J. Michael Johnson · Research Statement
University of California, Santa Barbara, California · Department of Geography

As a spatial data scientist I work with a range of engineers, geographers, computer scientists, hydrologists, and social scientists. I am interested in the data architecture supporting continental hydrologic modeling, and the development of methods to learn from, improve, and apply model output to research and application pursuits. When approaching the challenge of large-scale hydrologic modeling, there is an increasing need for a spatial perspective to inform process representation, analysis, and how output can be made accessible. Within these interests there are clear synergies with the faculty at Mines aligned with the *Hydrologic Science & Engineering* and *GIS & Geoinformatics* interdisciplinary programs related to hydrologic prediction, land cover disturbance and its impacts of hydrologic process and representation, building geospatial techniques for natural resource management, and leveraging data science techniques to answer novel questions related to the earth and environment.

Broadly, my work contributes to the United Nations Sustainable Development Goal of *Water and Sanitation*, the Open-Source Geospatial Foundation's mission to "promote collaborative development of open geospatial technologies and data," and the Semantic Web's mission to "allow everyone to find, share, and combine information more easily." In application, it contributes to the grand challenge of high-resolution continental scale hydrology; the US mission of becoming a more weather ready nation, and NOAA's aim of strengthening national water forecasting through the National Water Center. In the context of these societal goals, my research focuses are summarized below:

1) Supporting the Use of Big Earth Systems Data
(Data Science)

Large scale models provide a wealth of data that can transform research possibilities. However, the volume and structure of data outputs often limit usability. Therefore, I develop tools and software to facilitate flexible and programmatic ways to query remote geospatial data products for use in local computing environments. As a Consortium of Universities for Hydrologic Science (CUAHSI) HydroInformatics Fellow, I have been tackling the challenge of optimizing continental streamflow forecasts for time series extraction (paper in revision with *Nature Data Science*). Other data efforts have included: (1) improving the accessibility of the NHDPlus attribute and geometry data to support machine learning, statistical, and mapping applications (2) streamlining the data workflows and core data products for national flood forecasting using the Continental Flood Mapping Framework (CFIM), and (3) building tools for subseting the National Water Model (NWM) to support local model development, testing, and application.

2) Data Interoperability & Integration

(Computational Hydrology/Data Science)

To improve large scale modeling efforts, there is a need to integrate local and national data using a consistent data model to support dataset integration, feature addressing, and data assimilation. Much like we are able to provide a postal address for every home in the world, there is a need to assign hydrologic addresses that persist across scales. A recent collaboration with the USGS, NOAA, and Natural Resources Canada focused on the development of a logical model, called *Mainstems*, for multi-scale data integration. Building on this, I was selected as a USGS Pathways participant to help develop a continental dataset of mainstem identifiers, an initial set of community reference features (streamgages, dams, water quality sites), and methods for other agencies to annotate their data with mainstem identifiers. In conjunction, I was hired as a NOAA-Affiliate to contribute to the data model development for the next generation hydrologic modeling efforts. Collectively these efforts will make it easier to index disparate datasets and observation networks to a common addressing scheme, and the inclusion of mainstem identifiers in the next generation hydrologic modeling efforts will support model improvement, evaluation, intercomparison, and application.

3) Large-scale Modeling

(Computational Hydrology/Geoinformatics/Applied Machine Learning)

Water and energy simulations are often computed across grids representing the land surface that are analogous to the raster data model. The way grid cell properties are assigned can have cascading impacts on the accuracy of model results and are particularly difficult to get right in continental domains. While most high-resolution continental models operate on a ~1 km^2 grid, there are many modern resources that describe the variability of the land surface at finer scales. I am interested in how spatial properties can be characterized across scales to better inform model inputs and processes using novel applications of the raster data model. My initial work in the area focused on refining landcover maps and has sparked collaborations with members at San Diego State University focusing on advancing model base layers, and also at NCAR to explore how infiltration is represented in the NWM (WRF-Hydro). Equally, I am interested in quantifying the impacts of land cover change, both persistent and sudden, on water/energy simulations and the timescales at which change should be included in land surface models. This work relies heavily on integrating multiple-scale datasets and methods, including observation data from MODIS, historic simulations from NLDAS, feature datasets like fire burn perimeters and urban planning zones, and numerical models like NOAH-MP and urban growth models. Collectively these tools can help quantify what has happened in the past, evaluate how those processes were captured in historic simulations, highlight the capacity for new methods to capture these changes, and context for what we can expect in the future.

Aside from gridded model layers, many model elements are expressed in terms of parameters, including river channel geometry and reservoir operations. I am interested how hydrologic, hydraulic, and anthropogenic behavior can be learned from observation networks, and generalized using combinations of spatial data, network topologies, and machine learning. With early success in predicting rating curve relationships across the United States (in revisions at *Water Resources Research*), I am evaluating other hydraulic relations with the goal of providing a national dataset of idealized channel geometries that are necessary for hydrologic routing as well as many finer scale issues including sediment transport, species modeling, and flood forecasting.

4) Earth & Environment Information Infrastructures
(Data & Computer Science/Large-scale Modeling)

The last decade has seen unprecedented events including fires, floods, and a pandemic that have stressed aging infrastructure, revealed weaknesses in regional policies, and exposed the interconnected nature of social, ecological and technological systems. Within the NSFs focus on *"accelerating research to impact society at scale"* under the *"harnessing the big data revolution"* initiative, I have been part of the multi-phase, multi-university Urban Flooding Open Knowledge Network (UFOKN) project performing as a lead data scientist to formalize the connections between geodata (like roads and homes) and hydrologic models. These connections are expressed as linked data resources that rapidly relate forecasts to impacts in a knowledge graph infrastructure. Forecasting impacts in graph space is not only more efficient than traditional GIS approaches, but allows for integrative questions such as *"will there be flooding near me," "what substations will flood,"* or *"what communities rely on this road,"* that can be queried using search engines and semantic technologies.

As part of the project, we are required to develop a self-sustaining product through active industry partnerships with Google, Microsoft, Streamline Technologies, StormGeo, element84 and LeapAnalysis; academic partnerships with CUAHSI, Woods Hole Oceanographic Institution, Consortium for Ocean Leadership, the Internet of Water; and with federal partners like the EPA, USGS, and NOAA, and a number of local and county municipalities. Concurrently I am on also on the advisory board for a NOAA small business grant with Azavea Geographic Solutions which has provided more insights into how academia can support business. In all cases, these experiences have exposed ways for academic research to contribute to industry and entrepreneurial endeavors and cemented a number of partnerships that will help accelerate the impact and goals of my future research.