BIG DATA IN HYDROLOGY: FROM CONTINENTAL TO HILLSLOPE SCALES

Todd C. Rasmussen, James B. Deemy, and Seth E. Younger

AUTHORS: Warnell School of Forestry & Natural Resources, UGA, Athens GA 30602-2152

REFERENCE: *Proceedings of the 2017 Georgia Water Resources Conference*, held April 19–20, 2017, at the University of Georgia.

Abstract. Hydrologic data, information, and knowledge resolve differently depending upon the spatial and temporal scales of interest. A multi-scale hydrologic information system (HIS) is presented that can be designed and populated for a broad range of spatial (e.g., hillslope, local, regional, continental) and temporal (e.g., current, recent, historic, geologic) scales. Surface and subsurface hydrologic and transport processes are assumed to be scale-dependent, requiring unique governing equations and parameters at each scale. This robust and flexible framework is designed to meet the inventory, monitoring, and management needs of multiple federal agencies (Forest Service, National Park Service, Fish & Wildlife Service, National Wildlife Reserves). Multi-scale HIS examples are provided using Geographic Information Systems for the Southeastern US.

INTRODUCTION

Hydrologic models are routinely used for understanding and predicting water-resource behavior, including stores and fluxes along with their physical, chemical, and biological integrity. Fundamental to the development and application of these models is the grounding on observational data, usually from surface, airborne, and satellite platforms.

Both computational platforms and data acquisition and storage systems have expanded rapidly in recent decades, resulting in ever-greater spatial and temporal resolution and extent. Efforts to model and obtain data at desired spatial and temporal scales has been, and will continue to be, an important objective in studying and managing water resources (Price et al. 2014).

For example, Table 1 summarizes the range of spatial and temporal scales of interest when conducting hydrologic studies and assessments. It is important to note that the appropriate scale depends on the objectives of the specific project being investigated – there is no universal scale appropriate for all studies (Baffaut 2015). While regional and continental scales are useful for establishing a context for local studies, there may be limited need to apply continental-scale models for hillslope settings.

In fact, model parameters developed for one scale may not be relevant for applications at different scales or regions (Clark et al. 2014). For example, pore-scale models of fluid flow are of limited use in predicting flow and transport at continental scales, and vice versa.

Table 1. Spatio-temporal scales of hydrologic interest.

Attribute	**Spatial Scales**		
	Region	**State**	**Site**
Geology	Province	Units	Stratigraphy
Soil	Orders	Sub-orders	Series
HUC	2 or 4	6 or 8	10 or higher
DEM	300-m	30-m	10-m
Hydrography	Large	Medium	Small
Groundwater	Aquifers	Outcrops	DTW
LU/LC	Ecoregions	Land Use	Land Cover
	Temporal Scales		
	Decadal	**Annual**	**Synoptic**
Water levels	USGS	Academic (MS & PhD),	
Weather	NOAA	Federal Field Studies	
Ecology	LTER		

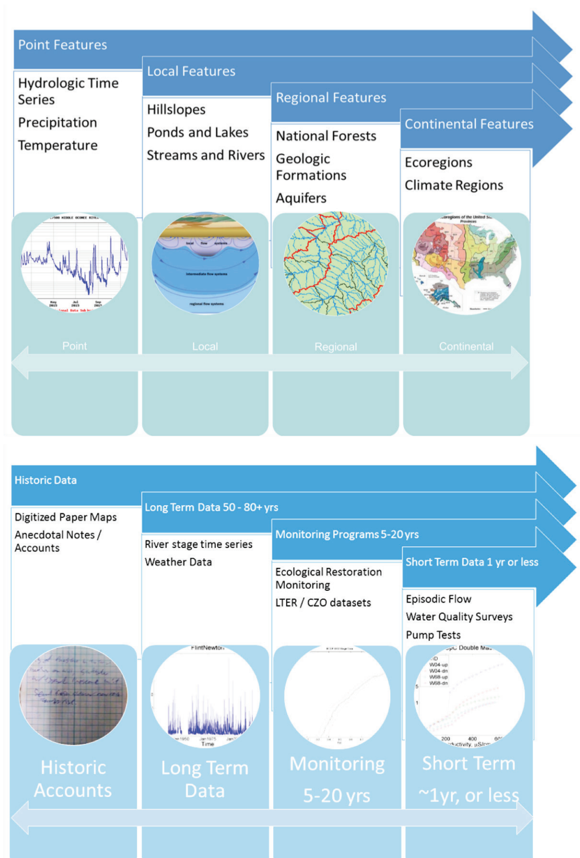

Figure 1. Multi-scale hydrologic information.

