



A post-processed NextGen framework reanalysis dataset



Savalan Naser Neisary¹, Md. Shahabul Alam², Josh Cunningham², James Halgren², Steven Burian^{1,2}

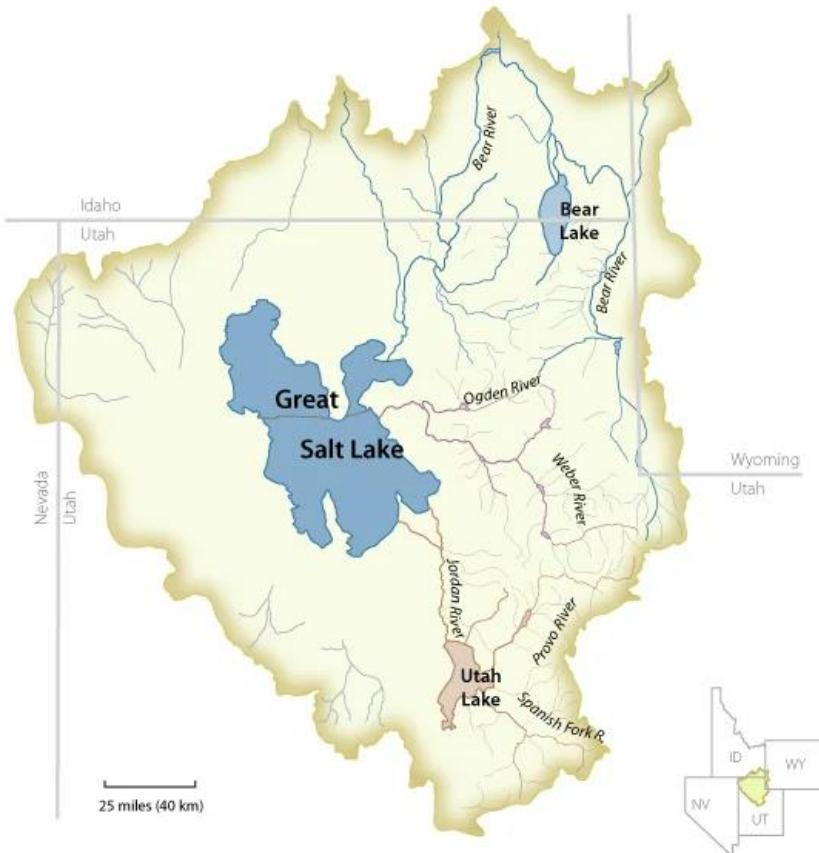
¹Department of Civil, Construction, and Environmental Engineering, The University of Alabama

