



A HIGH-RESOLUTION NATIONAL-SCALE HYDROLOGIC FORECAST SYSTEM FROM A GLOBAL ENSEMBLE LAND SURFACE MODEL¹

Alan D. Snow, Scott D. Christensen, Nathan R. Swain, E. James Nelson, Daniel P. Ames, Norman L. Jones, Deng Ding, Nawajish S. Noman, Cédric H. David, Florian Pappenberger, and Ervin Zsoter²

ABSTRACT: Warning systems with the ability to predict floods several days in advance have the potential to benefit tens of millions of people. Accordingly, large-scale streamflow prediction systems such as the Advanced Hydrologic Prediction Service or the Global Flood Awareness System are limited to coarse resolutions. This article presents a method for routing global runoff ensemble forecasts and global historical runoff generated by the European Centre for Medium-Range Weather Forecasts model using the Routing Application for Parallel computation of Discharge to produce high spatial resolution 15-day stream forecasts, approximate recurrence intervals, and warning points at locations where streamflow is predicted to exceed the recurrence interval thresholds. The processing method involves distributing the computations using computer clusters to facilitate processing of large watersheds with high-density stream networks. In addition, the Streamflow Prediction Tool web application was developed for visualizing analyzed results at both the regional level and at the reach level of high-density stream networks. The application formed part of the base hydrologic forecasting service available to the National Flood Interoperability Experiment and can potentially transform the nation's forecast ability by incorporating ensemble predictions at the nearly 2.7 million reaches of the National Hydrography Plus Version 2 Dataset into the national forecasting system.

(KEY TERMS: ECMWF; RAPID; Tethys Platform; CondorPy; HTCondor; CI-WATER; GloFAS; NFIE; flood prediction; streamflow prediction; forecast.)

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

¹Paper No. JAWRA-15-0080-P of the *Journal of the American Water Resources Association* (JAWRA). Received June 8, 2015; accepted April 13, 2016. © 2016 The Authors Journal of the American Water Resources Association published by Wiley Periodicals, Inc. on behalf of American Water Resources Association. This is an open access article under the terms of the Creative Commons Attribution-Non-Commercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made. **Discussions are open until six months from issue publication.**

²Research Civil Engineer (Snow), Coastal and Hydraulics Laboratory, US Army Engineer Research and Development Center, 3909 Halls Ferry Rd., Vicksburg, Mississippi 39180; Graduate Student (Christensen, Swain) and Professor (Nelson, Ames, Jones), Department of Civil and Environmental Engineering, Brigham Young University, Provo, Utah 84602; Product Engineer (Ding) and Lead Project Engineer (Noman), Esri, Redlands, California, 92373; Scientist (David), Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109; and Scientist (Pappenberger, Zsoter), European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, United Kingdom (E-Mail/Snow: alan.d.snow@usace.army.mil).

INTRODUCTION

Catastrophic floods impact tens of millions of people each year and cause significant infrastructure damage. Global statistics for the period of 2004-2014 indicate that more than 951 million people were impacted by floods, over \$324 billion in damage occurred, and there were approximately 66,000 deaths (Guha-Sapir *et al.*, 2015). Improvements in flood forecasting and the ability to communicate actionable information to emergency responders have a substantial lifesaving and monetary benefit (Pappenberger *et al.*, 2015). Because of this, one of the first priorities of the new National Water Center in Tuscaloosa, Alabama (<http://www.nws.noaa.gov/oh/nwc/>) is to engage the academic community in the National Flood Interoperability Experiment (NFIE) (Maidment, 2015). NFIE aims to address several critical science and technology questions including: (1) How can the National Hydrography Plus Version 2 (NHDPlus V2) dataset (Horizon Systems Corporation, 2011) be used to generate nationwide near-real-time hydrologic simulations at high spatial resolution? (2) Can such modeling lead to improved emergency response and community resilience? (3) What is a sustainable path from research to operations in terms of flood forecasting (CUAHSI, 2015)? This article begins to address these questions by presenting a computational forecast framework and a web-based visualization application that has the potential to be a part of the national forecasting system with near-real-time high-resolution ensemble flood forecasts. This system forms part of the foundation from which the NFIE can work to achieve the stated goals.

Advances in geospatial data, atmospheric and weather data, hydrologic modeling, and computing resources have led to an improved ability to make instream forecasts. There are several essential elements of the National Weather Service's (NWS) Hydrological Ensemble Forecasting Service (HEFS) that can be used as a model for a flood forecasting system. The main elements include: (1) a meteorological ensemble forecast; (2) a hydrologic processor that inputs the meteorological data into hydrologic, hydraulic, and reservoir models; (3) a hydrologic ensemble postprocessor to account for total hydrologic uncertainty; and (4) an ensemble verification service to identify the skill and error in the forecast (Demargne *et al.*, 2014). In this article, we demonstrate a method to improve the spatial resolution of the hydrologic routing portion of a streamflow prediction system.

The NWS hosts a web-based hydrologic prediction system known as the Advanced Hydrologic Prediction Service (AHPS). The predictions are created using the

Advanced Weather Interactive Processing System (AWIPS) which consists of automated gage data, satellite data, Doppler radars, weather observation stations, advanced computer models, and super computers. The AHPS can display the forecasted streamflow, the forecasted flood level, the probability of flooding, and maps of the areas potentially affected by the flooding at many of the 3,600 forecast stations nationwide. These predictions can range from hours to months in advance (Mcenery *et al.*, 2005; NOAA, 2015).

In addition, global weather forecasts and hindcasts are available from multiple sources. Dr. David Maidment's group at the University of Texas is working on developing a high-resolution forecasting system using United States (U.S.)-based models and datasets for NFIE (Salas *et al.*, 2014; Maidment, 2015). Therefore, in this article, we focus on the global runoff datasets generated by the European Centre for Medium-Range Weather Forecasts (ECMWF). The ECMWF global gridded runoff prediction dataset includes surface and subsurface runoff depth in meters derived from the Tiled ECMWF Scheme for Surface Exchanges over Land with a revised land surface Hydrology (HTESSEL) land surface model (Balsamo *et al.*, 2009; ECMWF, 2011). Modeling and forecasting is intrinsically uncertain (Buizza *et al.*, 2005; Pappenberger and Beven, 2006; Slingo and Palmer, 2011; Beven *et al.*, 2015). The ECMWF produces Numerical Weather Prediction (NWP) ensemble forecasts as a method for better representing and communicating the uncertainty in the forecast (Beven and Cloke, 2012). The ECMWF also produces a dataset for historical runoff that is an output of the global atmospheric reanalysis ERA-Interim and begins in 1979 and extends to the present with near-real-time updates (Dee *et al.*, 2011).

The Global Flood Awareness System (GloFAS), developed by ECMWF and the Joint Research Centre of the European Commission, is a coupled hydro-meteorological model that generates ensemble streamflow predictions for large-scale river basins globally for up to 15 days in advance (Alfieri *et al.*, 2013). The GloFAS grid cell size (0.1°) is too large for determining local impacts for watersheds smaller than $10,000 \text{ km}^2$; hence, high-resolution hydrologic, hydraulic, and flood impact models are required for more detailed forecasts capable of producing actionable information at the local level.

Although the AHPS provides national coverage and GloFAS provides global coverage, both are limited to a relatively coarse spatial resolution for streamflow predictions. The resolution gap can be bridged by routing the ECMWF runoff predictions through the NHDPlus V2 stream network using the Routing Application for Parallel ComputatIon of

Discharge (RAPID) model (David *et al.*, 2011; David, 2013). This ECMWF-RAPID integration generates streamflow predictions at a local scale corresponding to the U.S. NHDPlus V2 dataset. The ECMWF-RAPID integration has the potential to be incorporated into the national forecast system to increase the resolution to nearly 2.7 million forecast points with predictions as an ensemble. However, before this can occur, several improvements need to occur in the current ECMWF-RAPID system such as initialization, calibration, and adding reservoir routing. This system has been developed for use at the NFIE Summer Institute (Tuscaloosa, Alabama, summer 2015) where further testing, comparison, and application will continue to provide a pathway that transforms the spatial density of forecasting while incorporating ensemble forecasts that can better communicate uncertainties involved.

