as a service through the use of a web app has many benefits. Results can be displayed using open standards, while other functionality can be added to satisfy user needs from a simple web browser. Web apps can successfully link the back-end cyberinfrastructure needed to generate forecast results with state-of-the-art web development technologies to create dynamic environments where users from different levels can access information that is relevant to them by taking advantage of open standards.

1.2.3 Adoption

In general, adopting a new technology usually depends on the estimated benefits and costs of implementation. In the case of a large-scale streamflow prediction system, there are a number of general and specific factors that will determine such benefits and costs, and therefore influence implementation at the local level. Some of the general factors include the existence of a local system, and the disposition of the local community to incorporate or integrate a global system. In such a cases, the global system's value would most likely be in serving as a secondary tool to trigger action or to corroborate when the local model forecasts an extreme event. However, the greatest value of a large-scale system comes when there is no local system available.

More specific factors regarding the adoption of a large-scale forecasting system include the time it takes to adopt new technologies, and who will take responsibility for the performance of the model. Principles like the Technology Acceptance Model (TAM) suggest that the adoption of a new technology depends on the perceived ease of use and usefulness of the technology involved (Davis, 1985). In theory, a large-scale system offers a relative ease of use by eliminating the costs of producing the model and offering forecast results as web services that can be consumed by anyone. However, it is important to note that while a forecast can be easily provided using new communication technologies, model results still need to be interpreted by knowledgeable professionals, and decision-support systems to respond to forecasted events would also still be in the hands of the local community. Therefore, a reasonable depth of understanding of the model is required at the local level. In addition, each country/region that decides to implement a global prediction system will have a vested interest in the good performance of the model. To this end, a combined approach can be implemented to monitor model performance where a local implementation of the model can provide feedback and/or test the model based on local observed data (see section 1.2.4). Furthermore while improvements to the global model based on such local observations will certainly lag, biases identified can be accounted for and decisions properly adjusted locally in the interim. Close collaboration between the model developers and the local implementation is required to achieve this goal.

Success or failure of the model predictions imposes certain responsibility on the owner. But with a global/large-scale system, ownership may not be clear. While the developer of the model provides results, interpretation and response to the model fall at the local level. In practical terms, the weight of the decision support system developed from the model is of far more importance than the generation of a model. As a result, it is usually advised that a multi-criteria approach is used to support decisions. Examples of such systems usually include multiple models, or observation data (Ahmadisharaf, Kalyanapu, & Chung, 2016; Horita, Albuquerque, Degrossi, Mendiondo, & Ueyama, 2015; Niswonger, Allander, & Jeton, 2014; Svoboda, Fuchs, Poulsen, & Nothwehr, 2015; Wan et al., 2014). Based on these factors, users may welcome or reject ownership and therefore responsibility over certain aspects of a global model. To this end, there are a number of implementation levels that would depend on what is determined to work better at the local level by the local agency itself.

1. External model consumption through a web app: The model is accessed from a generic web app developed to display the complete large-scale model. Additional functionality in the app would allow for extraction and visualization of data for a specific area. This generic app could be hosted by an international organization working with different countries/regions.
2. Internal model consumption through a web app: The model is generated on-premise and displayed and accessed through the generic web app. Internal generation would allow for computation of areas of interest only.
3. External model consumption through web services: The model is accessed through open standards and a REST API, and displayed using a customizable web app or integrated into an existing visualization tool. This would allow for display of areas of interest only while avoiding the costs of an on-premise implementation. Further, once a proper understanding of the models value and limitations are understood for the decision at hand, a local agency can "brand" the service in such a way that lends confidence to the downstream users.

1.2.4 Validation

The accuracy and uncertainty of a model need to be quantified before forecasts can be useful for any decision-making. Traditionally, models are tested and calibrated for specific areas. This poses an additional challenge for a large-scale forecast system. Given the extent of such a model, validation and calibration would be a very arduous task. To this end, many large-scale models have instead carried over the uncertainty of their inputs by presenting an ensemble result that accounts for input uncertainty.

Another way the accuracy of the forecast can be evaluated is by comparing results to observed data in local areas post model deployment. Assuming a large-scale model has been adopted at a regional or local scale, the model could be easily compared to regional or local observed data. Moreover, a large-scale forecast system that uses open standards drastically improves the ease of comparisons with any other existing dataset. The results of the comparison analysis can then be shared with the model developer and proper corrections made.

A validation system that facilitates modeled and observed data comparison is necessary to ensure the consistency of such analyses so that they can be performed on different areas and by different groups. It is also critical for developing confidence for using simulated results at all levels.

1.3 Summary

Hydrologic modeling is an important tool for forecasting streamflow. A hyper-resolution model allows for prediction at "the street level," which in turn can be used for urban planning, decision-making, and emergency response that is useful at a local level. Hydrologic modeling results can be used in combination with other tools like flood maps or drought monitoring systems. A global streamflow prediction system is ideal for regions that lack a local model or areas where there is not available input data or human capacity to create a model at a local level.

Developing a large-scale hyper-resolution hydrologic model has been one of the main challenges of the hydrologic community over the past decade. Advances in technology have made possible the development of a number of large-scale hydrologic models by different organizations. Some of the main characteristics of these models include data assimilation, communication of uncertainty through ensembles, and flexible development frameworks. Currently, some of the main large-scale models include the US National Water Model (NWM), the Global Forecast Awareness System (GloFAS), and the family Land Data Assimilation System models derived from NASA's Land Information System (LIS).

Additional challenges have surfaced with the development of global models. Some of these challenges are big data, communication, adoption, and validation challenges. The big data and communication challenges have started to be answered by providing new ways to store and display data. Web applications offer a relatively easy way to display water data in a dynamic and intuitive way. However, multiple components are required to develop these types of applications, and the knowledge required to do so is often out of the expertise of a water resources expert. Web frameworks offer an easy way to develop web applications lowering the barrier for engineers and scientist.

This research seeks to answer the following questions: how can we empower scientists and decision makers so they can make a better use of hydrologic modeling and improve our overall ability to manage our water resource and respond to extreme events? And what improvements are needed to make a large-scale hydrologic model more useful to stakeholders?

Tackling these challenges and questions will increase the usefulness of existing and future global models while also providing regions that lack a forecasting system with data that is readily available and that can be incorporated to their existing systems in a robust way with the help of web services. This in turn has the potential to improve our decision-making ability regarding water resources management and to reduce losses due to extreme events such as floods.

A description of some of the most used large-scale hydrologic models, along with some of the main components necessary to make a streamflow prediction system operational is provided in the next chapter. A Case study using the US National Water Model is provided in chapter 3, with a focus on big data and data communication challenges. Chapter 4 uses the GloFAS-RAPID model to create a semi-global streamflow prediction system covering Africa, North America, South America, and South Asia. All the main hydroinformatic challenges are addressed. Finally, Chapter 5 provides a summary of the developed tools and workflows, along with future development opportunities and areas of improvement.

2 LARGE-SCALE HYDROLOGIC MODELING AND COMPONENTS REVIEW

2.1 Large-Scale Hydrologic Modeling Review

The development of a high-resolution global model to monitor surface water has been deemed a “grand challenges” of the hydrologic community (Wood et al., 2011). So far, this challenge has been addressed primarily from the development point of view; the creation of such a model requires a reevaluation of variable interactions, computational costs, and the incorporation of observed data. The last few years have seen the development of such large-scale models become more feasible. As a result, a number of large-scale hydrologic models have surfaced. A review of available large-scale hydrologic models and some of their main components is presented with a focus on cyberinfrastructure, and accessibility.

2.1.1 NWM

The National Water Model (NWM) is a hydrologic model that generates forecasts for multiple variables across the continental US (NOAA, 2016). It was released in 2016 by the National Weather Service (NWS) Office of Water Prediction (OWP) in collaboration with the National Center for Atmospheric Research (NCAR) and the National Center for Environmental Prediction (NCEP). The NWM simulates runoff conditions for the 2.7 million reaches of the National Hydrography Dataset (NHD) (USGS NHD, 2016), which represents a significant increase over the approximate 4000 locations forecasted by the NWS through the thirteen River Forecast Center’s operations. The model has four different configurations or forecast products, which differ in duration, time step, and frequency. All four configurations produce a unique forecast, with the exception of the long-range configuration, which is an ensemble forecast with four different members lagged by four six-hour time intervals for a total of 16 forecasts per day. The analysis and assimilation configuration is produced in near real-time and assimilates observation data from USGS gages. It serves as initialization for the other three configurations by providing an estimate of current conditions. In addition, the NWM produces results for three geospatial types or shapes: channel, land, and reservoir. The channel and reservoir types are based on the US NHD plus dataset, while the land type is based on a 1km^2 land surface grid. The outputs of the model are made available as netCDF files on the NOAA Operational Model Archive and Distribution System (NOMADS) and through an NCEP FTP server. These outputs are only persisted for two days, so they must be retrieved, processed and accessed from a separate site. The NWM produces forecasts for a number of hydrologic parameters that vary depending on the model's type and configuration. These variables include streamflow, streamflow velocity, total evapotranspiration, subsurface runoff, soil saturation, snow depth, and snow water equivalent. The latest streamflow forecast can be visualized from a dynamic map in the OWP main website (http://water.noaa.gov/map).

A new version of the NWM was released in May 2017. Besides a list of enhancements to the actual model that include stream connectivity refinements, and improved parameter calibration; this version significantly reduced the size of the outputs, allowing for an increase in previous forecast storage, and faster data retrieval.

The core of the NWM is the NCAR's Weather Research and Forecasting Hydrologic model (WRF-Hydro). The NWM inputs come from a variety of sources including Multi-Radar/Multi-Sensor System (MRMS) radar-gauge observed precipitation data, and High

Resolution Rapid Refresh (HRRR), Rapid Refresh (RAP), Global Forecasting System (GFS) and Climate Forecast System (CFS) Numerical Weather Prediction (NWP) forecast data. WRF-Hydro is configured to use the Noah and Noah-MP Land Surface Models (LSMs) to simulate land surface processes. Separate modules perform diffusive wave surface routing and saturated subsurface flow routing on a 250m grid, and then Muskingum-Cunge channel routing through the NHDPlusV2 stream network (NOAA Office of Water Prediction, 2017).

An additional method for visualizing and extracting NWM forecasts for extended periods is available through HydroShare's National Water Model Viewer web application. HydroShare is an online, collaborative resource for sharing hydrologic data (Tarboton et al., 2014). The NWM Viewer provides visualization and interaction with NWM forecasts. In addition, it convers NWM outputs from individual large-scale spatial files to localized time series files that can be downloaded manually or programmatically using a REST API developed as part of the web app (Souffront Alcantara, Kesler, et al., 2017).

2.1.2 WRF-Hydro

The Weather Research and Forecasting Model Hydrological Modeling System (WRF-Hydro) is a community-based model framework that allows coupling of meteorological models and terrestrial/hydrologic models. This model architecture, built using Fortran, was originally created to work with the WRF model, an atmospheric model for meteorological and numerical weather predictions; however, it has been expanded to provide a modularized approach that covers typical terrestrial and hydrological processes, and to allow modelers to use their own meteorological input files. The Noah and Noah-MP LSMs are the primary surface models in WRF-Hydro.

Noah and Noah-MP calculate vertical fluxes of energy, moisture, and soil states. The Noah model is a one-dimensional LSM that simulates soil moisture, soil temperature, skin temperature, snowpack, canopy water content, and surface energy flux and water flux (Ek et al., 2003). The Noah-MP model incorporates multiple-parameterization options allowing for improvements in simulation of runoff and other hydrological variables (Niu et al., 2011).

2.1.3 GloFAS

The Global Flood Awareness System (GloFAS) is an ensemble hydrologic model that generates 51 different runoff forecasts for the major rivers of the world on a global grid with a resolution of 16 km^2 on a continuous basis. A 52nd forecast is generated at a resolution of 8 km^2. GloFAS was released in 2011 by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the European Commission's Joint Research Centre (JRC), and has been operational since July 2011. The GloFAS system is composed of an integrated hydrometeorological forecasting chain and of a monitoring system that analyzes daily results and shows forecast flood events on a dedicated web platform (Alfieri et al., 2013). This model uses real-time and historical observations in combination with a Data Assimilation System (DAS) and a Global Circulation Model (GCM). The underlying framework used to create GloFAS is ECMWF's Integrated Forecasting System (IFS). GloFAS uses the HTESSEL model for its land surface scheme. HTESSEL is a hydrologically-revised version of the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL) model (Balsamo et al., 2009). This new land surface scheme corrected the absence of a surface runoff component in its predecessor, among other minor improvements. Finally, the LISFLOOD model is used to route the GloFAS results that are presented in their web application (http://globalfloods.jrc.ec.europa.eu/glofas-forecasting/).

LISFLOOD is a physically-based distributed model that provides one-dimensional channel routing using the kinematic wave method (Roo, Wesseling, & Deursen, 2000).

A predecessor to GloFAS, the European Flood Awareness System (EFAS) is an operational monitoring system for forecasting floods across Europe that follows the same structure as GloFAS, though at a higher resolution and with additional functionality due to the ability to connect through the European Union (EU) and several national hydrological and meteorological agencies for data assimilation, validation, and calibration. The aim of EFAS is to gain time for preparedness measures before major flood events strike, particularly for trans-national river basins both in the member states of the EU as well as on a general European level.

2.1.4 ERA Interim and ERA5

ERA-Interim is a global atmospheric reanalysis produced by ECMWF. This is a global dataset that spans January 1980 to December 2014 (35 years). One of the advantages of using reanalysis is that the data provides a global view that encompasses many essential climate variables in a physically consistent framework, with only a short time delay (Dee et al., 2011). This type of data becomes invaluable in areas where no actual observed data is available. A runoff derivative of this atmospheric reanalysis was produced on a 40km^2 global grid using a land surface model simulation in HTESSEL.

GloFAS forecasts can be visualized from their main website (http://globalfloods.jrc.ec.europa.eu/glofas-forecasting/), which combines the forecasts from GloFAS and the simulated historic run from the ERA Interim to provide an awareness system that displays warning points and the probability of an event occurring based on the ensemble forecasts and return periods extracted from the ERA data.

A web app exists for GloFAS that allows for visualization of forecasts. The underlying processes associated with this app also address GloFAS' density challenge by routing model results through a river network to provide results not only for major rivers, but for any potential river in the world. The Streamflow Prediction Tool (SPT) provides an intuitive user interface that allows for the easy lookup and visualization of results (Snow et al., 2016).

ERA-Interim is being updated by the ERA5 dataset; a new climate reanalysis dataset developed by ECMWF that will eventually cover a time window from 1950 to present. ERA5 provides estimates for a large number of atmospheric and land surface variables over the globe at an increased resolution of 30 square kilometers. Similar to the ERA-Interim/Land the ERA5 will be used in the HTESSEL land surface model to produce the global runoff land component, which in turn can then be run through the same river network to produce the simulated historical discharge data. Seven years of data have already been released. The rest of the data will be available in January of 2019 and is meant to replace the ERA-Interim dataset. Some of the improvements of the ERA5 dataset include a higher resolution, better global balance of precipitation and evaporation, and better soil moisture.

2.1.5 IFS

The Integrated Forecasting System is the framework where all the main ECMWF models run. It is a set of computer programs written in Fortran. The IFS was first implemented in 1994. The IFS was jointly developed and maintained by ECMWF and Meteo-France, the French national meteorological service. This model was the first to use 4D-Var data assimilation, which is a four-dimensional variational data assimilation technique. It performs a statistical interpolation in space and time between a distribution of meteorological observations and an a priori estimate of the model state (Andersson & Thépaut, 2008). This implementation represented a significant advance at the time.

2.1.6 HYPE

The Hydrological Predictions for the Environment model is a model fro small- and large-scale assessment of water resources and water quality (Lindström, Pers, Rosberg, Strömqvist, & Arheimer, 2010). This model has the peculiarity that it works with sub-basin areas as opposed to grids, and that it incorporates nutrient flow in addition to water flow.

The European Hydrological Predictions for the Environment (E-HYPE) is a multipurpose model currently used for water forecasting in Europe, and specific areas and research projects. It was developed by the Swedish Meteorological and Hydrological Institute (SMHI). The model is forced by daily precipitation and temperature and then calculates flow paths in the soil based on snow melt, evapotranspiration, surface runoff, infiltration, percolation, macropore flow, tile drainage, and lateral outflow to the stream from soil layers with water content above field capacity (Donnelly, Andersson, & Arheimer, 2015). E-HYPE distinguishes itself from other large-scale models in that it divides watersheds into sub-basins rather than using a regular grid. This model also has the ability to account for human influences like irrigation and hydropower. Results from E-HYPE can be visualized from SMHI's HypeWeb website (http://hypeweb.smhi.se/europehype/) and from SMHI's river info website (http://riverinfo.eu/). The former displays the different variable outputs per sub-basin while the latter presents streamflow forecasts together with historical data in a dynamic way through the use of an interactive map and a graph.

An EFAS-Hype initiative is currently been tested. This model uses forcing data from EFAS and is customize towards flood warning and transboundary flow.

2.1.7 GLDAS

The Global Land Data Assimilation System is a global land surface model developed jointly by NASA's Goddard Space Flight Center (GSFC) and NOAA's National Center for Environmental Prediction (NCEP). The main goal of GLDAS is to produce a high spatial resolution model in near-real time. GLDAS is a global, high-resolution, offline (uncoupled to the atmosphere) terrestrial modeling system that incorporates satellite- and ground-based observations in order to produce optimal fields of land surface states and fluxes in near–real time (Rodell et al., 2004). A series of other LDAS models have been developed for specific areas around the world. Some of these include the North American LDAS (NLDAS), the Famine Early Warning Systems Network LDAS (FLDAS) mainly for Africa's sub-saharan region, the ongoing South Asia LDAS (SALDAS), among others. LDAS models are creating using the Land Information System (LIS) framework.

2.1.8 LIS

LIS is a software framework that integrates the use of satellite and ground-based observational data along with advanced land surface models and computing tools to accurately characterize land surface states and fluxes (Kumar et al., 2006). LIS's available LSMs include the Community Land Model (CLM), Noah Land Surface Model, Variable Infiltration Capacity (VIC), and Hydrology with Simple SIB (HySSIB). This framework allows for the production of high-resolution results (as high as 1km), and the assimilation of earth observations from NASA's satellites (e.g. AQUA and TERRA). LIS was developed using Fortran and C.

VIC is a model that predicts surface runoff and energy fluxes. The model, first developed in the early 1990s, was created with the intention of representing land surface hydrology at a global scale for incorporation into General Circulation Models (GCMs) (Liang, Lettenmaier, Wood, & Burges, 1994).

2.1.9 Fortran

After reviewing some of the main modeling frameworks used in hydrological modeling it becomes clear that one of the main similarities between frameworks is that they all used Fortran to some extent. One of the oldest programming languages, Fortran is a High Performance Computing (HPC) language especially suited for numeric and scientific computations. It was first introduced in the 1950's, but despite its age remains one of the most used scientific HPC programing languages. By comparison, Python, which is a very popular interpreted language, is about 100 times slower than Fortran. Another factor in the use of Fortran is the existence of legacy code that has been optimized and can continue to be used for the development of new models.

2.2 Water Resources Communication and Visualization Components

Water data in all its forms needs to be communicated in effective ways to help accomplish our society goals, and to respond to extreme events in a timely manner. Hydroinformatics combines elements of hydrology, hydraulics, and Information and Communication Technology (ICT) to help provide the information necessary to solve water-related problems. Many different tools and technologies are used to analyze, interact, share, and communicate water data. This is due to the many uses of water, and the many ways in which water needs to be accounted for. A description of some of the main components used to communicate water data, specifically modeled data, and relative to our solution is provided in the next few paragraphs.

Modeled data has two main components: A geographic component, and a numeric component. The geographic component of water data refers to the way scientists and engineers represent water bodies in a Geographic Information System (GIS) environment. Water data is usually represented as vector or raster data. In order to communicate water information, a number of standards have been adopted (see section 1.2.2). These standards are stored and displayed using special servers.

2.2.1 Geospatial Servers

Geoserver is an open source server for sharing geospatial data (Iacovella, 2017). Vector or raster water data can be stored in a Geoserver to make it available online following any of the most common standards for sharing geospatial data.

ArcGIS Server is a licensed server that allows you to make geospatial data available to others as a web service. Other ways to store the geographic side of water data include the use of databases. PostgreSQL is an example of an open-source SQL database that can be geospatially enable to store geographic data.

In addition to geospatial data storing and sharing, web-mapping libraries provide a way in which this data can be consumed by providing visualization in a web map. Openlayers is an open-source web-mapping library for developing interactive web maps (Gratier, Spencer, & Hazzard, 2015). Openlayers can display most geospatial vector and raster formats. Leaflet is another example of an open-source web-mapping library. Leaflet also provides easy ways to display geographic data that is time enabled (Agafonkin, 2014). The ArcGIS API for JavaScript library is another open-source web-mapping library that provides ways to consume geospatial data using open-source standards and ArcGIS formats.

Swain et al. (2015) provided a more detailed review of geospatial databases and web-mapping libraries along with other commonly used software for developing water resources web applications.

2.2.2 Numeric Data Formats

While not the preferred way, the numeric component of water data can also be stored in geospatial and SQL databases. However, some of the most common ways to store modeled results include local storage using file and table formats. The netCDF file format is a standard for scientific research and offers good flexibility and structure (Rew & Davis, 1990). However, it has some limitations regarding data transfer and web accessibility for simple data (e.g., a time series). NetCDF files cannot be displayed directly through a web browser without the help of additional software, and usually need to be downloaded to access the data. Users are required to download a set of files and become familiar with the file structure to be able to extract meaningful information. On the other hand, WaterML, which is also a standard for water data sharing, offers a simple structure for sharing hydrologic time series data and its related metadata based on Extensible Markup Language (XML) (Zaslavsky, Valentine, & Whiteaker, 2007) that can be visualized in a browser.

Other formats such as the JavaScript Object Notation (JSON) and some of its variants such as GeoJSON have also found their way into some of the most used formats to store and share water-related data due to its compatibility with JavaScript and other web development languages to provide visualization and web tools that consume water data. JSON is a plain text data format that follows a dictionary-like pattern to store and access data (keys and values). Both geographic and numeric water data can be stored in JSON.

2.2.3 Water Data Archives

In addition to web servers hosting water data, other generalized water data archive systems provide ways to store, share, and retrieve data. Some of these systems include HydroShare, HydroServer, and IRODS.

HydroShare is an online, collaborative resource for sharing hydrologic data (Tarboton et al., 2014). It provides a "community of water data" necessary to comply with OWDI standards by providing a site where users can publish, share, manage, discover, visualize, and download water-related data and models. HydroShare provides ways to store and find all types of water-related data, from plain text results to actual models or even web applications.

The Integrated Rule Oriented Data System (iRODS) is open source data management software (Rajasekar et al., 2010). It virtualizes data storage resources allowing users to retrieve data regardless of where and on what device the data is stored through the use of a client.

HydroServer is a set of software applications for publishing hydrologic datasets on the Internet. It is a computer server that contains a collection of databases, web services, and software that allows data producers to store, publish, and analyze space-time hydrologic datasets (Horsburgh et al., 2010). HydroServer uses an Observations Data Model Database, which makes it ideal to store observed data or other water data that requires a relational structure. ODM is a relational model that defines the persistent structure of data, including the set of attributes that accompany the data, their names, their data type, and their context (Horsburgh, Tarboton, Maidment, & Zaslavsky, 2008).

There are a number of other water data archives that are available on the web. These include government projects designed to host water and weather related data produced by government agencies. An example of such a system is the NOAA Operational Model Archive and Distribution System (NOMADS), which provides a way to retrieve weather data produced by NOAA and other government funded organizations. NOMADS provides a way to access real-time, modeled, and historical data in a format-independent web-service-oriented way (Rutledge, Alpert, & Ebisuzaki, 2006).