²Alabama Water Institute, The University of Alabama

Drought impact in the western US

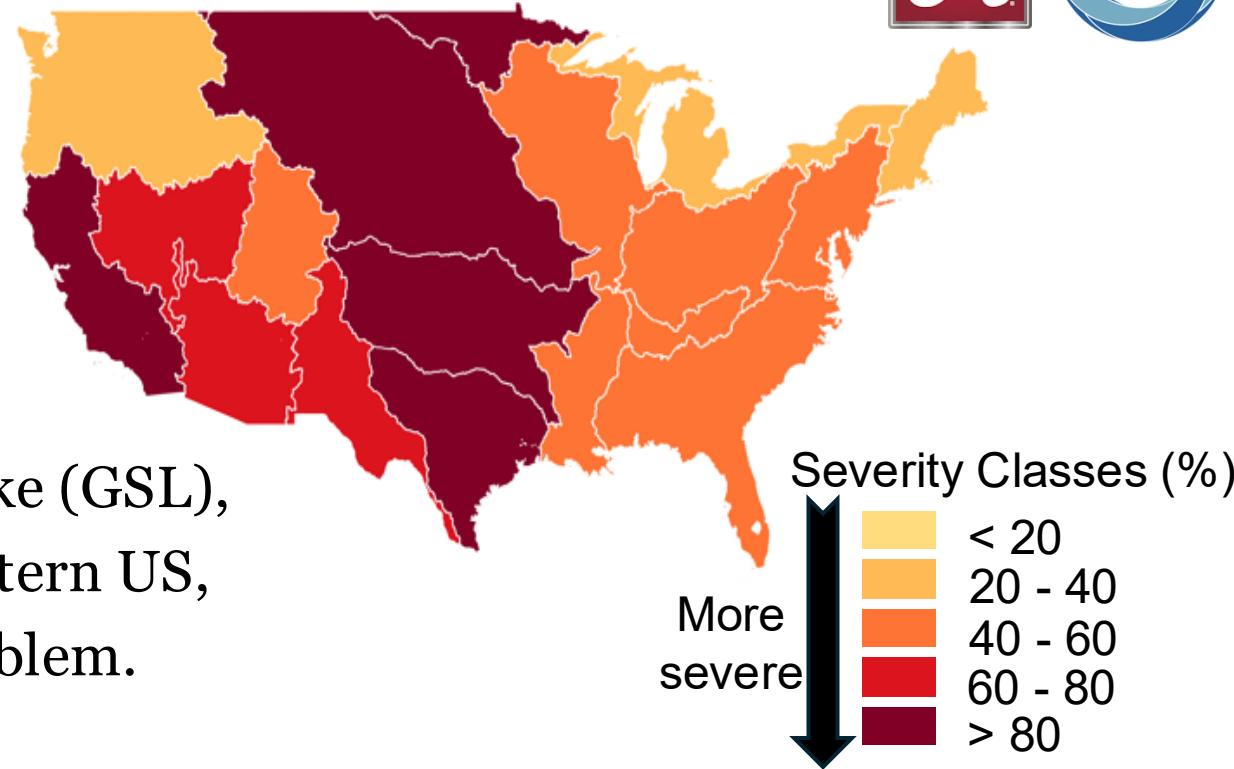


The US suffers from a drought.



Boyd, Eric S., et al. "Effect of salinity on mercury methylating benthic microbes and their activities in Great Salt Lake, Utah." *Science of the Total Environment* 581 (2017): 495-506.

The Great Salt Lake (GSL), located in the western US, has a drought problem.

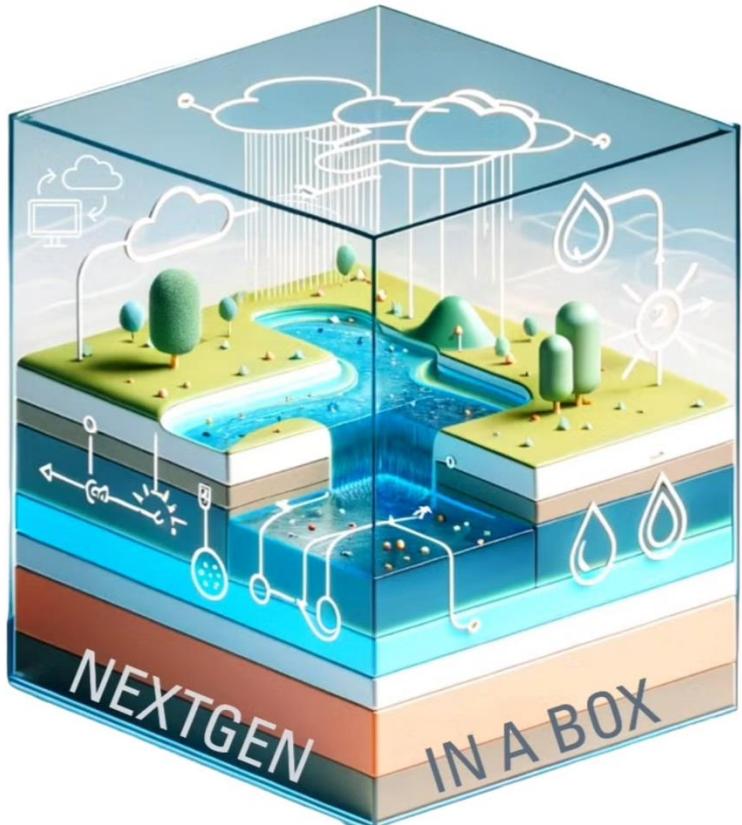


Accurate streamflow predictions are critical to assist water managers in preserving GSL.