METHODS

The goal of this work was to produce higher spatial resolution streamflow predictions and provide an intuitive method for viewing the predictions. This was achieved by developing a preprocessing method using ArcGIS tools to downscale both the ECMWF gridded runoff prediction dataset and ERA-Interim reanalysis gridded runoff dataset to the catchment level to create input data for RAPID. The RAPID model is run for the period 1980-2014 using the ERA-Interim runoff to generate approximate historical streamflow and return period data. Then, the RAPID model is run for each of the 52 ECMWF forecasts in the ensemble every 12 h using a distributed computational workflow. When the process repeats, future forecasts are initialized using the average of the forecasted streamflows from the previous forecast for each reach. Warning points are generated at the locations where the forecasted streamflow exceeds the estimated 20-year, 10-year, and 2-year return periods. Finally, a web app was created using Tethys Platform (Jones *et al.*, 2014; Swain, 2015) to manage and visualize the high-resolution stream forecasts and warning points for decision-makers in a standardized, intuitive format. The application also incorporates AHPS predictions and U.S. Geological Survey (USGS) observed streamflows for comparison and validation so that improvements can be made in subsequent iterations of the system.

The ECMWF global runoff forecast ensemble and ERA-Interim global reanalysis runoff dataset were used in this research. The ECMWF global runoff ensemble provides 52 separate predictions every 12 h that estimate cumulative runoff depths. The first 51

predictions represent an ensemble of equally probable conditions and are created at a lower resolution on a ~ 0.28 -degree grid cell (up to day 10, ~ 0.56 -degree grid cell thereafter) with a 6 h accumulated runoff time step and a 15-day lead time. The 52nd forecast is a deterministic “best estimate” solution produced at a higher resolution with a ~ 0.14 -degree grid cell and a varying time step of accumulated runoff with a 10-day lead time. The ERA-Interim dataset used has a T511 grid (~ 39 km grid cell), has a daily time step, and spans the years 1980-2014.

The RAPID model was used to route the ECMWF runoff through the NHDPlus V2 stream networks. RAPID is an open source model used to route runoff of surface and groundwater inflow to rivers downstream with any density stream network (David *et al.*, 2011). The NHDPlus V2 dataset combines the National Hydrography Dataset, the Watershed Boundary Dataset, and the National Elevation Datasets (NED) and adds attributes that define stream order and facilitates rapid stream network traversal and query (Horizons Systems Corporation, 2011). Reach routing with RAPID is based on the traditional Muskingum routing method which has two main parameters k and x , where k is a storage constant with a time dimension and x characterizes reach properties that contribute to wave diffusion, is dimensionless, and is stable from 0 to 0.5 (Cunge, 1969).

The high-resolution nature of the NHDPlus V2 stream network necessitates geoprocessing to convert the ECMWF runoff forecasts into the format required to route with RAPID. To this end, a collection of free and open source Python tools has been developed as geoprocessing tools for ArcGIS. The NHDPlus V2 dataset is conveniently available as an ArcGIS geodatabase. The ArcGIS tools are used to prepare inflow to each reach in the stream network by converting the ECMWF model forecast from a gridded runoff depth to runoff volume using the NHDPlus V2 catchments (we will refer to this process as “down-scaling”), and by generating other ancillary inputs for the RAPID flood routing model to ensure smooth and efficient data transfer between models. The following preprocessing operations are performed using the ArcGIS tools (version 10.3 or greater):

1. Create the stream network connectivity file by traversing the NHDPlus V2 network and considering upstream and downstream connectivity.
2. Calculate the Muskingum parameters (k and x) based on stream lengths and flow wave celerity input, and create the Muskingum parameter files for RAPID. The wave celerity is the speed at which the water flow wave propagates in a river channel. The k parameter in the Muskingum method can be computed based on the value of

the flow wave celerity using the equation $k = L/c$, where L is the length of the river reach, and c is the celerity of the flow wave going through it. More information on the relationship between flow wave propagation and the Muskingum method can be found in Cunge (1969) or in David *et al.* (2011).

3. Create a weight table by overlaying the NHDPlus V2 catchments on ECMWF runoff grids. The weight table describes the area (A_i) of the runoff grid cell (i) that overlays each catchment (Figure 1).
4. Create the inflow file for the stream network by computing the weighted average runoff volume from the ECMWF forecast at each time step for the catchment that corresponds to each stream reach.

$$V_j = \sum_{i=1}^n A_i \times R_{ji} \quad (1)$$

where V_j is the runoff volume in m^3 at time step j , A_i is the area of the catchment in m^2 in grid

cell i , and R_{ji} is the runoff depth in meters in the grid cell i at time step j , and n is the number of grid cells that contribute runoff to the catchment.

The ArcGIS preprocessing workflow Steps 1-3 produces static files, so they only need to be performed once for each watershed that is incorporated in the system. The resulting inflow file from Step 4 is a Network Common Data Form (NetCDF) file compatible with RAPID and contains the incremental runoff defined at each time step of the forecast for each catchment in the watershed.

Computational Forecast Framework

Before running the forecasts, the ERA-Interim data is downscaled and then routed in RAPID from 1980 to 2014 to produce 35 years of daily streamflow estimates. From this data, using a simple Weibull distribution (Benson, 1962) with the partial duration series method, estimates for return periods are

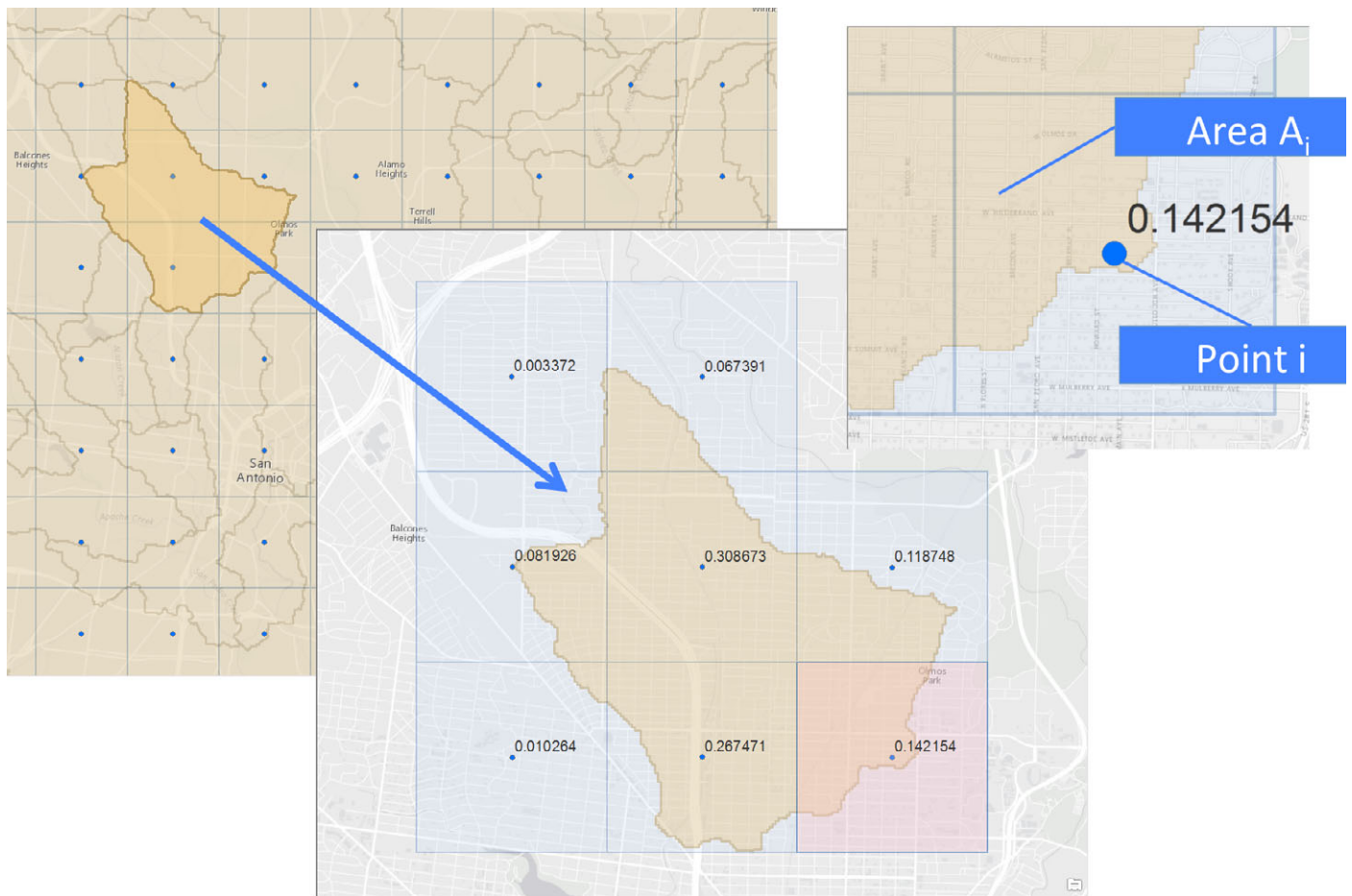


FIGURE 1. Weight Table Derivation Illustration with Runoff Grid and Catchment (Step 3).

generated for each of the reaches in the continental U.S. Due to time and computational constraints, the NFIE Mississippi Region return period data were not generated in this study.

