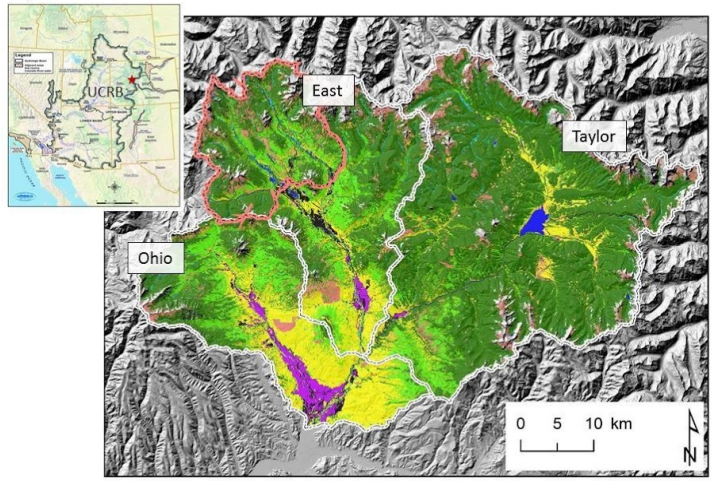
ExaSheds

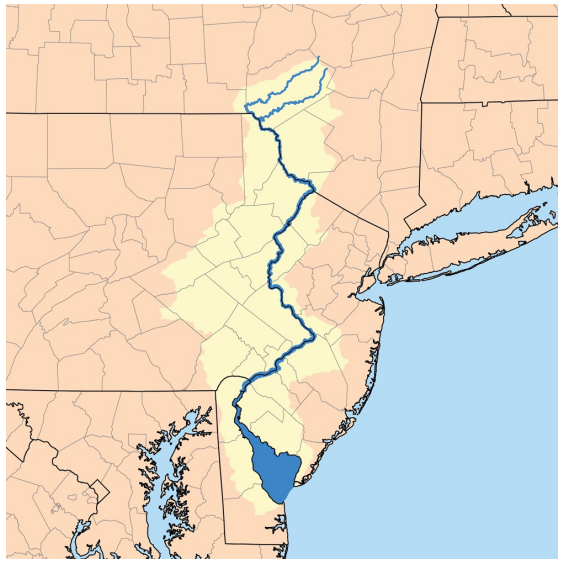
***Co-PIs*:** Carl Steefel and Scott Painter
Task Leads: Xingyuan Chen, Ethan Coon
Dipankar Dwivedi, David Moulton,

Advancing Watershed System Science using Machine Learning and Data-Intensive Extreme-Scale Simulation

- Interdisciplinary team (~15) across 4 labs
- Exploring strategies for learning-assisted simulation
  - development of model inputs from sparse, coarse, and indirectly related information
  - hybridization of process-resolving simulations and ML
- Working with data from
  - East River, Colorado, Watershed
  - Upper Colorado Water Resources Region
  - Continental US
  - Delaware River Basin
- Adapting DOE-developed watershed simulation tools to leadership-class computer architectures



