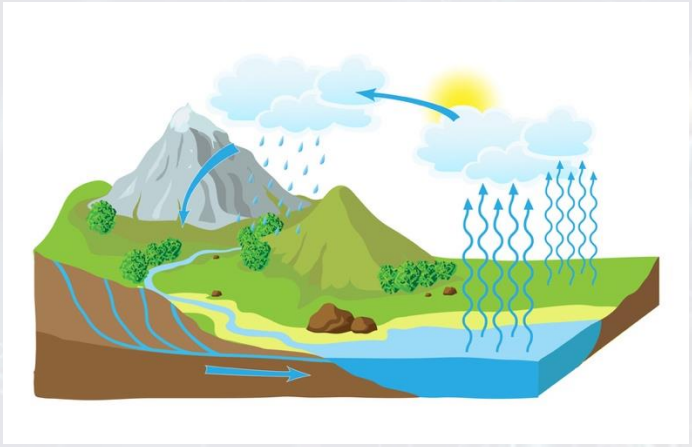
# Machine Learning for Post-Processing Hydrological Model Outputs

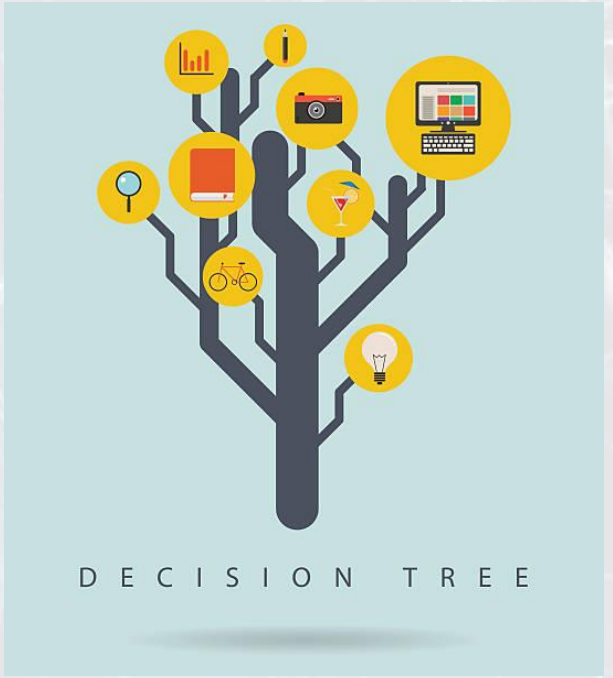
Savalan Naser Neisary
The University of Alabama

# Outline

- Workshop Goals
- Project Background
- Theoretical Background
- Dataset
- Model Development