The overall downscaling and routing process in prediction mode uses a parallel computational forecast framework illustrated in Figure 2. In Step 1 the ECMWF forecast ensemble is retrieved. The downscaling and RAPID routing are performed in Step 2 and the simulations are distributed across multiple machines in order to decrease computation time. The computation can be summarized as two processes: Step 2a the weight table preprocessed earlier by the Esri tools is used to downscale the ECMWF runoff forecast and Step 2b the forecasted runoff is routed through the reaches using RAPID. From the average of all of the forecasts generated from Step 2, in Step 3 the RAPID streamflow initialization file is created for the next prediction to then be used in Step 2b. In Step 4, the return period data generated from ERA-Interim data are used to generate warning points where the average or one standard deviation above the average of forecasted flow from Step 2 exceeds the return period. In Step 5 the instream forecasts and warning points are deposited in a CKAN data store. CKAN is an open source data

management portal that streamlines the process sharing and publishing data (Open Knowledge Foundation, 2013). Finally, in Step 6, a Tethys Platform web app downloads the forecasts and warning points from CKAN to display them to the user.

The ECMWF ensemble runoff forecasts are published every 12 h, introducing a constraint on the computation time of the downscaling and routing process. If the entire process were to require an execution time longer than 12 h, computations on updated datasets would be delayed, causing a lag. To prevent this situation on the large national-scale dataset, efficient computation methods are required to downscale the forecasts in a timely manner and enable the system to run operationally.

The method used to improve computational efficiency was to distribute the computations between computer processors. The distribution method can be applied to a single server with multiple processing cores or to a cluster of computers with shared processing cores as would be available via cloud services such as Amazon Web Services (AWS) or Microsoft Azure. In this study, both a local compute cluster and an AWS compute cluster, each with 52 cores, were used to perform computations. Using this compute cluster, it was possible to simultaneously downscale

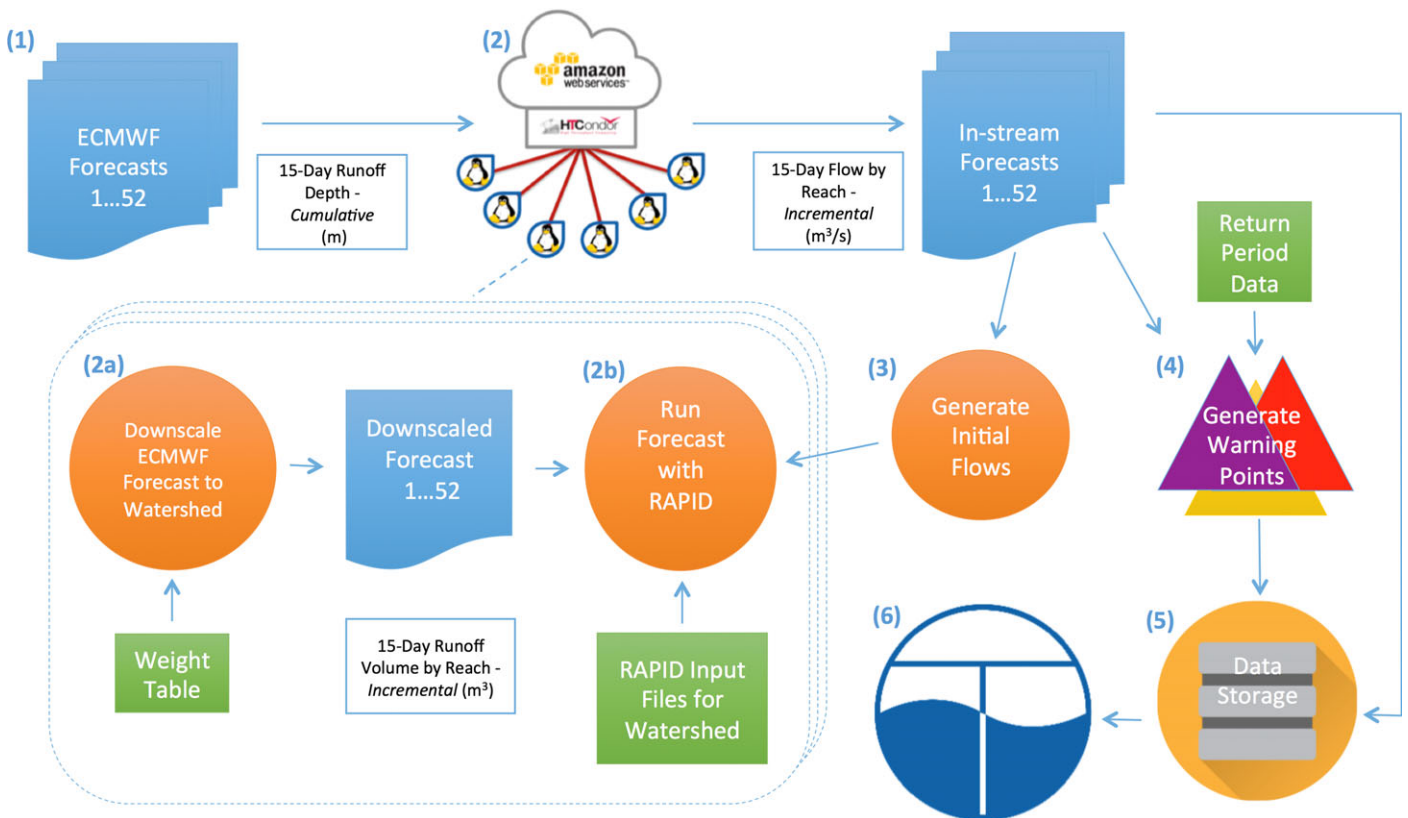


FIGURE 2. Schematic of ECMWF-RAPID Downscaling and Routing Process in Prediction Mode.

each of the 52 ECMWF runoff forecasts in the prediction ensemble for each individual watershed. HTCondor and a custom Python library named CondorPy (<https://pypi.python.org/pypi/condorpy>) were used to distribute the computations in Step 2 in the workflow. HTCondor is a batch-scheduling and resource management software that distributes jobs to computing resources based on resource availability (Buyya *et al.*, 2013). CondorPy interfaces with HTCondor and is used to facilitate programmatic job creation and submittal.

Streamflow Prediction Tool Web App

The final product of the downscaling and routing process is 52 NetCDF files, one for each member of the ensemble forecast, that contain the predicted hydrographs for each reach of the watershed. This information needs to be communicated to end-users in a comprehensible format so that it can inform decision-makers and the public at large (Pappenberger *et al.*, 2013). We addressed this need by developing a web application or web app called the “Streamflow Prediction Tool” using Tethys Platform. The web app medium is an effective way to share new developments in water resources modeling, informatics, and decision support. Tethys Platform bridges the divide that prohibits many water resources scientists and engineers from developing web apps by providing (1) a suite of free and open source software that addresses the unique data and computational needs common to water resources web app development, (2) a Python software development kit for incorporating the functionality of each software element into web apps, and (3) a customizable web site that can be used to deploy the finished web apps. Among the software projects included in the Tethys Platform are GeoServer, 52 North WPS, PostgresSQL with PostGIS, OpenLayers, Google Maps, Highcharts JS, and HTCondor (Jones *et al.*, 2014; Swain, 2015).

We designed the Streamflow Prediction Tool to consume the ensemble stream forecast produced in Step 2 and the warning points produced in Step 4 from the CKAN data store in Step 6 as shown in the workflow in Figure 2. This allows the computation and visualization to operate independently, offering flexibility in deployment of the system. The web app automatically maintains a cache of the most recent weeks’ worth predictions and warning points on the server to facilitate faster access to the data. The netCDF4-python and NumPy Python modules are used to extract, compile statistics, and analyze the stream forecast prediction ensemble.