East River

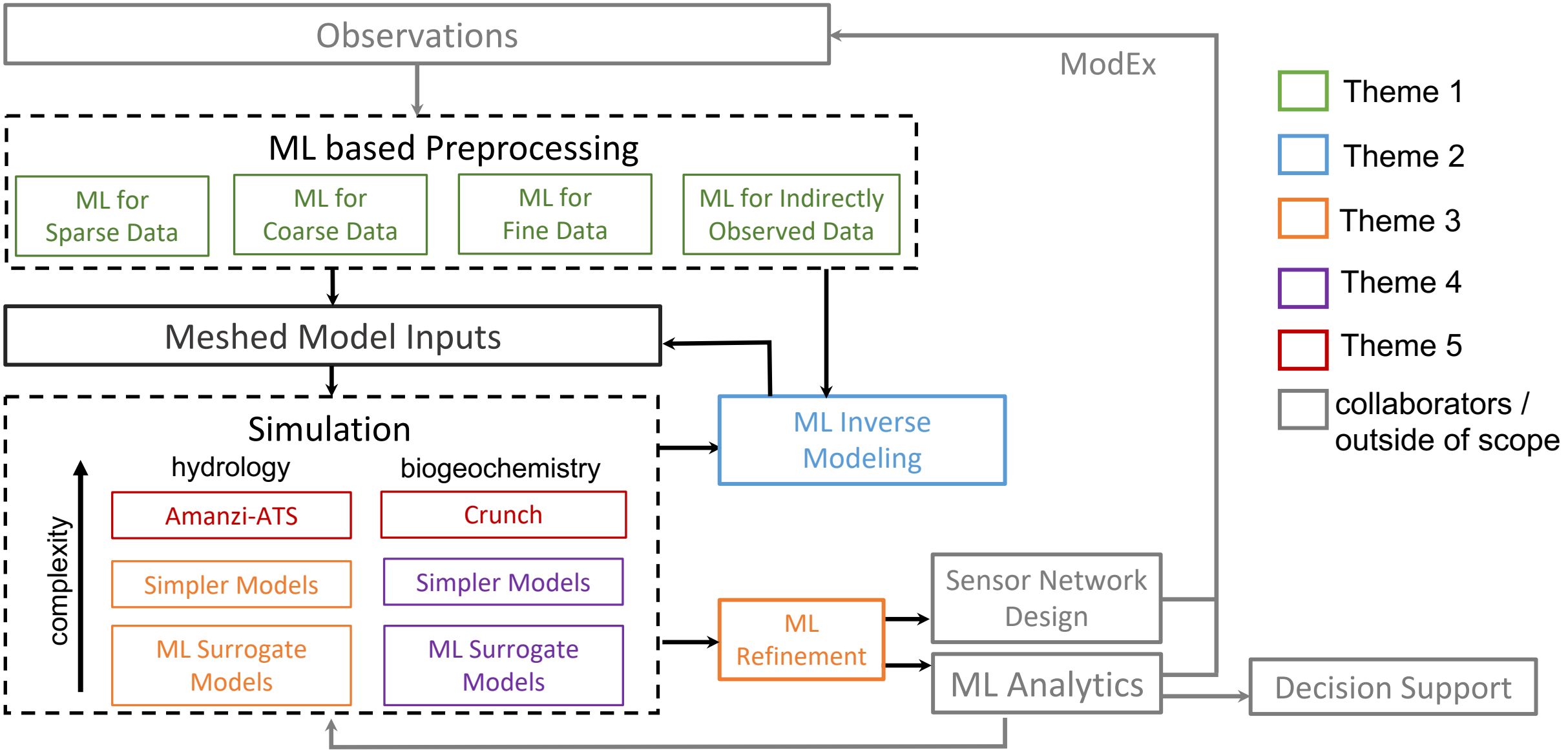


Delaware River

BERKELEY LAB · OAK RIDGE National Laboratory · Los Alamos NATIONAL LABORATORY EST. 1943 · Pacific Northwest NATIONAL LABORATORY

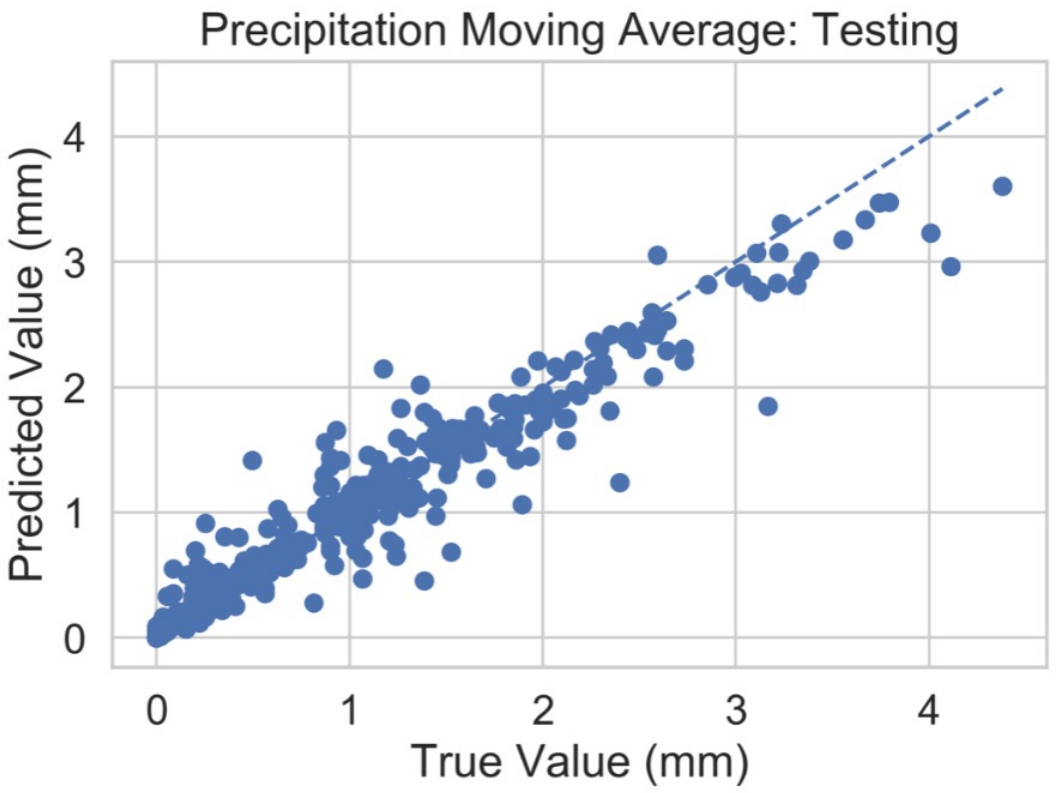
6 Year Vision: A novel multiscale strategy fusing process-resolving simulations and machine learning