2.2.4 Web Development Frameworks

The number of different components required to deploy a water resources web application makes the development process complex. Even though compatibility is guaranteed by using open standards, it is necessary to determine what specific software and formats an app will use. Web frameworks exist to facilitate this stage by providing a set of tools bundled together to address each of the requirements of a web app. This concept is common in computer science, however it is relatively new in water resources.

Two main web frameworks that focus on water resources are reviewed here. Tethys platform is a web framework for developing water resources web applications. It offers a suite of open-source software selected to address the unique development needs of water resources web apps (Swain et al., 2016). Tethys provides tools for the creation of an intuitive interface including interactive maps, forms, and graphs, while at the same time providing tools to connect this front-end-user interface to background processes usually required in water resources apps. The main goal of Tethys is to lower the barrier of web app development for water resource scientists and engineers with a working knowledge of web programming and specific capability in Python.

Similarly, The ArcGIS Enterprise Suite, which includes ArcGIS Portal, ArcGIS Server, and an ArcGIS Data Store, offers a licensed approach to easily deploy geospatially-enabled web applications. The development of background processes is leveraged using ArcGIS for Desktop combined with Python workflows, including Arcpy, a python package that includes most of the tools that form the core functionality of ArcGIS for Desktop.

A web framework facilitates the development of web application by providing an "out-of-the-box" solution to develop and deploy web applications. In the case of web frameworks oriented towards water resources, this is more so given the fact that the app developer is usually an engineer/scientist as opposed to a professional web developer.

3 IMPROVING DATA ACESSIBILITY FOR THE UNITED STATES NATIONAL WATER MODEL (NWM)

3.1 Background

United States national and local agencies responsible for forecasting have been trying to bridge the gap between meteorological models, hydrologic models, and end-user needs for the past several years (McEnery, Ingram, Duan, Adams, & Anderson, 2005; Xu, 1999). Some of the efforts to improve water resources forecasting include the following: the creation of an open water data initiative (OWDI) in 2014, which seeks to standardize and facilitate water data sharing; the creation of a National Water Center (NWC) in 2015 with the goal of creating a new National Water Model that covers the entire continental U.S. on regular time intervals; and the initiation of the NWC's National Flood Interoperability Experiment (NFIE) in collaboration with the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) starting in 2015. Since its creation, NFIE has served as an important connection between the academic community and the NWC for improvements in flood forecasting (Maidment, 2016). In 2015, NFIE focused on the creation of a prototype hydrologic model for the continental U.S. The success of this experiment was reflected in part by the release of the National Water Model (NWM) a year layer in August 2016. The NWM is a hydrologic model that generates forecasts for multiple hydrologic variables across the continental U.S. (NOAA, 2016). It was developed by the National Weather Service (NWS) Office of Water Prediction (OWP) in collaboration with the National Center for Atmospheric Research (NCAR), and the National Center for Environmental Prediction (NCEP). The NWM simulates conditions for the 2.7 million reaches of the National Hydrography Dataset (USGS NHD, 2016). This represents a significant increase in the NWS flood forecasting ability, which was previously available at approximately 4,000 locations served by a number of models and using different modeling techniques. The massive amount of new information being produced by the NWM presents a new challenge in water data management.

The NWM has four different configurations or forecast products, which differ in duration, time step, and frequency (Figure 3-1). All four configurations produce a unique forecast, with the exception of the long-range configuration, which is an ensemble forecast with four different members produced over four staggered six-hour time blocks (16 total forecast members per day). The analysis and assimilation configuration is produced in near real time and assimilates observation data from U.S. Geological Survey (USGS) gages. It serves as initialization for the other three configurations by providing an estimate of current conditions at the outset of a new forecast. In addition, the NWM produces output for three geospatial types or shapes: channel, land, and reservoir. The channel and reservoir types are based on the U.S. NHD Plus dataset, while the land type is based on a 1km^2 grid system covering the continental U.S. The outputs of the model are made available as netCDF files on the NOAA Operational Model Archive and Distribution System (NOMADS) and through an NCEP FTP server. However, these outputs are only stored for two days. While streamflow is the primary output from the NWM, the forecasts also provide a number of other hydrologic parameters that can provide useful information for validation and decision support. These variables include the following: streamflow velocity, total evapotranspiration, subsurface runoff, soil saturation, snow depth, and snow water equivalent.

With the advent of new and more comprehensive hydrologic models, the scientific community is now starting to close the second gap: providing timely, continuous, and freely accessible water intelligence. An important part of achieving this goal requires addressing big data and data communication issues such as storage, accessibility, distribution, visualization, and relevancy. In 2016, NFIE was renamed as the National Water Center Summer Institute, and focused on complimenting the NWM by exploring how to better generate decision support products derived from the NWM forecasts such as flood maps. Flood maps offer an effective way to convert water information into water intelligence by providing it in a context (spatially and temporally enabled forecasts) that is relevant to users.

3.2 NWM Storage and Data Access Improvements

I developed an interface to easily access NWM forecast outputs was developed as part of the activities leading up to the 2016 Summer Institute. I also developed a cyberinfrastructure for storing and retrieving NWM forecasts in collaboration with the Renaissance Computing Institute (RENCI), a leader in data science support. The web interface consisted of a series of web applications created using the open source Tethys Platform for accessing and visualizing the forecasts via the web. These web applications serve as a gateway to the NWM forecasts (Souffront Alcantara et al., 2017). In addition, other web apps were derived from the data-access apps for comparing forecasts to observed data, and demonstrating the capability to provide dynamic flood maps from the NWM forecasts for specific areas using a REST API. These additional apps serve as examples of what can be done with the output from the NWM.

In the next few paragraphs, I will introduce these web apps, which served as a starting point for the development of a more comprehensive toolset to validate the NWM while improving the overall national ability to access, visualize, and develop additional resources such as flood maps using NWM forecasts. The developed cyberinfrastructure and app designs, storage and performance metrics, and the deployed apps are presented.

Configuration	Cycling Frequency	Forecast Duration	Forecast Step
Analysis & Assimilation	Hourly	-3 hours	1 Hour
Short Range	Hourly	0-15 hours	1 Hour
Medium Range	Daily	0-10 days	3 Hours
Long Range	4xDaily (16 mem)	0-30 days	6 Hours

Figure 3-1. National Water Model Configuration (adapted from http://water.noaa.gov/about/nwm)

Table 3-1. Deployed Web Applications

Functionality	Web App
Data Access	NWM Viewer NWM Explorer
Utility	USGS/AHPS Gauge Viewer West Virginia Flood Map Tuscaloosa Flood Map

3.2.1 A new Cyberinfrastructure for the NWM

As part of the preparation for the NWC summer institute we developed and deployed two web applications along with the accompanying back-end processes and storage architecture to expose the volumes of data generated daily to support the NWM.

An NWM netCDF output file ranges from 72MB to 2GB in size depending on model configuration and type. The number of files per forecast varies by configuration depending on frequency, duration, and forecast time step. One netCDF file includes a single time step value for the set of stream reaches, grid cells, or reservoir points modeled. Therefore, a single forecast is composed of many large netCDF files. For example, one complete streamflow forecast for the short-range configuration (the shortest configuration) would include fifteen ~72MB netCDF files for a total of 1GB uncompressed. The total size of a complete forecast for the other configurations scales up quickly. The medium-range forecast for the simulated land results includes 80 files of approximately 2GB each. The two examples above represent the total size for a single forecast. To provide a sense of how much data is being produced daily, there are twenty-four 15-hour short-range forecasts, one medium-range forecast, and one 16-member ensemble long-range forecast that collectively result in about 2TB of data per day. This high volume of daily storage emphasizes the need for a cyberinfrastructure designed specifically to facilitate forecast extraction for a specific feature or group of features (stream reach, grid-cell area, or reservoir point) without requiring the formidable task of downloading the entire set of files.

To develop this cyberinfrastructure, we established a process that searches NOMADS and automatically downloads the NWM outputs using a secure server within RENCI's data center each time a new forecast is discovered. The compressed files are extracted to make the NWM forecast results more accessible to the public as direct file downloads or through apps designed to mine selected time series for reaches, reservoirs, or land-grid cells. Figure 3-2 shows the file structure used to store the processes files. The files were stored by forecast range type and then by date.

Figure 3-3 shows the cyberinfrastructure for preparing, storing, and communicating NWM forecasts with web apps. Once the NWM forecasts are copied to RENCI, subsequent scripts are used to unzip and process the forecasts into georeferenced netCDF files by adding geographic coordinates. The georeferenced files are then moved into corresponding directories where the short, medium, and long-range forecasts are persisted for two weeks and the Analysis & Assimilation forecast is persisted since the inception of the NWM in May of 2016. Finally, the naming convention for each file included parameters like date, time, configuration, and geometry to facilitate data querying.

A Jenkins automation server within RENCI's data center manages this entire workflow (Figure 3-4). Jenkins is an automation server written in Java that helps automate processes.

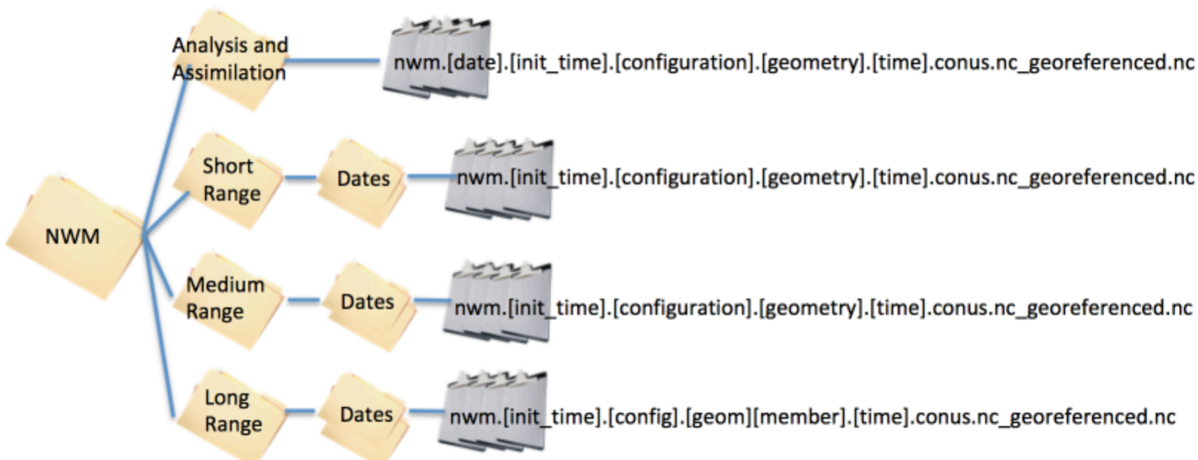


Figure 3-2. National Water Model Forecast File Directory Structure.

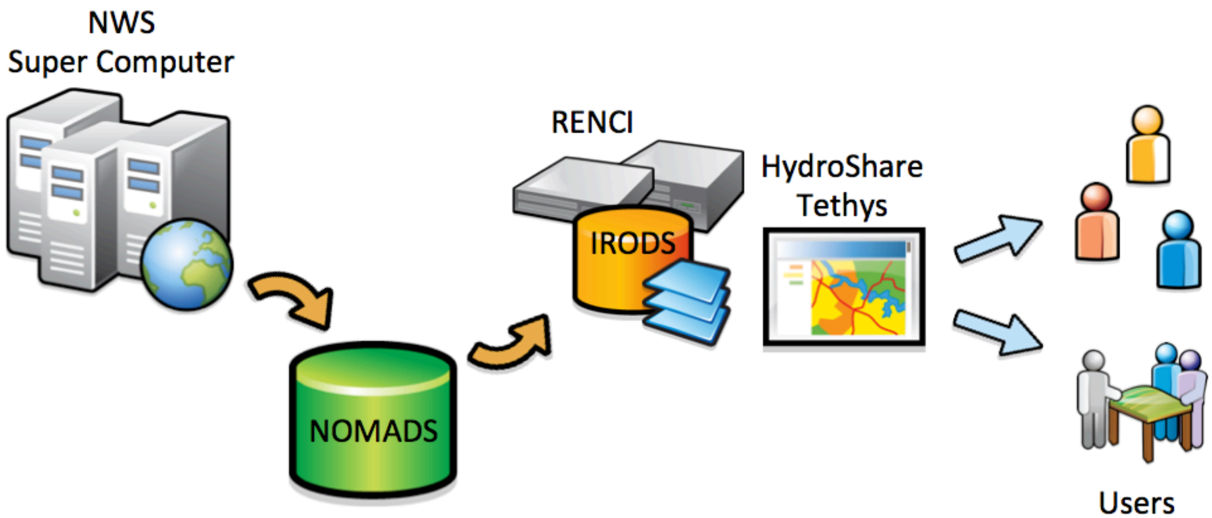


Figure 3-3. Web Apps Cyberinfrastructure.

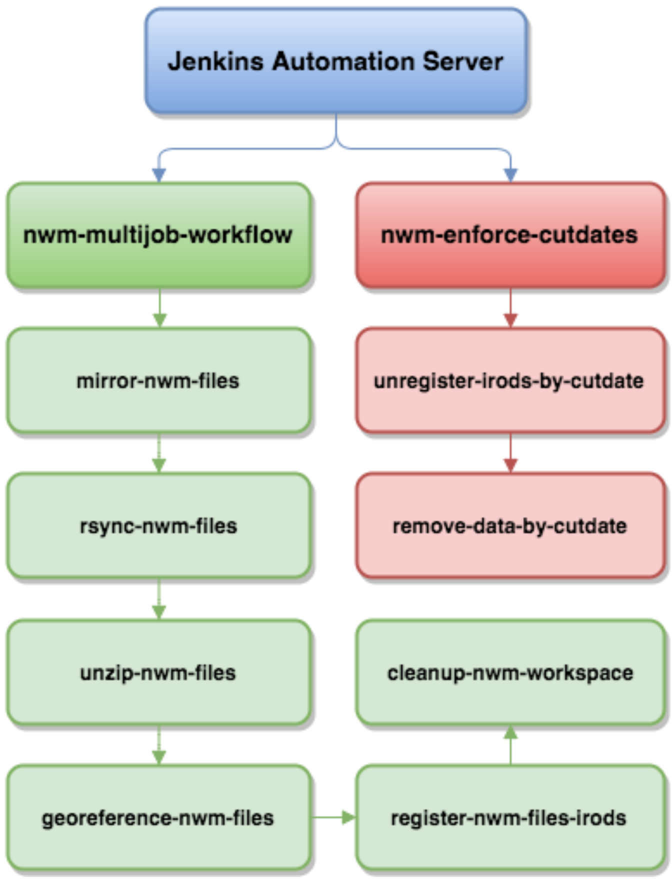


Figure 3-4. Jenkins workflow for preparing NWM forecast data.

3.2.2 Data-Access Apps

Two web applications were developed for accessing the NWM data. These apps access the NWM outputs by connecting directly with the cyberinfrastructure at RENCI. The data-access apps operate on top of the NWM output files as well as the inputs (forcing files) used to create them, and were designed to extract specific forecasts for a stream, reservoir, or land area provided by the user.

The NWM Explorer app was designed to search, discover, and download individual NWM forecast files, or entire sets of zipped forecast files within RENCI's data storage directories (Figure 3-5). Filters can be used to further drill-down access to forecasts based on geometric feature (stream, reservoir, land-grid cell) or time. The forcing files used as input for the NWM are also available to enable scientists to run their own models using the same inputs as the NWM and possibly compare their results to the NWM. Data can be downloaded via three different methods.

1. The file system explorer allows users to directly browse RENCI's directories
2. The iRODS explorer, which can be accessed separately from HydroShare or from a command prompt
3. The REST API, which facilitates download automation and the retrieval of selected model output

Other useful functionality includes options for getting a list of available forecasts and file metadata for a specific file.

The NWM Viewer app was developed to provide visualization and extraction of output for the four configurations of the NWM, and all of the variables associated with the grid, stream, and reservoir geometries. The app is composed of a map for spatial interaction (i.e., selecting streams, grid cells, or reservoirs), a form for forecast parameter specification (i.e., configuration or date), and a graph area for displaying the forecasts in the form of a hydrograph (Figure 3-7). Besides interaction and visualization, another important functionality from this app is that it converts the NWM forecast netCDF files into a time series format that can be displayed in a hydrograph, and can also be downloaded as a WaterML file. The NWM Viewer extracts information from multiple netCDF files given a time range and other basic parameters and assembles them in a time series (Figure 3-6). In addition, the NWM Viewer app has the option to convert NWM forecasts from their default metric units to U.S. customary units and to extract the time series data as a CSV file.

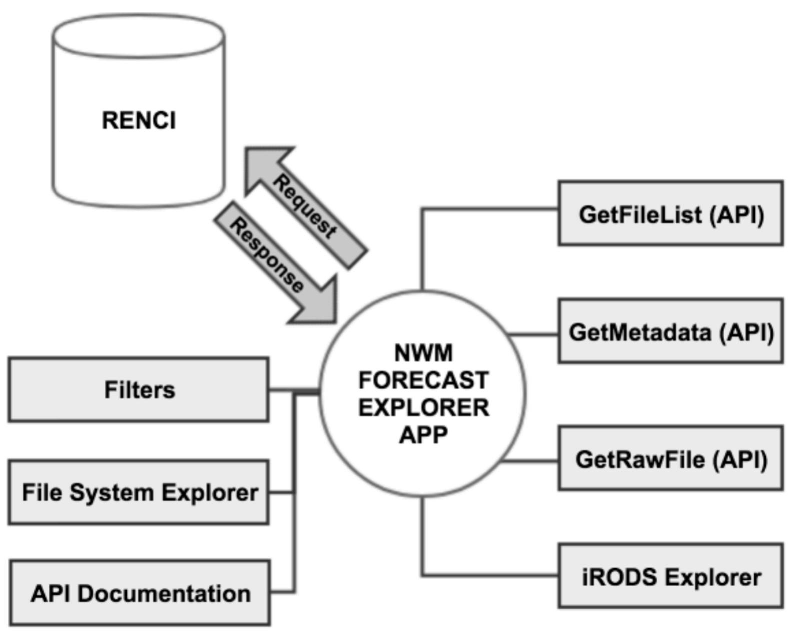


Figure 3-5. National Water Model Explorer App Structure.

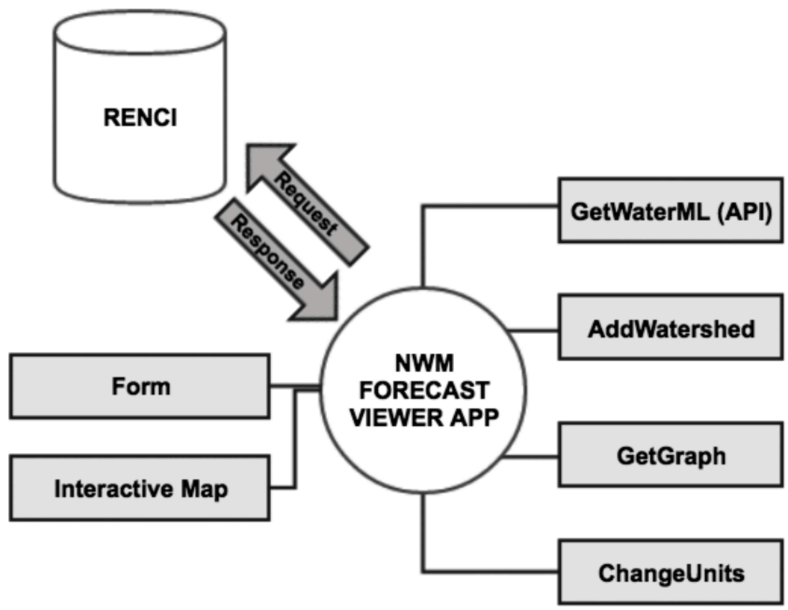


Figure 3-6. National Water Model Forecast Viewer Structure.

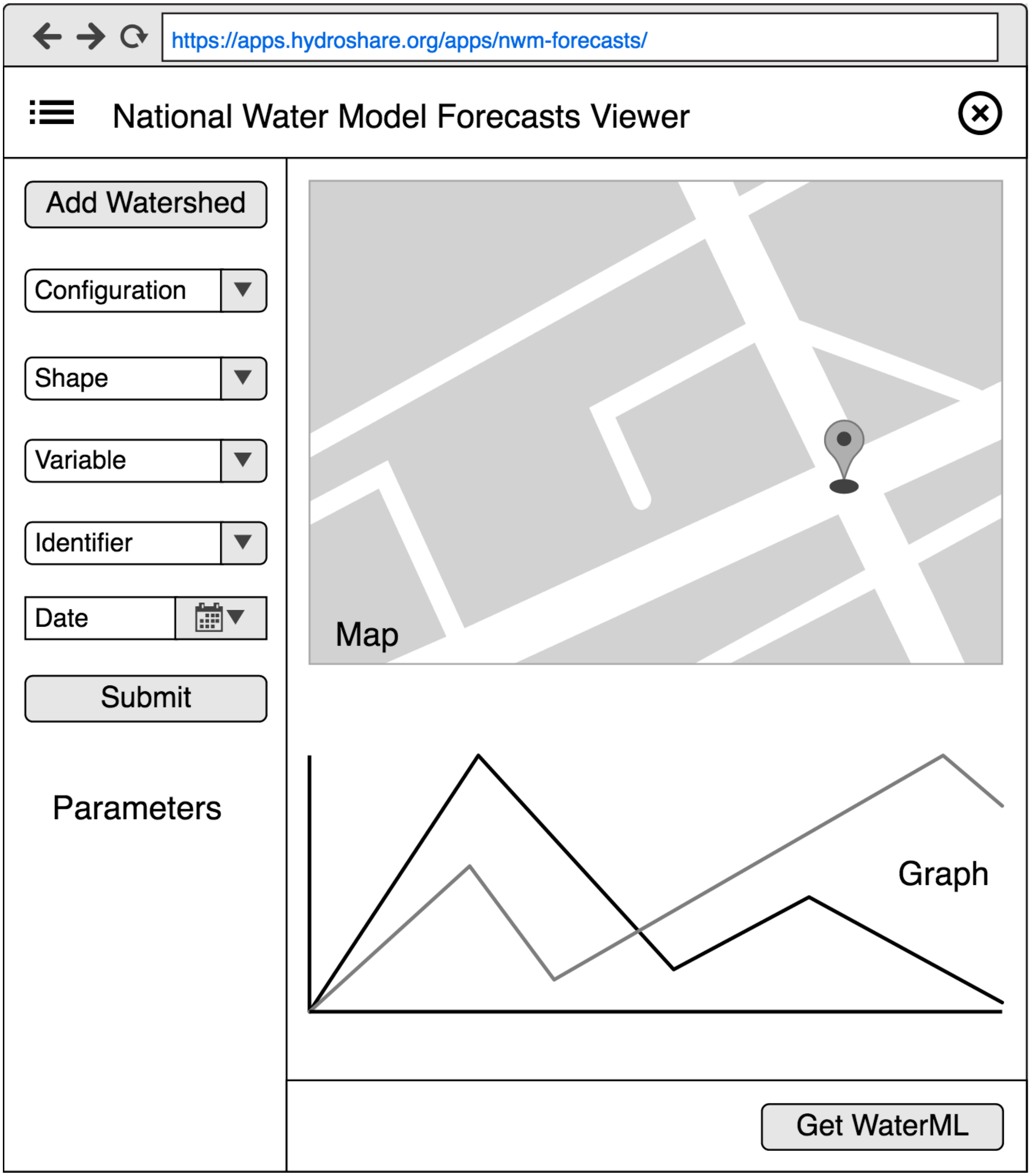


Figure 3-7. National Water Model Viewer Interface Design.

