Improving Short Term Predictability of Hydrologic Models with Deep Learning

Ryan King, Ariel Miara, Andrew Glaws, Jordan Macknick
National Renewable Energy Laboratory

Focal Area

Focal Area 2. Predictive modeling through the use of AI techniques and AI-derived model components; the use of AI and other tools to design a prediction system comprising of a hierarchy of models (e.g., AI driven model/component/parameterization selection).

Science Challenge

A major challenge exists in the lack of reliable predictive modeling of water-cycle extremes using macro-scale hydrologic models that are driven by atmospheric climate data. This causes a critical knowledge gap in understanding the magnitudes, probabilities, and severity of droughts, rainfall, and flooding. Improved understanding of these extreme events requires more accurate modeling at high spatial and temporal resolutions. This shortcoming is especially apparent when complex couplings between atmospheric quantities and engineered systems emerge, such as during extreme precipitation or drought events and at intersections of atmospheric-land-fluvial-ocean systems. Added complexity around representing operations of engineered infrastructure leads to a limited capability to analyze the temporal and spatial implications of compounding extreme events across inland and coastal regions. This results in profound challenges for water and power systems operators, agriculture production, and regional infrastructure planning. The science challenge can be summarized as

How can we use AI/ML to improve short term predictability of extreme water-cycle events and associated risk, especially compounding events and extreme rainfall patterns that result in flooding, to inland and coastal communities under different climate scenarios?

Rationale

This workshop, and its research outcomes, are necessary to address the multi-faceted, grand challenges outlined in the Department of Energy's (DOE) Climate and Environmental Sciences Division (CESD) Strategic Plan. In particular, the "data-model integration scientific grand challenge", the "drivers and responses in the earth system scientific grand challenge" and, in particular, the "integrated water cycle scientific grand challenge" that aims to: *advance understanding of the integrated water cycle."* Specifically, the predictability of extreme events in the water cycle has been hampered by poor empirical models of river flows, insufficient spatiotemporal resolution of atmospheric data like rainfall in both inland and coastal areas, and incomplete understanding of engineered systems such as dams and reservoir operations. The dynamics of the water cycle are tightly intertwined with engineered systems, such as man-made dams and agriculture, which transform the spatiotemporal characteristics of runoff and river discharge fluxes as well as resulting flooding or water scarcity events. Better predictability of extreme events is crucial to ensure resilient operation of built infrastructure in future climates.

We posit that emerging AI/ML capabilities can improve the accuracy and predictability of extreme events in the water cycle over large domains, from inland to coastal regions, by addressing the following research gaps:

1. Enhancing the accuracy and resolution of atmospheric data driving hydrologic models.

DOI: https://doi.org/10.2172/1769722
DOE BER, AI4ESP 2021 #AI4ESP1068 (https://www.ai4esp.org)

2. Improving model performance and calibration at high temporal resolution by capturing sensitivities to multivariate climate-fluvial-ocean phenomena.
3. Understanding the effects of coupled natural-human systems on the water cycle through improved representations of engineered riverine systems, including reservoir operations and land use.

Narrative

To break through these barriers, we propose three AI/ML research thrusts:

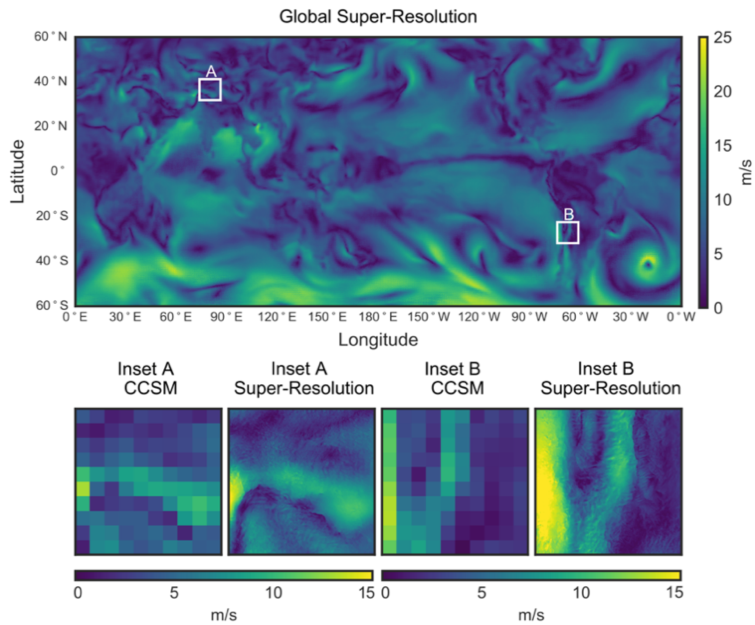

Figure 1 *Wind velocity field from a global climate model super resolved by 50x. Reproduced from Stengel et al[4].*

<u>Addressing Research Gap 1: Super Resolution of Atmospheric Quantities Affecting the Water Cycle</u>

Climate data and macro-scale hydrologic models are typically unable to accurately capture short-term perturbations of the water cycle, despite having the computational capability to simulate water resource conditions at a daily time step and at high spatial resolutions. Results are reported at monthly or annual timescales, which is the level at which calibration is typically done, for reasons including a lack of observational data and unreliable climate input data at higher temporal resolutions. Climate datasets, such as those from Global Circulation Models (GCM), are rarely available at a higher spatial (e.g., 1-km) scales also needed for accurately capturing extremes.

