AI-Directed Adaptive Multifidelity Modeling of Water Availability and Quality at River Basin Scales

Scott L. Painter[1], Ethan Coon[1], and Dan Lu[1]

[1]*Oak Ridge National Laboratory, Oak Ridge, TN, USA*

Focal Area

(2) Predictive modeling using AI techniques and AI-derived model components; use of AI and other tools to design a prediction system comprising of a hierarchy of models

Science Challenge

Earth system changes and ever-increasing demands are placing unprecedented stresses on global water supplies. River-basin scale models are essential tools to help understand changes in water availability and quality due to wildfire, land use change, extreme weather, and climate change. Advances in AI/ML offer great potential for dramatically improving models at river basin scales, with the potential to fundamentally improve predictive understanding of water availability and quality in a changing climate.

Rationale

Most hydrological models that are applicable at the river basin scale lump large swaths of the land surface into single hydrological response units. These semi-distributed models are computationally expedient but are severely limited in their ability to represent responses to changing climate, and especially extreme weather and disturbances. Fully distributed integrated surface/subsurface hydrological models that resolve the land surface at high resolution provide a powerful alternative to semi-distributed models and have great potential to improve predictive understanding. However, integrated hydrological models are computationally demanding and struggle to simulate even a single watershed of the many required to conduct basin-scale simulations. Advances in AI/ML offer great potential for making high-resolution state-of-the-art integrated hydrology models viable at river basin scales, thus fundamentally improving the predictive understanding of water availability and quality in a changing climate.

Narrative

Globally, water resources are under increasing pressure from Earth system change and ever-increasing demands for clean water, food, and energy driven by human activity (e.g., Vorosmarty et al. 2010). Those stresses on our water supply and quality are transmitted through watersheds. Watersheds are the fundamental functional units of the Earth's surface, within which vegetation, fluvial systems, soils, and the subsurface interact to govern water quality and availability.

Quantitative modeling at the river basin scale is essential to address changes in water availability and quality due to wildfire, land use change, extreme weather, and climate change. Most hydrological models that are applicable at the river basin scale lump large swaths of the land surface into single hydrological response units. These semi-distributed models are computationally expedient but are severely limited in their ability to represent responses to changing climate and especially extreme weather and disturbances. Fully distributed integrated surface/subsurface hydrological models that resolve the land surface at high resolution have great potential to improve predictive understanding of water availability in a changing climate with an intensifying hydrological cycle. However, integrated hydrological models are computationally demanding and

DOI: https://doi.org/10.2172/1769669
DOE BER, AI4ESP 2021 #AI4ESP1101 (https://www.ai4esp.org)

struggle to simulate a single watershed, whereas predictive capabilities are required for large river basins that include a great number of smaller watersheds.

Advances in AI/ML offer great potential for making high-resolution state-of-the-art integrated hydrology models viable at river basin scales and thus fundamentally improving the predictive understanding of water availability and quality in a changing climate. We envision an AI-controlled data-driven multiscale and multifidelity modeling system that simulates large river basins in their entirety, with high-fidelity, high-resolution, physics-based models representing selected sub-basins, and fast-running ML-derived surrogate models or data-driven ML models representing the remaining sub-basins. Multifidelity and hybrid ML-physics models are under development within the ExaSheds project, but the scope of that project is only an initial step toward what is ultimately possible with AI-controlled multifidelity modeling. In particular, we envision a system that builds surrogate models from computationally demanding physics-based models *on the fly* while adaptively switching sub-basins between high- and low-fidelity models as needed based on user-defined responses of interest, simulated system behavior, and changing land use and climate drivers. In this approach, high-fidelity models would be used when and where the quality of lower-fidelity surrogate models deteriorates, weighted by the influence on the basin-scale responses of interest as determined adaptively by a controlling AI. Thus, as the simulations project into a future climate regime in which surrogate models initially informed from historical observations become unreliable, a subset of the sub-basins could be switched to physics-based models both to improve the overall model and to update surrogate models. Then the simulations could return to the surrogate models after they are updated to sufficient accuracy. Such a system would provide stakeholders with profound flexibility to carry out high-resolution simulations to support scenario-based planning, capability that is not available in current and near-term anticipated modeling systems.

The envisioned capability would exploit the nested multiscale spatial structure of river basins. Watersheds are inherently hierarchical, with each unit composed of weakly connected subunits that can be delineated a priori. Multiscale algorithms would be used to decompose a river basin into successively smaller (and computationally more tractable) subdomains. At the fine scale, AI control would choose from the available representations—process-based or surrogate model—for each subdomain based on error estimates. A coarse-scale model would be used to integrate each subdomain into projections of the entire river basin. Initial efforts to better exploit that hierarchical structure through domain decomposition algorithms are being explored in the ExaSheds project and the National Water Model's Next Generation Hydrofabrics concept; this concept would build on those approaches to leverage expensive process-based models across large spatial extents.

The modeling pipeline required would include data analysis, automated ML, meta-learning, explainable AI, and transfer learning. Data analysis is needed to determine which subbasins require high-fidelity, physics-based models and which subbasins can be simulated using ML-based surrogate or data-driven ML models, depending on the physics knowledge and data availability. Automated ML techniques are needed to select an appropriate neural network architecture and its hyperparameters for constructing the data-driven ML model based on the data features. We envision meta-learning algorithms to ensemble the sub-basin simulations from the hybrid model ensembles (including the physics-based models and the data-driven ML models) and give each individual model a performance-based weight according to its influence on the basin-scale responses of interest. Explainable AI is needed to diagnose the strengths and weaknesses of the models based on their contributions to advancing model understanding. Finally, as the simulations project into a future climate, we envision using transfer learning to update the hybrid modeling system to incorporate and reflect the change.