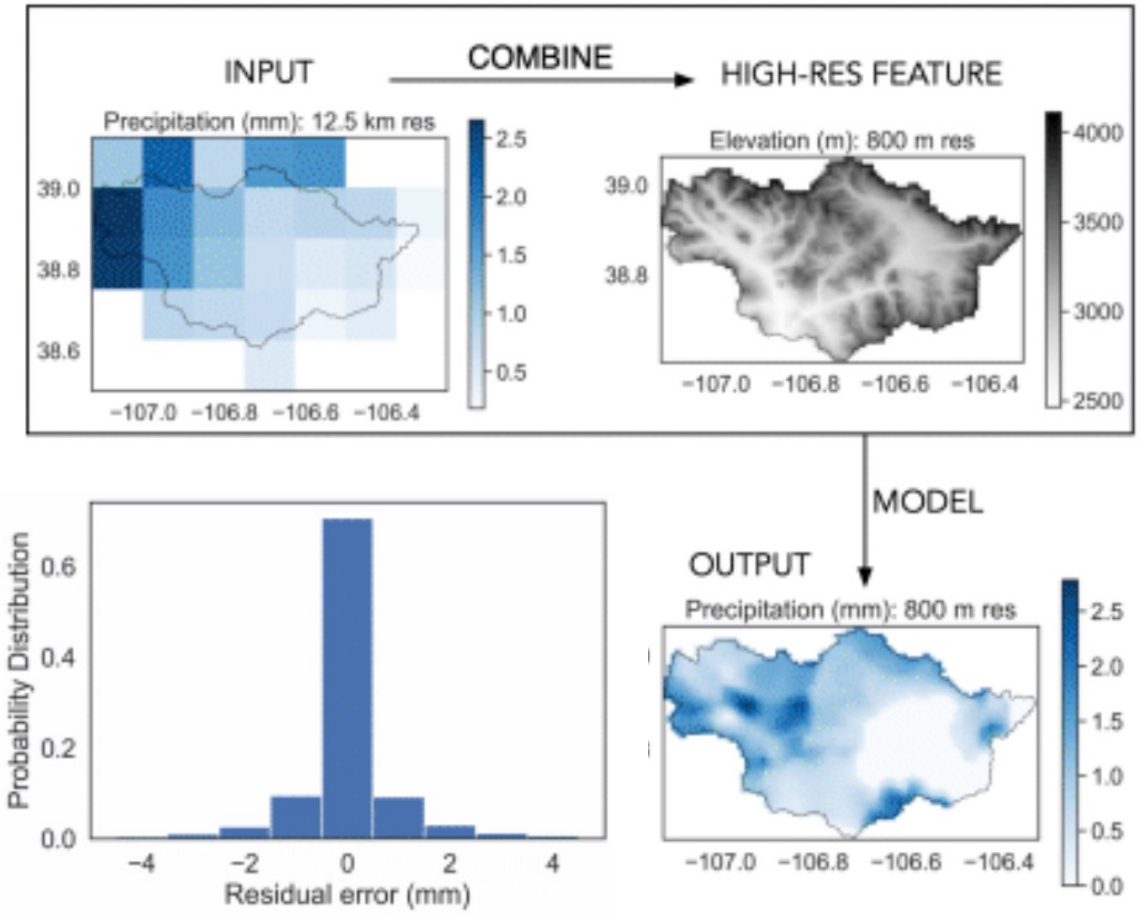
- Tightly integrated role for machine learning
  - Synthesizing spatially distributed model inputs from diverse data streams
  - Inverse modeling/parameter identification
  - Surrogate models trained on process-rich simulations
  - Machine learning used to “refine” the output of process-based simulations
  - Feedback of modeling to distributed sensor networks
- Process-explicit integrated surface/subsurface flow and reactive transport codes
  - Represent biogeochemical processes and their hydrologic controls at their native scales
  - Adapt to heterogeneous leadership-class architectures, providing path to exascale
  - High throughput on leadership-class facilities will facilitate model-data integration
- Open source community resources for ML-assisted high-resolution simulation
  - GPU-capable versions of ATS and Crunch
  - Application-specific, python-based ML tools
  - Workflows and tutorials