3.2.3 NWM REST API

A REST API was developed to facilitate NWM data retrieval for a single feature, and programmatic manipulation using languages like Python, R, or MatLab. A REST API is a web service or a set of functions that can be used to access data without a web interface. Our REST API communicates with RENCI's server using the HTTP protocol. A request is sent to the server using a URL, which contains the criteria for the desired forecast. The parameters given in the URL are the same parameters used in the interface (i.e. configuration, geometry, variable, identifier). A response from the RENCI server is returned as a WaterML file, which contains a time series with the desired forecast and basic metadata. The WaterML file can be viewed in a browser, downloaded, or read and parsed as a text file using a programming language of choice. Other methods available through our API allow users to query, explore, and download raw NWM forecasts and the inputs used to produce the forecasts without ever having to open the actual apps. Four methods are included in the API.

- GetFileList: Returns a list with the names of the NWM files available for a specific query.
- GetFile: Returns a streaming download of the specified file in NetCDF format.
- GetFileMetadata: Returns key-value pairs with the available metadata for the file specified.
- GetWaterML: Returns a WaterML file with the time-series data for the specified forecast query.

A load test was designed to determine how much data could be requested from the API before it overloaded. The results of the test along with information about the deployed apps are described in the Results section.

3.3 Results

Our data-access apps allow users to easily visualize, and download NWM forecasts in WaterML, an intuitive format to use and transfer time series data. The apps achieve this by accessing raw NWM forecasts that are previously downloaded and processed in a server from RENCI. This cyberinfrastructure handles an average of 4 TB of files per day, with forecasts currently being stored for 14 days. A metrics test revealed that between October 2017 and October 2018, more than a hundred users have accessed our web services with just over 900 hits for our web apps, and more than 42500 calls to our REST API. Figure 3-8 shows estimated user locations for our NWM apps and API. These numbers show that our apps have successfully exposed the NWM to a community of water data users not only in the United States but around the world.

NWM forecast files can be downloaded using the NWM Explorer app. Raw forecast products can also be downloaded as groups based on filters that query the files by configuration, shape, and time. There are three modes in which data can be accessed: the file system explorer, iRODS, and the API.

This app provides a mechanism for scientists wanting to access raw NWM files for research purposes. Furthermore, all the variables available for the different configurations of the NWM as well as the input forcing files used to create the NWM forecasts can be browsed and downloaded directly.

The NWM Viewer app simplifies the NWM forecast for a specific feature. It converts NWM output files from individual large-scale spatial files to localized time series files (Figure 3-9). In addition, it offers visualization and data extraction from the more complex and binary-based netCDF format to a simpler, XML-based format. Parameters are provided by the user through a form, a specific feature (stream reach, grid cell, or reservoir point) is selected through an interactive map, and the resulting time series is displayed in the graph area, which can also be downloaded as WaterML or CSV.

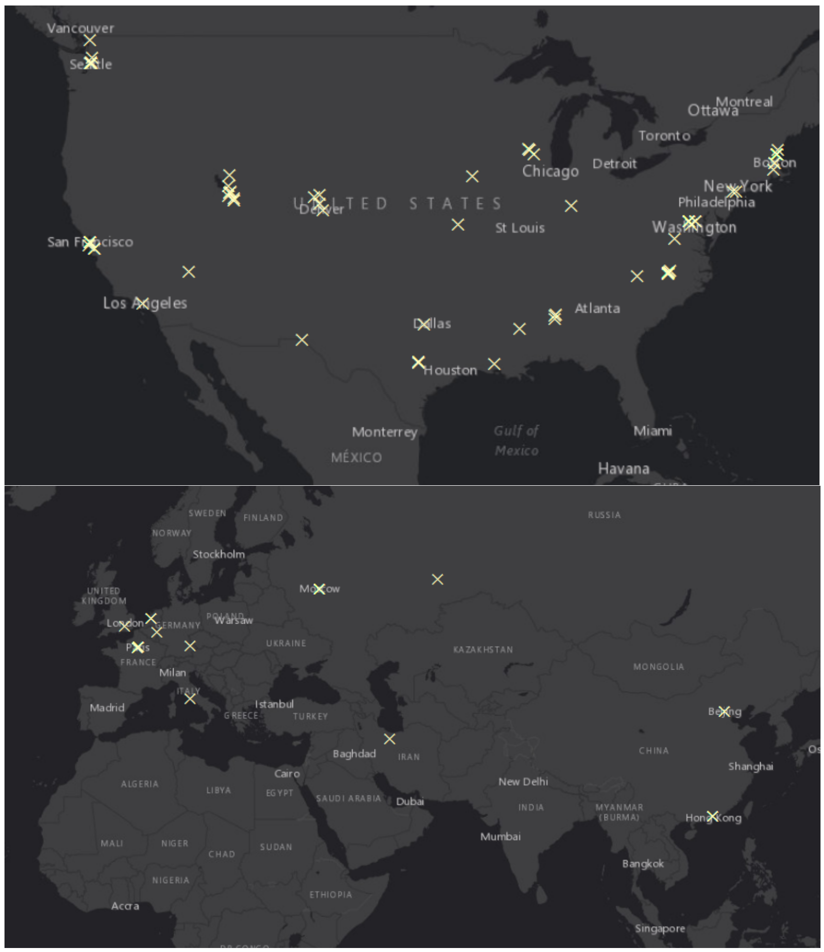


Figure 3-8. National Water Model App User Estimated Locations for the U.S. and Eurasia.

A time series is created for a selected river reach, reservoir point, or land grid, containing a subset of data from multiple netCDF files that vary in number depending on given parameters. This app solves the problem of relevancy that comes with a national-scale hydrologic forecast by allowing users to access data for a particular river reach, reservoir, or area as opposed to all the rivers, reservoir points, or land grid cells within the U.S. The NWM Viewer app successfully translates spatially based NWM forecasts into time-based files containing data for a specific feature. This process reduces the amount of storage required considerably. For example, the raw netCDF files require download of the complete files even if the user is only interested in a single stream reach. A complete set of netCDF files for the short-range configuration of the NWM amounts to 1GB while a complete forecast containing a time series for one reach is about 5.5MB. Table 3-2 shows a summary of the storage space used on the RENCI server with raw and processed files.

Besides downloading data, a user may only be interested in monitoring a particular reach without ever needing to retrieve any data. The NWM viewer produces a time-series graph for most of the variables available in the NWM allowing for instantaneous visual inspection of any of the 2.7 million streams, the reservoir points, or the grid cells within the US.

The NWM Viewer app produces a time series graph for most of the variables available in the NWM allowing for instantaneous visual inspection without the need to download any data.

The REST API is based on the OWDI concept of providing water data as a service. The NWM API facilitates automation of data retrieval and the integration of our apps with third party projects. Multiple web applications have been developed that use the NWM API and further analyze or provide additional information such as comparing NWM forecast to observed data or generating dynamic flood maps for specific areas within the United States. These apps serve as examples of how our NWM API can be used to support third-party applications.

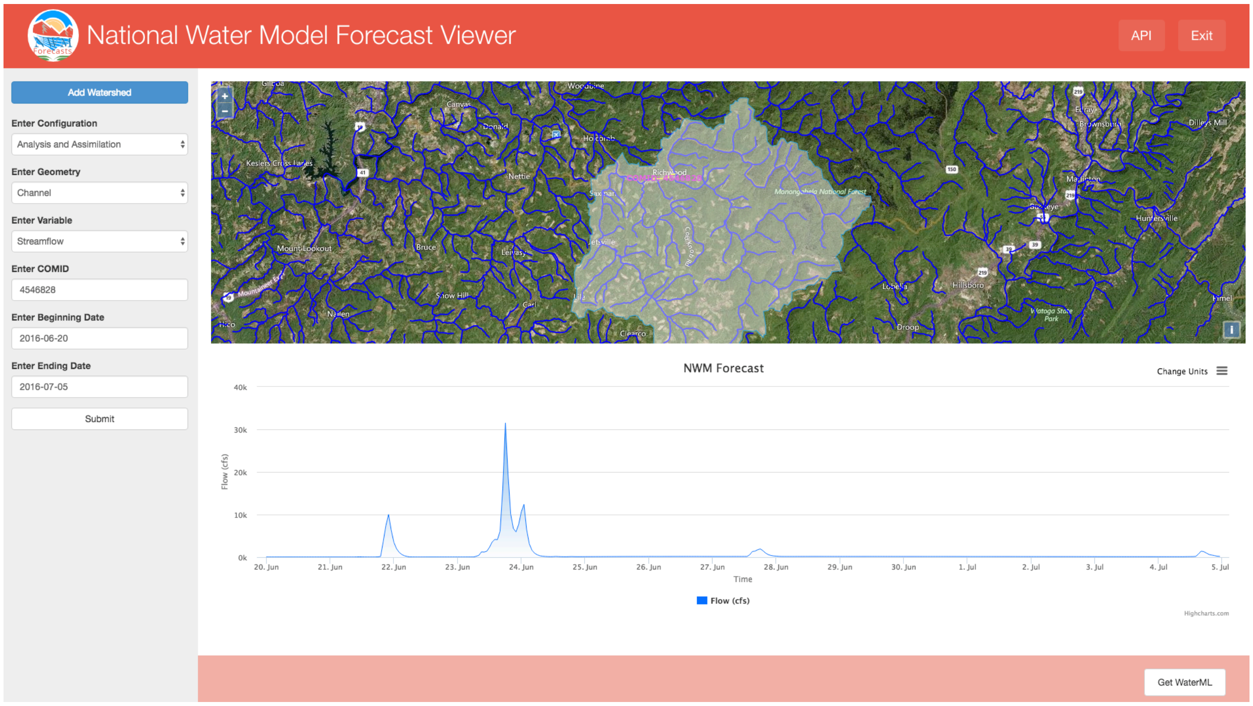


Figure 3-9. NWM Forecast Viewer App Displaying 2016 West Virginia Floods.

Table 3-2. Average NWM data storage increase per day.

Model Configuration Inputs and Outputs	Raw Files (GB/day)	Processed Files (GB/day)
Analysis and Assimilation	6.2	63
Short Range	94.6	802
Medium Range	81.7	340
Long Range Member 1	42.1	82
Long Range Member 2	42	82
Long Range Member 3	42.1	82
Long Range Member 4	42	82
Analysis and Assimilation Input	8.6	13
Short Range Input	131.3	191
Medium Range Input	92.9	128
Total	583.5	1865

I performed a test to determine how much data can be extracted simultaneously using our API. I ran 100 different identifiers, one at a time with a 1.5-second delay between requests. Ninety-nine of these requests succeeded. I also tried 40 different identifiers without any delay between requests with a success rate of 34 out of 40. Finally, I tried 100 identifiers with 20 concurrent requests at a time in five batches, but they all failed. From this test, I determined that our API could take a high number of requests as long as they do not come at exactly the same time. A one-second interval between requests should be sufficient for them to work. Smaller times may result in an overload of the API.

3.4 Updates to the NWM and Discussion

A new version of the United States National Water Model was released in May 2017. Some of the enhancements to the model include an improved forecasting cycling, which extended the short-range forecast from 15 to 18 hours, and increased the frequency of medium-range forecasts from once to four times per day; parameter updates, which reduced hydrologic biases and improved precipitation blend from various sources for the short-range forecast; upgraded netCDF output format from netCDF3 to netCDF4, and metadata and data structure for compatibility with netCDF file readers.

The main improvement in terms of storage for the latest release of the NWM was the change to integer values with a scaling factor to store variable values within the file structure as opposed to the use of floats. This reduced the size of individual netCDF files significantly, and most importantly, most netCDF readers automatically recognize this method because it is a common practice for storing scientific data.

Accordingly, the NWM Viewer app was updated to work with the newer version of the NWM. A document detailing some of the main areas of improvement for the NWM Viewer was drafted as part of this research. The suggestions presented in this document included the addition of grids to complement the hydrography and extract gridded data from the NWM such as forcing data; the addition of subsetting capabilities to allow users to extract data from multiple reaches within an area of interest as opposed to having to retrieve the data from each individual river reach or grid and having to know the unique identifiers for these elements ahead of time; the improvement of code and how data is served to increase the efficiency and speed of data delivery. These changes have slowly been incorporated into the NWM Viewer and the cyberinfrastructure behind it.

3.5 Conclusion

The US National Water Model is a hydrologic model that generates forecasts for multiple hydrologic variables across the continental U.S. It was first introduced in 2016, and then upgraded in 2017. Aside from model improvements, the main challenge for the NWM is data communication. The results from the NWM are exposed on the NOAA Operational Model Archive and Distribution System (NOMADS) for two days only, after which results are deleted. Water data initiatives such as the Open Water Data Initiatives in the U.S. seek to standardize water data use and sharing. The OWDI has four key functionalities for providing water data in an efficient way: Water Data Catalog, Water Data as a Service, Enriching Water Data, and Community for Water Data. Following OWDI principles we provided tools that allows users to access NWM data and provide a way to cover the main four key functionalities described in this initiative, while at the same time providing a cyberinfrastructure that allows for an extended period of data to be stored.

Hydrologic forecasts need to be translated into water intelligence that end users can easily understand. Data accessibility and visualization are important components necessary to achieve this goal. We created a set of web applications that facilitate interaction with the NWM. The NWM Forecast Viewer app provides an interactive way to access NWM forecast through the use of a dynamic web map and a form that can be used to query forecast and extract relevant data. The NWM Viewer app also has functionality to facilitate the extraction of large amounts of data both spatially and temporarily. The app allows for spatial queries through the use of polygons (usually watersheds) so that users can retrieve all the elements within the area of interest. Furthermore, the NWM Viewer app converts the raw NWM forecasts from spatially-based one-time-step netCDF files to a complete time series queried according to the element (river reach, reservoir point, or grid) or area of interest. These apps provide accessibility and an example of how NWM forecasts can be exposed so other third-party web applications, and analyses can easily consume NWM data. An example of how this can be done is through the use of the NWM Viewer REST API.

I developed a REST API that allows for programmatic NWM data retrieval. This API facilitates data retrieval in an efficient way without ever having to open the web interface. This is especially useful for derivative apps or analysis that requires constant data retrieval. The REST API has methods to retrieve the main forecast as well as other methods to retrieve metadata about the forecast and available data.

We developed a cyberinfrastructure to store and retrieve NWM forecasts for an extended period of time as opposed to two days, which is the time forecasts were stored in the NOMADS server. The current cyberinfrastructure is hosted at the Renaissance Computing Institute (RENCI) where the main forecasts were initially stored for two weeks, and then with the release of the newer version of the NWM in 2017 this was increased to a month. A corrected forecast that assimilates gage station data is also stored indefinitely.

Enabling the NWM by providing a better data communication scheme that includes a cyberinfrastructure, a web app, and a REST API served as a first step to working with other large-scale models and led to the development of a global streamflow forecasting application based on the Global Flood Awareness System (GloFAS).

4 GLOBAL STREAMFLOW PREDICTION USING THE GLOFAS-RAPID MODEL: ADDRESSING PRACTICAL CHALLENGES

4.1 Background

A global streamflow prediction system that provides water information relevant at a local scale has emerged as a priority to the hydrologic community over the past few years where a functional high-resolution hydrologic model is seen now as a “grand challenge” in hydrology (Wood et al., 2011). As a result, the creation of a high-resolution global forecast has been the focus of research and development by the hydrologic sciences community.

In recent years a number of large-scale models have surfaced (Alfieri et al., 2013; Lindström et al., 2010; NOAA Office of Water Prediction, 2017; Rodell et al., 2004). The development of such models has been possible because of the evolution of computational hydrology, which includes a number of internal advances, but also a vertical expansion where elements from meteorology have been integrated into improved land surface and routing models. As a result, we have increased our ability to predict floods by developing better hydro-meteorological models. Recent technologies have also made larger-scale models that provide value to local decision-makers possible. For example, probabilistic forecasts now offer an alternative to incorporating the model uncertainty by providing an ensemble forecast that includes multiple possible scenarios as predicted outcomes. Nevertheless, despite the rapidly improving models and computational infrastructure, major challenges remain. For example, the inherent uncertainty introduced by the model itself is often neglected, but can be significant (Butts et al., 2004). In addition, the amount of data produced by large-scale models presents new challenges in water data management. Furthermore, integrating and communicating model results has historically been a major challenge due to the evolving nature of hydrologic models (Beran & Piasecki, 2009). In general, communicating water data to different groups such as scientists, emergency responders, decision-makers, and the general public has been a major challenge because of the broad range of backgrounds, understanding of hydrologic model output, and needs (Souffront Alcantara, Crawley, et al., 2017). This challenge has begun to be answered with the adoption of standards, a push to create Earth Observation Systems (EOS) and model results that can be accessed as a service via the Internet, and the creation of tools that facilitate the interpretation and validation of data.

Other barriers to large-scale modeling are Big Data, Adoption or confidence in model output, and validation (see chapter 1 section 1.2). The realization that having a high-resolution model does not make it at once useful has prompted the need to address these significant challenges (Figure 4-1).

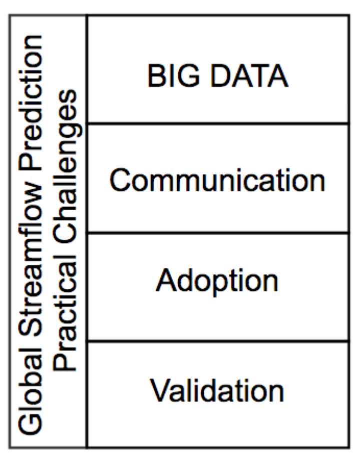


Figure 4-1. Global streamflow prediction practical challenges.

The Global Forecast Awareness System (GloFAS) developed by the European Centre for Medium-ranged Forecasts (ECMWF) is a hydrologic model that generates daily forecasts and makes them available through a web interface. However, the relatively coarse resolution makes the forecast useful only on large river basins. In an effort to create a higher density version of GloFAS that would include streams of smaller basins (Snow et al., 2016) developed a methodology to map the gridded runoff of GloFAS to basins and route them through a river network using the River Application for Parallel Computation of Discharge (RAPID) routing model at a national level covering the main hydrologic regions within the United States. This groundbreaking research laid the foundation for the development of a global streamflow prediction system that can provide actionable information on every mapped stream in the world using the same principle. Building on this work to overcome important barriers relative to the use of a global system I have developed a more complex workflow so that forecasts can be produced in a seamless and efficient way (see chapter 4).

A workflow to generate high-density forecasts as well as a cyberinfrastructure hosted on the Cloud are two important requirements that address the big data challenges for such a global streamflow prediction system. In a similar manner to the US National Water Model (Souffront Alcantara, Crawley, et al., 2017), data communication mechanisms are necessary to allow users to access data that is relevant and efficiently extract information. Furthermore, these tools need to be socialized, documented, and developed in a way that addresses the specific needs of different groups so that adoption of these new tools is seamless and easy to integrate with other existing tools.

The GloFAS-RAPID method developed by Snow et al. (2016) was used as a guideline to develop a workflow that generates daily forecasts for Africa, North America, South America, and South Asia in a cloud environment. In addition, other large-scale models were tested in combination with the RAPID method to create high-density vector-based results that help provide a solution to the data communication challenges faced by these models, while at the same time giving an alternative that demonstrate the flexibility of our method and the ability to adapt to other models that provide gridded hydrologic runoff. The primary challenges described above were also addressed using the resulting routed forecast.

4.2 Global Streamflow Prediction

The GloFAS-RAPID model was used to create a forecast prediction system covering Africa, the Americas, and South Asia. GloFAS-RAPID forecast are the result of routing GloFAS gridded runoff over a pre-computed stream network using the RAPID model. The runoff from the ERA-Interim historic simulation is also routed to provide context to the forecast in the form of return periods and awareness points if the flow from a specific river reach surpasses any of this return periods. The main components of GloFAS-RAPID are described below.

4.2.1 The Global Flood Awareness System

GloFAS is based on the ECMWF 51-member ensemble hydrologic forecast model that generates runoff on a global grid with a resolution of $16km^2$. The gridded runoff is routed over a stream network, but because of the relative coarse resolution, streamflow discharges are only available for large basins. A 52^{nd} forecast is generated at a resolution of $8km^2$. GloFAS was first released in 2011 by ECMWF and the European Commission's Joint Research Centre (JRC), and has been operational since July 2011. The GloFAS system is composed of an integrated hydrometeorological forecasting chain and of a monitoring system that analyzes daily results and shows forecast flood events on a dedicated web platform (Alfieri et al., 2013). This model uses real-time and historical observations in combination with a Data Assimilation System (DAS) and a Global Circulation Model (GCM).

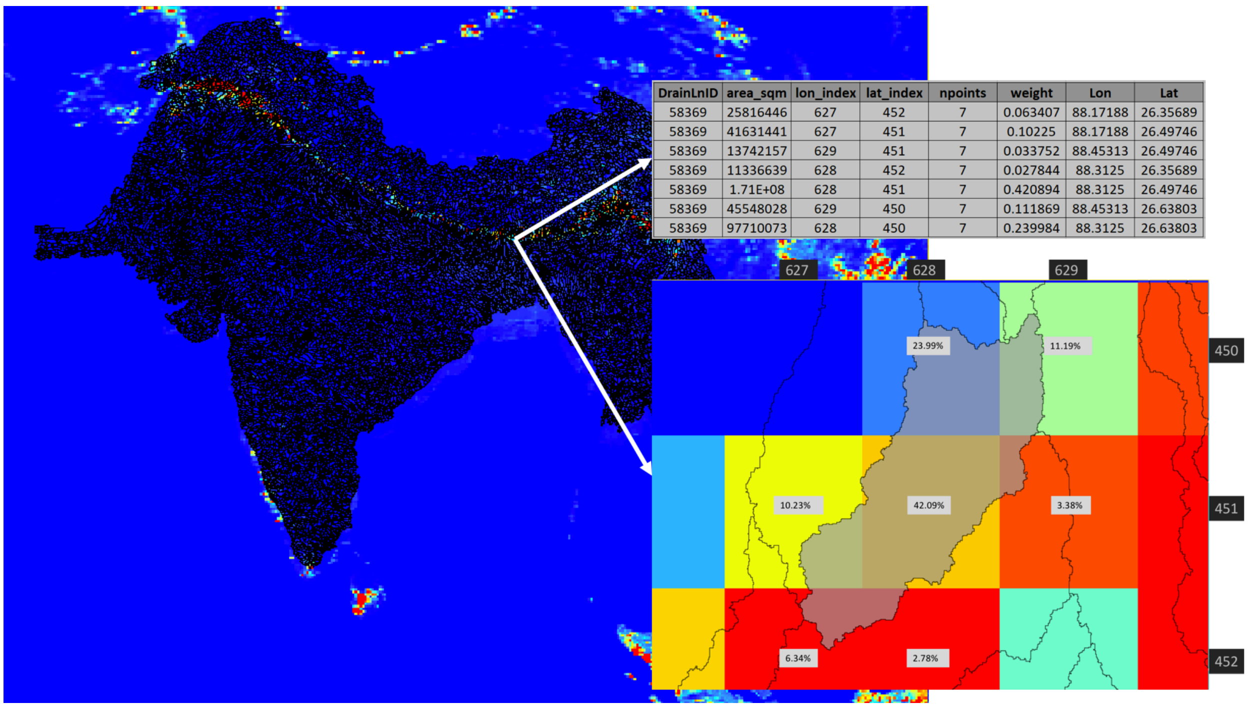
4.2.2 The ERA-Interim Historic Simulation

The ERA-Interim data is the result of a global atmospheric reanalysis produced also by ECMWF. This data covers from January 1980 to December 2014 (35 years) for the entire globe. One of the advantages of using reanalysis is that the data provides a global view that encompasses many essential climate variables in a physically consistent framework, with only a short time delay (Dee et al., 2011). This type of data becomes invaluable in areas where no actual observed data is available. A runoff derivative of this atmospheric reanalysis was produced on a $40km^2$ global grid using a land surface model simulation in HTESSEL. Return periods and awareness points are calculated from this dataset.