Next Generation Water Resources Modeling Framework (NextGen)

- NextGen is an open-source, flexible, and modular framework.
- NextGen is the future version of the National Water Model.
- NextGen In A Box (NGIAB) is a containerized, ready-to-run package of NextGen.
- NextGen has low accuracy downstream due to extensive human infrastructure.



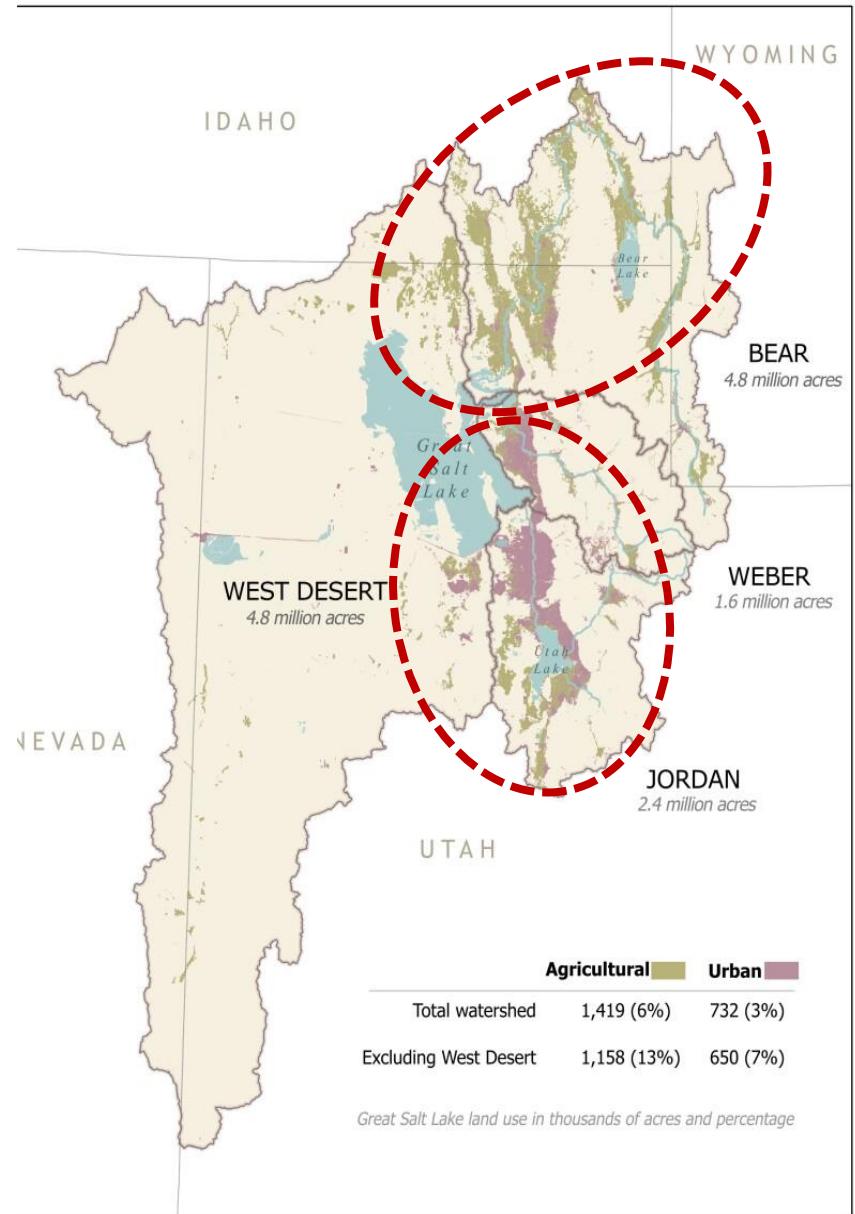
Objective

- How can we improve NextGen results in a watershed with extensive water resources?
- Create an ML framework to enhance NextGen flow simulations in the GSL watershed by accounting for water resources infrastructure and capturing dominant hydrological processes.