The Streamflow Prediction Tool provides an intuitive user interface that allows the easy lookup and visualization of results (Figure 3). GIS visualization of the stream network and other spatial layers was accomplished through a coupling of GeoServer 2.7.0 spatial data publishing and OpenLayers 3.2.1 spatial mapping systems. Stream layers are served as an Open Geospatial Consortium Web Feature Service (OGC-WFS) and the other spatial layers are served by GeoServer as OGC Web Map Services (OGC-WMS) (Michaelis and Ames, 2012; Open Geospatial Consortium, 2015a, b). OpenLayers is used to query GeoServer using OGC-WFS and OGC-WMS and display the layers in an interactive map. On the map, clicking on a reach will look up the forecast for that reach.

The ensemble forecast is summarized and displayed on a Highcharts plot (www.highcharts.com), which provides interactive visualization. In the plot, the black line represents the high-resolution forecast and the green bands represent the uncertainty of all 52 members of the prediction ensemble. Where USGS stream gages exist, that serve observed data, the data are added to the chart for preceding days on selected stream reaches, which can be particularly useful for evaluating performance when viewing older forecasts for which now the observed data exists. If available, AHPS stations provide streamflow predictions or observed data for some of the reaches. The observed USGS gage data for the station will appear as a blue dashed line and the AHPS station data will appear as a purple dashed line on the plot. Additionally, the estimated return period data is also shown on the chart with colored bands. The yellow band represents the range of flows between the 2-year and 10-year return period streamflows. The red band represents the range of flows between the 10-year and 20-year return period streamflows. And the purple band represents flows exceeding the 20-year return period streamflow. Many of these elements are demonstrated in the chart in Figure 3.

The app demonstrates a novel approach to displaying high-density stream networks, using stream order to dynamically load streams based on the zoom level, similar to how the pyramiding technique is applied to high-resolution images. When the zoom level is set to the full extent of the watershed, only reaches with higher stream orders are displayed as illustrated by Figure 4a. A mid-range zoom level will result in reaches with mid-range stream orders being added to the display as shown in Figure 4b. On the last zoom level, reaches of all stream orders are displayed as shown in Figure 4c.

The app also presents a new method for displaying high-resolution warnings at both an overview

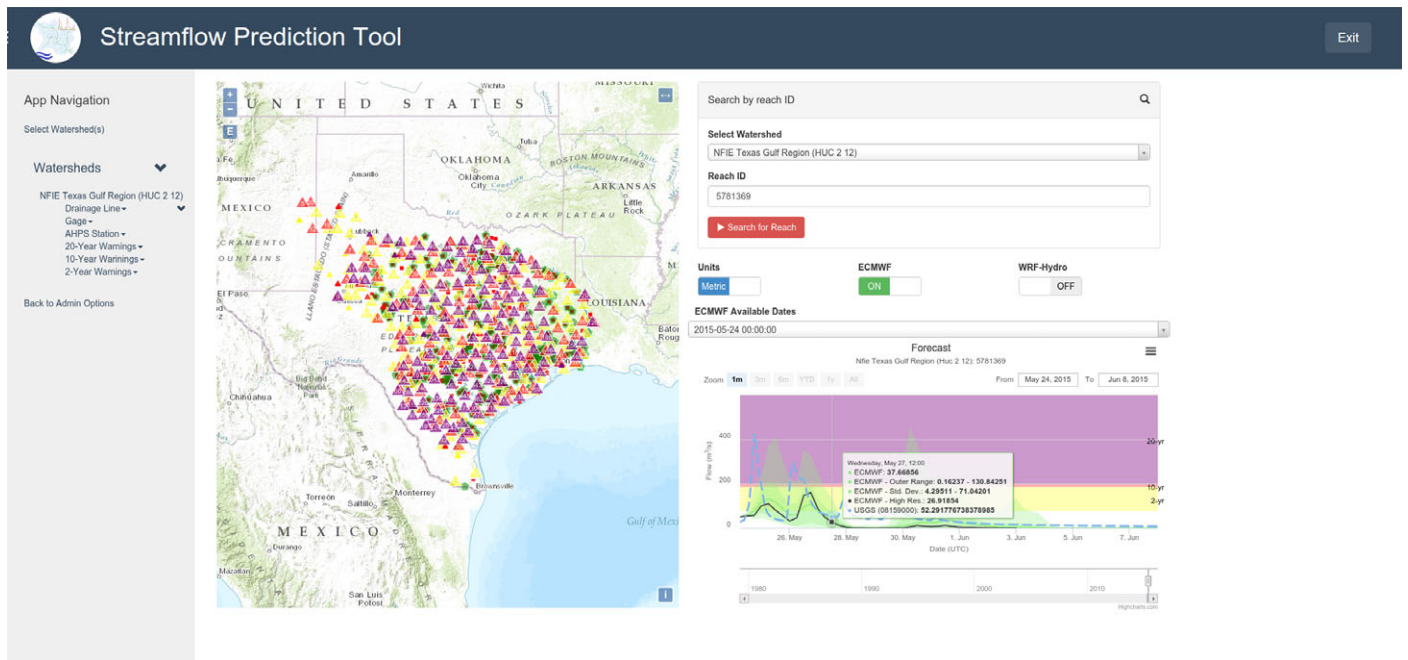


FIGURE 3. Screenshot of Streamflow Prediction Tool App.

and at the level of each individual river reach. It facilitates display at the NFIE Region scale by combining the points within a close proximity together and representing the group as a single icon with a count of the number of points included. Additionally, it divides the warning points into three main groups corresponding to the peak flow of the average (represented by larger triangles) or the standard deviation above the average (represented by smaller triangles) of the ensemble forecast that exceeded the return period. The warning with the highest return period is the only one generated for the reach if any warnings exist. Yellow triangles represent exceedance of the 2-year flow, red triangles represent exceedance of the 10-year flow, and purple triangles represent exceedance of the 20-year flow. Examples are shown for the warnings generated for each return period during the May 2015 flooding in Texas in Figure 5.

VALIDATION

Setup

We used the watersheds from the NFIE to validate the performance of the downscaling and routing computation framework and the Streamflow Prediction Tool app. We performed timing tests using a local compute cluster as well as using the Amazon Web

Service computing cloud. A map of all of the NFIE regions is shown in Figure 6.

COMPUTATIONAL FORECAST FRAMEWORK

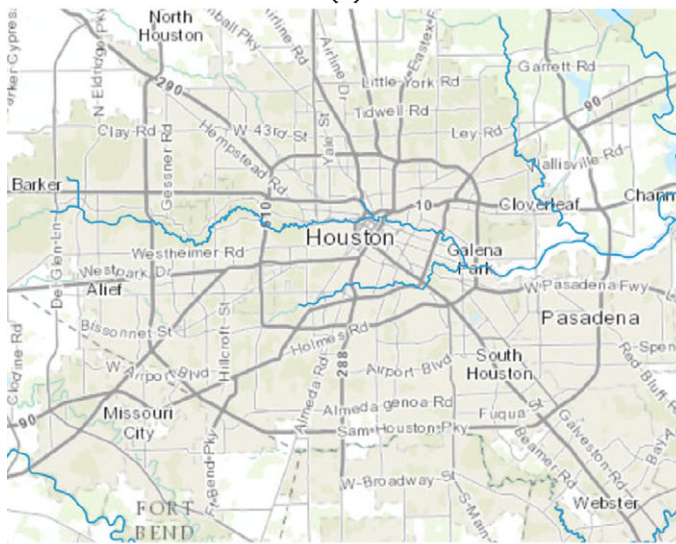
We compared the computational performance of computing using a local compute cluster and the Amazon Web Services (aws.amazon.com). Additionally, we estimated the results computing serially based off of the individual computation times as a means of measuring the time saved using distributed computing.

It is apparent from Table 1 that the use of the distributed computing significantly reduces computation time in all cases. In fact, distributed computing is essential for meeting the 12-h (43,200 s) constraint for an operational system that covers the entire U.S.