The goal of the research reported here is to support water-resource modeling that answers scientific and management questions over a broad range of spatial and temporal scales using a prototype information management and modeling framework that:

- Addresses hydrologic questions that span multiple temporal and spatial scales; and
- Provides the ability for management agencies to seamlessly transfer modeling results of scale-dependent processes.

Because hydrologic models require data for each application, users often make simplifying assumptions, such as a uniform, homogeneous landscape. Yet, many important hydrologic flow and transport processes are sensitive to spatio-temporal variation, a continental-scale model using high-resolution information is not yet possible because of the enormous computational burden needed to assemble and process these data. The questions we pose are:

- How can we create accurate regional models using local data?
- Can we parameterize models using pseudo-data that have been transformed to provide information at regional scales?

These questions are key components of the CUAHSI Water Data Center (cuahsi.org/dataservices), which provides a portal for hydrologic data publication, discovery, and access to hydrologic databases. The CUAHSI HIS platform supports a wide range of time-series and geospatial data, along with services to store, inventory, and serve hydrologic data via a web interface.

A clear need exists for the development of tools and services to inventory and manage these temporal and spatial data (Buytaert et al. 2008).

METHODOLOGY

Our proposed methodology is based on first specifying the following components:

- Identify management objectives, such as minimum cost or risks; maximum resilience or protection
- Specify analytic approach
- Identify available information
- Assemble spatial and temporal information at appropriate scales
- Provide model outcomes for a range of alternative scenarios
- Estimate model and parameter uncertainties
- Identify additional information needs

Our framework incorporates information from multiple spatial scales (hillslope, local, regional, continental) and dimensions (raster, point, line, polygon). We also incorporate information across a spectrum of temporal scales (historic, continuous, decadal, intermittent, event) for both soft (anecdotal, descriptive, qualitative) and hard (measured, computed, quantitative) data. Figure 1 summarizes the scope of our multi-scale modeling framework.

Each data type has a unique data scale — topographic features may vary dramatically from local to regional, while geology may have greater variation at regional and continental scales. Soils can also vary across local scales, as do recharge and depth-to-water. Because of these contrasting behaviors as a function of scale, each variable must be quantified in a manner that captures its information content at each scale.

APPLICATIONS

We first demonstrate our methodology for an intrinsic groundwater vulnerability assessment based on the EPA DRASTIC methodology. The DRASTIC groundwater vulnerability index, V, uses seven factors to describe each hydrogeologic setting (karst, coastal sediments, fractured bedrock)

$$\mathbf{V = w_D\, D + w_R\, R + w_A\, A + w_S\, S + w_T\, T + w_I\, I + w_C\, C}$$

where w_i are the weights for each factor. Factor weights and ratings are based on elevations, soils, geology, precipitation, and other data sources.

D = Depth to water: Shallow water tables pose a greater chance for the contaminant to reach the groundwater surface as opposed to deep water tables.

R = Net recharge: This is the principal vehicle that transports the contaminant to the groundwater; the vulnerability increases with greater recharge.

A = Aquifer media: The aquifer material influences contaminant mobility.

S = Soil media: The soil media determines the amount of percolating water that reaches the watertable.

T = Topography: Steeper slopes lower the vulnerability due to greater runoff and lower infiltration.

I = Impact of the vadose zone: The texture of the vadose zone materials determines the pollutant travel time through it.

C = Conductivity: The soil hydraulic conductivity determines the rate of percolation to groundwater.

Factors are created as layers in our geo-spatially explicit model, and then assigned to a range or category depending upon whether quantitative (numeric) or qualitative (descriptive) information are available, respectively.

A rating between one and ten is assigned to each range or category, depending upon its importance. A weight is assigned to each factor to define the relative importance of that factor to affect pollution transport. The final DRASTIC index is divided into four relative categories: Low, Moderate, High, and Very High.

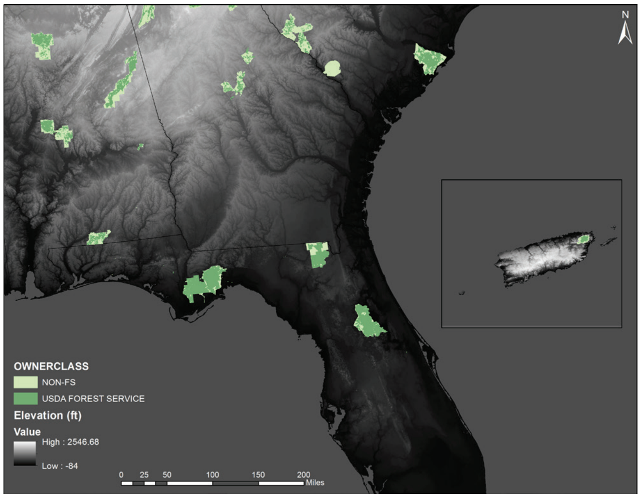

Figure 2. National Forest Units (green) in Alabama, Georgia, South Carolina, Florida, and Puerto Rico.