4.2.3 River Application for Parallel Computation of Discharge

RAPID is a numerical model that simulates the propagation of water flow waves in networks of rivers composed of tens to hundreds of thousands of river reaches (David, Famiglietti, Yang, Habets, & Maidment, 2016). The RAPID model is based on the Muskingum method, which has a time and a dimensionless parameter as its main variables. RAPID successfully created a way to efficiently adapt the Muskingum method to any river network.

We created a river network and weight tables for Africa, North America, South America, and South Asia following the methodology presented by Snow et al. (2016) (Figure 4-2). A river network for a specific area is created using the HydroSHEDS dataset, which is a hydrographic dataset based on elevation data from the Shuttle Radar Topography Mission (SRTM) that provides data at a global scale (Lehner, Verdin, & Jarvis, 2008). In addition to generating hydrography, this preprocessing also generates weight tables, and Muskingum/RAPID parameters for converting the gridded results from GloFAS to a forecast on every reach in the river network.



DrainLnID	area_sqm	lon_index	lat_index	npoints	weight	Lon	Lat
58369	25816446	627	452	7	0.063407	88.17188	26.35689
58369	41631441	627	451	7	0.10225	88.17188	26.49746
58369	13742157	629	451	7	0.033752	88.45313	26.49746
58369	11336639	628	452	7	0.027844	88.3125	26.35689
58369	1.71E+08	628	451	7	0.420894	88.3125	26.49746
58369	45548028	629	450	7	0.111869	88.45313	26.63803
58369	97710073	628	450	7	0.239984	88.3125	26.63803

Figure 4-2. River network and subbasin generation example for the South Asia region.

4.3 Addressing Hydroinformatic Challenges

Hydrologic modeling ultimately has the potential to save lives and reduce the cost of damages caused by water-related extreme events. However, there are many challenges that prevent hydrologic models from being used to their full potential. I have divided some of these main challenges into four main categories described below

4.3.1 Big Data

The amounts of data downloaded and produced by the GloFAS-RAPID model require a cybeinfrastructure that allows for daily computations. This cyberinfrastructure includes downloading tasks to obtain the raw GloFAS gridded runoff that serves as the main input to RAPID. The average size of the GloFAS gridded runoff for the entire globe is 6.5GB. In addition, the 35 years of simulated data used to calculate return periods and alerts amount to a static 80GB.

In the past, local servers have been used to develop proof of concept results for specific regions around the globe. However, such an approach would require not only a focus on the deployment and maintenance of a global streamflow prediction system, but also the administration and maintenance of a local cyberinfrastructure (see section 1.2.1).

The Microsoft Azure Cloud Service was selected to develop a cloud-based cyberinfrastructure. A 64-core 256GB virtual machine manages the main task of generating forecasts continually. A smaller second virtual machine was deployed to manage geospatial data visualization. The main virtual machine includes tasks to download the GloFAS gridded runoff, create inflow files for RAPID, and route the runoff. In addition, it includes workflows to clear the raw inputs once they are used and to move results to the Windows virtual machine and other data visualization servers (Figure 4-3). The ArcGIS suite is used in combination with PostGreSQL to store forecast results both as tables and as spatial services. An additional workflow is also run on this machine to create a time-enabled streamflow animation service by combining the tabular and spatial datasets (see the end of section 4.3.2). The resulting services can then be consumed from any web application or portal such as ArcGIS Online or Tethys Platform.

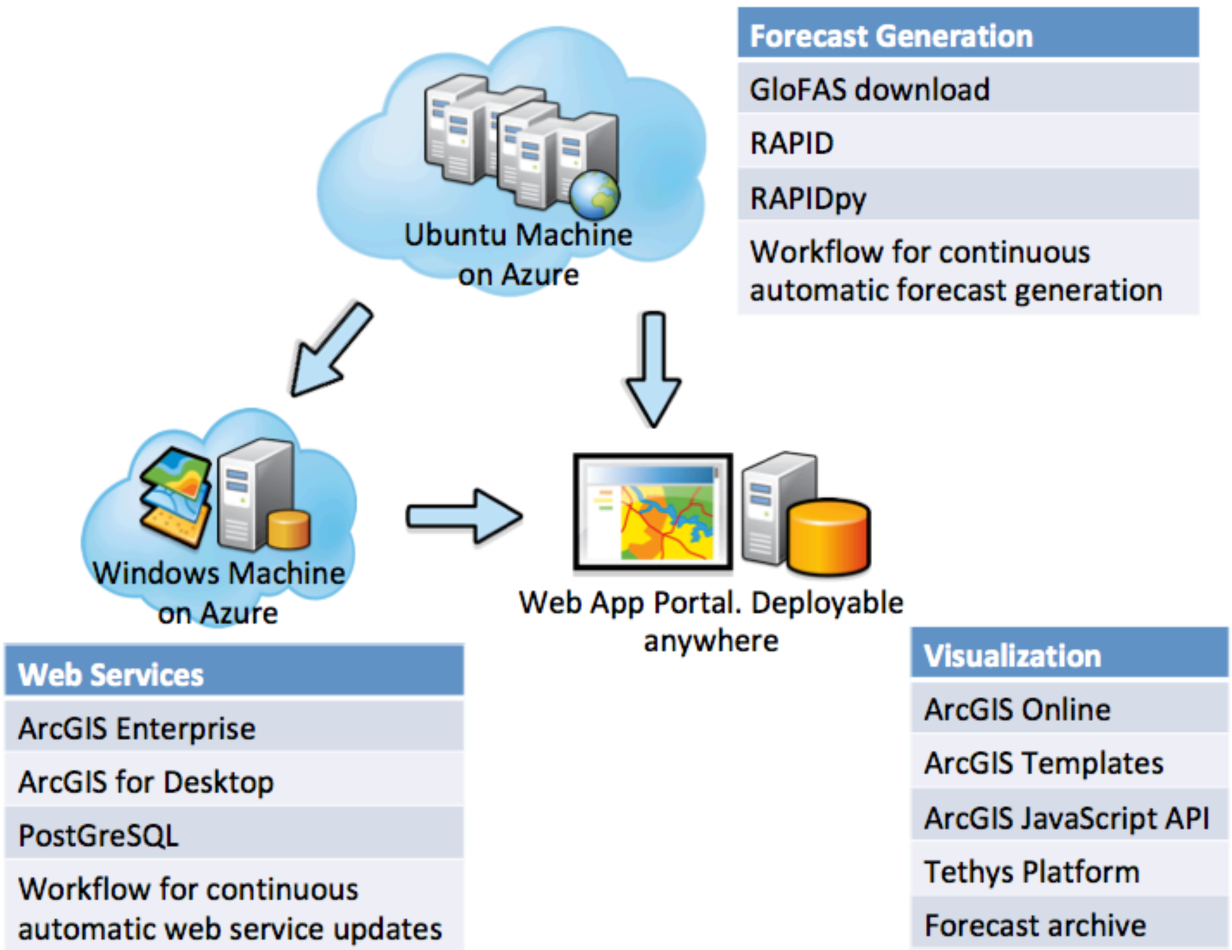


Figure 4-3. Global streamflow prediction cyberinfrastructure on MS Azure Cloud.

4.3.2 Data Communication

A data communication scheme that facilitates the extraction of relevant information is necessary for a large-scale ensemble forecast due to the large amount of data generated. In the absence of such a scheme, users would be required to download all the data and then try to filter it to extract relevant information. This type of approach is not efficient, and can result in potentially useful information not been used due to the difficulty of extracting it. Following this premise, a web application approach was chosen over a desktop approach given the increasing availability of the Internet around the world, and the ability to access data easily through a browser without requiring any other software. Relevant information can be visualized from its original location, and evaluated or download for further processing, thus facilitating and improving data consumption. Tethys Platform was used to create two web applications (see section 2.2.4). The Streamflow Prediction Tool (SPT), first developed by Snow et al. (2016), was created with the main goal to visualize GloFAS-RAPID results. This app has continued to improve in terms of front-end design and data visualization and access. The main addition to the SPT is the inclusion of a REST API to enable programmatic retrieval of data.

A REST API is a web service that can be used to access data without the need of a web interface. REST APIs use the http protocol to request data where parameters are passed through a URL. This development facilitates integration of our forecast results with third-party web apps, or any other workflow; the automation of forecast retrievals using any programing language; and the development of derived applications that consume these results through the API and further process them as opposed to incurring the same computational costs of generating their own forecast results. This last use, allows for the development of lightweight applications that provide complex results by relying on APIs from other apps.

The developed REST API for the SPT includes the following methods (see Appendix A for detailed descriptions of the methods and their arguments):

- **GetForecasts**: a method to extract forecast statistics from the 51 different ensembles available from the GloFAS-RAPID results. The available statistics are mean, max, min, and standard deviation. A high resolution 52nd ensemble result is also available.
- **GetEnsemble**: a method to extract individual ensembles. Each ensemble can be retrieved separately, or a range of ensemble can be selected.

- **GetHistoricData**: a method to extract the 35 years of historic simulated data for a specific river reach.
- **GetReturnPeriods**: a method to extract the 2, 10, and 20 year return periods for a specific river reach calculated using the historic simulation.
- **GetAvailableDates**: a method for extracting the available forecasted dates.
- **GetWarningPoints**: a method that returns the center of a river reach along with information about the forecasted flow and if it is greater than any of the calculated return periods for that reach.

The GloFAS-RAPID results can be provided as a service through the use of the methods available in the API.

The SPT tool is a generic tool that provides visualization for different regions that can be selected by the user. However, additional regions are not necessary or relevant in an operational environment targeted for a specific region or country. A second web application called the HydroViewer was developed to solve this issue by providing a customizable interface and presenting results only for a specific region or country.

The HydroViewer app is a lightweight web application. It was designed to visualize streamflow forecasts for specific regions using not only GloFAS-RAPID, but also different model alternatives, which can be added to the app in a relatively easy way. So far the app includes the aforementioned GloFAS-RAPID model, the South Asia Land Assimilation System (SALDAS), and the High Intensity Weather Assessment Toolkit (HIWAT) model for monitoring intense thunderstorms. This app relies on the use of the SPT REST API to retrieve and visualize water data and has the goal of providing a country, region, or watershed management authority the ability to have their own customized application that draws on the global streamflow prediction system results. The HydroViewer app was also designed to allow easy customization of functionality according to the needs of the stakeholders and decision-makers using it. It allows users to rebrand the interface of the web app and integrate it into their system, an idea that is important for the adoption of a global system (see section 4.3.3).

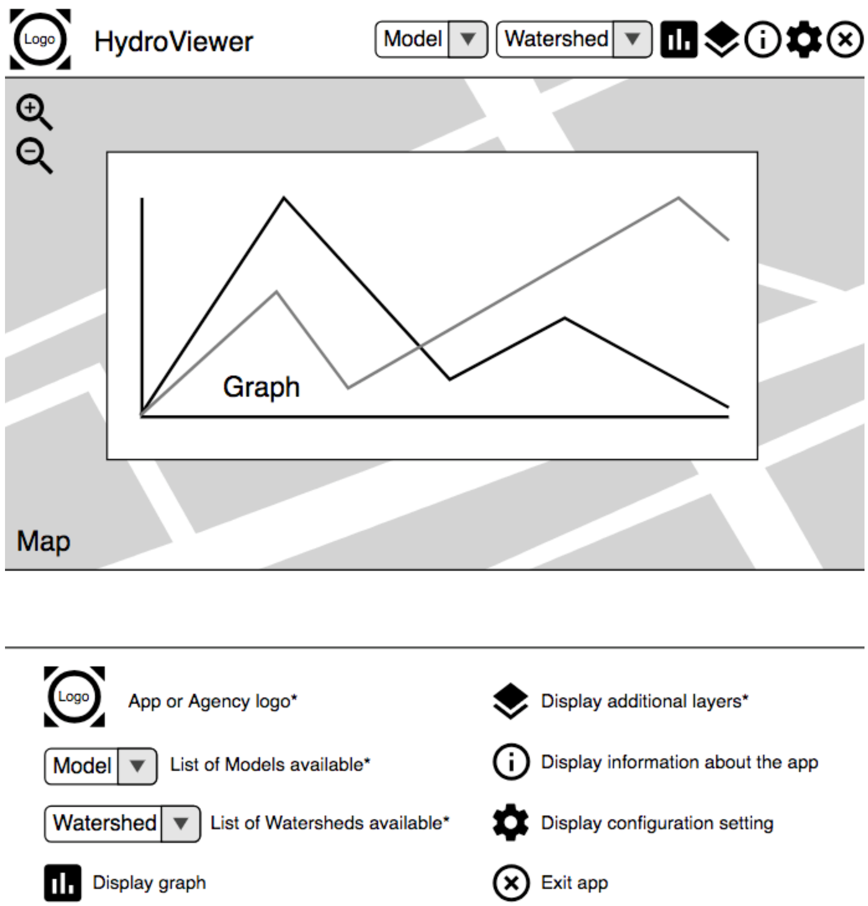


Figure 4-4. HydroViewer app design.

An improved way to display results visually has also been developed. A time-enabled geospatial web service has been created using ArcGIS technology. This service presents a dynamic river hydrography at different scale levels with a symbology based on predicted forecast volume and color-coded based on return period exceedance (Figure 4-5). This represents a significant improvement from the time-static service used in the SPT and the color-coded triangles used to identify reaches exceeding a specific return period.

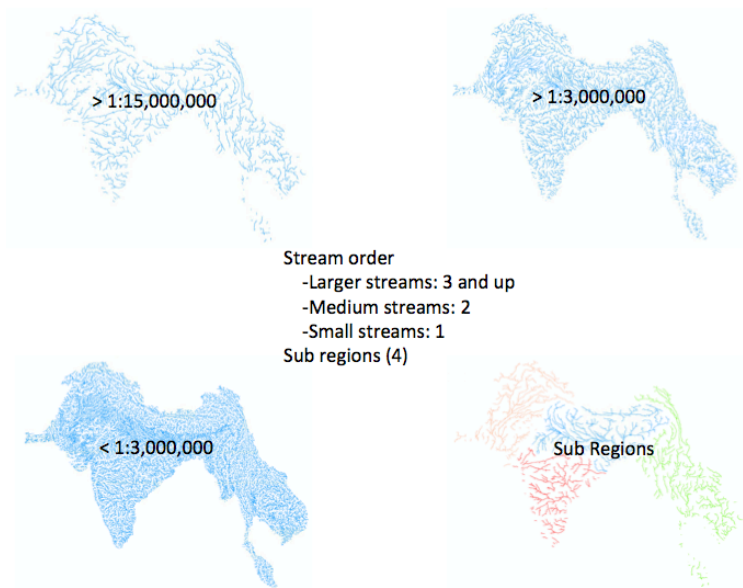


Figure 4-5. Example of scale levels and sub regions used to dissolve the South Asia Region. Other parameters used in the dissolving process are time, return period, and flow magnitude.

The streamflow animation workflow consists of a number of scripts created to calculate forecast statistics and convert the results from netCDF files to the simpler CSV format; dissolve spatial features based on forecast values such as stream order, region, and flow magnitude, and a Server Object Interceptor (SOI) add-on to redirect geospatial queries to the undissolved bottom-level stream feature that contains the complete set of results. The forecast tabular results from the

CSV file are combined with the geospatial features using a geospatially-enabled enterprise database, where the dissolving steps are performed daily as new forecasts become available (Figure 4-6). The resulting service is hosted using an ArcGIS Server where the SOI functionality is enabled to improve display and forecast access speed

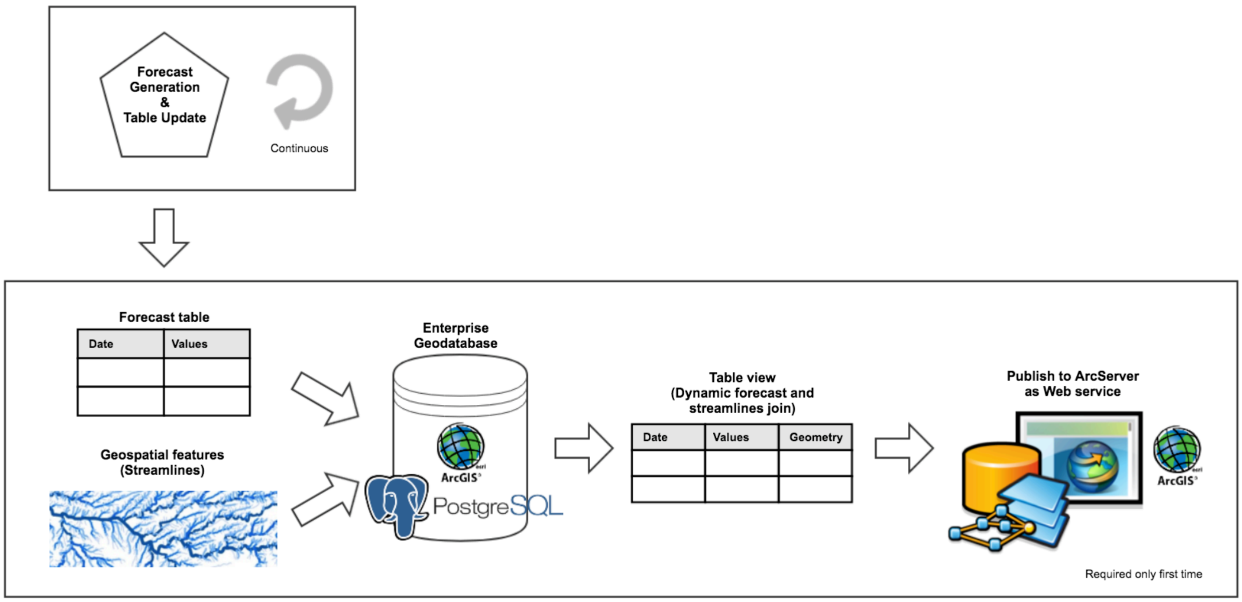


Figure 4-6. Streamflow animation main workflow.

4.3.3 Exposure and Adoption

A recurrent issue faced with hydrologic models is finding ways in which results can be useful and models can transition into operational systems that benefit stakeholders. This is mainly due to the basic science approach traditionally taken regarding hydrologic modeling, where the main goal is to better understand hydrologic processes and improve model performance. While this is an essential part of hydrologic modeling and the main drive to improve our models, there is also an applied science approach that has the potential to help communities facing extreme events.

Some of the other practical challenges faced by hydrologic predictions directly affect the degree to which modeled results can be adopted by local communities. For example, elements of the data communication challenge such as data accessibility and compatibility are a big factor in determining the ease with which results can be adopted or integrated into existing systems.

The benefits of adopting a forecasting system to aid in preparedness steps before extreme events have been discussed by many (Carsell, Pingel, & Ford, 2004; Godschalk et al., 2009; Pappenberger et al., 2015). On the other hand, the lack of actionable hydrological information worldwide has been identified by many international organizations and national governments (Goodall, Horsburgh, Whiteaker, Maidment, & Zaslavsky, 2008; Hilderbrand & Lead, 2014).

A successful global streamflow prediction systems needs to be backed by various organization that can provide the necessary exposure needed for local communities to become interested in the system. In addition, the system should be flexible enough to meet the different needs at the local level with alternatives in the way results are consumed and can be integrated. Finally, the model needs to perform well, and include mechanisms to improve performance in areas where it fails to do so, be it by data validation/calibration techniques or by assimilating local observed data.

4.3.4 Validation of a Global Streamflow Prediction System

A question that is commonly asked before a streamflow prediction system is adopted at any level is “how accurate is it?” A global streamflow prediction system presents a unique challenge regarding model validation. The geospatial extent of the results makes it nearly impossible to calibrate the model using traditional methods. For example, the total number of streams within our system is just over 200000 reaches.

A number of validation tests have been performed to:

1. Assert that our high-density routed forecasts yield, in essence, is the same result as the GloFAS and ERA Interim result at the same locations where both systems can be evaluated.
2. The choice of a given resolution to create the global network at does not change the results at a common watershed outlet point.
3. Model results are close to observed data at different locations around the world.

Data validation is essential for determining the value and limitations of the data and in determining systematic biases that could be accounted for even while waiting for improvements to model formulations. Jackson (2018) compiled a number of commonly used error metrics that can be used to compare hydrologic modeled data to observed data. Some of these metrics include the Root Mean Square Error (RMSE) and derivatives, Coefficient of Determination, Coefficient of Correlation, Anomaly Correlation Coefficient, Nash-Sutcliffe Efficiency (NSE), and the Spectral Angle. Most of these error metrics have been compiled in a Python package called HydroStats (https://github.com/BYU-Hydroinformatics/Hydrostats).

Using HydroStats, we compared our modeled results to observed data from several locations where we could obtain local observations. We describe the general results and process using examples from Colombia, and Nepal. We analyzed eight stations for the former, and 12 stations for the latter (Figure 4-7 and Figure 4-8).

In our analysis, we used a number of different metrics. We used the anomaly correlation coefficient, the root mean square error, the interquartile range normalized root mean square error, the Nash-Sutcliffe Efficiency metric, the Pearson correlation coefficient, the Spearman correlation coefficient, the spectral angle metric, the improved Kling-Gupta efficiency, and the refined index of agreement. We chose to use this suite of metrics to give a more complete picture of how well the simulated data correlates to the observed data (Krause, Boyle, & Bäse, 2005).

I performed a comparison between our high-density routed results with the gridded result from GloFAS at selected locations. Data was collected from six GloFAS locations found in Nepal including Chatara, Chepang, Chisapani, Devghat, Kusum, and Parigaun. Our assumption is that if our result had similar trends and values to those of the native GloFAS then it meant that our RAPID processing did not introduce significant errors by converting the gridded GloFAS results to a higher density set of basins and routing through the river network. In addition, we also assumed that the results of this comparison could be applied to other areas outside of the locations used for the comparison. Data was collected every day for 9 weeks and summarized weekly. We used the mean flow of both datasets to perform the comparison as the best representation from all the ensembles.

I routed the ERA-Interim historic simulation using RAPID at three different resolutions to determine the effects of varying resolutions on flow at a given location. Five watersheds in the United States where used: the Meramec River watershed in Missouri, the East Delaware River watershed in New York, the Alsea River watershed in Oregon, the White River watershed in Arizona, and the North Fork Clearwater River watershed in Idaho. These watersheds were selected because of their unimpaired flows and relative unaltered state.

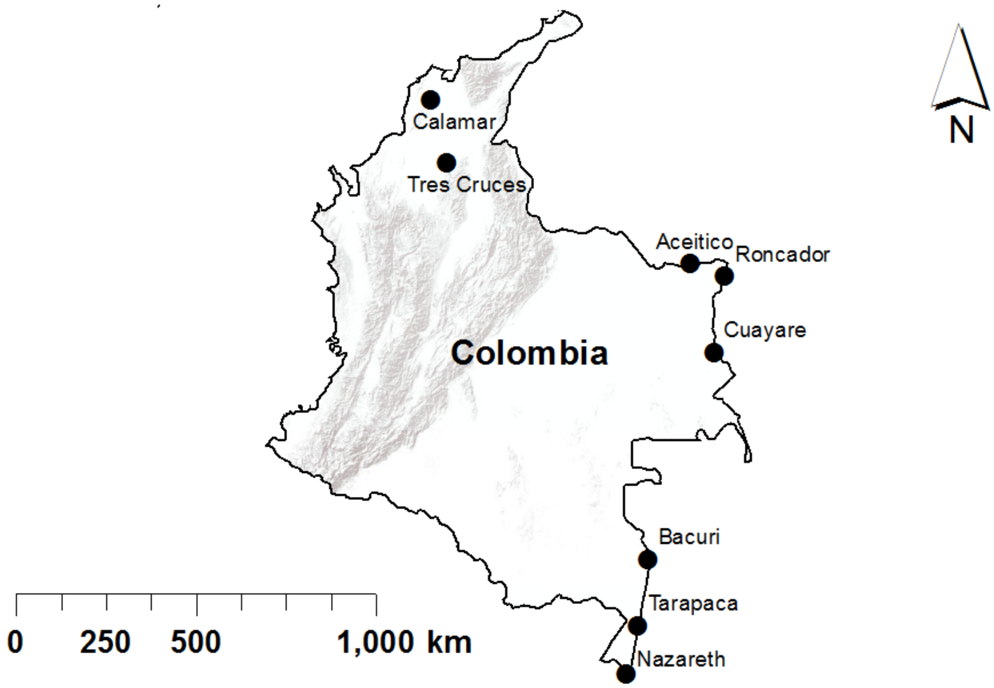


Figure 4-7. Station locations in Colombia.