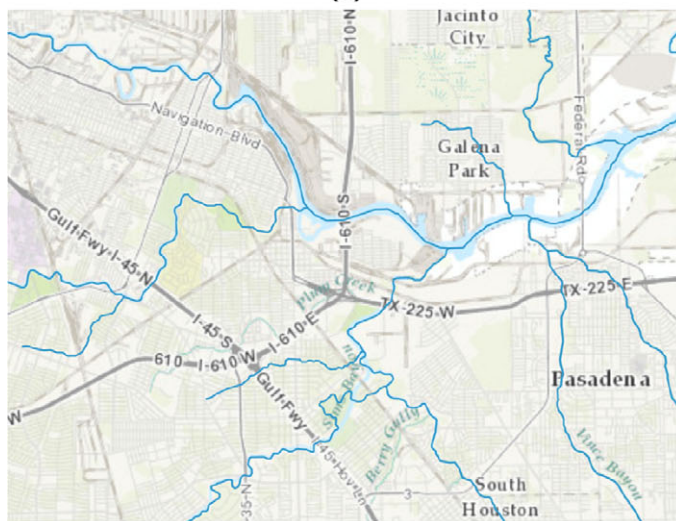
The computation time *vs.* the number of reaches from Table 1 is shown in Figure 7 with polynomial order 2 trend lines having coefficients of determination of 0.99. As expected, the computing time increases with the size of the computing problem (i.e., the number of river reaches). Note that the good fit with a second order polynomial order suggests that the solving procedure might include two inner loops. This is likely due to the default RAPID option for solving linear systems (an iterative Richardson method) being used in this study. Alternative approaches using noniterative solvers (David *et al.*,



(a)

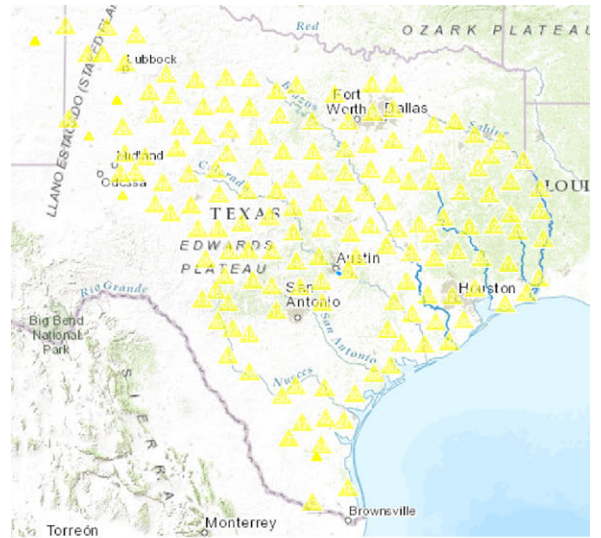


(b)

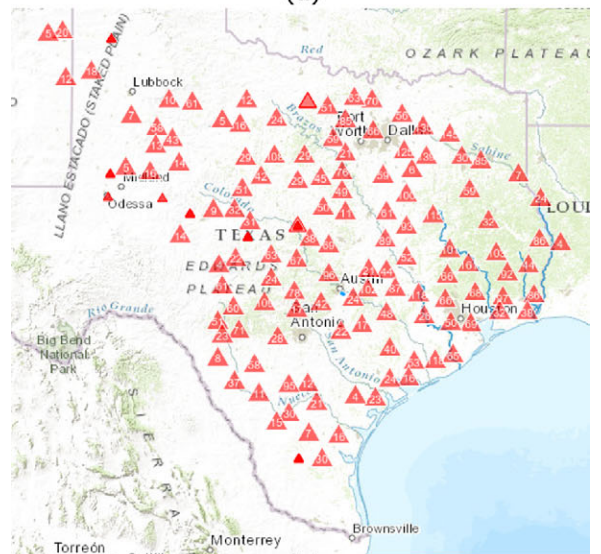


(c)

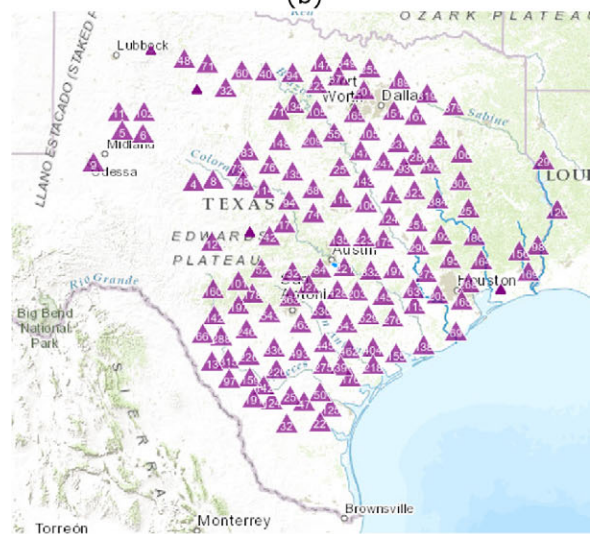
FIGURE 4. Stream Network Zoom Levels with Dynamic Stream Densities: (a) High Stream Order, (b) Mid-Level Stream Order, (c) Complete Stream Network.



(a)



(b)



(c)

FIGURE 5. Warning Points for Texas Floods for May 2015: (a) 2-Year Warnings, (b) 10-Year Warnings, (c) 20-Year Warnings.



FIGURE 6. National Flood Interoperability Experiment (NFIE) Regional Divisions.

TABLE 1. Results of Computation Time Based on Area and the Number of Reaches.

Watershed Name	Area (sq km)	Number of Reaches	Compute Time (seconds)		
			Serial	Amazon Web Services	Local Cluster
NFIE Souris-Red-Rainey Region	213,488	29,053	7,343	49	141
NFIE Rio Grande Region	564,840	55,854	9,083	64	175
NFIE New England Region	169,445	65,858	10,906	63	210
NFIE Texas-Gulf Region	464,493	66,373	10,417	68	200
NFIE Great Basin Region	367,058	96,269	8,589	43	165
NFIE Great Lakes Region	324,434	104,645	15,873	115	305
NFIE Mid-Atlantic Region	277,755	125,398	16,900	109	325
NFIE California Region	421,995	140,759	22,367	164	430
NFIE Colorado Region	660,454	187,010	28,105	210	540
NFIE Pacific Northwest Region	814,493	231,806	24,325	180	468
NFIE South Atlantic-Gulf Region	675,734	360,175	41,083	319	790
NFIE Mississippi Region	3,302,913	1,242,008	316,930	1,558	6,095

2015) could help in decreasing computing time for this application.

This is especially true of the local cluster, where the rate of computation of the Mississippi region is half that of the Colorado region. The slowdown is not likely caused by the processing algorithms, because the Amazon cloud and local compute cluster curves do not slow at the same rate. There are various factors at play, but the slowdown may be caused by differences in hardware or possibly overhead of the computing environments. The efficiency of the system running on the local cluster could be improved significantly by

dividing the Mississippi watershed into two or three watersheds with 400,000-600,000 reaches each.

Streamflow Prediction Tool Web App

The GIS visualization capabilities of the Streamflow Prediction Tool app successfully displayed all of the NFIE regions. The loading time is relatively fast, averaging around 7 s to load. Displaying large high-density stream networks is currently limited to stream networks defined by the NHDPlus V2 dataset

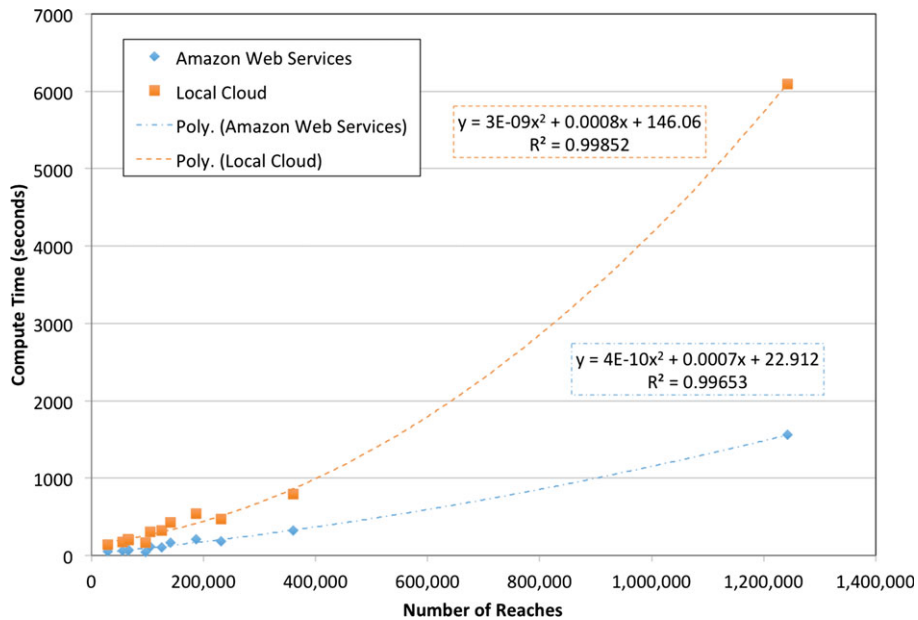


FIGURE 7. Computation Time vs. Number of Reaches.