Observations
ModEx
ML based Preprocessing
ML for Sparse Data
ML for Coarse Data
ML for Fine Data
ML for Indirectly Observed Data
Meshed Model Inputs
Simulation
hydrology
biogeochemistry
complexity
Amanzi-ATS
Crunch
Simpler Models
Simpler Models
ML Surrogate Models
ML Surrogate Models
ML Inverse Modeling
ML Refinement
Sensor Network Design
ML Analytics
Decision Support
Theme 1
Theme 2
Theme 3
Theme 4
Theme 5
collaborators / outside of scope

Machine learning can assist high-resolution simulation by helping to develop model inputs

Gap filling missing meteorological forcing data

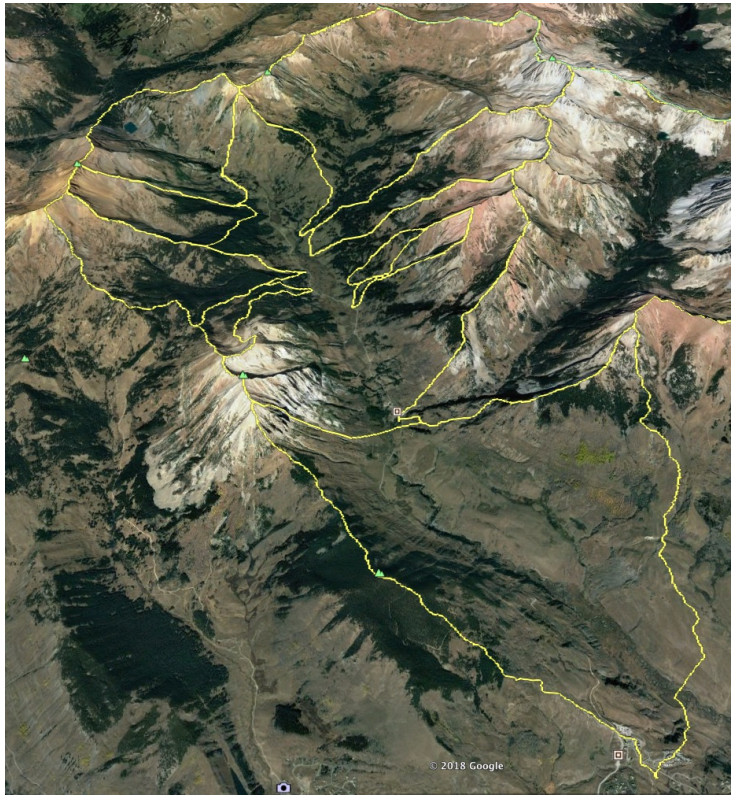


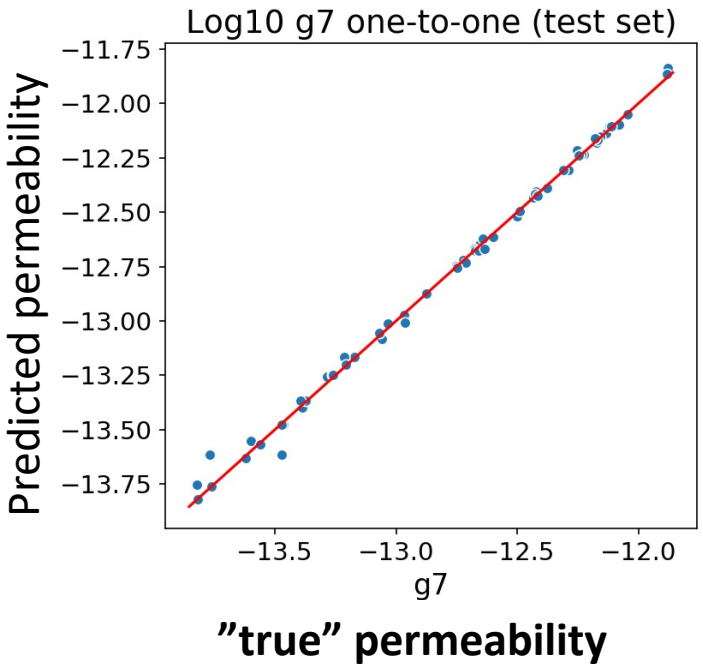


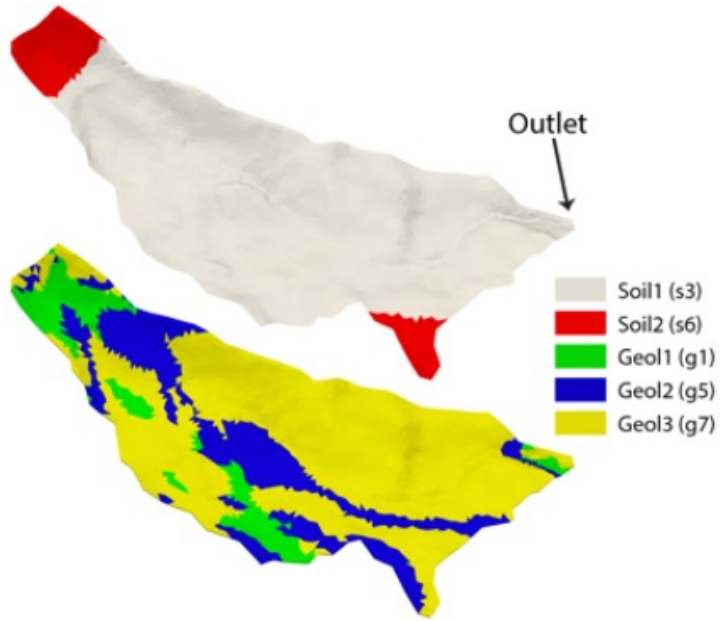
New approach to spatial downscaling of low-res precipitation

~12.5 km to model resolution ~100 m

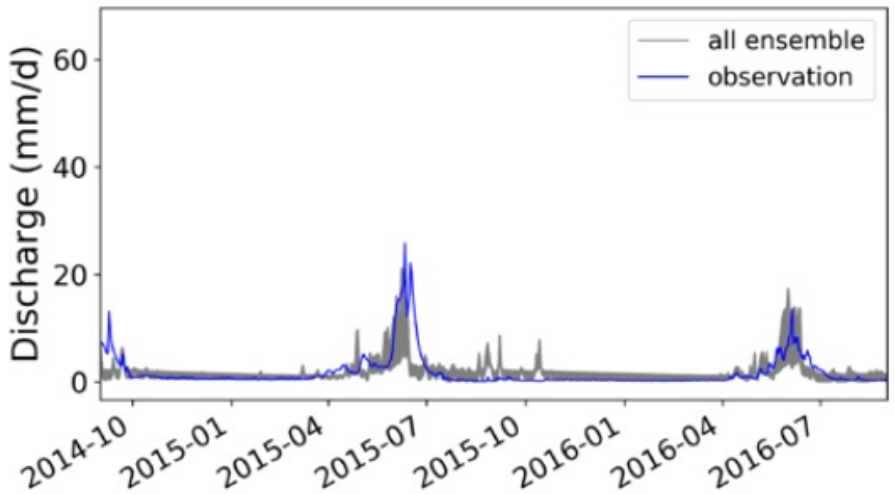
Machine learning can assist high-resolution simulation by replacing inverse modeling