# What are the Workshop Goals?

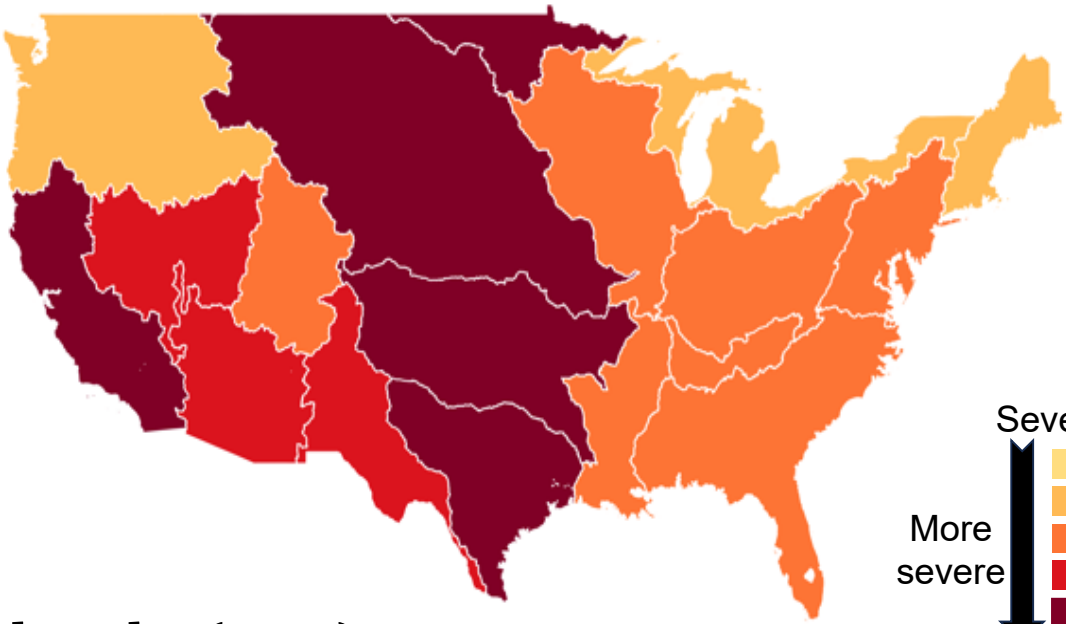


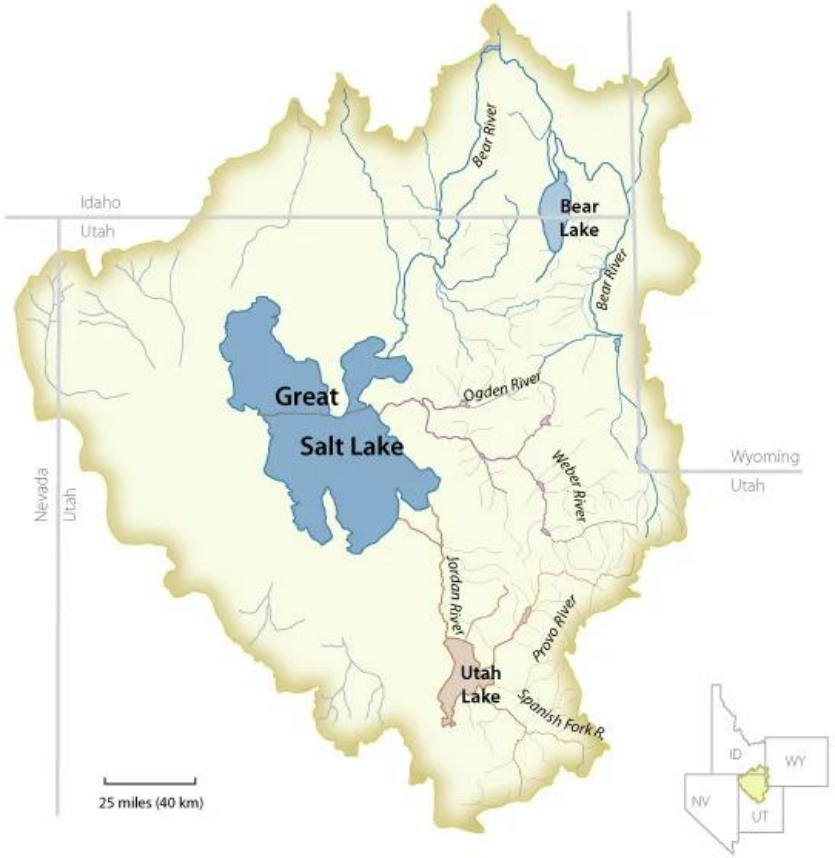



- Gain a general understanding of post-processing and machine learning algorithms.
- Learn how to develop Long Short-Term Memory (LSTM) models for post-processing a real-world problem.
- Learn how to tune model hyperparameters and their importance.

# Drought in the Western US

The US suffers from a drought.