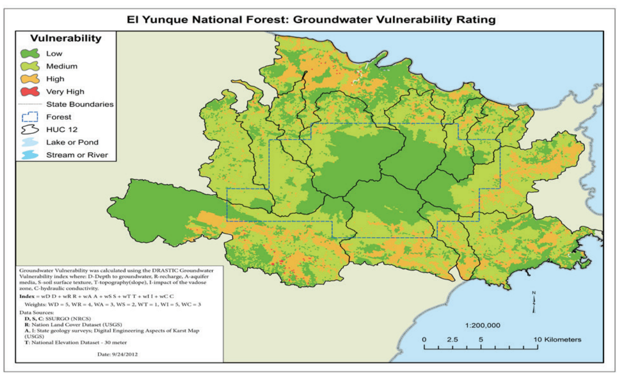

Figure 3. Intrinsic groundwater vulnerability for El Yunque National Forest, Puerto Rico, using 10-m raster data. HUC-12 watersheds also shown.

Figure 2 shows some of the National Forest units in the Southeastern United States. Figure 3 summarizes the DRASTIC groundwater vulnerability index for El Yunque National Forest in Puerto Rico. Note that most groundwater is at low to moderate risk within the National Forest, but is at greater risk as the land use transitions to agricultural and urban uses at lower elevations.

As noted above, data resources include topographic information, such as those presented in Figure 4 for the entire Southern Region. These data are useful for regional scales, but become too coarse at local scales, which are challenging to resolve at continental scales.

Depth-to-groundwater is a key layer that is critically lacking in existing information systems. Efforts to estimate this data layer are ongoing using autocorrelations between groundwater elevation in space, as well as correlations with surface topography. Figure 5 presents a preliminary estimate for the Oconee National Forest in Georgia.

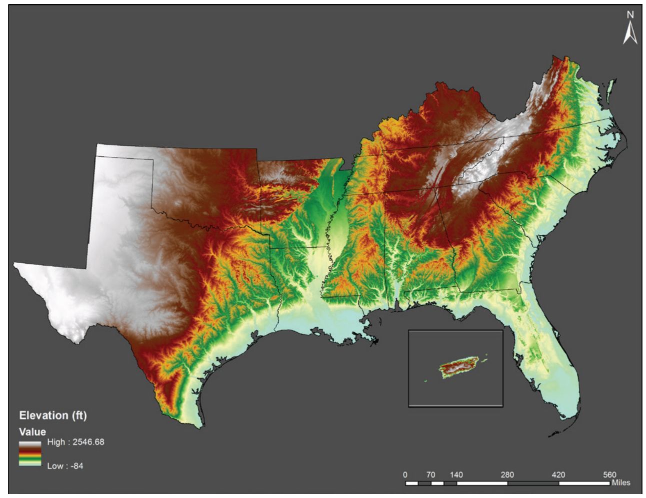

Figure 4. Regional topographic layer based on Digital Elevation Model (DEM) raster at 300-m scale.

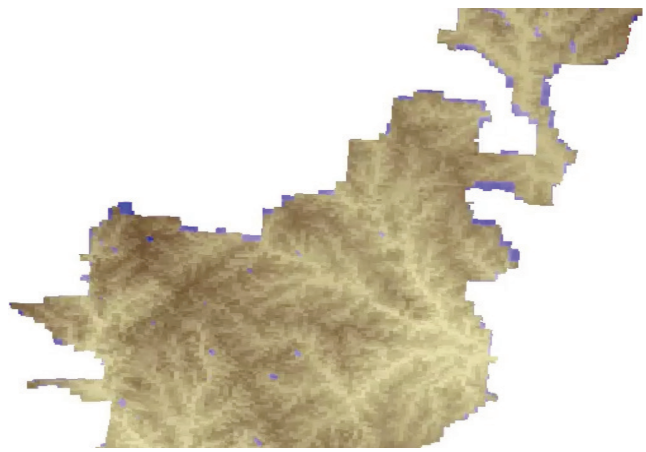

Figure 5. Depth-to-groundwater raster for Oconee National Forest. LIDAR provides 1-m resolution.

The DRASTIC methodology neglects potential sources of contamination (industrial, municipal, agricultural, transportation) as well as the presence-absence of populations (human, endangered species) that may be at risk.

Landscapes with high intrinsic vulnerability may not pose a risk if risk sources and vulnerable populations are absent. To account for risk sources and receptors, we add land use and cover, as well as hazardous waste use, storage, transport, and disposal.