Great Salt Lake

- The GSL watershed includes Bear, Weber, and Jordan sub-basins.
- Evaporation is the only outflow, and precipitation, groundwater, and streamflow are inflows.
- Snow melt (Snow Water Equivalent) supplies 95% of streamflow.
- GSL has an extensive water infrastructure.



Post-Processing

- Post-processing corrects biases by transforming model outputs based on the relationship between observations and the model.
- XGBoost stands for Extreme Gradient Boosting.
- Builds models sequentially
- It is fast, efficient, and robust in handling sparse data.



Post-Processing Machine Learning-Based (PP-ML) Framework



Data Collection

- SWE
- NextGen Data
- Catchment Characteristics
- Reservoir Storage
- Meteorological Variables
- Seasonality Index
- Streamflow Characteristics

Hyperparameter Tuning

- Bayesian Optimization
- Cross-validation

Feature Analysis

- XGBoost Feature Importance

Model Evaluation

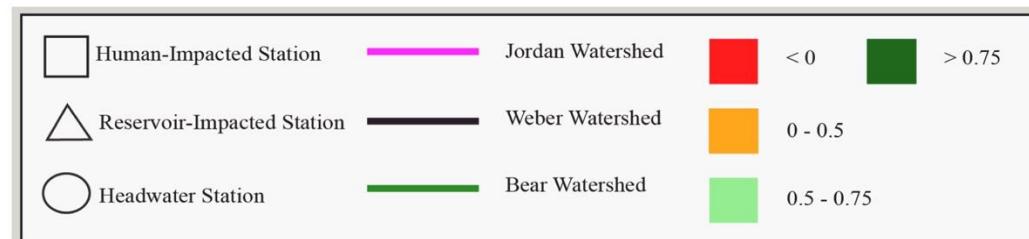
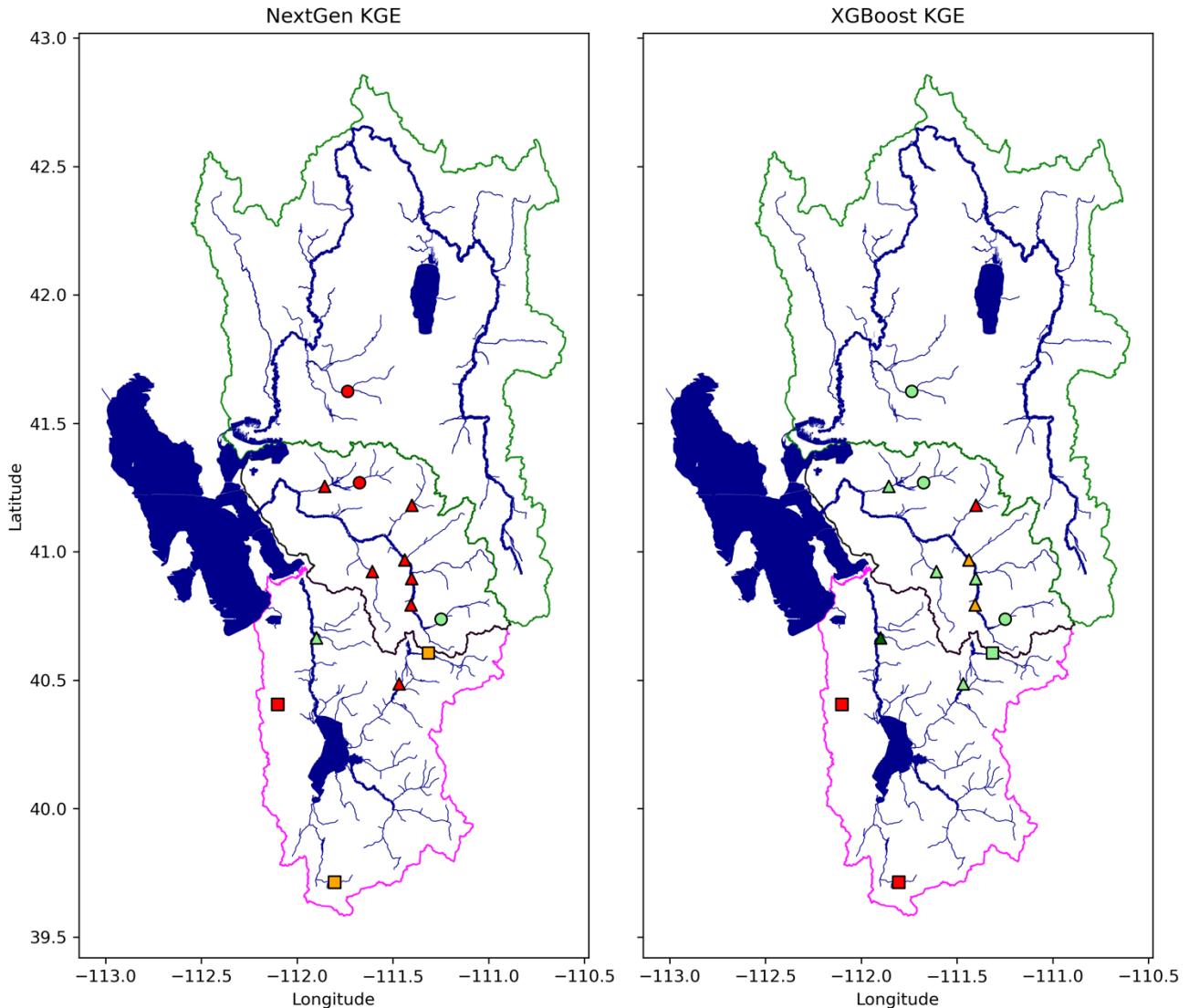
- KGE, PBias, and RMSE
- Hydrological Signatures