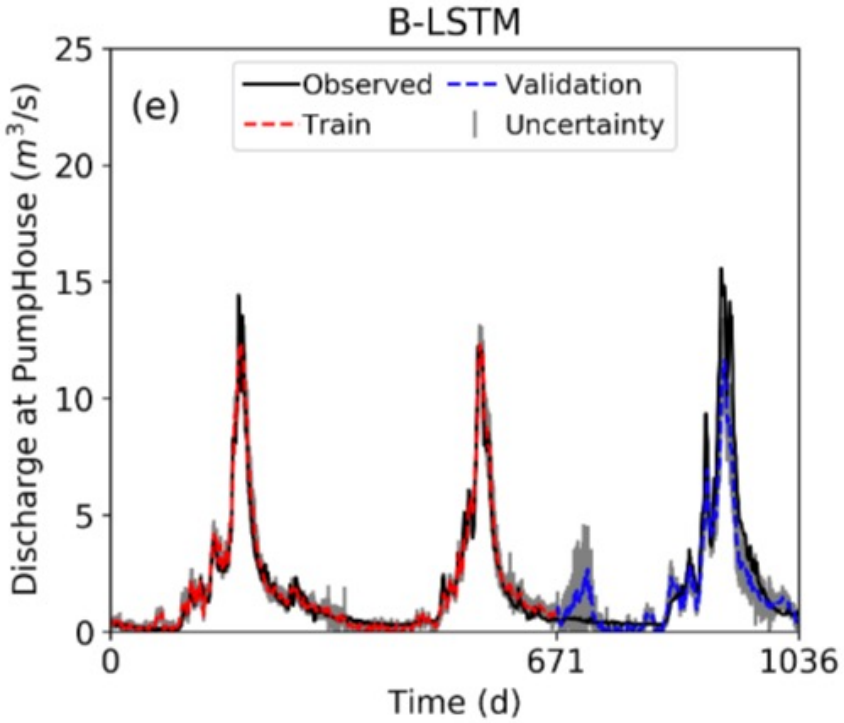
(b) Simulated vs observed discharge



Estimation of model parameters from stream discharge: Machine learning as an alternative to inverse modeling

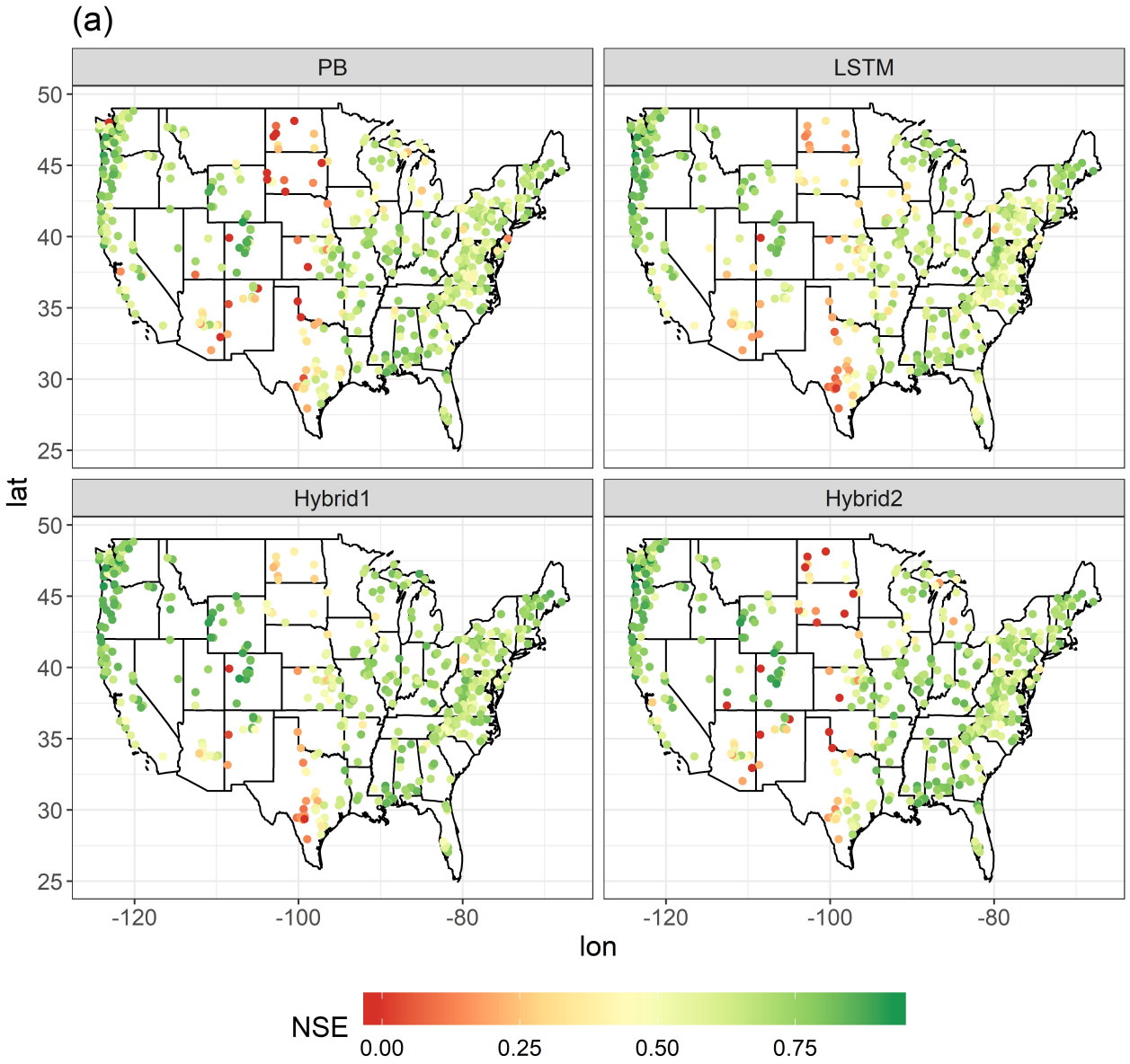
Hybrid simulation capability where machine learning refines output of process-based simulations