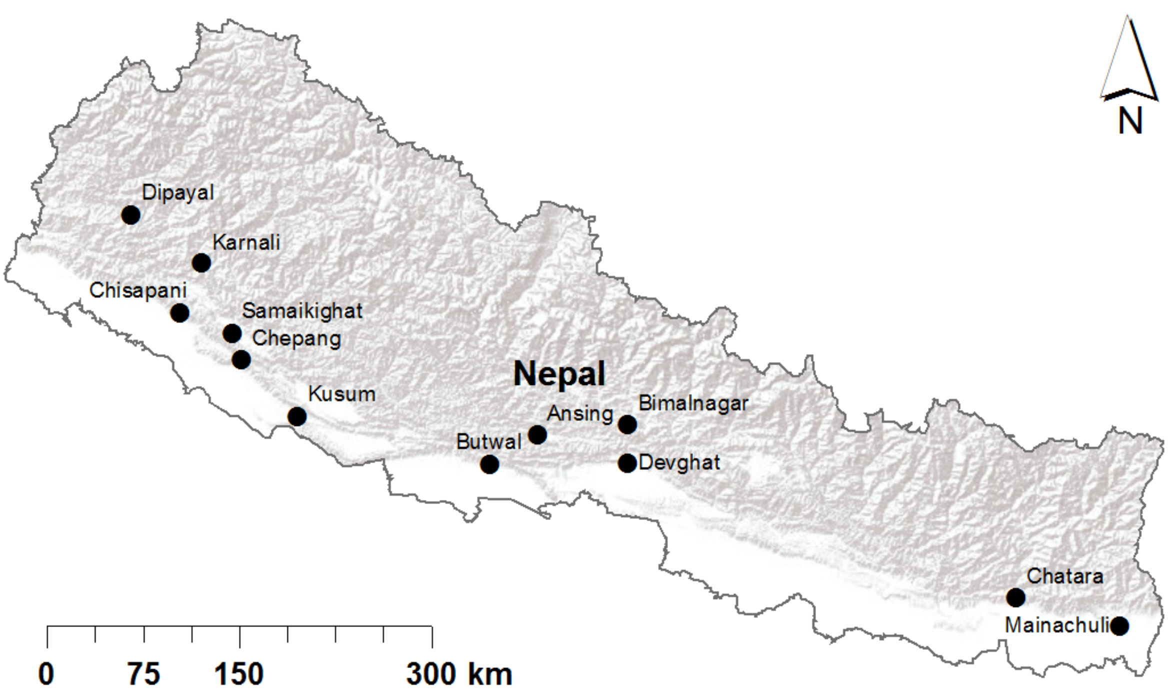


Figure 4-8. Station locations in Nepal.

4.4 Results

The computational resources necessary to produce a high-density global streamflow prediction forecast represents a major challenge. A cloud cyberinfrastructure has been created using the MS Azure cloud platform. Forecasts using the GloFAS-RAPID model are generated daily for Africa, North America, South America, and South Asia.

The new Streamflow Prediction Tool provides visualization of our GloFAS-RAPID results as well as data retrieval in CSV and WaterML formats (Figure 4-9). Previous forecast results are available in the app for one week, after which they are removed and archived in a dedicated server. Forecasts for a specific reach can be accessed by selecting the desired reach, which performs the appropriate data mining of the large datasets and delivers the single result of the selected stream. A pop-up window displays the hydrograph for that reach, which includes common interactions like zoom in our out, and data download as an image of CSV file. The hydrograph identifies the two, ten, and twenty year return periods to provide context of when a given forecast might be extreme relative to past flows. The 51-member ensemble forecast for the reach is displayed using statistics that show the mean, min, max, and standard deviation. A percent exceedance table also displays the probability of a specific flow value surpassing a return period based on the forecast (Figure 4-10). Other information available for each reach, are the 35-year historical simulation from the ERA Interim, a flow duration curve and seasonal averages derived from the historical simulation. Both the forecast and simulated historical data can be downloaded from the interface (Figure 4-12).

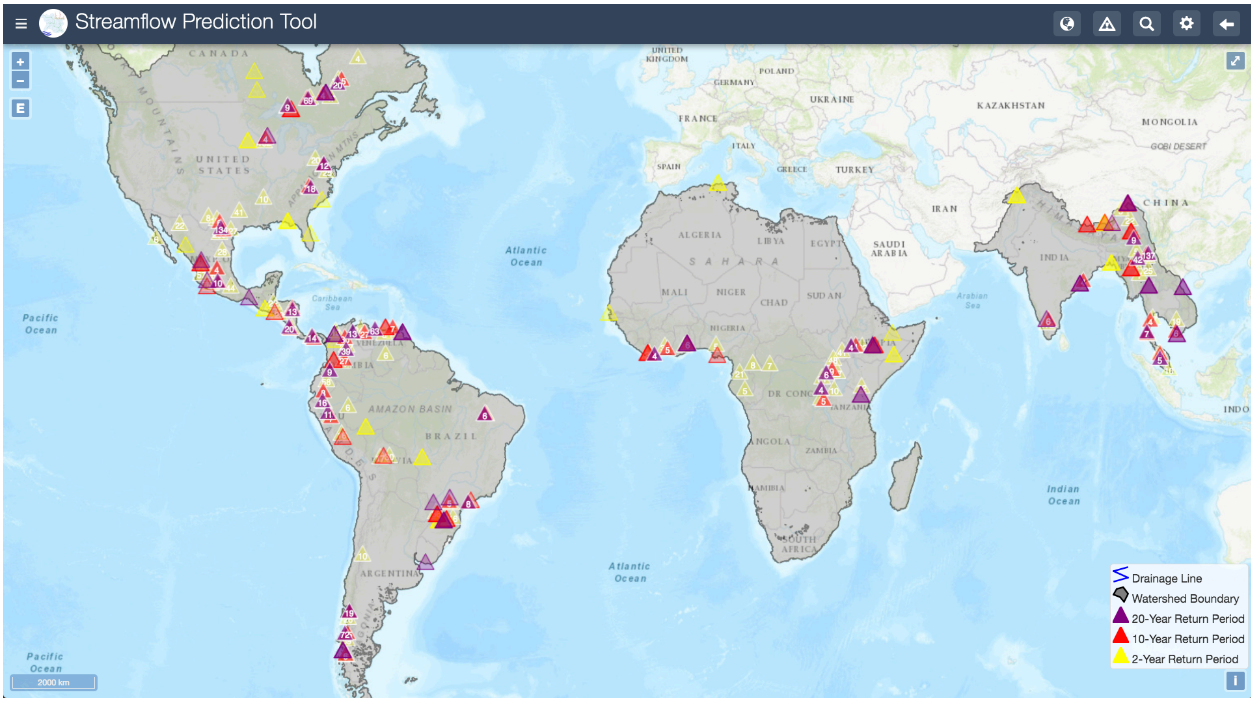


Figure 4-9. Streamflow Prediction Tool Interface.

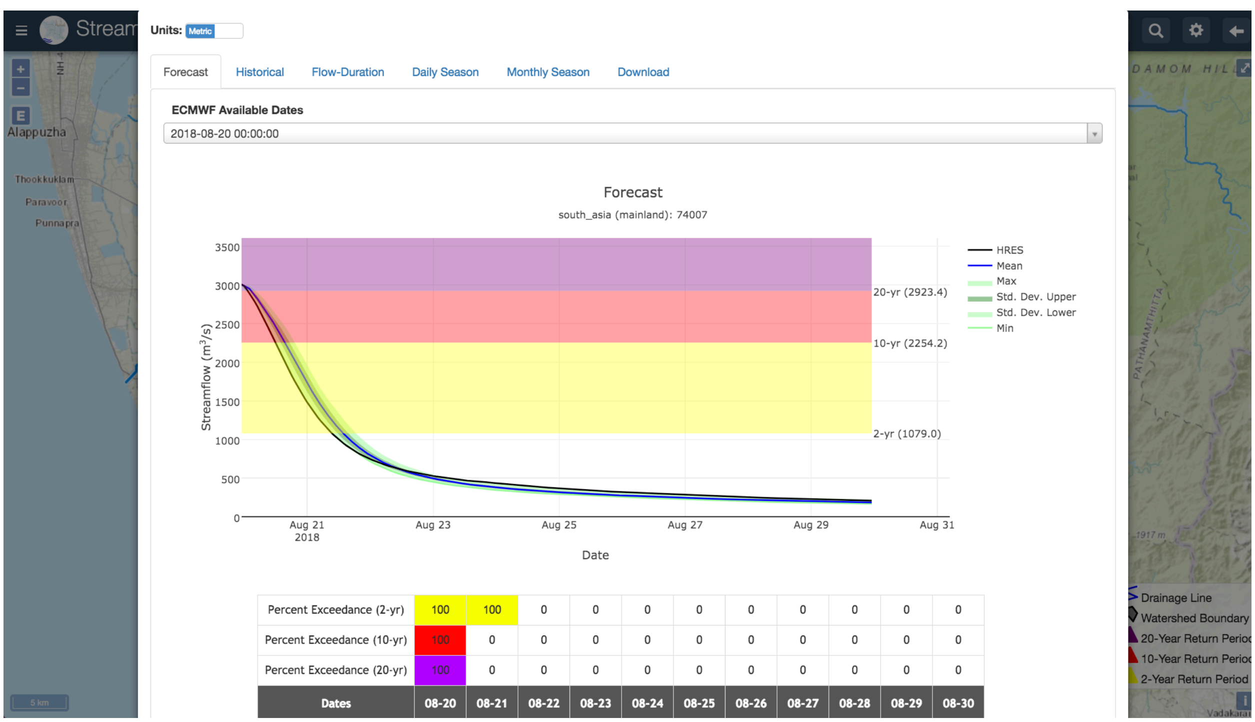


Figure 4-10. Streamflow Prediction Tool hydrograph and percent table.

The SPT REST API includes methods to programmatically retrieve forecast statistics such as mean, min, and max, as well as individual forecast ensembles. It also provides methods to retrieve the return periods, historic simulation, and warning levels for a specific river reach. Figure 4-11 shows an example of the REST API response in WaterML format.


```
This XML file does not appear to have any style information associated with it. The document tree is shown below.

▼<timeSeriesResponse xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.cuahsi.org/waterML/1.1/">
 ▶<queryInfo>...</queryInfo>
 ▼<timeSeries>
   ▶<sourceInfo xsi:type="SiteInfoType">...</sourceInfo>
   ▶<variable>...</variable>
   ▼<values>
      <value censorCode="nc" dateTime="2018-10-15T00:00:00" timeOffset="0" dateTimeUTC="2018-10-15T00:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">56.372234</value>
      <value censorCode="nc" dateTime="2018-10-15T03:00:00" timeOffset="0" dateTimeUTC="2018-10-15T03:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">60.402584</value>
      <value censorCode="nc" dateTime="2018-10-15T06:00:00" timeOffset="0" dateTimeUTC="2018-10-15T06:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">69.562325</value>
      <value censorCode="nc" dateTime="2018-10-15T09:00:00" timeOffset="0" dateTimeUTC="2018-10-15T09:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">78.93196</value>
      <value censorCode="nc" dateTime="2018-10-15T12:00:00" timeOffset="0" dateTimeUTC="2018-10-15T12:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">90.39336</value>
      <value censorCode="nc" dateTime="2018-10-15T15:00:00" timeOffset="0" dateTimeUTC="2018-10-15T15:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">108.03064</value>
      <value censorCode="nc" dateTime="2018-10-15T18:00:00" timeOffset="0" dateTimeUTC="2018-10-15T18:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">128.47995</value>
      <value censorCode="nc" dateTime="2018-10-15T21:00:00" timeOffset="0" dateTimeUTC="2018-10-15T21:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">149.96243</value>
      <value censorCode="nc" dateTime="2018-10-16T00:00:00" timeOffset="0" dateTimeUTC="2018-10-16T00:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">171.1761</value>
      <value censorCode="nc" dateTime="2018-10-16T03:00:00" timeOffset="0" dateTimeUTC="2018-10-16T03:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">191.38385</value>
      <value censorCode="nc" dateTime="2018-10-16T06:00:00" timeOffset="0" dateTimeUTC="2018-10-16T06:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">210.89426</value>
      <value censorCode="nc" dateTime="2018-10-16T09:00:00" timeOffset="0" dateTimeUTC="2018-10-16T09:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">230.17268</value>
      <value censorCode="nc" dateTime="2018-10-16T12:00:00" timeOffset="0" dateTimeUTC="2018-10-16T12:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">250.0806</value>
      <value censorCode="nc" dateTime="2018-10-16T15:00:00" timeOffset="0" dateTimeUTC="2018-10-16T15:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">272.1158</value>
      <value censorCode="nc" dateTime="2018-10-16T18:00:00" timeOffset="0" dateTimeUTC="2018-10-16T18:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">297.37216</value>
      <value censorCode="nc" dateTime="2018-10-16T21:00:00" timeOffset="0" dateTimeUTC="2018-10-16T21:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">322.91232</value>
      <value censorCode="nc" dateTime="2018-10-17T00:00:00" timeOffset="0" dateTimeUTC="2018-10-17T00:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">348.03592</value>
      <value censorCode="nc" dateTime="2018-10-17T03:00:00" timeOffset="0" dateTimeUTC="2018-10-17T03:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">372.58786</value>
      <value censorCode="nc" dateTime="2018-10-17T06:00:00" timeOffset="0" dateTimeUTC="2018-10-17T06:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">396.66043</value>
      <value censorCode="nc" dateTime="2018-10-17T09:00:00" timeOffset="0" dateTimeUTC="2018-10-17T09:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">420.24554</value>
      <value censorCode="nc" dateTime="2018-10-17T12:00:00" timeOffset="0" dateTimeUTC="2018-10-17T12:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">443.87552</value>
      <value censorCode="nc" dateTime="2018-10-17T15:00:00" timeOffset="0" dateTimeUTC="2018-10-17T15:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">469.67795</value>
      <value censorCode="nc" dateTime="2018-10-17T18:00:00" timeOffset="0" dateTimeUTC="2018-10-17T18:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">496.58203</value>
      <value censorCode="nc" dateTime="2018-10-17T21:00:00" timeOffset="0" dateTimeUTC="2018-10-17T21:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">520.9754</value>
      <value censorCode="nc" dateTime="2018-10-18T00:00:00" timeOffset="0" dateTimeUTC="2018-10-18T00:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">544.37866</value>
      <value censorCode="nc" dateTime="2018-10-18T03:00:00" timeOffset="0" dateTimeUTC="2018-10-18T03:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">567.2402</value>
      <value censorCode="nc" dateTime="2018-10-18T06:00:00" timeOffset="0" dateTimeUTC="2018-10-18T06:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">589.15076</value>
      <value censorCode="nc" dateTime="2018-10-18T09:00:00" timeOffset="0" dateTimeUTC="2018-10-18T09:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">610.20386</value>
      <value censorCode="nc" dateTime="2018-10-18T12:00:00" timeOffset="0" dateTimeUTC="2018-10-18T12:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">630.4986</value>
      <value censorCode="nc" dateTime="2018-10-18T15:00:00" timeOffset="0" dateTimeUTC="2018-10-18T15:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">651.709</value>
      <value censorCode="nc" dateTime="2018-10-18T18:00:00" timeOffset="0" dateTimeUTC="2018-10-18T18:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">673.95966</value>
      <value censorCode="nc" dateTime="2018-10-18T21:00:00" timeOffset="0" dateTimeUTC="2018-10-18T21:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">695.73413</value>
      <value censorCode="nc" dateTime="2018-10-19T00:00:00" timeOffset="0" dateTimeUTC="2018-10-19T00:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">717.28796</value>
      <value censorCode="nc" dateTime="2018-10-19T03:00:00" timeOffset="0" dateTimeUTC="2018-10-19T03:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">738.0066</value>
      <value censorCode="nc" dateTime="2018-10-19T06:00:00" timeOffset="0" dateTimeUTC="2018-10-19T06:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">757.94366</value>
      <value censorCode="nc" dateTime="2018-10-19T09:00:00" timeOffset="0" dateTimeUTC="2018-10-19T09:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">777.36176</value>
      <value censorCode="nc" dateTime="2018-10-19T12:00:00" timeOffset="0" dateTimeUTC="2018-10-19T12:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">796.5868</value>
      <value censorCode="nc" dateTime="2018-10-19T15:00:00" timeOffset="0" dateTimeUTC="2018-10-19T15:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">815.8663</value>
      <value censorCode="nc" dateTime="2018-10-19T18:00:00" timeOffset="0" dateTimeUTC="2018-10-19T18:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">835.49146</value>
      <value censorCode="nc" dateTime="2018-10-19T21:00:00" timeOffset="0" dateTimeUTC="2018-10-19T21:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">855.6351</value>
      <value censorCode="nc" dateTime="2018-10-20T00:00:00" timeOffset="0" dateTimeUTC="2018-10-20T00:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">876.4694</value>
      <value censorCode="nc" dateTime="2018-10-20T03:00:00" timeOffset="0" dateTimeUTC="2018-10-20T03:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">898.0543</value>
      <value censorCode="nc" dateTime="2018-10-20T06:00:00" timeOffset="0" dateTimeUTC="2018-10-20T06:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">920.4696</value>
      <value censorCode="nc" dateTime="2018-10-20T09:00:00" timeOffset="0" dateTimeUTC="2018-10-20T09:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">943.7679</value>
      <value censorCode="nc" dateTime="2018-10-20T12:00:00" timeOffset="0" dateTimeUTC="2018-10-20T12:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">967.9515</value>
      <value censorCode="nc" dateTime="2018-10-20T15:00:00" timeOffset="0" dateTimeUTC="2018-10-20T15:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">992.9593</value>
      <value censorCode="nc" dateTime="2018-10-20T18:00:00" timeOffset="0" dateTimeUTC="2018-10-20T18:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1018.7063</value>
      <value censorCode="nc" dateTime="2018-10-20T21:00:00" timeOffset="0" dateTimeUTC="2018-10-20T21:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1045.0374</value>
      <value censorCode="nc" dateTime="2018-10-21T00:00:00" timeOffset="0" dateTimeUTC="2018-10-21T00:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1071.8223</value>
      <value censorCode="nc" dateTime="2018-10-21T06:00:00" timeOffset="0" dateTimeUTC="2018-10-21T06:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1097.7643</value>
      <value censorCode="nc" dateTime="2018-10-21T12:00:00" timeOffset="0" dateTimeUTC="2018-10-21T12:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1151.9052</value>
      <value censorCode="nc" dateTime="2018-10-21T18:00:00" timeOffset="0" dateTimeUTC="2018-10-21T18:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1205.1272</value>
      <value censorCode="nc" dateTime="2018-10-22T00:00:00" timeOffset="0" dateTimeUTC="2018-10-22T00:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1256.4407</value>
      <value censorCode="nc" dateTime="2018-10-22T06:00:00" timeOffset="0" dateTimeUTC="2018-10-22T06:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1305.3385</value>
      <value censorCode="nc" dateTime="2018-10-22T12:00:00" timeOffset="0" dateTimeUTC="2018-10-22T12:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1351.6895</value>
      <value censorCode="nc" dateTime="2018-10-22T18:00:00" timeOffset="0" dateTimeUTC="2018-10-22T18:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1395.6526</value>
      <value censorCode="nc" dateTime="2018-10-23T00:00:00" timeOffset="0" dateTimeUTC="2018-10-23T00:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1437.5516</value>
      <value censorCode="nc" dateTime="2018-10-23T06:00:00" timeOffset="0" dateTimeUTC="2018-10-23T06:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1477.8319</value>
      <value censorCode="nc" dateTime="2018-10-23T12:00:00" timeOffset="0" dateTimeUTC="2018-10-23T12:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1516.8706</value>
      <value censorCode="nc" dateTime="2018-10-23T18:00:00" timeOffset="0" dateTimeUTC="2018-10-23T18:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1555.1042</value>
      <value censorCode="nc" dateTime="2018-10-24T00:00:00" timeOffset="0" dateTimeUTC="2018-10-24T00:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1593.137</value>
      <value censorCode="nc" dateTime="2018-10-24T06:00:00" timeOffset="0" dateTimeUTC="2018-10-24T06:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1630.1892</value>
      <value censorCode="nc" dateTime="2018-10-24T12:00:00" timeOffset="0" dateTimeUTC="2018-10-24T12:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1665.3821</value>
      <value censorCode="nc" dateTime="2018-10-24T18:00:00" timeOffset="0" dateTimeUTC="2018-10-24T18:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1700.663</value>
      <value censorCode="nc" dateTime="2018-10-25T00:00:00" timeOffset="0" dateTimeUTC="2018-10-25T00:00:00" methodCode="1" sourceCode="1" qualityControlLevelCode="1">1736.3276</value>
    ▶<qualityControlLevel qualityConsourcetrolLevelID="1">...</qualityControlLevel>
    ▶<method methodID="1">...</method>
    ▶<source sourceID="1">...</source>
    ▶<censorCode>...</censorCode>
    </values>
  </timeSeries>
</timeSeriesResponse>
```


Figure 4-11. Example of an API response as seen from a web browser.

The HydroViewer app is a lightweight web application that allows users to display only relevant data and customize the web app according to stakeholder needs. This app makes use of web services to display results as opposed to replicating the hardware, software, and modeling expertise to generate their own hydrological forecasts. The app uses the SPT REST API to display forecast results and geospatial web services to display hydrographic data. The interface of the app can be customized to display the colors and logo of the organization it is deployed for, thus branding it as their own and lending confidence to those that use it. In addition, the HydroViewer app was designed with the principle of visualizing hydrologic results from different models, not only the GloFAS-RAPID model.

Additional models can be added to the app in a relatively easy way as long as the results of the model are provided using a web service. Models such as the South Asia Land Assimilation System (SALDAS), and the High Intensity Weather Assessment Toolkit (HIWAT) model for monitoring intense thunderstorms have been already added to the app.

Customizations for different organization also include the addition of hydrographs displaying observed data, data comparison displays, or the inclusion of other important geospatial data such as districts or country boundaries.

Instances of the HydroViewer have been deployed for the following countries: Argentina, Bangladesh, Brazil, Colombia, La Hispaniola (The Dominican Republic, and Haiti), Nepal, Peru, and Tanzania.

