WATER DATA STANDARDS ENABLE NEW MODELS AND APPLICATIONS
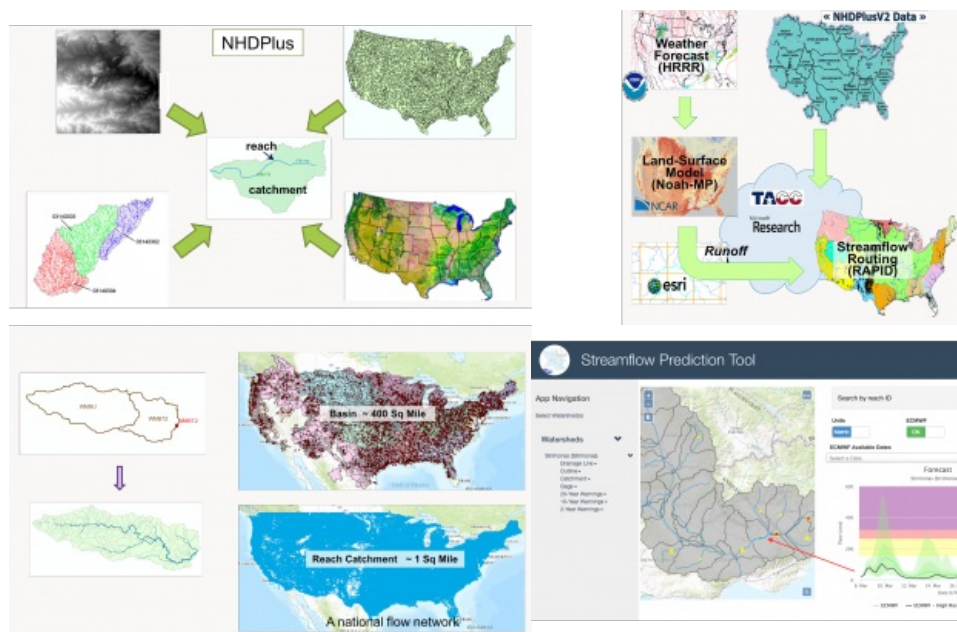
Hydrography for Continental Streamflow Modelling

Land surface hydrography now supports the modelling of continental-scale streamflow at high resolution and in near real time. The US National Hydrography Dataset contains about 2.7 million stream segments averaging 2km in length. This database establishes a national topological network of streams and their drainage areas. The network is now being used in computational models of streamflow that can be run in just 12 minutes for the entire USA, and the approach is also starting to be applied globally.

Imagine a continental-scale hydrographic dataset being used for local-scale operational forecasting of streamflow and flooding in near-real-time. This article outlines the surface water hydrography, datasets, standards and models that have become key parts of the National Water Model (NWM) for the United States National Weather Service (NWS), as well as global outreach to bring this capability to other countries. This starts with a discussion of the National Flood Interoperability Experiment (NFIE).

How It Started: NFIE 2015

As a contribution to the US Open Water Data Initiative, NFIE began organising a collaboration of academia, commercial software and engineering firms, and federal agencies in the fall of 2014, for improving our ability to model the nation's streamflow and flooding (Maidment 2015). NFIE brought together experts and tools in hydrographic and other representations of the nation's landscape, land-surface modelling of hydrometeorological dynamics, and streamflow routing of runoff from rainfall. These tools built on each other, leading to an operational workflow running every three hours, every day. The NFIE Summer Institute 2015, conducted at the National Water Center (NWC) in Tuscaloosa Alabama, was a 7-week laboratory bringing together dozens of researchers from around the world, to start tackling flood mapping, data integrity, calibration, uncertainty, and other topics. Their work will appear soon in a special theme issue of the Journal of American Water Resources Association (JAWRA). The next NFIE Summer Institute 2016 will be held again in June-July at the NWC.

Hydrologic Geospatial Fabric

The foundation data for this process is the *hydrologic geospatial fabric*, or just *geofabric*, a term first introduced in Australia. In the US, this refers to the hydrography, land cover, and surface elevations of the landscape. The hydrography component of the geofabric in NFIE is based on the US National Hydrography Dataset (NHD) and Watershed Boundary Dataset (WBD), developed by the US Geological Survey (USGS) to portray surface water features on The National Map. The NHD was first generated at a scale of 1:100,000, then later densified to 1:24,000. The NHDPlus (now in version 2) is the NHD at 1:100,000 scale, augmented to incorporate the unique drainage area (catchment) associated with each stream segment (reach). At this scale there are about 2.7 million reach-catchment pairs in the continental US, averaging 2km in length and 3 square km in area. The surface elevation started with the National Elevation Dataset (NED) which has 10-metre resolution, but this will transition to the USGS 3D Elevation Program (3DEP), which started collecting 1-metre and 3-metre resolution elevations by Lidar in 2015. The computational models can work with varying resolutions of landscape, but the high-resolution elevation is more critical for flood mapping than it is for streamflow routing; 10-metre resolution or better is needed. The geofabric's land cover is from the National Land Cover Database (NLCD). These are illustrated in Figure 1.