Model Training

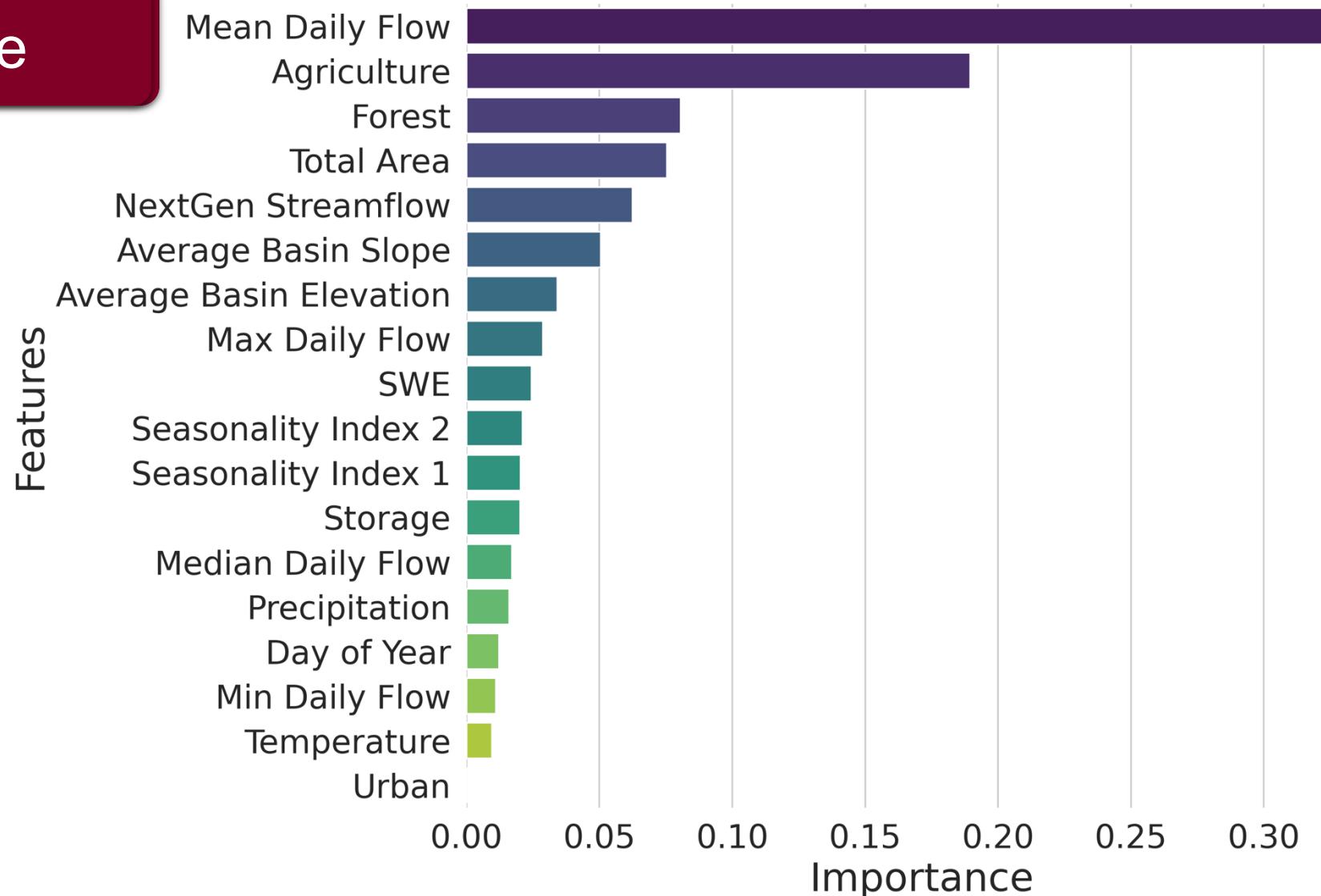
- XGBoost
- Training: 2007 – 2020
- Testing: 1990 – 2006
- 18 stations.