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

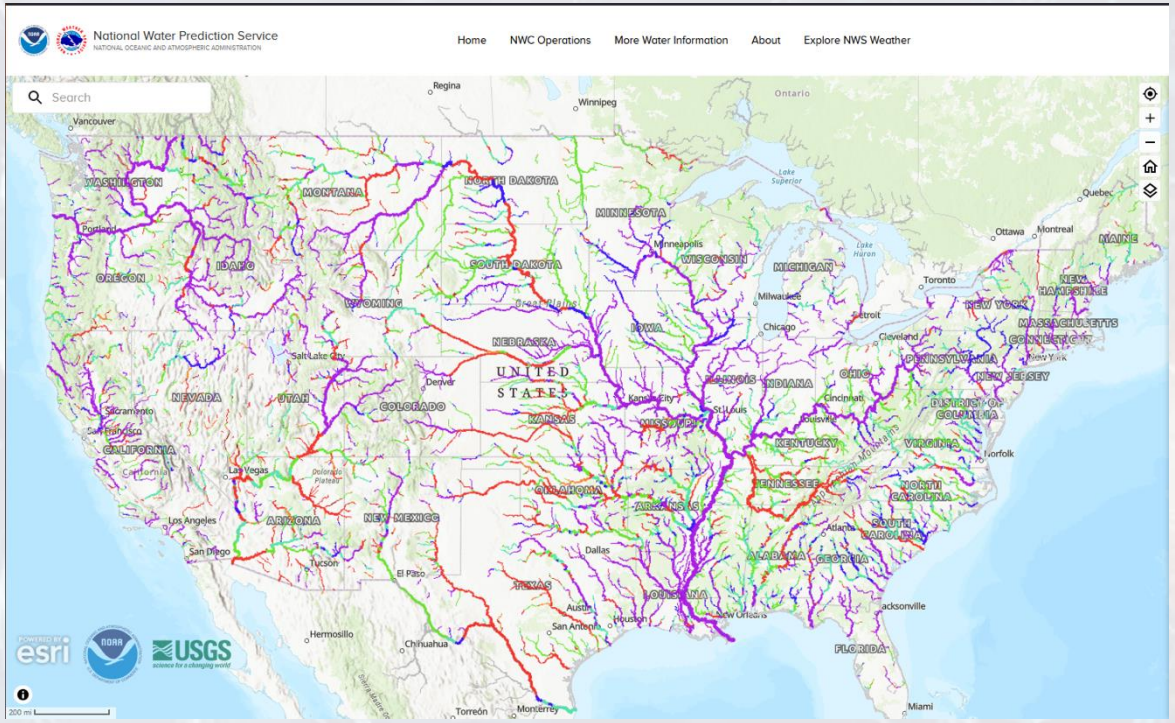
https://www.upr.org/programs/2019-09-23/revisiting-the-disappearing-great-salt-lake-with-wayne-wurtsbaugh-on-mondays-access-utah

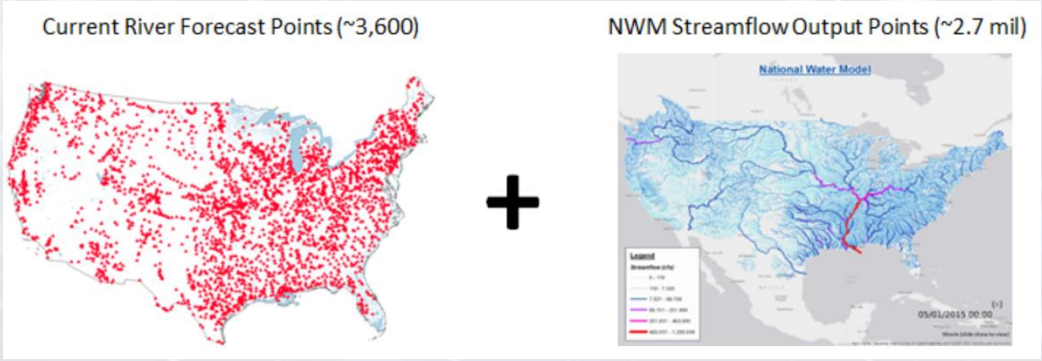
GSL is shrinking due to droughts and increased demand, which will cause health and economic problems.

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 National Water Model

- Addresses the need for a consistent, large-scale forecast.
- Created by NOAA's Office of Water Prediction.
- Developed based on WRF-Hydro.
- Provides predictions for 2.7 million reaches.
- Our evaluation showed it has low accuracy downstream due to extensive human infrastructure.