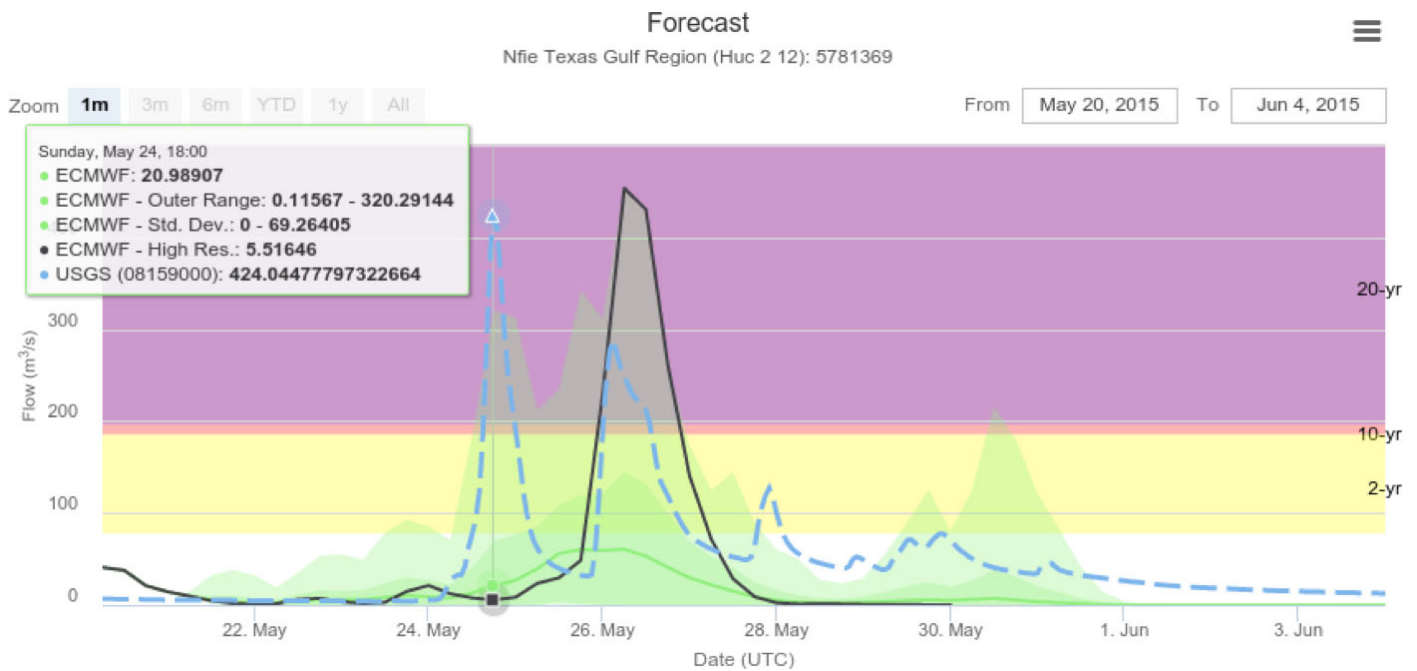


FIGURE 8. Forecast Comparison Beginning May 20, 2015 (four days before flood).

as it has the stream order defined, making dynamic display of stream networks possible.

POTENTIAL NFIE COMPARISON

From May 24 to May 28 of 2015 major flooding occurred in the Austin, Texas region. In this section,

we will demonstrate how participants at NFIE can compare the forecasted results from ECMWF-RAPID with USGS stations using an example at the reach with COMID 5781369 in Onion Creek (coordinates 30.175347, -97.656148) and USGS station 08159000.

Four days before the beginning of the flood event, the ECMWF-RAPID forecast shows that there is an event that will occur as shown in Figure 8. However, the predicted average of the magnitude and timing of the event is significantly off. In addition, the two

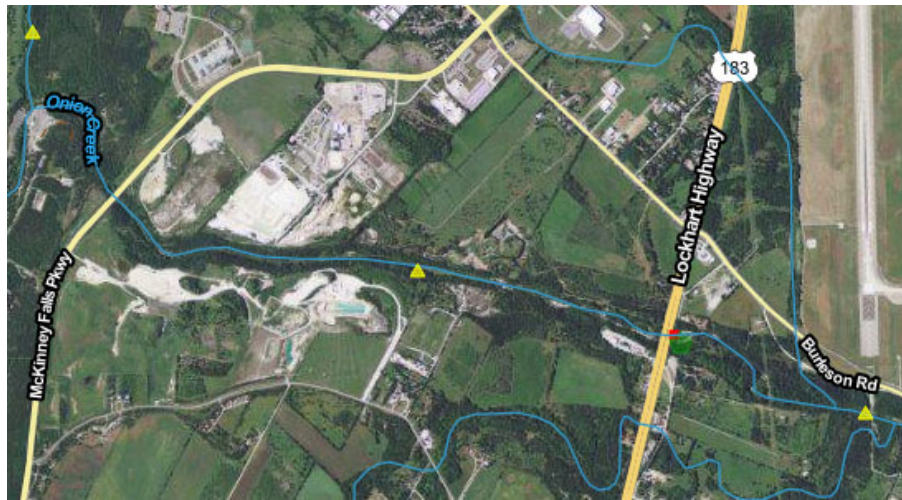


FIGURE 9. Onion Creek Warning Point Produced, May 20, 2015.

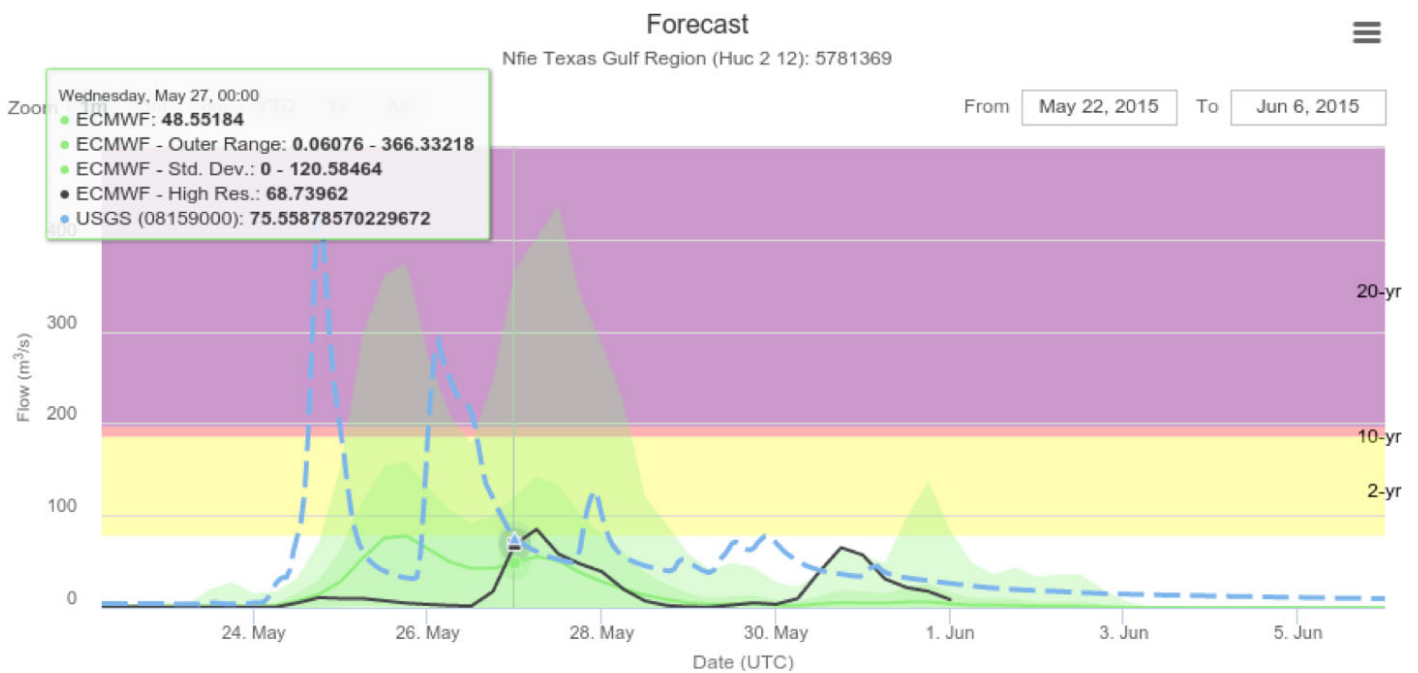


FIGURE 10. Forecast Comparison Beginning May 22, 2015 (two days before flood).

events seem to be merged into one event. Nonetheless, the high-resolution ensemble seems to capture the timing of second event, though the peak flow is overestimated. Due to the under prediction of the magnitude of both storms, only a minor 2-year warning was developed as shown by the small yellow triangle in the middle of the stream in Figure 9.