Improvement Across Most Stations

- Most stations reach higher than 0.5 KGE.
- The number of stations with negative KGE decreased to 3.
- Improved all the stations impacted by reservoirs.

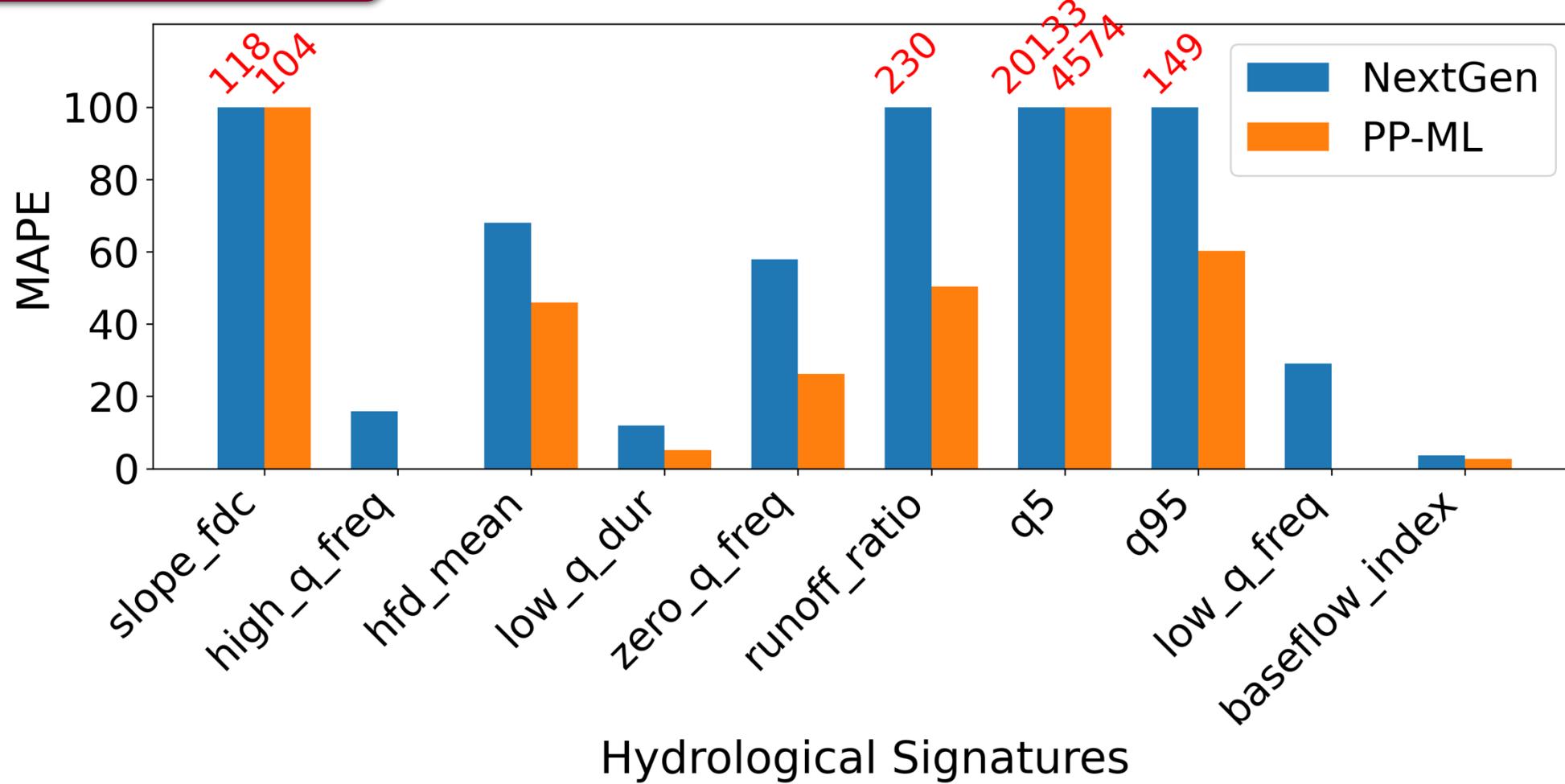


Feature Importance



- SWE, storage, and NextGen streamflow have high importance.
- Catchment characteristics are critical for the model.

Hydrological Signatures



- Improvements in all the hydrological signatures.
- Signatures related to Low-flow show significant improvement, but still not good!

Key Findings

PP-ML can incorporate water regulation and dominant hydrological processes without requiring coding within the model.

Catchment characteristics, SWE, and NextGen streamflow are the most important features.

Modified reanalysis data assist long-term water resources planning by the Utah Water Division.