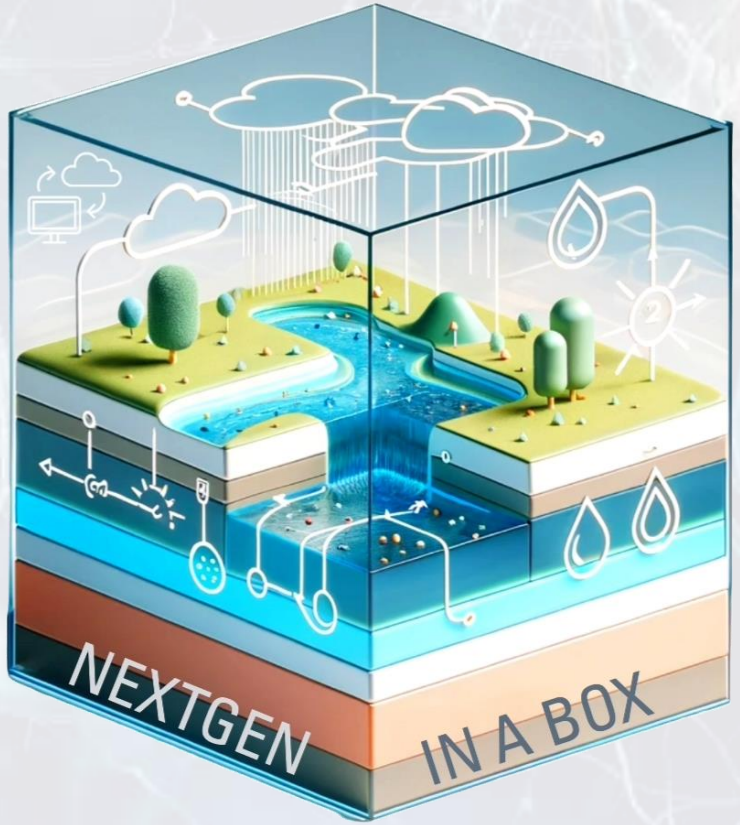



# NextGen Framework

- It will serve as the next version of the NWM.
- A modular hydrologic modeling framework developed by NOAA and partners.
- Enables rapid integration of new models, datasets, and machine learning.
- We used the CFE v.1.0 module of NextGen Framework.
- NGIAB provides a containerized and user-friendly solution for running the NextGen framework, allowing you to control inputs, configurations, and execution on your local machine.

