Using Machine Learning to Extend High-resolution Integrated Hydrology Models to River Basin Scales

Ethan Coon,[1] Dan Lu,[1] Scott Painter,[1] Carl Steefel,[2] Dipankar Dwivedi,[2] Xingyuan Chen[3], David Moulton[4] & the ExaSheds Team

[1]Oak Ridge National Laboratory, Oak Ridge, TN;
[2]Lawrence Berkeley National Laboratory, Berkeley, CA;
[3]Pacific Northwest National Laboratory, Richmond, WA;
[4]Los Alamos National Laboratory, Los Alamos, NM

Contact: coonet@ornl.gov

Project Lead Principal Investigator (PI): Carl Steefel and Scott Painter

BER Program: Other (Data Management)

Project: ExaSheds Project

Project Website: https://exasheds.org/

Project Abstract: Many of the impacts of climate and land use change are felt through the water cycle. Water availability, quantity, and quality are currently volatile, and ongoing change only reduces predictability. A predictive understanding of the impacts of these changes requires capability that operate across scales; while many of the processes governing the water cycle occur in watersheds at scales from 10s-100s of meters (macrotopography, land cover and land use variation), these processes must be integrated across full river basins (100s of kilometers) to determine their effect on society and the water cycle.

The ExaSheds project is developing a novel capability to address this scaling problem and thereby advance watershed system understanding. We have embraced a hybrid methodology, bringing together Machine Learning (ML) and process-rich simulations using high performance computing. The combined capability is fundamentally hierarchical, multiscale, and organized by watersheds, the natural unit for integrating water impacts. The goal is a general simulation capability that can be used by scientists and their stakeholder partners to advance understanding of regional-scale water cycles under future climates and conditions.

This talk describes progress toward ExaShed's vision of an ML-enabled strategy for scaling process-based models to river basin scales, focusing initially on stream discharge. We leverage ML to build surrogate models for high-resolution ATS models of individual watersheds within the larger basin. These surrogate models take as input time series of meteorological forcing and incoming flow from watersheds upstream as well as static attributes of the watershed. We are using current-day and future meteorological forcings downscaled by ML techniques from climate projections in forward simulations to train the surrogate models. Once trained using a

few years forcing data, the surrogate models can be used to invert for uncertain model parameters, extend the simulations to decadal time scales over large basin scales, and address scenario uncertainty associated with the future climate and land-use changes. The resulting modeling framework leverages the strengths of both data-driven and process-based modeling approaches – computationally feasible, large-scale predictions that maximize predictive power under current conditions while being robust to shifts to unprecedented future conditions.