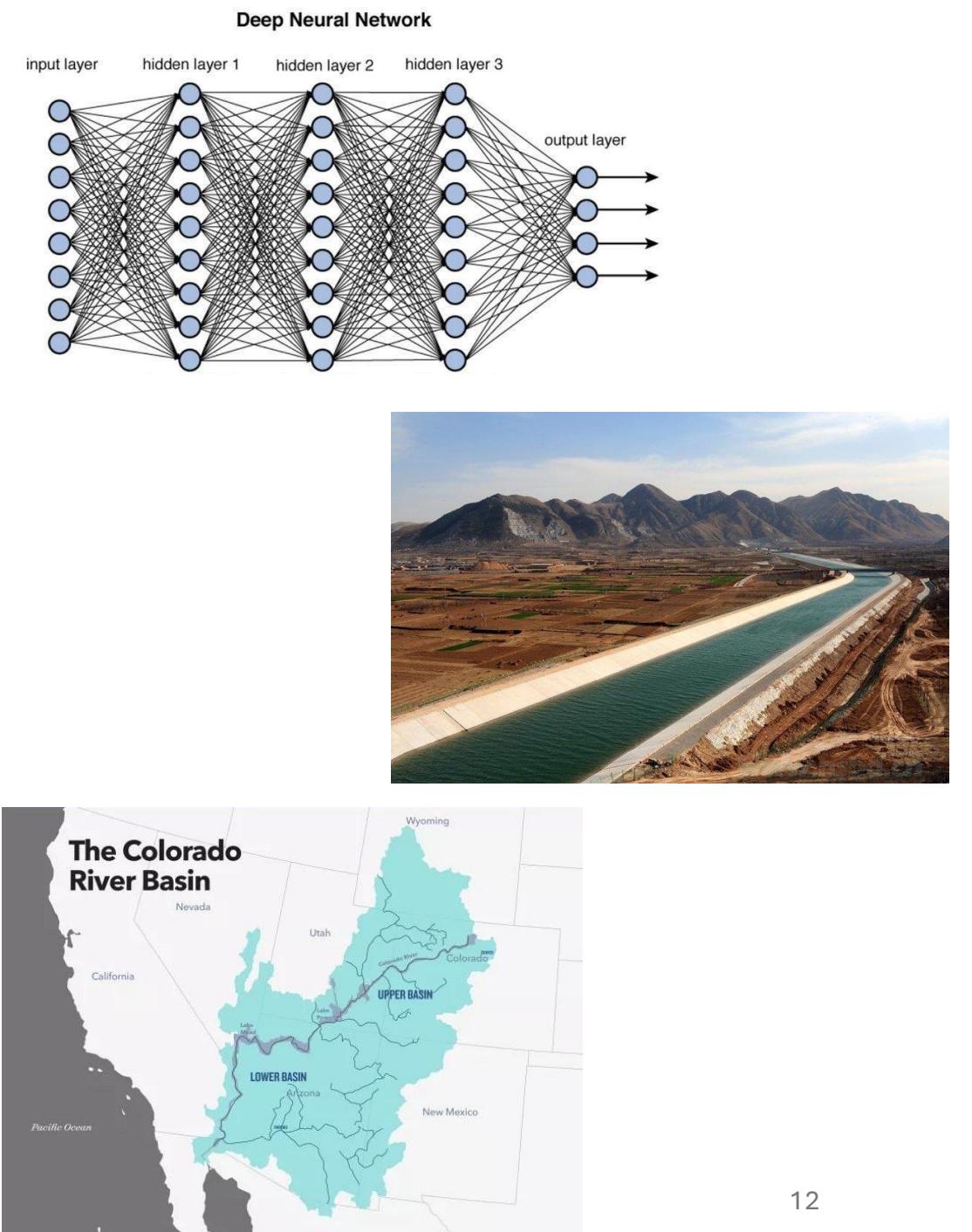
PP-ML facilitates the integration of NextGen into water supply forecasting and Great Salt Lake management and decision-making.

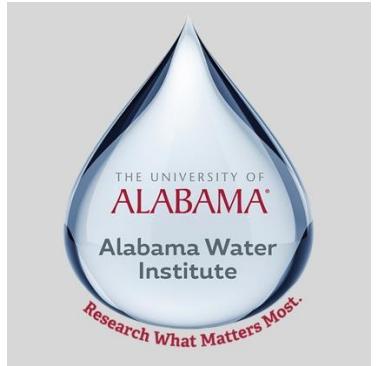
Future Work

Develop Deep Neural Networks.

Investigate other natural and anthropogenic variables as inputs.

Expanding the research area to other regions, including Alabama.

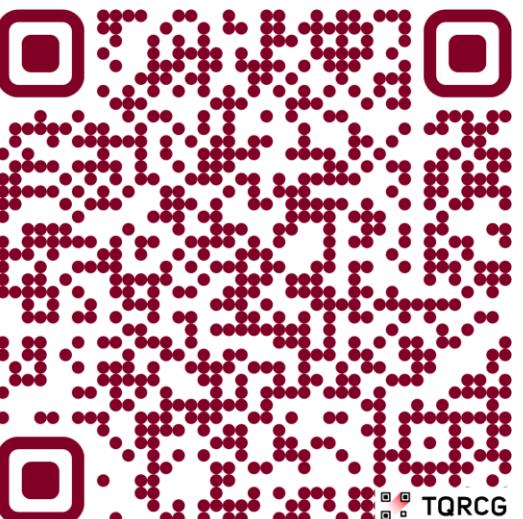




Thank You!



snaserneisary@crimson.ua.edu



NGIAB paper



PP-ML paper