NGIAB

Data Preprocessor



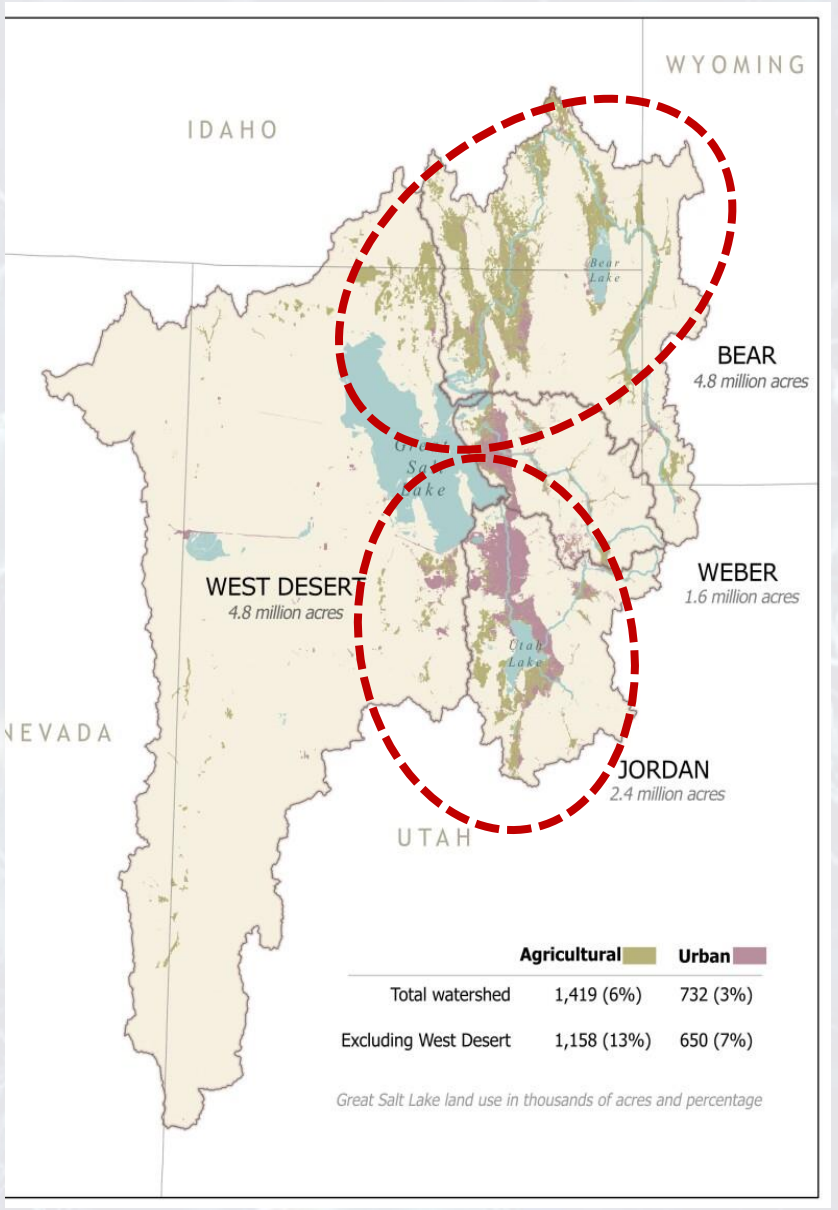
# Objective

- How can we improve NextGen results in a watershed with extensive water resources like GSL?

- Create an ML framework to enhance NextGen flow simulations in the GSL watershed by accounting for water resources infrastructure and bypassing reservoir-release parameterizations.

# Great Salt Lake

- GSL watershed includes Bear, Weber, and Jordan sub-basins.
- Evaporation is the only outflow, and precipitation, groundwater, and streamflow are inflows.
- The Bear River is the largest tributary (~55% of flow volume)
- High irrigation and urbanization.