New hybrid model outperforms process-based hydrology model and pure data-driven approach



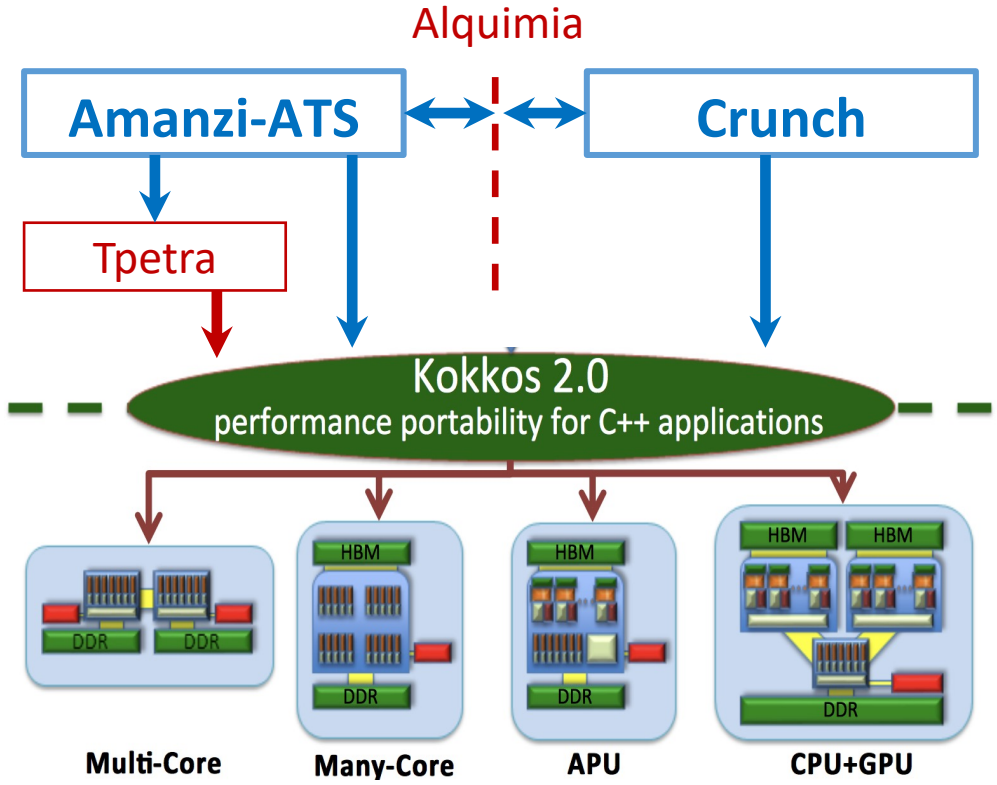
New Bayesian ML model greatly improves streamflow predictions using short training records

Improvement in streamflow prediction compared with hydrology code



Hydrobiogeochemical simulation capability for heterogeneous architectures

- ML-based data-model integration at scale requires higher computational throughput
- Prototyping Amanzi-ATS implementation using Kokkos abstraction layer
  - Required simultaneous adoption of Kokkos data and execution models
  - Proof-of-concept simulations solving Richards equation onGPU-based Summit supercomputer
- Preparing Crunch biogeochemical reaction engine for heterogeneous architectures
  - Profiling, refactoring, and analyzing performance of existing version
  - Design and preliminary implementation of C++ version