Then, at two days out, the ECMWF-RAPID forecast mean is beginning to show two distinct events as shown in Figure 10. However, while the timing and magnitude of the events are closer, they are still predicting flood peaks lagging behind what actually happened with peaks captured only in the upper extremes of the

prediction. Similar to the forecast from 20 May, only a minor 2-year warning was developed as shown by the small yellow triangle in the middle of the stream in Figure 11.

On the day of the beginning of the flood event, the forecast more closely aligns with the timing and magnitude of the event. However, for both peaks, the mean predicted streamflow is significantly below the actual streamflow as shown in Figure 12. Because the predicted series that was a standard deviation above the average had a peak flow above the 20-year threshold, a small purple warning point was generated. Also, because the peak flow of the average

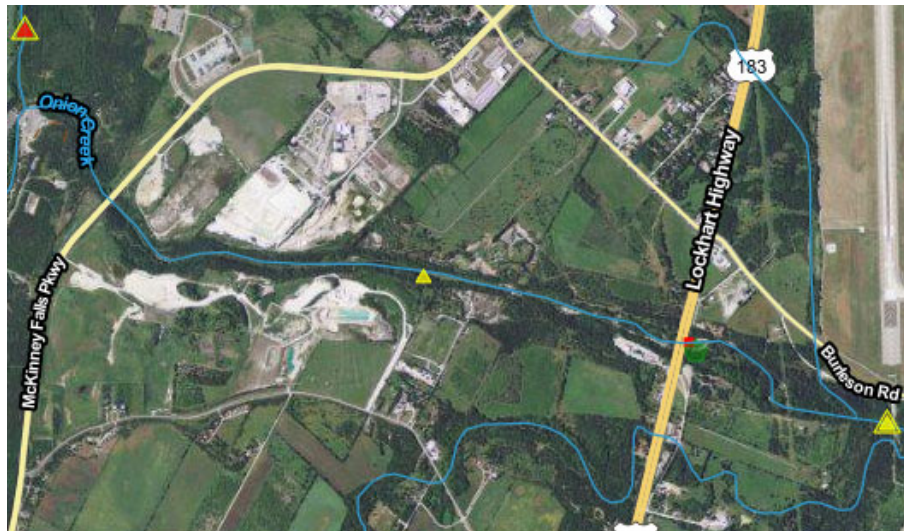


FIGURE 11. Onion Creek Warning Point Produced, May 22, 2015.

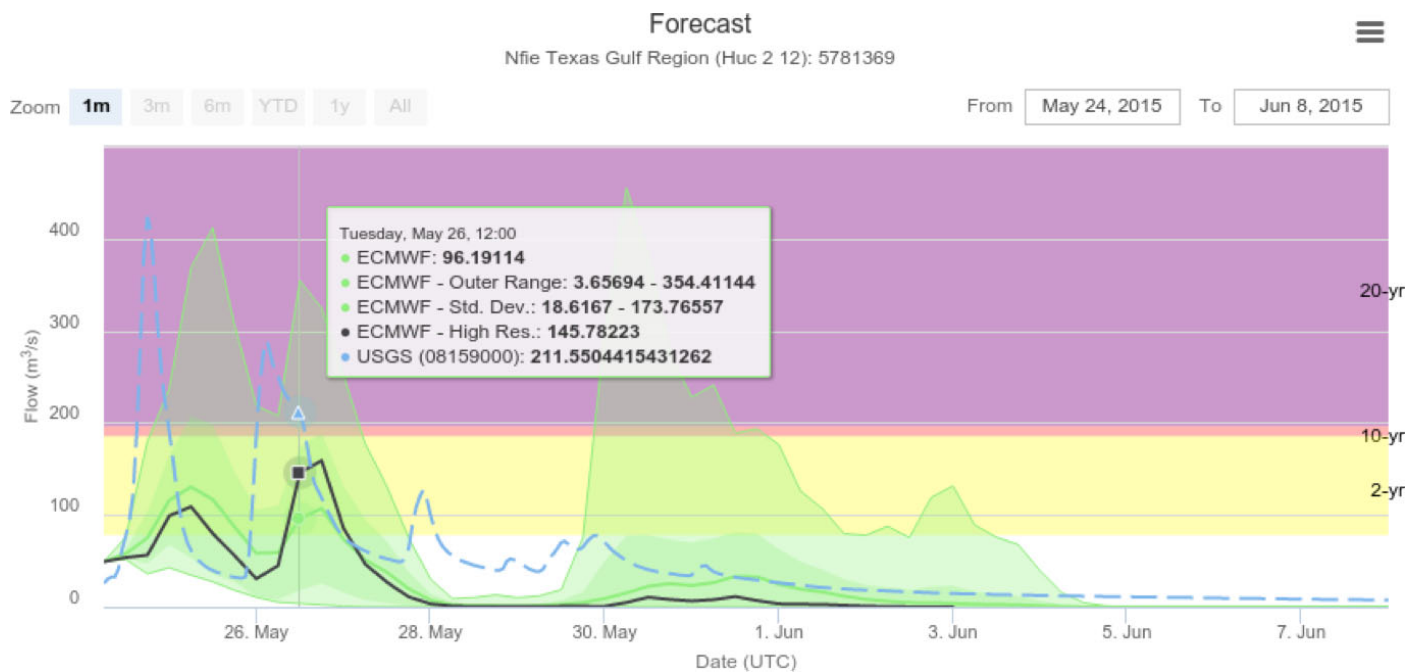


FIGURE 12. Forecast Comparison Beginning May 24, 2015 (day of flood).

series was above the 2-year threshold a large yellow triangle was also generated. Both triangles are shown in the center of the stream in Figure 13.

As expected, the case study illustrates that the predictions become more accurate, the closer to the time of the flooding event. It is clear that the system needs further evaluation and improvement to accurately capture events over the broad stream network. However, the framework does provide participants at NFIE and others with the ability to begin providing a more widespread evaluation of the value of this downscaled, high-density hydrologic forecast.

CURRENT LIMITATIONS AND POTENTIAL IMPROVEMENTS

Using the method and tools developed by this research and the evaluations that can be performed through NFIE, the creation of additional tools and improvements to the current system is needed. The most pressing need is improving the streamflow initialization. The current method begins at zero flow for the very first prediction and then initializes the next prediction from the average of previously



FIGURE 13. Onion Creek Warning Point Produced, May 24, 2015.

predicted flows. This method requires multiple forecast runs in order for the model to “spin-up” the streamflow in the rivers. Additionally, initializing from predicted streamflows may not render the best results for what the actual streamflow will be. As such, data assimilation methods will need to be incorporated into the model to provide better estimates for initial streamflow at the beginning of each prediction cycle. This can include incorporating real-time stream gage data or running the model with post-processed reanalysis meteorological data to improve the streamflow predictions.

Additional known improvements include items such as methods for calibrating the model as well as a method for including reservoir releases in the modeling process. New tools could be created based on the high-resolution stream forecasts. These new tools could involve ideas such as predictive flood index maps derived from the instream forecasts or improved analysis methods to determine how likely a flood will occur or when and how to warn the public of oncoming floods.

CONCLUSIONS

The creation of a flood warning system that can provide predictive information for floods days and even weeks in advance at a high spatial resolution at a national-scale is within reach. The implementation developed by this research provides an important contribution to the national aspirations to create such a system. By downscaling runoff forecasts generated by the ECMWF using Esri’s RAPID toolbox and

routing the runoff using the RAPID model, we were able to produce high-density ensemble national-scale stream forecasts. However, for this system to be fully functional, improvements in initialization and calibration as well as the addition of reservoir operations need to be incorporated.

With Tethys Platform, we developed an interface to display the high-density streamflow forecasts to decision-makers that includes the ability to compare to existing NWS forecasts and observations at USGS gages where these data streams exist. This tool gives decision-makers information from NetCDF datasets containing streamflow forecasts for hundreds of thousands of reaches, including a statistical summary of the potential streamflow up to 15 days in advance. Additionally, the tool displays warning points where the 20-year, 10-year, and 2-year thresholds have been exceeded in the predictions for each individual reach. These warnings make the catchment level predictions applicable at a national-scale. As such, the app simplifies the data access and interpretation for decision-makers.

SOFTWARE AVAILABILITY

Esri’s RAPID Toolbox is available under the Apache License Version 2.0 and the source code is available on GitHub at the repository: <https://github.com/Esri/python-toolbox-for-rapid>.