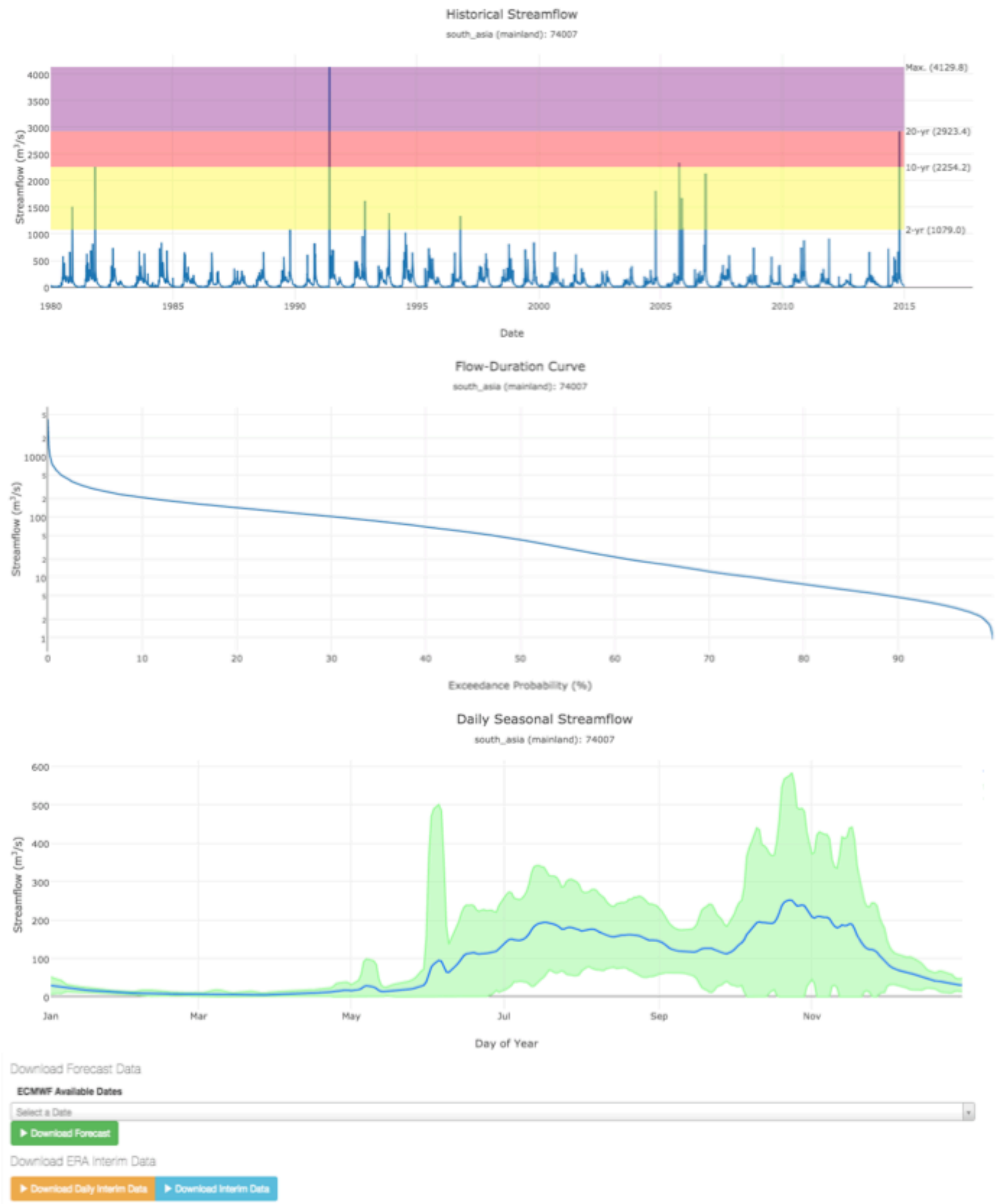


Figure 4-12. Streamflow Prediction Tool addition tabs displaying historic simulation, flow duration curve, seasonal averages, and download buttons.

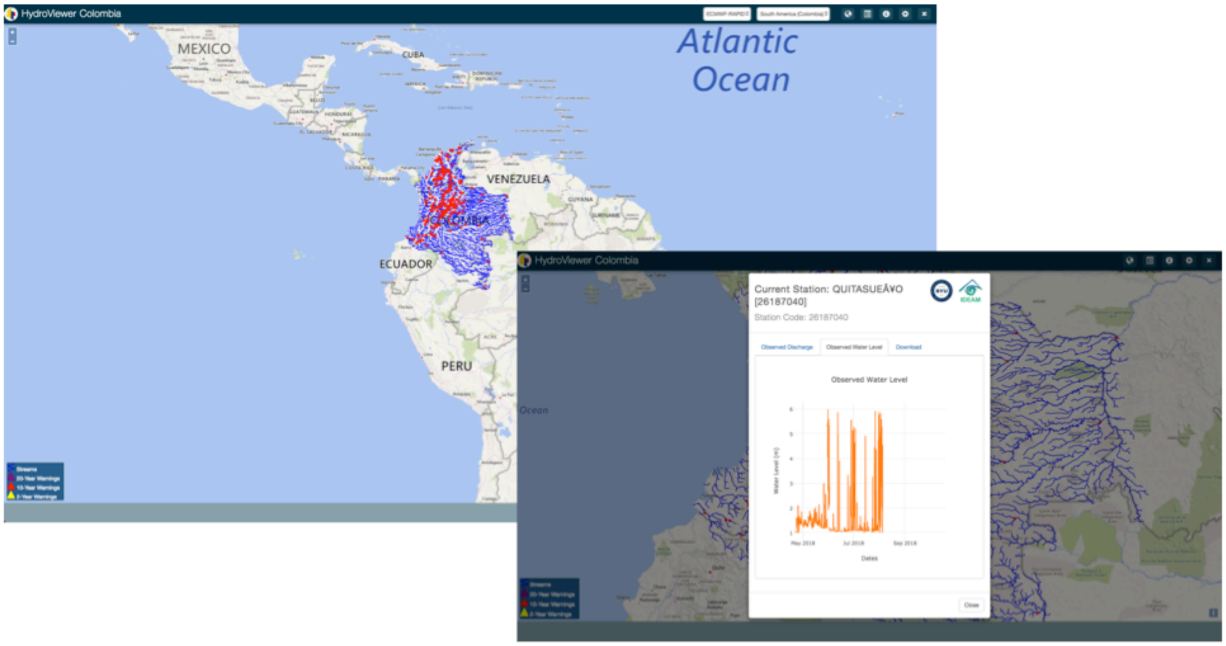


Figure 4-13. HydroViewer Colombia displaying observed data customization.

The developed streamflow animation service successfully addresses one of the main interpretation issues of the SPT. The SPT displays warning levels based on return periods, and using color-coded triangles at the center of the river reach. However, because the geospatial service is not time enabled these triangles are displayed all at the same time. For example, if the forecasted flow exceeds the two-year return period on the second day of the forecast and then exceeds the twenty-year return period on the sixth day of the forecast, both warning levels are displayed as overlapping triangles color-coded yellow and purple to identify the two-year and twenty-year return periods, respectively.

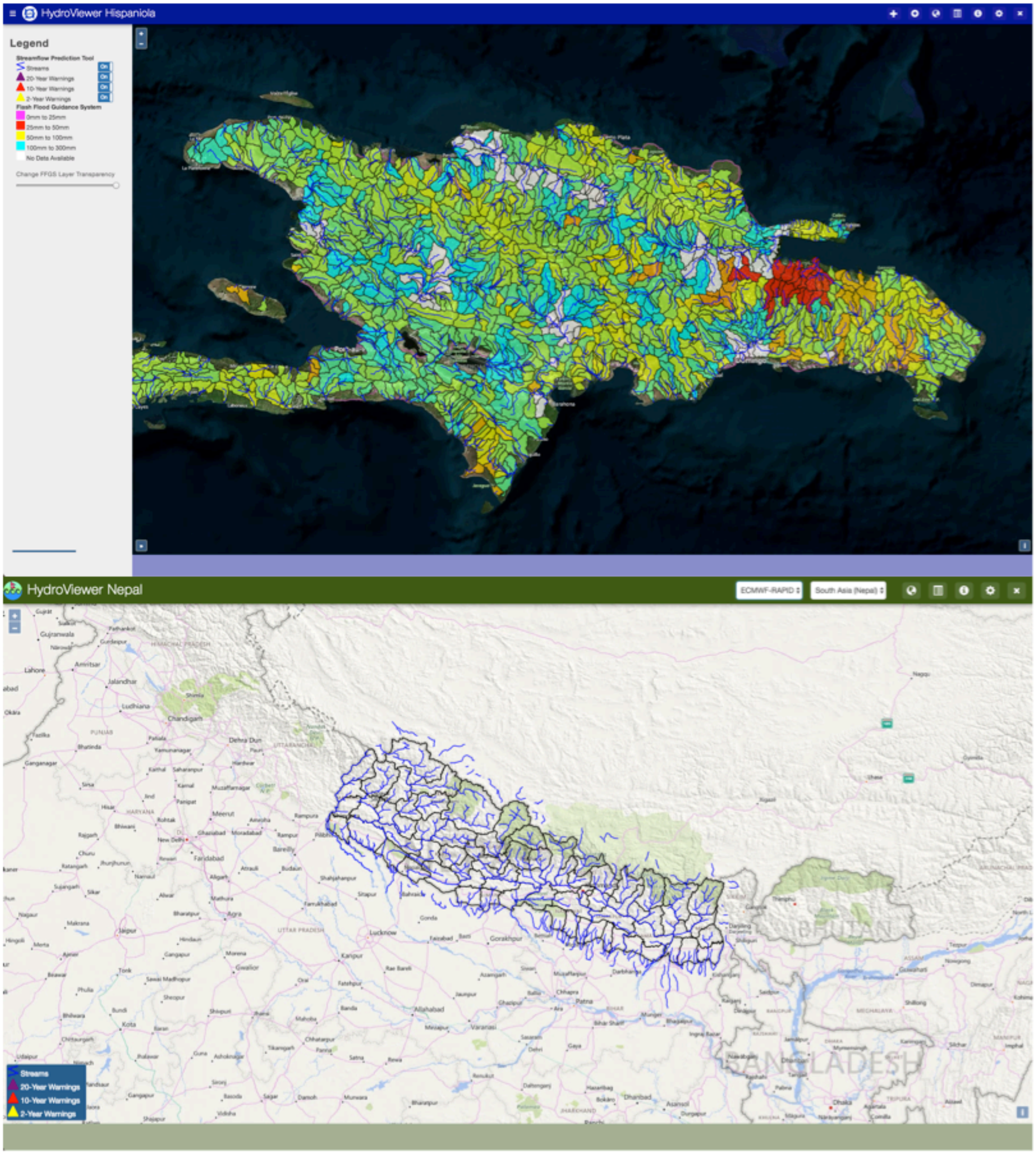


Figure 4-14. HydroViewer La Hispaniola and Nepal displaying customizations.

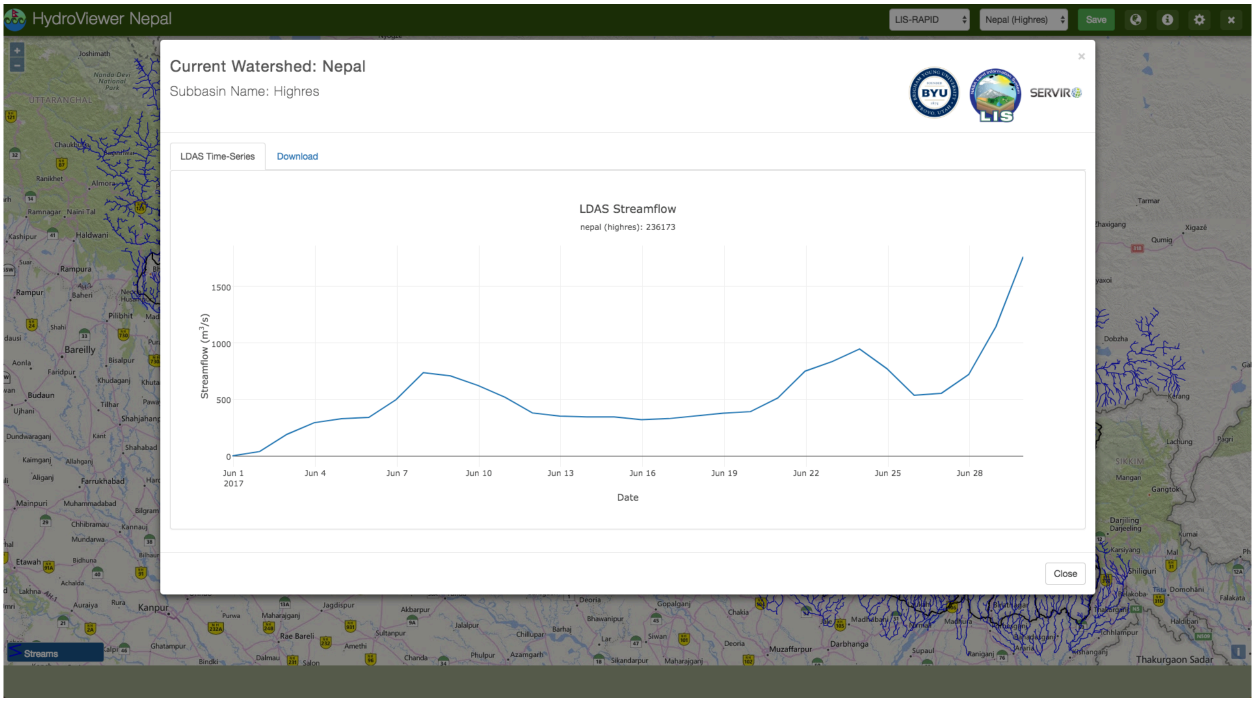


Figure 4-15. HydroViewer Nepal displaying SALDAS hydrograph.

It is the nature of forecasts to better predict the immediate future; therefore the warning level on the second day would be more likely than the one on the sixth day. However, this warning would be overlapped. Similarly, by displaying all warning levels at the same time, without considering time or space, the visual result usually gives a false sense that an extreme event may be imminent at the area and quite certain as opposed to ten days out with much greater uncertainty. Figure 4-16 shows a comparison between the two visualization methods. The SPT shows all the warning levels at the same time and at all zoom levels, including those for minor streams, while the new streamflow animation method displays warning levels only at a specific time and zoom level and it is much easier to distinguish whether the threat is for a minor tributary or the main stem of a river.

While the streamflow animation service is hosted using the proprietary ArcGIS Server, it follows OGC standards, allowing the service to be accessed by any system. The procedure used to display streams at different zoom levels based on stream order together with the warning level display symbology on the streams themselves as opposed to triangles and the Server Object Interceptor (SOI) to always query the bottom layer independently of the zoom level drastically improves the display time and responsiveness of the streamflow animation service.

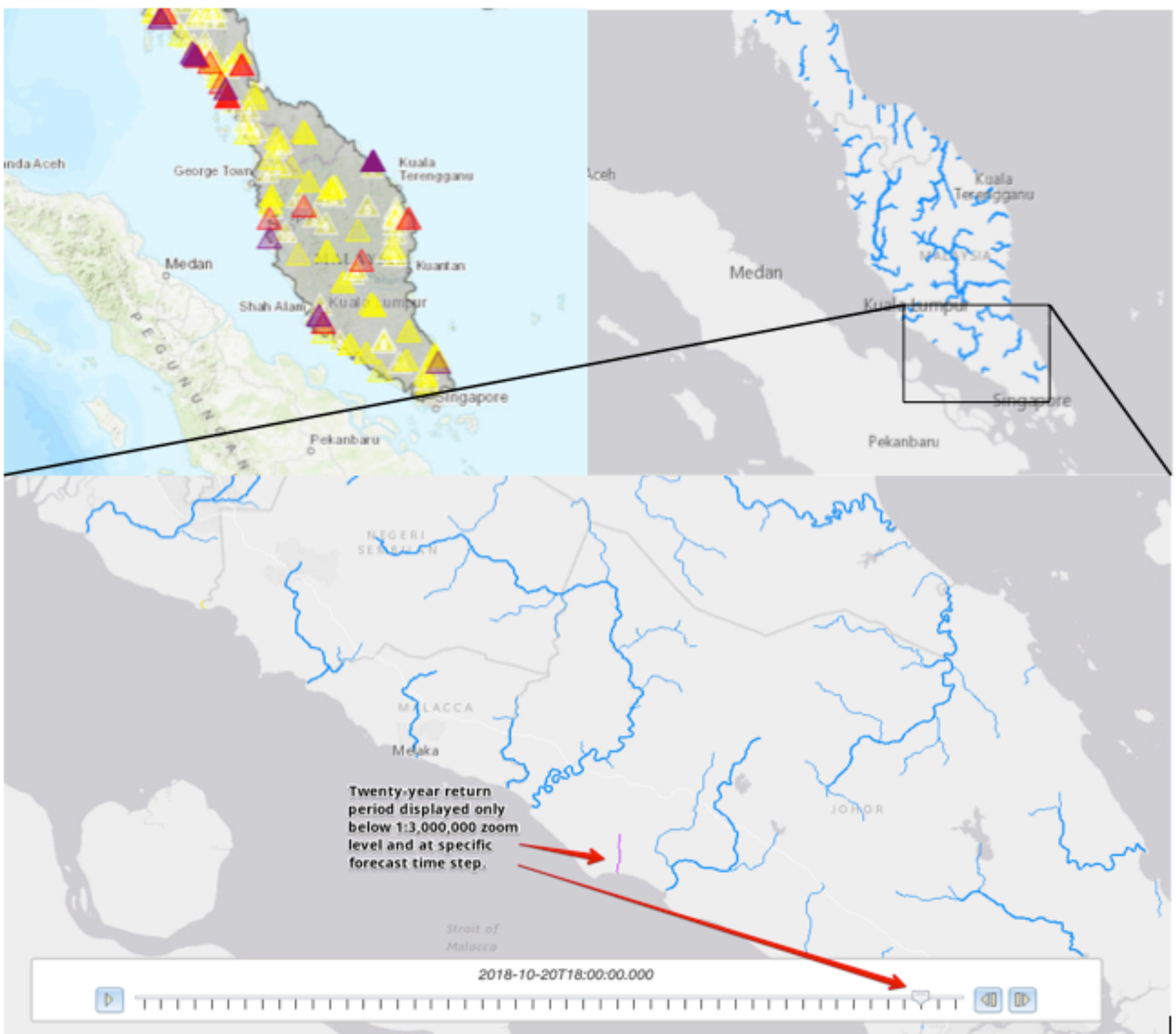


Figure 4-16. Streamflow visualization comparison between the SPT and the new time-enabled streamflow animation.

The total number of stream features displayed for all the regions comes to about 200000. Considering that these features are time-enabled and change their symbology display depending on a specific forecast prediction they, and that there are 85 time steps in the forecast, the amount of unique features displayed by this service comes to about 17000000 features. All these features are loaded on a map in a matter of seconds thanks to the aforementioned techniques used to improve the display.

Some of the different efforts to increase the exposure and the adoption of the GloFAS-RAPID model include a focus on the development of decision support tools, and collaborations with different organizations and countries around the globe. Hydrology is a foundational science that affects many other earth sciences and disciplines, as such hydrologic data can be incorporated into other disciplines to help provide a more informed answer to some of the main problems we faced as a society. The HydroViewer app is an example of a web application that can be used to help inform decision makers. However, thanks to the service-oriented approach in which the GloFAS-RAPID results are presented, it is relatively easy to develop additional web applications to address specific problems. Some examples include a flood mapping app and reservoir management app. These apps can consume the GloFAS-RAPID results through its REST API and use the resulting forecast as part of additional processes that are important for the given location or stakeholder. Figure 4-17 shows a flood mapping tool developed for the Dominican Republic that uses the GloFAS-RAPID API to generate a dynamic flood map specific to the predicted forecast.

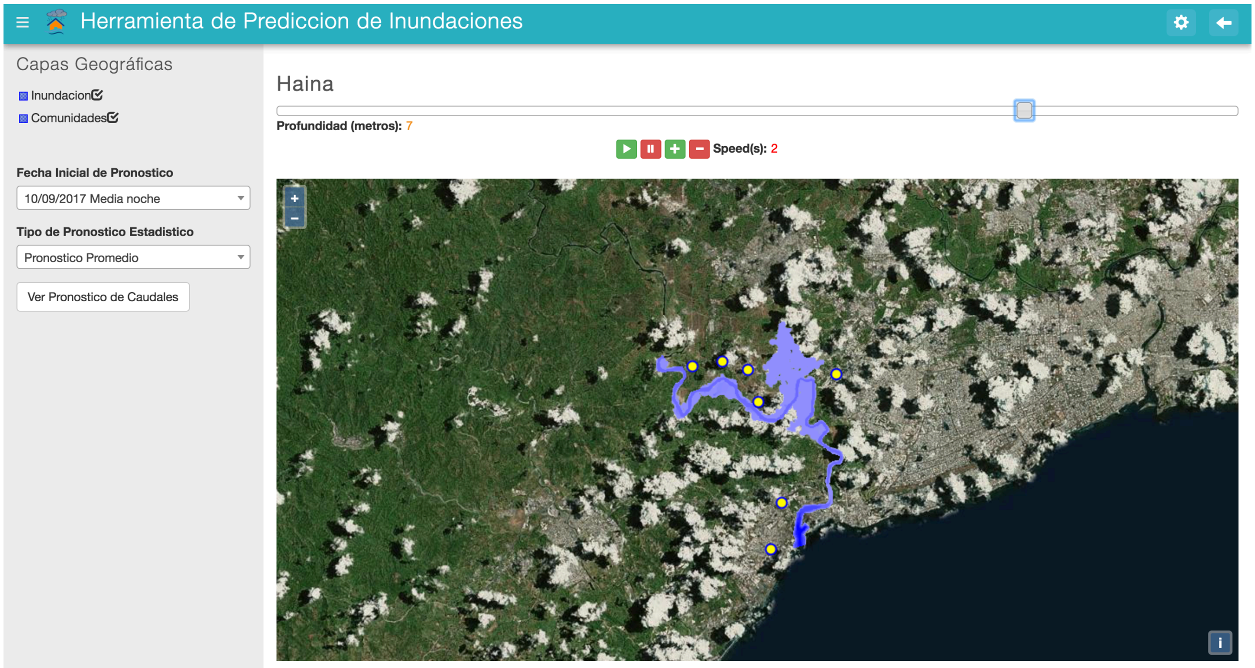


Figure 4-17. Flood Mapping Tool for Dominican Republic.

Different instances of web applications that make use of the GloFAS-RAPID model results through its API have been developed for many countries around the world. Some of these countries include Argentina, Bangladesh, Brazil, Colombia, Nepal, Peru, Tanzania, and the Dominican Republic. The development of these apps was coordinated through collaboration with local agencies and the NASA SERVIR Program, whose primary motivation is to help provide environmental decision-making tools that are adopted and fully functioning by stakeholders with the responsibility for water resources and emergency management.

In addition to collaborations with SERVIR, this global streamflow prediction system has been one of the main projects supported by the Group on Earth Observations Water Sustainability Initiative (GEOGLOWS). These organizations help provide the necessary exposure for this system to be useful. Multiple trainings and capacity building efforts have taken place thanks to the support of these agencies. Collaboration with local agencies also enabled validation efforts by providing access to local observed data. We compared our historic simulation results to observed data from 20 different locations in Nepal and Colombia. Figure 4-18 shows that the routed historic simulation successfully follows the same pattern as the observed data and captures most events with a tendency to under-predict.