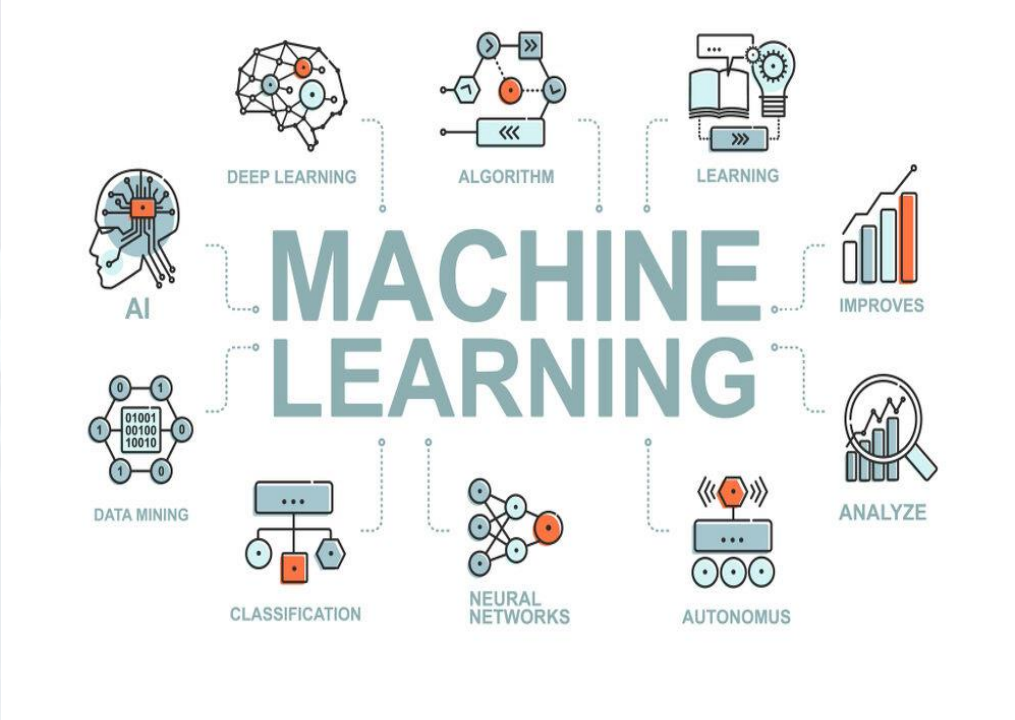

*Richter, Brian D., et al. "Reducing irrigation of livestock feed is essential to saving Great Salt Lake." Environmental Challenges 18 (2025): 101065.*

# Post-processing Hydrological Predictions

- There are different ways to improve the predictions, including post-processing.
- Post-processing corrects biases by transforming model outputs based on the relationship between observations and the model.
- ML models proved to be useful in post-processing.

# Machine learning

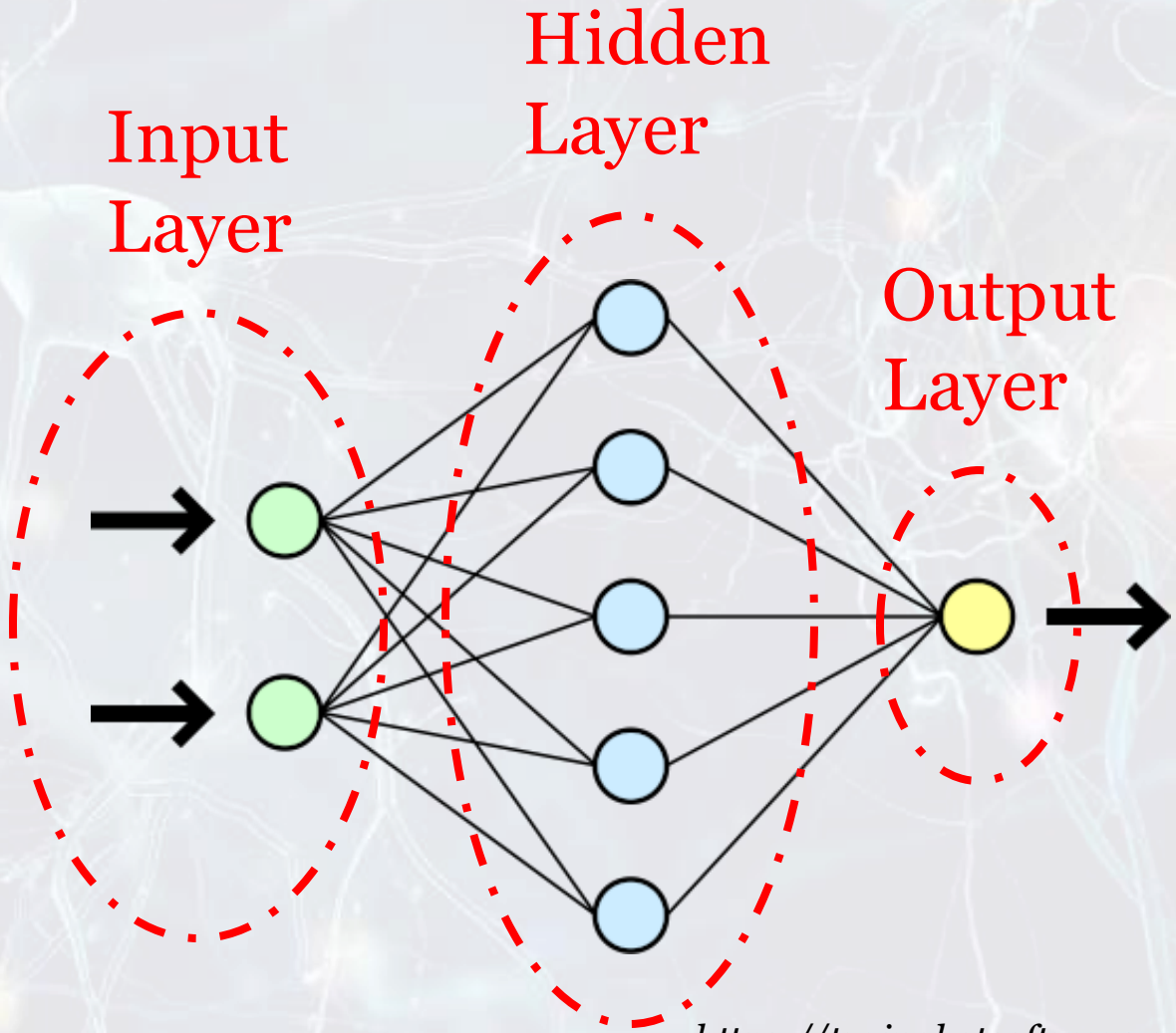
- A method of teaching computers to learn patterns from data.
- Can be supervised, unsupervised, or reinforcement-based.
- Often used for classification, regression, and clustering tasks.
- Improves automatically with more data and experience.
- Powers applications from spam detection to flood forecasting.