The code for automating the process of downscaling the ECMWF prediction datasets using Esri’s RAPID toolbox and RAPID is available under the

BSD 3-Clause License and the source code is available on the GitHub repository: https://github.com/erdc-cm/spt_ecmwf_autorapid_process.

The Streamflow Prediction Tool app is available under the BSD 3-Clause License and the source code is available on GitHub at the repository: https://github.com/erdc-cm/tethysapp-streamflow_prediction_tool. The Streamflow Prediction Tool app was developed using Tethys Platform. The source code for Tethys Platform is distributed under the BSD 2-Clause license and is available through GitHub at <https://github.com/tethysplatform/tethys>. Documentation and a live demo of Tethys Platform can be accessed at <http://www.tethysplatform.org>.

The source code for RAPID is available under a Berkeley Software Distribution (BSD) 3-clause license on the RAPID GitHub repository at <https://github.com/c-h-david/rapid> and further information can be found on the RAPID website at <http://rapid-hub.org>.

ACKNOWLEDGMENTS

This research is based upon work supported by the National Science Foundation under Grant No. 1135483. Cédric H. David was supported by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We would also like to express appreciation to Curtis Rae for his assistance in the GIS preprocessing of the NFIE regions.

LITERATURE CITED

- Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger, 2013. GloFAS—Global Ensemble Streamflow Forecasting and Flood Early Warning. *Hydrology and Earth System Sciences* 17(3):1161-1175.
- Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A.K. Betts, 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.
- Benson, M.A., 1962. Plotting Positions and Economics of Engineering Planning. *Journal of the Hydraulics Division* 88(6):57-71.
- Beven, K., H. Cloke, F. Pappenberger, R. Lamb, and N. Hunter, 2015. Hyperresolution Information and Hyperresolution Ignorance in Modelling the Hydrology of the Land Surface. *Science China Earth Sciences* 58(1):25-35.
- Beven, K.J. and H.L. Cloke, 2012. Comment on “Hyperresolution Global Land Surface Modeling: Meeting a Grand Challenge for Monitoring Earth’s Terrestrial Water” by Eric F. Wood et al. *Water Resources Research* 48(1), DOI: 10.1029/2011WR010982.
- Buizza, R., P. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005. A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Monthly Weather Review* 133(5):1076-1097.
- Buyya, R., C. Vecchiola, and S.T. Selvi, 2013. *Mastering Cloud Computing: Foundations and Applications Programming*. Morgan Kaufmann, Waltham, Massachusetts.
- CUAHSI, 2015. “National Flood Interoperability Experiment.” The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI). <https://www.cuahsi.org/NFIE>, accessed February, 2015.
- Cunge, J.A., 1969. On the Subject of a Flood Propagation Computation Method (Muskingum Method). *Journal of Hydraulic Research* 7(2):205-230.
- David, C.H., 2013. RAPID v1.4.0, Zenodo, DOI: 10.5281/zenodo.24756.
- David, C.H., J.S. Famiglietti, Z.-L. Yang, and V. Eijkhout, 2015. Enhanced Fixed-Size Parallel Speedup with the Muskingum Method Using a Trans-Boundary Approach and a Large Sub-Basins Approximation. *Water Resources Research* 51:1-25, DOI: 10.1002/2014WR016650.
- David, C.H., D.R. Maidment, G.-Y. Niu, Z.-L. Yang, F. Habets, and V. Eijkhout, 2011. River Network Routing on the NHDPlus Dataset. *Journal of Hydrometeorology* 12(5):913-934.
- Dee, D.P., S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer, L. Haimberger, S.B. Healy, H. Hersbach, E.V. Hölm, L. Isaksen, P. Källberg, M. Köhler, M. Matricardi, A.P. McNally, B.M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart, 2011. The ERA-Interim Reanalysis: Configuration and Performance of the Data Assimilation System. *Q.J.R. Meteorological Society* 137:553-597, DOI: 10.1002/qj.828.
- Demargne, J., L. Wu, S.K. Regonda, J.D. Brown, H. Lee, M. He, D. Seo, R. Hartman, H.D. Herr, M. Fresch, J. Schaake, and Y. Zhu, 2014. The Science of NOAA’s Operational Hydrologic Ensemble Forecast Service. *Bulletin of the American Meteorological Society* 95:79-98, DOI: 10.1175/BAMS-D-12-00081.1.
- ECMWF (European Centre for Medium Range Weather Forecasts), 2011. “IFS Documentation – Cy37r2 Operational Implementation 18 May 2011 - Part II: Data Assimilation.
- Guha-Sapir, D., R. Below, and P. Hoyois, 2015. EM-DAT: International Disaster Database. Université Catholique de Louvain, Brussels, Belgium.
- Horizons Systems Corporation, 2011. Weaving the National Hydrologic Geospatial Fabric. NHDPlus Version 2. (April 6, 2015).
- Jones, N., J. Nelson, N. Swain, S. Christensen, D. Tarboton, and P. Dash, 2014. Tethys: A Software Framework for Web-Based Modeling and Decision Support Applications. In: D.P. Ames, N.W.T. Quinn, and A.E. Rizzoli (Editors), *Proceedings of the 7th International Congress on Environmental Modelling and Software*, June 15-19, San Diego, California, ISBN: 978-88-9035-744-2.
- Maidment, D., 2015. A Conceptual Framework for the National Flood Interoperability Experiment. https://www.cuahsi.org/Files/Pages/documents/13623/nfieconceptualframework_revised_feb_9.pdf, accessed March 2015.
- Mccenery, J., J. Ingram, Q. Duan, T. Adams, and L. Anderson, 2005. NOAA’s Advanced Hydrologic Prediction Service: Building Pathways for Better Science in Water Forecasting. *Bulletin of the American Meteorological Society* 86:375-385, DOI: 10.1175/BAMS-86-3-375.
- Michaelis, C. and D. Ames, 2012. Considerations for Implementing OGC WMS and WFS Specifications in a Desktop GIS. *Journal of Geographic Information System* 4(2):161-167, DOI: 10.4236/jgis.2012.42021.
- NOAA, 2015. Advanced Hydrologic Prediction Service. National Oceanic and Atmospheric Administration’s (NOAA) National Weather Service (NWS). <http://water.weather.gov/ahps/about/about.php>, accessed February, 2015.
- Open Geospatial Consortium, 2015a. Web Map Service. <http://www.opengeospatial.org/standards/wms>, accessed June 2015.

- Open Geospatial Consortium, 2015b. Web Feature Service. <http://www.opengeospatial.org/standards/wfs>, accessed June 2015.
- Open Knowledge Foundation, 2013. DataStore Extension. CKAN v2.2. <http://docs.ckan.org/en/ckan-2.2/datastore.html>, accessed June 2015.
- Pappenberger, F. and K.J. Beven, 2006. Ignorance Is Bliss: Or Seven Reasons Not to Use Uncertainty Analysis. *Water Resources Research* 42, W05302, DOI: 10.1029/2005WR004820.
- Pappenberger, F., H.L. Cloke, D.J. Parker, F. Wetterhall, D.S. Richardson, and J. Thielen, 2015. The Monetary Benefit of Early Flood Warnings in Europe. *Environmental Science & Policy* 51:278-291, DOI: 10.1016/j.envsci.2015.04.016.
- Pappenberger, F., E. Stephens, J. Thielen, P. Salamon, D. Demeritt, S.J. Andel, F. Wetterhall, and L. Alfieri, 2013. Visualizing Probabilistic Flood Forecast Information: Expert Preferences and Perceptions of Best Practice in Uncertainty Communication. *Hydrological Processes* 27(1):132-146.
- Salas, F., M. Somos, D. Gochis, D. Maidment, C. David, K. Tolle, C. Navarro, and R. Corby, 2014. Building Cyberinfrastructure to Support a Real-Time National Flood Model. *In: AGU Fall Meeting Abstracts*, San Francisco, December 15–19, Vol. 1, p. 1207. <https://agu.confex.com/agu/fm14/webprogram/Paper25860.html>, accessed May 2016; https://www.cuahsi.org/Files/Pages/documents/13623/poster_salasetal_buildingcyberinf.pdf, accessed May 2016..
- Slingo, J. and T. Palmer, 2011. Uncertainty in Weather and Climate Prediction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369 (1956):4751-4767.
- Swain, N.R., 2015. Tethys Platform: A Development and Hosting Platform for Water Resources Web Apps. All Theses and Dissertations, Paper 5832. <http://scholarsarchive.byu.edu/etd/5832>, accessed May 2016.