Emerging ML/AI tools have achieved significant levels of super-resolution for atmospheric data like windspeed, solar irradiance, and temperature (up to 50x spatially or 24x temporally)[1]. Additionally, conditional generation extensions of these methods can recover probabilistic super-resolved data that can inform extreme event analysis. This same methodology could be used to downscale other atmospheric variables relevant to the water cycle, such as rainfall, cloud cover, and humidity. Although previous applications were intended for renewable energy studies, they could be used to improve the fidelity of hydrologic modeling and enable a spatial resolution sensitivity analysis beyond scales analyzed by E3SM and the macro-scale modeling community. By improving the spatial resolution of these quantities to reflect watershed boundaries, this approach would symbiotically benefit improved river flow models that incorporate the graph structure of river networks. Furthermore, the improved temporal resolution would assist in short term forecasting accuracy. Finally, super resolving data from different climate scenarios would enable a consideration of additional extreme events.

<u>Addressing Research Gap 2: Learning on Graphs for River Flows</u>

Hydrological modeling of river flows and reservoir management can be improved by taking advantage of the physical topological structure of river flows as directed graphs and using this to inform the inductive bias of the learning method[2]. Recent publications demonstrate that this spatial structure can be effectively modeled using graph convolutional neural networks[3] and temporal patterns of river flows can be captured with recurrent neural networks[4]. By incorporating the river network structure into the learning process, plus including land surface characteristics, we believe the empirical river flow models can be replaced or augmented with faster and more accurate data-driven surrogates.

Addressing Research Gap 3: Reinforcement Learning to Enhance Reservoir Operations Modeling
Detailed information on reservoir operations, specifically inflow-outflow relationships and the decision-making behind them, is lacking. Various research efforts have attempted to formulate reservoir operation schemes with a range of success. These challenges, related to model-free controls, sensor placement, and state estimation, can be addressed with AI/ML techniques such as reinforcement learning (RL) and data assimilation. Macro-scale hydrologic models have validated reservoir operations for regional-to-global scales. Yet, high-resolution simulations, both spatial and temporal, and the resulting effects on downstream river discharge, flooding and freshwater fluxes to coastal zones sufficient for decision-making rarely exist. Aligning with the MODEX framework, RL controllers can learn short-term corrections to coarse-scale control models. Real-time data streams (potentially enhanced by the super-resolution techniques mentioned earlier) can informed the RL models and provide insights into risk assessment and responses to extreme conditions, especially for detecting situations with compounding events. For instance, extreme precipitation in an inland region, combined with certain reservoir operations that result in inundation at a downstream coastal area.

Each of the proposed AI/ML research thrusts require learning algorithms adapted to specific characteristics of the water cycle, e.g., directed graph river topologies, while also preserving known physics in a hybrid approach that balances data-driven and first principles-based knowledge to improve overall water cycle predictability. Research improvements proposed through the workshop will facilitate the dissemination of products (model results, underlying data, diagnostic calculations/methods) that would be publicly available to ensure transparency and enable public and stakeholder engagement.

The proposed AI/ML capabilities can also leverage a number of pre-existing DOE resources and investments. Super-resolved atmospheric quantities can be trained on existing high-resolution datasets such as NREL's WIND Toolkit and Nation Solar Radiation Database using high performance computing capabilities located at DOE national labs and the National Energy Research Scientific Computing Center. Graph neural nets for river flow data can be trained using USGS river flow gauges and the National Water Model. Finally, enhanced atmospheric data sets, improved river flow models and reservoir operations models can all be implemented in E3SM to make further study of coupled extreme events accessible to the broader community.

Suggested Partners Experts
Vincent Tidwell, Sandia National Laboratory
Charles Vorosmarty, City University of New York

References

[1]Stengel et al, *Adversarial super-resolution of climatological wind and solar data*, Proceedings of the National Academy of Sciences, 2020. https://doi.org/10.1073/pnas.1918964117.
[2]Battaglia et al, *Relational inductive biases, deep learning, and graph networks*, https://arxiv.org/abs/1806.01261.
[3]Zhao et al, *Joint Spatial and Temporal Modeling for Hydrological Prediction*, IEEE Access, 2020. https://doi.org/10.1109/ACCESS.2020.2990181.
[4]Jia et al, *Physics-Guided Recurrent Graph Model for Predicting Flow and Temperature in River Networks*, https://arxiv.org/abs/2009.12575.