Machine Learning–enabled Model-Data Integration for Predicting Subsurface Water Storage

Dan Lu (lud1@ornl.gov)

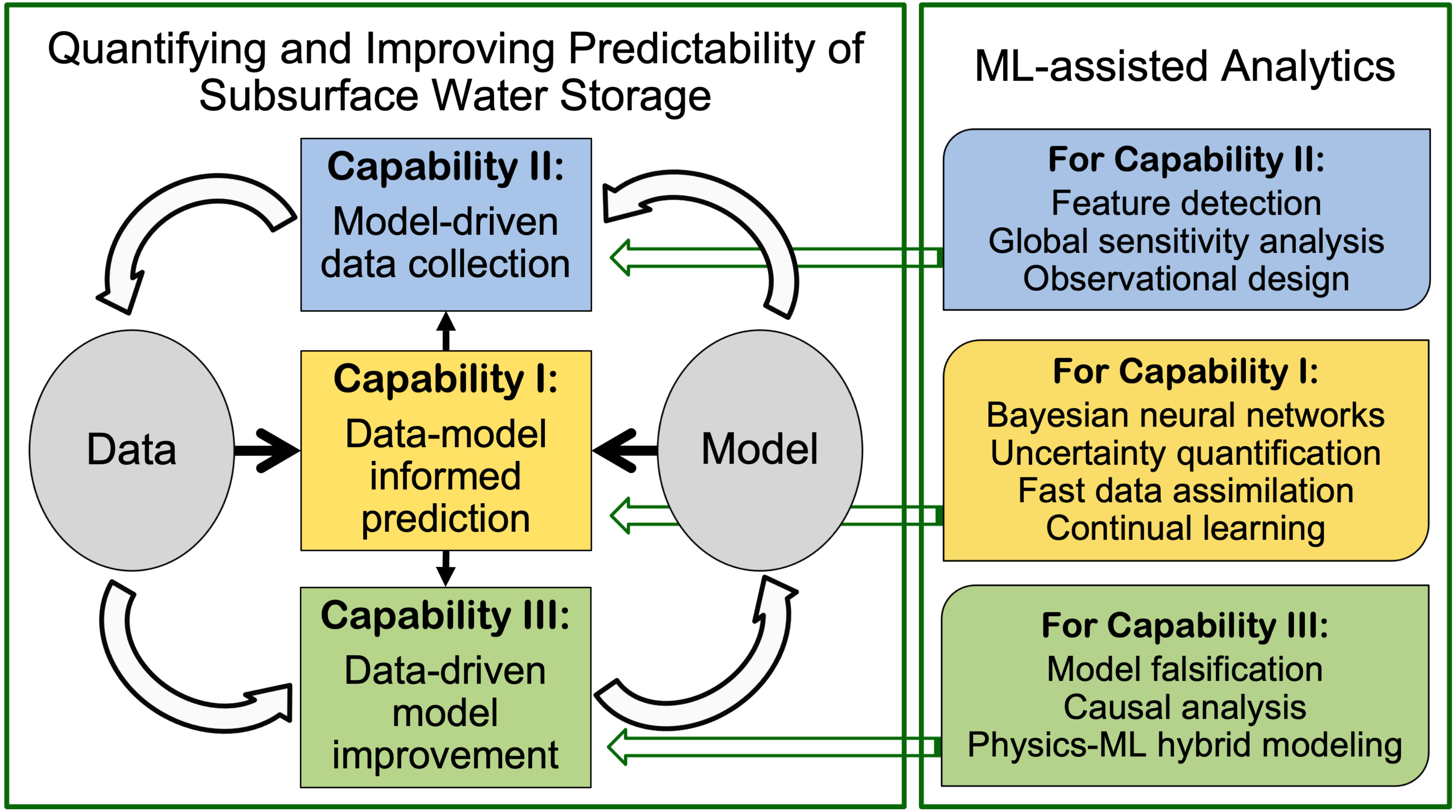
Eric Pierce (ORNL); Shih-Chieh Kao (ORNL); David Womble (ORNL)

Li Li (PSU); Daniella Rempe (UT Austin)

Summary of this work

- Subsurface water storage (SWS) plays a critical role in climate-change projections and can mitigate the impacts of climate change on ecosystems.
- However, because of the difficult accessibility of the underground, hydrologic properties and dynamics of SWS are poorly known. Direct observations of SWS are limited, and accurate incorporation of SWS dynamics into Earth system land models remains challenging.
- We propose a machine learning-enabled model-data integration framework to improve the SWS prediction at local to conus scales in a changing climate by leveraging all the available observation and simulation resources, as well as to inform the model development and guide the observation collection.

Machine Learning–enabled Model-Data Integration



Machine Learning–enabled Model-Data Integration

The proposed framework consists of three interconnected capabilities

- (I) a data-model informed prediction that links model and data and sufficiently extracts their information for prediction with considering various sources of uncertainty;
- (II) a model-driven data collection that analyzes data limits to predictability, identifies informative data, and guides data investment to enhance predictive skill;
- (III) a data-driven model improvement that analyzes model limits to predictability, identifies model deficiency, and complements missing physics with ML models to advance model development.

Capability I: A Novel Data-Model Informed Prediction

- The proposed prediction framework focuses on leveraging ML techniques to learn a relationship between data and prediction variables, and then deploys this learned ML model for direct prediction based on the actual observations.
  - Bayesian deep neural networks to learn the data-prediction relationship;
  - Surrogate modeling to accelerate the forward simulation;
  - Dimension reduction and feature detection to extract sample information;
  - Continual learning to assimilate data streams.

Capability II: Model-Driven Data Collection

- We propose to use feature detection and sensitivity analysis to guide the spatiotemporal data acquisition.
  - Feature detection techniques to identify where, what type and how much data are needed to improve the prediction.
  - A two-way global sensitivity analysis to identify key data variables and locations that constrain those uncertain parameters and processes that have a vital impact on predictions.
  - A value of information analysis for the cost-effective observational design.

Capability III: Data-Driven Model Improvement

- Model falsification and casual analysis will be used to inform the SWS dynamics implementation in physics-based land models.
  - Model falsification to analyze the consistency between the model generated data samples and the actual observations.
  - Causal analysis to explore the underlying variable interconnections from the data and generate new hypothesis.
  - A data-driven ML model from the hypothesis generation to compensate the missing SWS dynamics in the physics-based model.

Impacts and Future work

- SWS is a key variable of the climate system. It constrains plant transpiration and photosynthesis, with consequent impacts on the water, energy and biogeochemical cycles.
- SWS is involved in several feedbacks at the local, regional and global scales, and plays a major role in climate-change projections.
- We identified four intensively studied watersheds with diverse geology and climate for demonstration of the proposed idea.
- Diverse data sources at these four sites will provide inputs for the ML analysis. After testing and refining the techniques on the local scale, we will extend the framework to a continental scale.