Weather and Land-Surface Models

Streamflow, typically measured in cubic feet (or metres) per second, is controlled by a number of factors, mainly precipitation, runoff, elevation change, and upstream contributions, which might be affected by dams and other flood control structures. The NWS generates a range of forecasts, from near-term (0-15 hours) to long-term (up to 30 days). For the NFIE 2015, the High Resolution Rapid Refresh (HRRR) 15-hour forecast was used. This estimated rainfall and snowfall, divided over a seamless grid of 3km x 3km cells across the country, updated every 3 hours.

For NFIE 2015, this forecast was processed to estimate surface and subsurface runoff to the stream network, by a land-surface model called Noah Multi-Physics (Noah-MP). Based on land-surface descriptions from elevation and land cover, Noah-MP predicted runoff, soil moisture, and many other parameters. The runoff data and NHDPlus stream network were then fed into the streamflow routing model, RAPID (**R**iver **A**pplication for **P**arallel **C**omputat**I**on of **D**ischarge), which estimated the streamflow for each of the 2.7 million NHDPlus stream reaches. Noah-MP and RAPID were integrated through the WRF-Hydro (Weather Research and Forecast – Hydrological) framework, managed by the National Center for Atmospheric Research (NCAR), and run on a high-performance platform at the University of Texas Advanced Computing Center (TACC). Figure 2 shows the overall workflow.

What a Difference This Made

Before NFIE, the national weather forecasts were developed into streamflow forecasts by thirteen Regional Forecast Centers (RFC). Each RFC had its own streamflow models, which had much coarser spatial resolution than NHD. There are 3600 forecast points and 6600 basins across the country, averaging 400 square miles per basin. Many coastal areas were not included in these forecast points and basins, totaling about 100 million in population. With NFIE, this densified to 2.7 million forecast points and catchments, fully coast-to-coast, for an improvement in forecast points by a factor of 750 (see Figure 3). Moreover, the overall quality of predictions was generally good, particularly when reservoir releases could be taken into account.

The NFIE procedure took a combination of efforts and accomplishments including extensive cloud usage, high-performance computing, consultation with XSEDE experts, and a lot of attention to performance. As a result, depending on the number of processors used, the workflow from ingest of HRRR precipitation forecasts to output from the combined WRF-Hydro/RAPID system could be run in 10-12 minutes, for the 2.7 million stream reaches in the continental US, for each 15-hour forecast. This is the first-ever continental-scale simulation of the national river system as a single network.

Data Exchange Standards

Underlying the workflow are data exchanges to feed the weather forecast (rainfall gridded at 3km x 3km) and NLCD to the Noah-MP land surface model; then to feed the land surface model output (runoff gridded at 3km x 3km, converted to reach-catchment boundaries) to the RAPID streamflow routing model. The data payloads for these exchanges were encoded using the Open Geospatial Consortium (OGC) international standards, network Common Data Form (netCDF) and Climate and Forecast (CF) Convention. NetCDF defines an application programming interface (API) to multi-dimensional array-based datasets, while the CF convention defines rules for handling Earth observation data specifically within netCDF. This is a compact binary data format that is portable across computing platforms.

The project also collected daily observations at about 9000 USGS stream gages operating in real-time across the country, through the National Water Information System (NWIS). These observations were published in another OGC standard encoding WaterML 2, converted to netCDF-CF, then compared with the model forecasts to calibrate and improve the model predictions (Lin et al, 2016). WaterML 2 is preferred by the USGS for publishing water observations because of its rich metadata. NetCDF-CF however, is much more compact, taking just kilobytes of storage vs. megabytes of WaterML, with the same timeseries data content. Both forms have important roles.

Global Outreach

Fast, accurate streamflow and flood forecasts are goals throughout the world, and the need is particularly acute in developing countries. An important programme started by the European Centre for Medium-Range Weather Forecasting (ECMWF) in Reading, UK, and the Joint Research Centre (JRC) in Ispra, Italy, is called the Global Flood Awareness System (GloFAS), which is an extension of the European Flood Awareness System (EFAS). EFAS provides real-time flood forecasting throughout Europe. GloFAS predicts floods elsewhere up to 2 weeks in advance using ECMWF's longer-term forecast. A team based at Brigham Young University (BYU) has developed an application called the Streamflow Prediction Tool that uses the same ECMWF 2-week forecasts of rainfall and runoff used in GloFAS, but applies them to the RAPID model for streamflow routing, instead of the model used in GloFAS. This was built using BYU's Tethys Platform. Figure 4 shows an example of using this tool to predict recent high streamflow in the Strimonis River basin which crosses Bulgaria and Greece. The yellow and red triangles in this figure are warning symbols, and the red arrow indicates the reach being graphed. Note the reach-catchment hydrography in Figure 4, which was necessary for using RAPID. This hydrography was developed locally, and the model workflow could be run on a desktop computer; did not require high-performance computing.

Acknowledgements

More Information

- Maidment, D.R., 2015. *A Conceptual Framework for the National Flood Interoperability Experiment.* URL: https://www.cuahsi.org/Files/Pages/documents/13623/nfieconceptualframework_revised_feb_9.pdf
- Lin, P., Rajib, M.A., Somos-Valenzuela, M.A., Yang, Z.L., Merwade, V., Maidment, D.R., Wang,Y., Chen, L., 2016. *Spatio-Temporal Evaluation of Evapotranspiration and Streamflow in the NFIE Modeling Framework.* Submitted to JAWRA featured collection, Open Water Data Initiative & National Flood Interoperability Experiment.