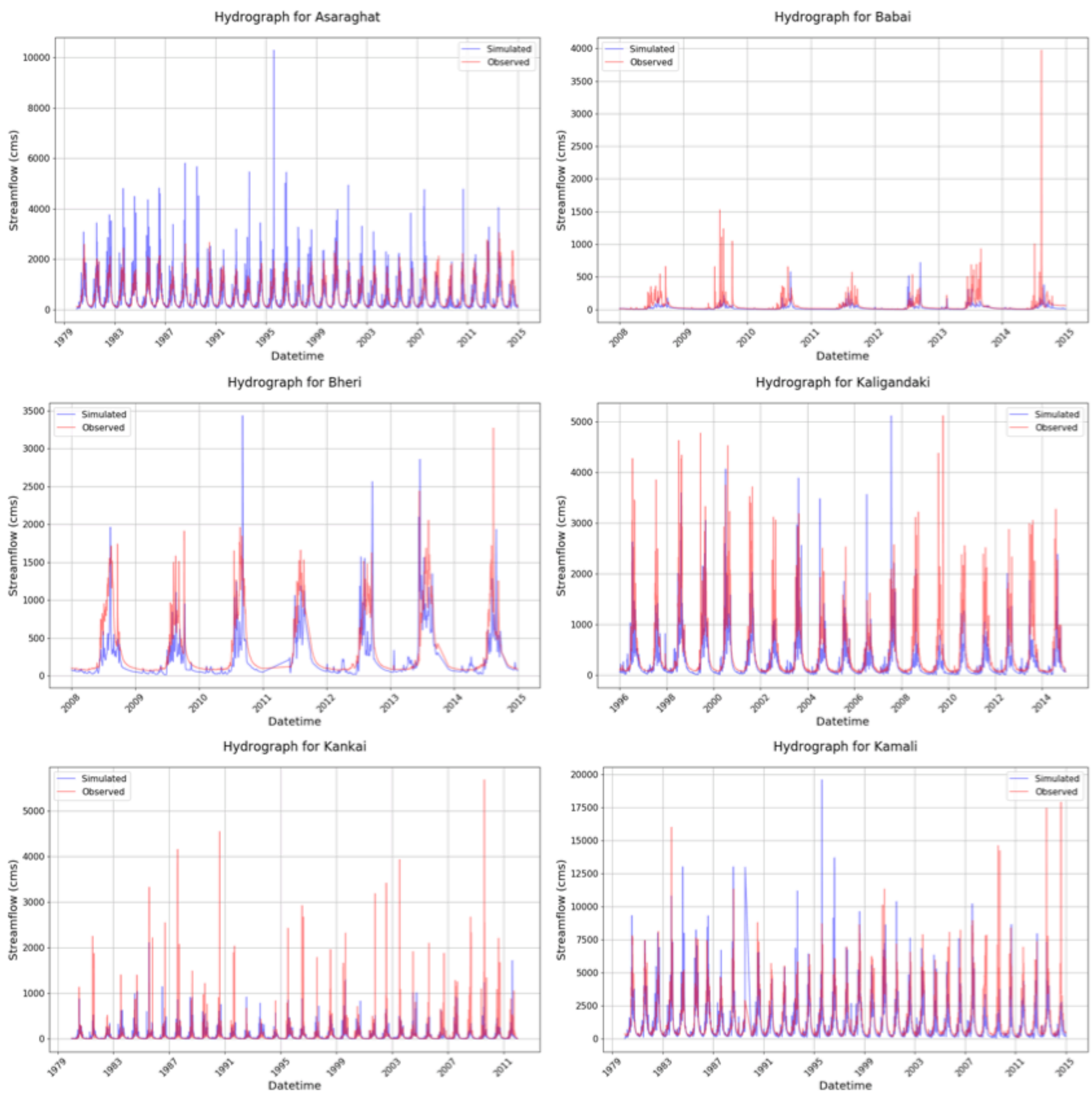


Figure 4-18. Simulated vs Observed data for different locations in Nepal.

We also compared the daily average of our historic simulation for those of the observed data at the selected locations. Figure 4-19 shows that the historic simulation also captures seasonal averages.

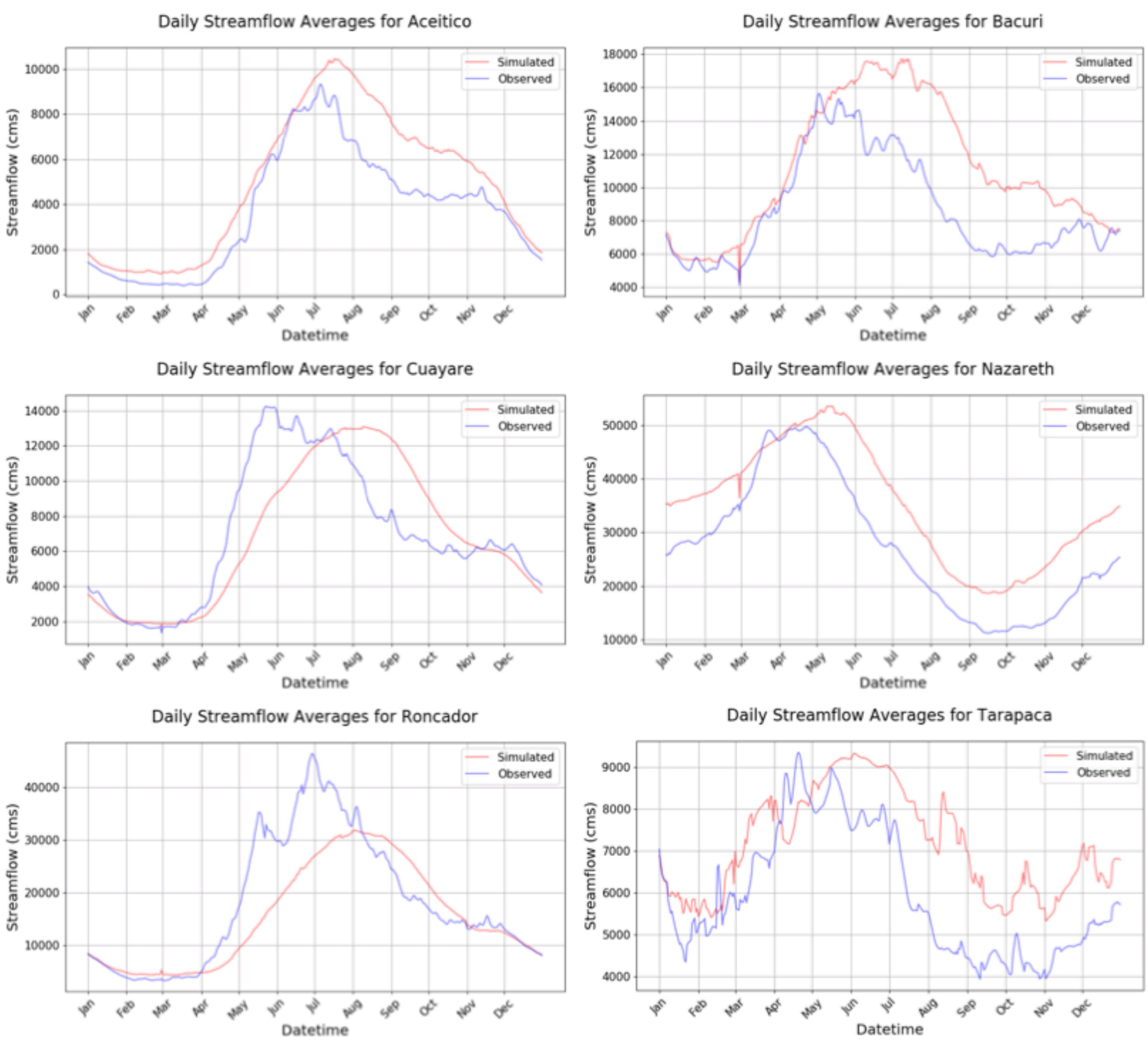


Figure 4-19. Simulated vs observed daily averages for different locations in Colombia.

We performed an analysis to determine if our GloFAS-RAPID routed results were similar to the coarser GloFAS results. Data was collected for nine weeks during the summer of 2017 and summarized weekly.

Figure 4-20. GloFAS-RAPID vs GloFAS forecasts.

Figure 4-20 shows that GloFAS-RAPID provides a very similar result to the original GloFAS and follows trends with very similar shapes. This information demonstrates that even though GloFAS-RAPID is routing results over smaller watersheds, results from the same locations are still very similar in volume, with the main differences being the initialization methods used with each model, and the differences in the terrain and hydrography used for the routing.

Finally, we performed an analysis to determine if our selected watershed size for routing results had any effect or introduced any variability on forecasted results. This was done by comparing forecasted results at the mouth of a watershed using three different spatial decompositions of the watershed upstream. Figure 4-21 shows an example of how the watersheds were subdivided.

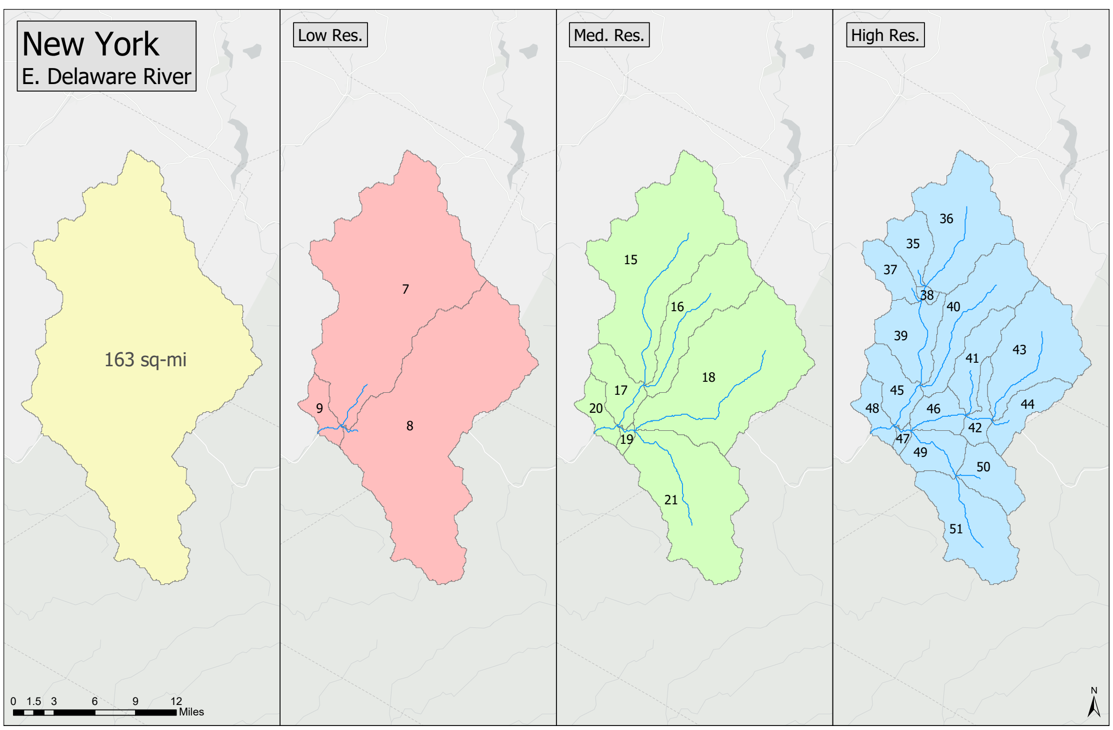


Figure 4-21. Varying resolutions used for the East Delaware River.

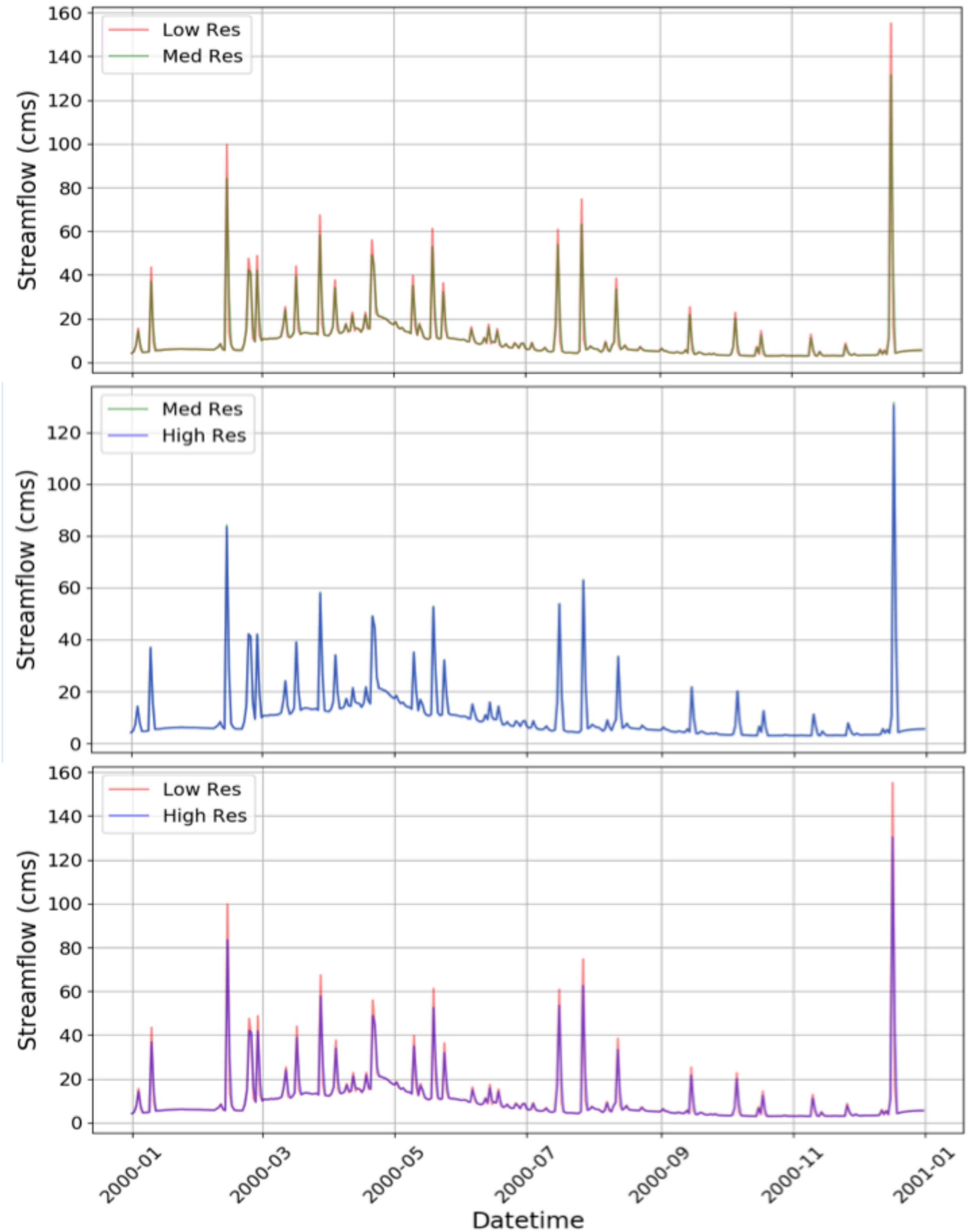


Figure 4-22. East Delaware River comparison between different resolutions.

As expected, the results from varying resolutions at the mouth of all the tested watersheds did not yield any significant differences in the results. Figure 4-22 shows three hydrographs comparing different resolutions. These results are consistent with the fact that the RAPID preprocessing methodology assigns a percentage of the total runoff volume to each sub-basin. The sum of these volumes at the mouth of a watershed should always be about the same.

Aside from initial validation, data validation for a large-scale forecast prediction system at specific locations is a complicated task. This is in part due to the extent covered by the model. Local involvement is necessary to validate results and to provide feedback about the model. The collaboration efforts described above, as well as the development of validation tools, and accessibility tools such as REST APIs that facilitate forecasted and observed data analyses, provide a long term approach to validating and improving overall model results at the local level.

4.5 Conclusion

The development of a large-scale streamflow prediction system based on the ECMWF ensemble global runoff forecast presents a series of new challenges to run the system in an operational environment and to make the resulting streamflow information useful at the local scale. These hydroinformatic challenges were divided into four categories: big data, data communication, adoption, and validation. The GloFAS-RAPID model was used as the main model to address these challenges. GloFAS-RAPID provides a high-density result by routing runoff volume from GloFAS using the RAPID routing model. A solution for each of the main challenges was provided (Figure 4-23). A cloud cyberinfrastructure was developed to host model workflows, inputs, and outputs. Web applications were deployed to expose results over the Internet. Web services such as a REST API and geospatial services were created to provide accessibility to our forecasted results. Additional web applications were created with the main goal to allow customizations and provide flexibility for local agencies to use results according to their specific needs. These projects were demonstrated in different countries around the world. Some of these countries include: Argentina, Bangladesh, Brazil, Colombia, Haiti, Peru, Nepal, Tanzania, the Dominican Republic, and the United States. We tested our results by comparing our forecasts to observed data. We determined that the GloFAS-RAPID results are in essence the same as the GloFAS results, but in a higher density. We also determined that the GloFAS-RAPID result is usually close to observed values and is able to capture most extreme events. For areas where biases exist, we have developed a way to understand those biases and adjust, even while paving the way for a feedback mechanism that could lead to improved model results. Finally, we analyzed the effect of density variations on GloFAS-RAPID, and determined that sub-basin sizes do not significantly affect results at the mouth of the watershed.

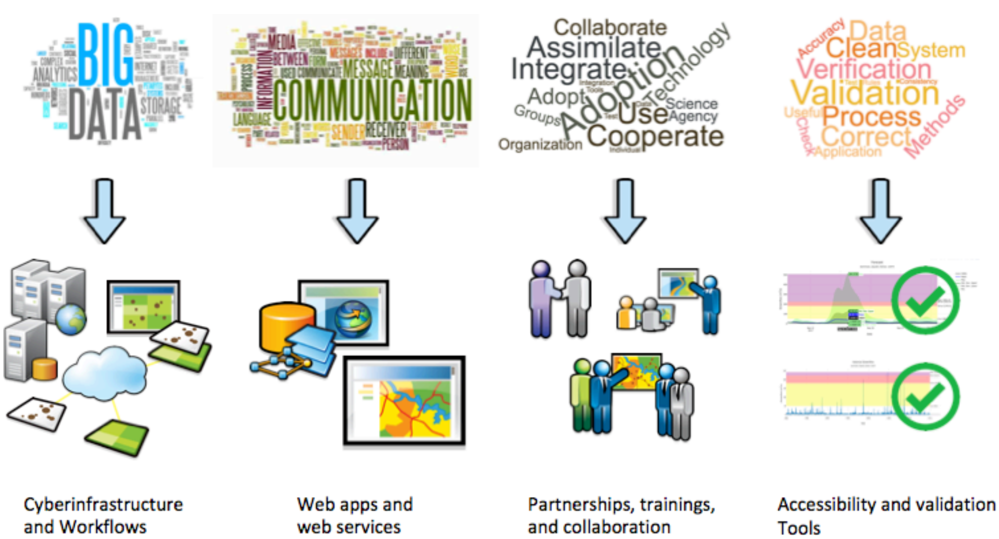


Figure 4-23. Hydroinformatic challenges and solutions.

5 SUMMARY AND FUTURE WORK

5.1 Summary

The development of a global streamflow prediction system that produces forecasts on a regular basis can answer many of the difficulties our society faces when dealing with water resources problems. Different initiatives exist with the goal to attain water security from a global perspective. The United Nation's Sustainable Development Goals is a collection of 17 goals to help improve our society. More than half of these goals include a water component. In addition, the SENDAI Framework, named after the Japanese city where it was agreed, constitutes an agreement endorsed by the UN to reduce disaster risk, and subsequently the losses of lives, livelihoods, and environmental assets at the individual, community and country scale due to natural disasters. These two components: water for attaining our goals as a society, and an acceptable level water-related risk is what defines the term water security.

An understanding of future water availability is necessary to attain water security. Hydrologic modeling can us help better predict, understand, and manage water resources by providing the necessary water intelligence required. To this end, the development of a large-scale high-resolution hydrologic model was deemed a grand challenge for the hydrologic community. Various models have surfaced during the past few years. However, these new models have also brought the realization that having a good hydrologic model does not make it automatically useful to stakeholders.

As part of my work with some of the more recognized large scale hydrologic models, including the US National Water Model, and GloFAS, I determined that there were additional challenges that needed to be overcome in order to create a successful operational global streamflow prediction system, namely big data, communication, adoption, and validation.

A case study working with the US National Water Model (NWM) was used to focus on big data and communication challenges. I created a cyberinfrastructure to store streamflow forecast results for the continental US on a daily basis using a semi-cloud environment in cooperation with the Renaissance Computing Institute (RENCI). I also developed the NWM Viewer web application to allow users to interact with the NWM results. This app allows for data visualization, querying, and extraction. Metrics tests between October 2017 and October 2018 revealed that the NWM Viewer app has been accessed more that 42,500 times to retrieve streamflow data.

I then developed our own streamflow prediction system using the GloFAS-RAPID model covering Africa, North America, South America, and South Asia. A cloud cyberinfrastructure using the Microsoft Azure platform was used to answer big data issues such as the computational power, and large amounts of data required to run a model for these major regions.

I helped improve the Streamflow Prediction Tool, a web application designed to visualize data from the GloFAS-RAPID model. I also created a REST API to improve programmatic data retrieval and the incorporation of our forecast results into third-party applications using a service-oriented approach. I also created a time-enabled geospatial service to display forecast results on each river reach within our covered regions.

To answer some of the adoption issues faced by hydrologic models, we worked in coordination with international agencies such as NASA, and NOAA; as well as local agencies around the world to provide training and customized solutions for end users. Some of the countries where we worked include Argentina, Bangladesh, Brazil, Colombia, Nepal, Peru, Tanzania, and the Dominican Republic. The HydroViewer web app was developed to facilitate the usage of forecasts results. The HydroViewer app is a lightweight web application designed to visualize streamflow forecasts for specific regions using not only the GloFAS-RAPID model, but also different model alternatives, which can be added to the app in a relatively easy way.

Finally, we conducted different tests to validate our modeled results following a collaboration approach using the HydroStats analytical package to facilitate model validation using observed data from collaborators where the HydroViewer app was deployed. Figure 4-23 (see section 4.5) shows our main answer to each one of the hydroinformatic challenges.

Overall, I have provided a methodology to operationalize a large-scale streamflow prediction system and provide meaningful results at the local level. This research has the potential to allow decision makers to focus on solving some of the most pressing water-related issues we face as a society by providing the cyberinfrastructure necessary to generate this type of water data, the tools necessary to access, consume, and manipulate data, the exposure and flexibility necessary to engage local communities, and validation and feedback mechanisms to instill confidence in its use and to improve results for specific areas.

5.2 Future Work

A global streamflow prediction system needs to cover the entire globe. One of the main improvements our GloFAS-RAPID streamflow prediction system needs is to set up other regions currently missing. In addition, improvement to our initialization methods and Muskingum parameter selection are needed. Improved Digital Elevation Model (DEMs) would also help improve the overall quality and resolution of the model. Machine learning and other Artificial Intelligence (AI) can help provide an answer to some of these issues. Some work has already been done regarding this. For example, AI has already been used to help estimated Muskingum parameters (Bai, Wei, Yang, & Huang, 2018; Farzin et al., 2018).

Countries like the United States usually possess a hydrography service with the country's river network and a unique identifier for each river reach within it. The US National Hydrography Dataset (NHD) is an example of this and the US National Water Model uses it as the foundational stream network for generating forecasts. Developing countries do not always possess such a service. Moreover, an official global hydrography dataset that integrates the existing hydrography of every country does not exist. The current GloFAS-RAPID system arbitrarily assigns an identifier for every region. A global streamflow prediction service would benefit from the consistency that a global hydrography dataset service would offer.

The developed REST API can be improved by adding functions to retrieve data based on a coordinate location as opposed to a river reach ID. In this way, users would not be required to know the specific ID for their river of interest.

Adoption and validation challenges are part of a paradigm that is meant to continue. New technologies in forecast prediction and data visualization need to be incorporated and pertinent training needs to be provided. A feedback approach is necessary to help validate a model system covering such an extensive area. Ways to facilitate model validation such as the development of web applications or an API that allow for modeled and observed data comparisons would help further this ongoing validation challenge.

REFERENCES

Agafonkin, Vladimir. (2014). Leaflet: an open-source JavaScript library for mobile-friendly interactive maps. *accessed, 28*, 2016.