We next demonstrate using time-series data that also contain information at multiple scales, including daily, seasonal, annual, and decadal fluctuations. Risks to aquatic ecosystems from hydroperiod and water quality alterations (e.g., sediments, nutrients) may be manifested at each of these scales.

Long-term (20-year), high-resolution (15-min) data from USGS (Figure 6) and NOAA (Figure 7) sensors in Apalachicola Bay must be assimilated to understand the complex dynamics of ecosystem function.

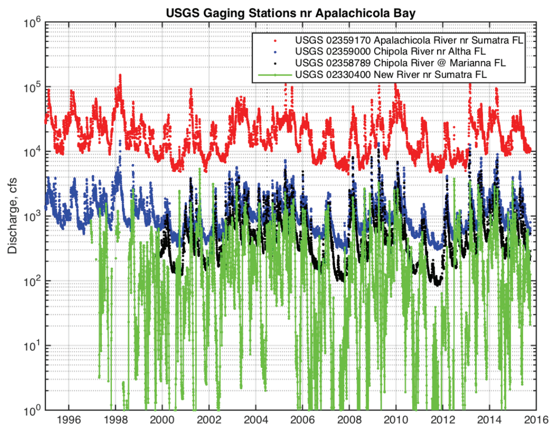

Figure 6. USGS discharge data for Apalachicola Bay, Florida. Data is collected at 15-min intervals and displays rapid changes at daily, monthly, seasonal, annual and decadal scales.

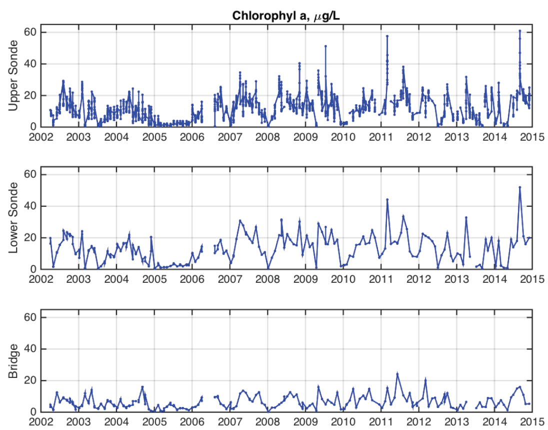

Figure 7. NOAA water-quality 15-min time-series data for two sondes placed in East Bay, and embayment within Apalachicola Bay, Florida. Data show dynamic behavior at multiple temporal scales.

DISCUSSION

Our ability to generate, store, and access water-resource information is growing rapidly, and challenges our ability to assimilate and apply these data to solve management problems. Our proposed framework is intended to facilitate hydrologic modeling by organizing information along a space-time continuum that can be tailored to management needs.

Databases are resolved at multiple temporal and spatial scales. These data are often disjoint at jurisdictional boundaries (counties, states, countries). Also, data are fraught with problems (missing, mis-mapped, overlapping) – data assimilation and reconciliation are critical to reconcile these issues. Our current lack of capability to manage large databases means that extracting data at appropriate scales must be performed on a case-by-case basis, often resulting in additional uncertainty in understanding, prediction, and management of water resources.

Acknowledgements. Many thanks to our federal sponsors (US Fish and Wildlife Service, US Forest Service, US National Park Service, US National Nuclear Security Administration, US Geologic Survey) as well as the UGA Hydrology Laboratory.

REFERENCES

Baffaut, C., S.M. Dabney, M.D. Smolen, M.A. Youssef, J.V. Bonta, M.L. Chu, J.A. Guzman, V.S. Shedekar, M.K. Jha, J.G. Arnold. 2015. Hydrologic and water quality modeling: Spatial and temporal considerations. *Transactions of the ASABE*. 58(6):1661-1680

Buytaert, W., D. Reusser, S. Krause, J.-P. Renaud. 2008. Why can't we do better than Topmodel. *Hydrological Processes*. 22:4175-4179

Clark, M.P., B. Nijessen, J.D. Lundquist, D. Kavetski, D. Rupp, R.A. Woods, J.E. Freer, E.D. Gutmann, A.W. Wood, L.D. Brekke, J.R. Arnold, D.J. Gochis, R.M. Rasmussen. 2014. A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*. 51:2498-2514

Price, K., S.T. Purucker, S.R. Kraemer, J.E. Babendreier, C.D. Knightes. 2014. Comparison of radar and gauge precipitation data in watershed models across varying spatial and temporal scales. *Hydrological Processes*. 28:3505-3520

Sivapalan, M. 2009. The secret to 'Doing better hydrological science': Change the question! *Hydrological Processes*. 23:1391-1396