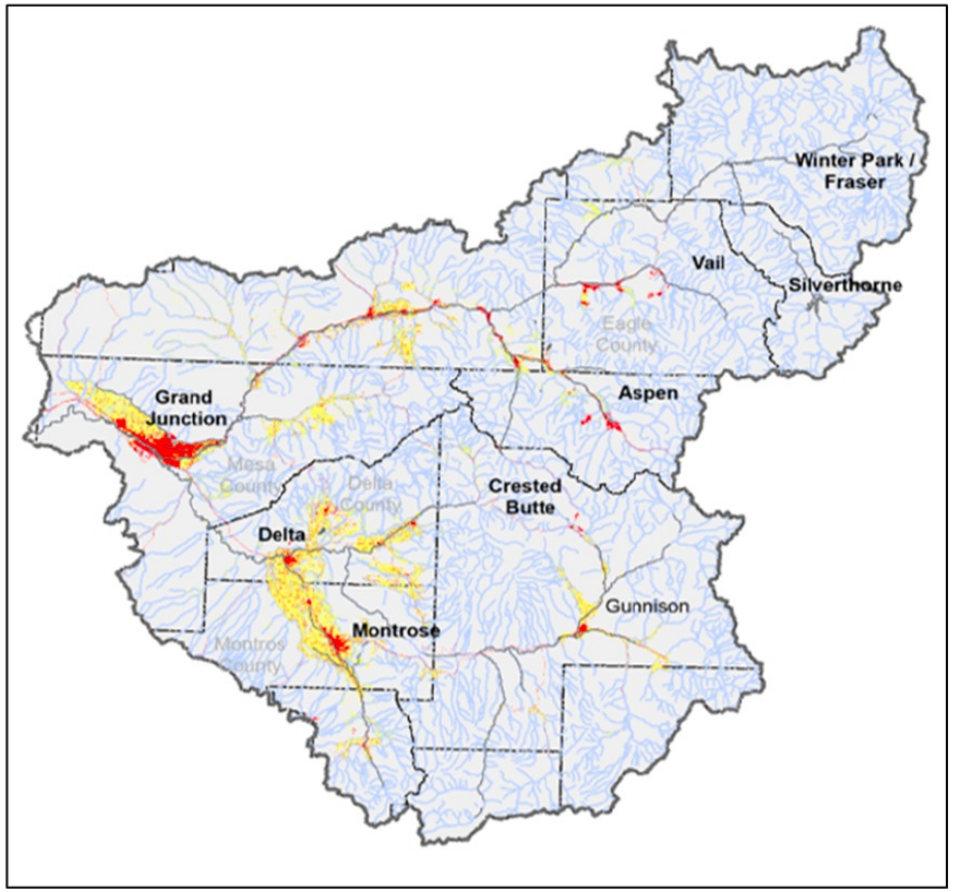


Looking forward – renewal proposal in preparation

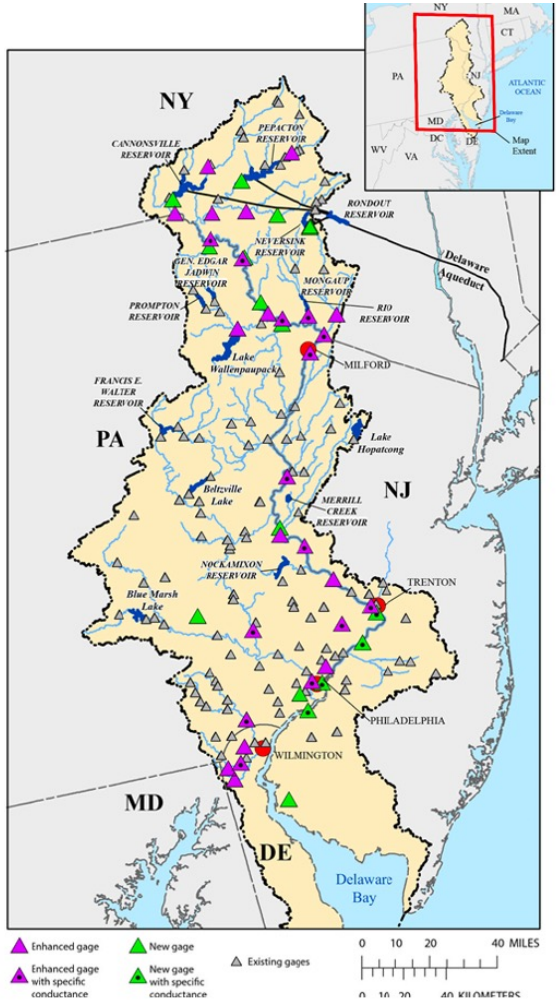
- Use cases will make use of USGS NGWOS and BER-supported data
  - Water availability in UCWRR
  - Water temperature and salinity intrusion in Delaware River Basin
- Refine and continue testing ML-based downscaling and inverse modeling, including extension to water temperature
- Continue testing our hybrid simulation capability
  - ATS-LSTM hydrology simulation capability focusing on non-stationary climate
  - Develop analogous hybrid capability for reactive transport
- Performance-portable ATS and CrunchFlow
- Scale to basin scales using multiscale algorithms that exploit watershed-based domain decomposition

Looking forward – renewal proposal in preparation

- Use cases will make use of USGS NGWOS and BER-supported data in basin-scale simulations
  - Water availability in UCWRR
  - Water temperature and salinity intrusion in Delaware River Basin



A



B