Ahmadisharaf, Ebrahim, Kalyanapu, Alfred J., & Chung, Eun-Sung. (2016). Spatial probabilistic multi-criteria decision making for assessment of flood management alternatives. *Journal of Hydrology, 533*, 365-378. doi: https://doi.org/10.1016/j.jhydrol.2015.12.031

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS - global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences, 17*(3), 1161-1175. doi: 10.5194/hess-17-1161-2013

Andersson, E, & Thépaut, JN. (2008). ECMWF's 4D-Var data assimilation system–the genesis and ten years in operations. *ECMWF Newsletter, 115*, 8-12.

Bai, Tao, Wei, Jian, Yang, Wangwang, & Huang, Qiang. (2018). Multi-Objective Parameter Estimation of Improved Muskingum Model by Wolf Pack Algorithm and Its Application in Upper Hanjiang River, China. *Water, 10*(10), 1415. doi: 10.3390/w10101415

Balsamo, Gianpaolo, Beljaars, Anton, Scipal, Klaus, Viterbo, Pedro, van den Hurk, Bart, Hirschi, Martin, & Betts, Alan K. (2009). A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. *Journal of Hydrometeorology, 10*(3), 623-643. doi: 10.1175/2008jhm1068.1

Beran, Bora, & Piasecki, Michael. (2009). Engineering new paths to water data. *Computers & Geosciences, 35*(4), 753-760. doi: 10.1016/j.cageo.2008.02.017

Blodgett, David, Read, Emily, Lucido, Jessica, Slawecki, Tad, & Young, Dwane. (2016). An Analysis of Water Data Systems to Inform the Open Water Data Initiative. *JAWRA Journal of the American Water Resources Association, 52*(4), 845-858. doi: 10.1111/1752-1688.12417

Butts, Michael B., Payne, Jeffrey T., Kristensen, Michael, & Madsen, Henrik. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology, 298*(1-4), 242-266. doi: 10.1016/j.jhydrol.2004.03.042

Carsell, Kim M., Pingel, Nathan D., & Ford, David T. (2004). Quantifying the Benefit of a Flood Warning System. *Natural Hazards Review, 5*(3), 131-140. doi: doi:10.1061/(ASCE)1527-6988(2004)5:3(131)

David, Cédric H., Famiglietti, James S., Yang, Zong - Liang, Habets, Florence, & Maidment, David R. (2016). A decade of RAPID—Reflections on the development of an open source geoscience code. *Earth and Space Science, 3*(5), 226-244. doi: 10.1002/2015EA000142

Davis, Fred D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results.* Massachusetts Institute of Technology.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., . . . Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society, 137*(656), 553-597. doi: 10.1002/qj.828

Donnelly, Chantal, Andersson, Jafet C. M., & Arheimer, Berit. (2015). Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrological Sciences Journal, 61*(2), 255-273. doi: 10.1080/02626667.2015.1027710

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., . . . Tarpley, J. D. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres (1984–2012), 108*(D22). doi: 10.1029/2002JD003296

Emerton, Rebecca E., Stephens, Elisabeth M., Pappenberger, Florian, Pagano, Thomas C., Weerts, Albrecht H., Wood, Andy W., . . . Cloke, Hannah L. (2016). Continental and global scale flood forecasting systems. *Wiley Interdisciplinary Reviews: Water, 3*(3), 391-418. doi: 10.1002/wat2.1137

Farzin, Saeed, Singh, Vijay, Karami, Hojat, Farahani, Nazanin, Ehteram, Mohammad, Kisi, Ozgur, . . . El-Shafie, Ahmed. (2018). Flood Routing in River Reaches Using a Three-Parameter Muskingum Model Coupled with an Improved Bat Algorithm. *Water, 10*(9), 1130. doi: 10.3390/w10091130

Godschalk, David R, Rose, Adam, Mittler, Elliott, Porter, Keith, & West, Carol Taylor. (2009). Estimating the value of foresight: aggregate analysis of natural hazard mitigation benefits and costs. *Journal of Environmental Planning and Management, 52*(6), 739-756.

Goodall, Jonathan L, Horsburgh, Jeffery S, Whiteaker, Timothy L, Maidment, David R, & Zaslavsky, Ilya. (2008). A first approach to web services for the National Water Information System. *Environmental Modelling & Software, 23*(4), 404-411.

Gratier, Thomas, Spencer, Paul, & Hazzard, Erik. (2015). *OpenLayers 3: Beginner's Guide*: Packt Publishing Ltd.

Grey, David, & Sadoff, Claudia W. (2007). Sink or swim? Water security for growth and development. *Water policy, 9*(6), 545-571.

Guha-Sapir, D., Below, R., & Hoyois, P. (2017). The CRED/OFDA International Disaster Database. Retrieved 2017, from Le Université Catholique de Louvain http://www.emdat.be

Hashem, Ibrahim Abaker Targio, Yaqoob, Ibrar, Anuar, Nor Badrul, Mokhtar, Salimah, Gani, Abdullah, & Ullah Khan, Samee. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems, 47*, 98-115. doi: https://doi.org/10.1016/j.is.2014.07.006

Hilderbrand, Douglas, & Lead, Weather-Ready Nation Ambassador. (2014). Building a Weather-Ready Nation. *WMO Bulletin, 63*(2), 27.

Horita, Flávio E. A., Albuquerque, João Porto de, Degrossi, Lívia C., Mendiondo, Eduardo M., & Ueyama, Jó. (2015). Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks. *Computers & Geosciences, 80*, 84-94. doi: https://doi.org/10.1016/j.cageo.2015.04.001

Horsburgh, Jeffery S, Tarboton, David G, Maidment, David R, & Zaslavsky, Ilya. (2008). A relational model for environmental and water resources data. *Water Resources Research, 44*(5).

Horsburgh, Jeffery S, Tarboton, David G, Schreuders, Kimberly AT, Maidment, David R, Zaslavsky, Ilya, & Valentine, David. (2010). HydroServer: A platform for publishing space-time hydrologic datasets.

Iacovella, Stefano. (2017). *GeoServer Beginner's Guide: Share geospatial data using Open Source standards*: Packt Publishing Ltd.

Jackson, Elise K. (2018). *An Analysis of Using Error Metrics to Determine the Accuracy of Modeled Historical Streamflow on a Global Scale.* Brigham Young University. Retrieved from https://scholarsarchive.byu.edu/etd/6750

Kelman, Ilan. (2013). Disaster mitigation is cost effective.

Krause, Peter, Boyle, DP, & Bäse, Frank. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in geosciences, 5*, 89-97.

Kull, Daniel, Mechler, Reinhard, & Hochrainer ‐ Stigler, Stefan. (2013). Probabilistic cost ‐ benefit analysis of disaster risk management in a development context. *Disasters, 37*(3), 374-400.

Kumar, S., Peterslidard, C., Tian, Y., Houser, P., Geiger, J., Olden, S., . . . Dirmeyer, P. (2006). Land information system: An interoperable framework for high resolution land surface modeling. *Environmental Modelling & Software, 21*(10), 1402-1415. doi: 10.1016/j.envsoft.2005.07.004

Lehner, Bernhard, Verdin, Kristine, & Jarvis, Andy. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union, 89*(10), 93-94.

Liang, Xu, Lettenmaier, Dennis P., Wood, Eric F., & Burges, Stephen J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres, 99*(D7), 14415-14428. doi: 10.1029/94JD00483

Lindström, Göran, Pers, Charlotta, Rosberg, Jörgen, Strömqvist, Johan, & Arheimer, Berit. (2010). Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrology Research, 41*(3-4), 295. doi: 10.2166/nh.2010.007

Maidment, David R. (2016). A Conceptual Framework for the National Flood Interoperability Experiment. *NFIE Conceptual Framework Pap*.

McEnery, J., Ingram, J., Duan, Q., Adams, T., & Anderson, L. (2005). NOAA'S Advanced Hydrologic Prediction Service: Building Pathways for Better Science in Water Forecasting. *Bulletin of the American Meteorological Society, 86*(3), 375-385. doi: 10.1175/bams-86-3-375

Nam, UNDP Viet. (2015). Transforming our world: The 2030 agenda for sustainable development.

Niswonger, Richard G., Allander, Kip K., & Jeton, Anne E. (2014). Collaborative modelling and integrated decision support system analysis of a developed terminal lake basin. *Journal of Hydrology, 517*, 521-537. doi: https://doi.org/10.1016/j.jhydrol.2014.05.043

Niu, Guo-Yue, Yang, Zong-Liang, Mitchell, Kenneth E., Chen, Fei, Ek, Michael B., Barlage, Michael, . . . Xia, Youlong. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research, 116*(D12). doi: 10.1029/2010jd015139

NOAA. (2016). NOAA launches America’s first national water forecast model. from http://www.noaa.gov/media-release/noaa-launches-america-s-first-national-water-forecast-model

NOAA Office of Water Prediction. (2017). Information About the National Water Model. Retrieved October 31, 2017, from http://water.noaa.gov/about/nwm

Pappenberger, Florian, Cloke, Hannah L., Parker, Dennis J., Wetterhall, Fredrik, Richardson, David S., & Thielen, Jutta. (2015). The monetary benefit of early flood warnings in

Europe. *Environmental Science & Policy, 51*, 278-291. doi: http://dx.doi.org/10.1016/j.envsci.2015.04.016

Rajasekar, Arcot, Moore, Reagan, Hou, Chien-yi, Lee, Christopher A, Marciano, Richard, de Torcy, Antoine, . . . Gilbert, Lucas. (2010). iRODS Primer: integrated rule-oriented data system. *Synthesis Lectures on Information Concepts, Retrieval, and Services, 2*(1), 1-143.

Rew, Russell K, & Davis, Glenn P. (1990). *The unidata netCDF: Software for scientific data access.* Paper presented at the Sixth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., . . . Toll, D. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society, 85*(3), 381-394. doi: 10.1175/BAMS-85-3-381

Roo, A. P. J. De, Wesseling, C. G., & Deursen, W. P. A. Van. (2000). Physically based river basin modelling within a GIS: the LISFLOOD model. *Hydrological Processes, 14*(11‐12), 1981-1992. doi: 10.1002/1099-1085(20000815/30)14:11/12<1981::AID-HYP49>3.0.CO;2-F

Rutledge, Glenn K, Alpert, Jordan, & Ebisuzaki, Wesley. (2006). NOMADS: A climate and weather model archive at the National Oceanic and Atmospheric Administration. *Bulletin of the American Meteorological Society, 87*(3), 327-342.

Snow, Alan D., Christensen, Scott D., Swain, Nathan R., Nelson, E. James, Ames, Daniel P., Jones, Norman L., . . . Zsoter, Ervin. (2016). A High-Resolution National-Scale Hydrologic Forecast System from a Global Ensemble Land Surface Model. *JAWRA Journal of the American Water Resources Association, 52*(4), 950-964. doi: 10.1111/1752-1688.12434

Souffront Alcantara, Michael A., Crawley, Shawn, Stealy, Michael J., Nelson, E. James, Ames, Daniel P., & Jones, Norm L. (2017). Open Water Data Solutions for Accessing the National Water Model. *Open Water Journal, 4*(1), 12.

Souffront Alcantara, Michael A., Kesler, Christian, Stealey, Michael J., Nelson, E. James, Ames, Daniel P., & Jones, Norm L. (2017). Cyberinfrastructure and Web Apps for Managing and Disseminating the National Water Model. *JAWRA Journal of the American Water Resources Association*. doi: 10.1111/1752-1688.12608

Svoboda, Mark D., Fuchs, Brian A., Poulsen, Chris C., & Nothwehr, Jeff R. (2015). The drought risk atlas: Enhancing decision support for drought risk management in the United States. *Journal of Hydrology, 526*, 274-286. doi: https://doi.org/10.1016/j.jhydrol.2015.01.006

Swain, Nathan R., Christensen, Scott D., Snow, Alan D., Dolder, Herman, Espinoza-Dávalos, Gonzalo, Goharian, Erfan, . . . Burian, Steven J. (2016). A new open source platform for lowering the barrier for environmental web app development. *Environmental Modelling & Software, 85*, 11-26. doi: 10.1016/j.envsoft.2016.08.003

Swain, Nathan R., Latu, Kilisimasi, Christensen, Scott D., Jones, Norman L., Nelson, E. James, Ames, Daniel P., & Williams, Gustavious P. (2015). A review of open source software solutions for developing water resources web applications. *Environmental Modelling & Software, 67*, 108-117. doi: 10.1016/j.envsoft.2015.01.014

Tarboton, DG, Idaszak, R, Horsburgh, JS, Heard, J, Ames, D, Goodall, JL, . . . Arrigo, J. (2014). *HydroShare: advancing collaboration through hydrologic data and model sharing.* Paper presented at the International Environmental Modelling and Software Society (iEMSs) 7th International Congress on Environmental Modelling and Software, San Diego, California, USA. www. iemss. org/society/index/php/iemss-2014-proceedings.

USGS NHD. (2016). USGS: National Hydrography Dataset. US Department of Interior: United States Geological Survey.

van Alphen, J., Martini, F., Loat, R., Slomp, R., & Passchier, R. (2009). Flood risk mapping in Europe, experiences and best practices. *Journal of Flood Risk Management, 2*(4), 285-292. doi: 10.1111/j.1753-318X.2009.01045.x

Wan, Zhanming, Hong, Yang, Khan, Sadiq, Gourley, Jonathan, Flamig, Zachary, Kirschbaum, Dalia, & Tang, Guoqiang. (2014). A cloud-based global flood disaster community cyber-infrastructure: Development and demonstration. *Environmental Modelling & Software, 58*, 86-94. doi: https://doi.org/10.1016/j.envsoft.2014.04.007

Wehrli, A, Herkendell, J, & Jol, A. (2010). Mapping the impacts of natural hazards and technological accidents in Europe: EEA Technical report 13/2010, European Environment Agency, Copenhagen, Denmark.

Wood, Eric F., Roundy, Joshua K., Troy, Tara J., van Beek, L. P. H., Bierkens, Marc F. P., Blyth, Eleanor, . . . Whitehead, Paul. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research, 47*(5). doi: 10.1029/2010wr010090

Xu, Chong-yu. (1999). Climate Change and Hydrologic Models: A Review of Existing Gaps and Recent Research Developments. *Water Resources Management, 13*(5), 369-382. doi: 10.1023/a:1008190900459

Zaslavsky, Ilya, Valentine, David, & Whiteaker, Tim. (2007). CUAHSI WaterML. *Open Geospatial Consortium Discussion Paper OGC*.

APPENDIX A. STREAMFLOW PREDICTION TOOL REST API DOCUMENTATION

A REST API is a web service or a set of methods that can be used to produce or access data without a web interface. REST APIs use the http protocol to request data. Parameters are passed through a URL using a predetermined organization. A REST API has been developed to provide access to the Streamflow Prediction Tool (SPT) forecasts without the need to access the web app interface. This type of service facilitates integration of the SPT with third party web apps, and the automation of forecast retrievals using programing languages like Python, or R. The available methods and a description of how to use them are shown below.

GetForecast for Forecasts Statistics

Parameter	Description	Example
watershed_name	The name of watershed or main area of interest.	Nepal
subbasin_name	The name of the sub basin or sub area.	Central
reach_id	The identifier for the stream reach.	5
forecast_folder	The date of the forecast (YYYYMMDD.HHHH). (Optional) (YYYMMDD.H)	20170110.1200 20170110.0 most_recent
stat_type	The selected forecast statistic. (high_res, mean,	mean

Parameter	Description	Example
	std_dev_range_upper, std_dev_range_lower, max, min).	
units	Set to 'english' to get ft3/s. (Optional)	english
return_format	Set to 'csv' to get csv file. (Optional)	csv

Example

```
>>> import requests
>>> request_params = dict(watershed_name='Nepal', subbasin_name='Central', reach_id=5,
forecast_folder='most_recent', stat_type='mean')
>>> request_headers = dict(Authorization='Token asdfqwer1234')
>>> res = requests.get('[HOST Portal]/apps/streamflow-prediction-
tool/api/GetForecast/', params=request_params, headers=request_headers)
```

GetEnsemble (1 - 52)

Parameter	Description	Example
watershed_name	The name of watershed or main area of interest.	Nepal
subbasin_name	The name of the sub basin or sub area.	Central
reach_id	The identifier for the stream reach.	5
forecast_folder	The date of the forecast (YYYYMMDD.HHHH). (Optional) (YYYMMDD.H)	20170110.1200 20170110.0 most_recent
ensemble	The selected forecast ensemble(s). The value can be a	Number: 52

Parameter	Description	Example
	number, a list, or a range. Accepted values go from 1 to Leave empty or ensemble=all for retrieving all.	List: 1,3,6,9 Range: 1-15
units	Set to ‘english’ to get ft3/s. (Optional)	english

Example

```
>>> import requests

>>> request_params = dict(watershed_name='Nepal', subbasin_name='Central', reach_id=5,
forecast_folder='most_recent', ensemble='52')

>>> request_headers = dict(Authorization='Token asdfqwer1234')

>>> res = requests.get('[HOST Portal]/apps/streamflow-prediction-
tool/api/GetEnsemble/', params=request_params, headers=request_headers)
```

GetHistoricData (1980 - Present)

Parameter	Description	Example
watershed_name	The name of watershed or main area of interest.	Nepal
subbasin_name	The name of the sub basin or sub area.	Central
reach_id	The identifier for the stream reach.	5
units	Set to ‘english’ to get ft3/s. (Optional)	english
return_format	Set to ‘csv’ to get csv file. (Optional)	csv

Example

```
>>> import requests

>>> request_params = dict(watershed_name='Nepal', subbasin_name='Central', reach_id=5)
```

```
>>> request_headers = dict(Authorization='Token asdfqwer1234')

>>> res = requests.get('[HOST Portal]/apps/streamflow-prediction-
tool/api/GetHistoricData/', params=request_params, headers=request_headers)
```

GetReturnPeriods (2, 10, and 20 year return with historical max)

Parameter	Description	Example
watershed_name	The name of watershed or main area of interest.	Nepal
subbasin_name	The name of the sub basin or sub area.	Central
reach_id	The identifier for the stream reach.	5
units	Set to ‘english’ to get ft3/s. (Optional)	english

Example

```
>>> import requests

>>> request_params = dict(watershed_name='Nepal', subbasin_name='Central',
return_period=2)

>>> request_headers = dict(Authorization='Token asdfqwer1234')

>>> res = requests.get('[HOST Portal]/apps/streamflow-prediction-
tool/api/GetReturnPeriods/', params=request_params, headers=request_headers)
```

GetAvailableDates

Parameter	Description	Example
watershed_name	The name of watershed or main area of interest.	Nepal
subbasin_name	The name of the sub basin or sub area.	Central
reach_id	The identifier for the stream reach.	5

Example

```
>>> import requests
>>> request_params = dict(watershed_name='Nepal', subbasin_name='Central', reach_id=5)
>>> request_headers = dict(Authorization='Token asdfqwer1234')
>>> res = requests.get('[HOST Portal]/apps/streamflow-prediction-tool/api/GetAvailableDates/', params=request_params, headers=request_headers)
```

GetWatersheds

This method takes no parameters and returns a list of the available watersheds.

Example

```
>>> import requests
>>> request_headers = dict(Authorization='Token asdfqwer1234')
>>> res = requests.get('[HOST Portal]/apps/streamflow-prediction-tool/api/GetWatersheds/', headers=request_headers)
```

GetWarningPoints

Parameter	Description	Example
watershed_name	The name of watershed or main area of interest.	Nepal
subbasin_name	The name of the sub basin or sub area.	Central
return_period	The return period that the warning is based on.	(2,10, or 20)
forecast_folder	The date of the forecast (YYYYMMDD.HHHH). (Optional)	20170110.1200

Example

```
>>> import requests
>>> request_params = dict(watershed_name='Nepal', subbasin_name='Central', return_period=20, forecast_folder='20170802.0')
```

```
>>> request_headers = dict(Authorization='Token asdfqwer1234')
>>> res = requests.get('[HOST Portal]/apps/streamflow-prediction-tool/api/GetWarningPoints/', params=request_params, headers=request_headers)
```

APPENDIX B. US NATIONAL WATER MODEL REST API

GetWaterML

Supported Methods	GET				
Returns	A WaterML file of the specified forecast.				
Params	**Name**	**Description**	**Valid Values**	**Required**	**Default if omitted**
	archive	The archive (data source) of the forecast.	Indicate the archive (data source) used for query; Accepted value: "rolling", "harvey" and "irma"	No	Default value: "rolling"
	config	The configuration of the forecast.	One and only one of the following strings: "short_range", "long_range", "medium_range", or "analysis_assim".	Yes	Cannot be ommitted.
	geom	The geometry of the forecast.	One and only one of the following strings: "channel_rt", "land", "reservoir" or "forcing".	Yes.	Cannot be ommitted.
	variable	The variable of the forecast.	One and only one of the following strings, depending on the specified configuration and geometry. analysis_assim + channel_rt: "streamflow" or "velocity". analysis_assim + reservoir: "inflow" or "outflow". analysis_assim + land: "SNOWH", "SNEQV", "FSNO", "ACCET", "SOILSAT_TOP", or "SNOWT_AVG". analysis_assim + forcing: "RAINRATE", "LWDOWN", "PSFC", "Q2D", "SWDOWN", "T2D", "U2D", "V2D". short_range + channel_rt: "streamflow" or "velocity". short_range + reservoir: "inflow" or "outflow". short_range + land: "SNOWH", "SNEQV", "FSNO", "ACCET", "SOILSAT_TOP", or "SNOWT_AVG". short_range + forcing: "RAINRATE",	Yes.	Cannot be ommitted.

			"LWDOWN", "PSFC", "Q2D", "SWDOWN", "T2D", "U2D", "V2D". medium_range + channel_rt: "streamflow" or "velocity". medium_range + reservoir: "inflow" or "outflow". medium_range + land: "SNOWH", "SNEQV", "FSNO", "ACCET", "SOILSAT_TOP", "SNOWT_AVG", "UGDRNOFF", "ACCECAN", "SOIL_T", "SOIL_M", or "CANWAT". medium_range + forcing: "RAINRATE", "LWDOWN", "PSFC", "Q2D", "SWDOWN", "T2D", "U2D", "V2D". long_range + channel_rt: "streamflow". long_range + reservoir: "inflow" or "outflow". long_range + land: "SNEQV", "ACCET", "SOILSAT", "UGDRNOFF", "SFCRNOFF", "CANWAT". long_range + forcing: N/A (long_range has no forcing files.)		
	COMID	The identifier of the stream reach, reservoir, or grid cell for the forecast.	A numeric string. e.g. "12345678". If geometry=land, enter the grid south_north index followed by a comma and then the grid west_east index. e.g. "1357,2468"	Yes.	Cannot be ommitted.
	lon	Longitude	A numeric string with a longitude coordinate in decimal degrees".	No.	Empty string.
	lat	Latitude	A numeric string with a latitude coordinate in decimal degrees".	No.	Empty string.
	startDate	The beginning date of the forecast.	A string of the form "YYYY-MM-DD"	Yes.	Cannot be ommitted.
	endDate	*Only applicable/valid if config=analysis_assim.* The ending date of the analysis assimilation.	A string of the form "YYYY-MM-DD" representing any date between "2016-06-09" and the current date.	No.	The endDate is the startDate plus one day.
	time	*Only applicable/valid if config=short_range or medium_range.* The UTC time of day at which the forecast is initialized, represented by an hour from 00 to 23. Time 00 corresponds to 12:00AM, and so forth up to time 23 for 11:00PM.	Numeric string from 00 to 23. e.g. "00" short_range: 00, 01, 02 ...23. medium_range: 00, 06, 12, 18.	No.	"00"

	lag	*Only applicable/valid if config=long_range.* The time lag of the long range ensemble forecast.	The following strings: t00z, t06z, t12z. e.g. "t00z"	No	"t00z"
	member	*Only applicable/valid if config=long_range.* Represents the desired ensemble member of the long range forecast.	Numeric string between 1 and 4. e.g. "1".	No	"1"