*https://jmesgray.medium.com/weka-software-for-machine-learning-6d1114a76143*

# Neural Networks

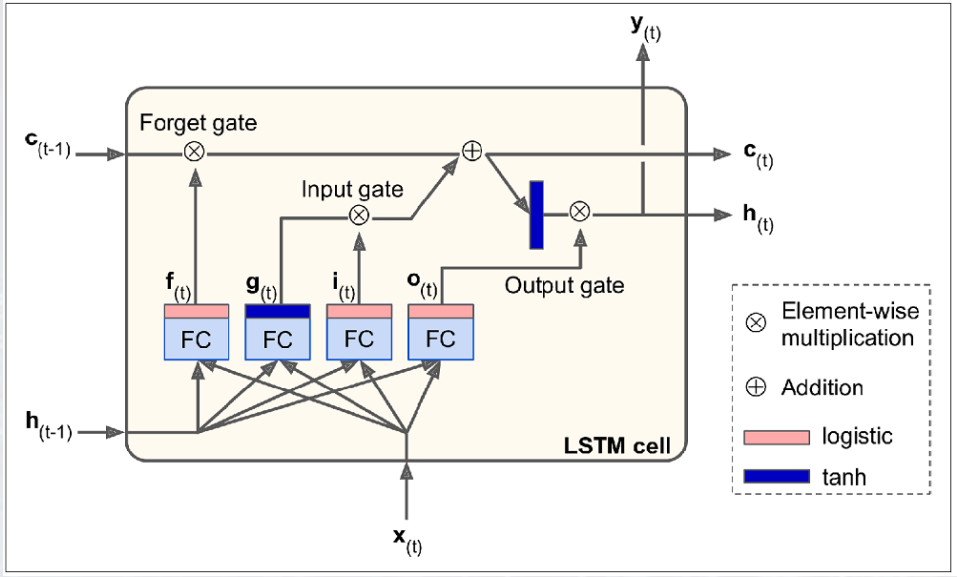
- Inspired by the structure of the human brain.
- Made up of layers of interconnected nodes (neurons).
- Learn complex relationships by adjusting weights during training.
- Feedforward networks are the most basic type.
- Form the foundation of deep learning.




*https://turingbotsoftware.com/blog/neural-network-alternatives/*

# Long Short-Term Memory (LSTM)

- Long Short-Term Memory network: a type of Recurrent Neural Network (RNN).
- Designed to handle sequences and time-series data.
- Remembers long-term dependencies using gated memory cells.
- Useful in applications like speech recognition and time series forecasting.
- Bidirectional LSTM processes data in both forward and backward directions for improved context understanding.



# How do we use LSTM?

## Data Set

- We collected three NHD reaches with USGS monitoring stations.
- 2007 to 2020 Training
- 1990 to 2006 Testing

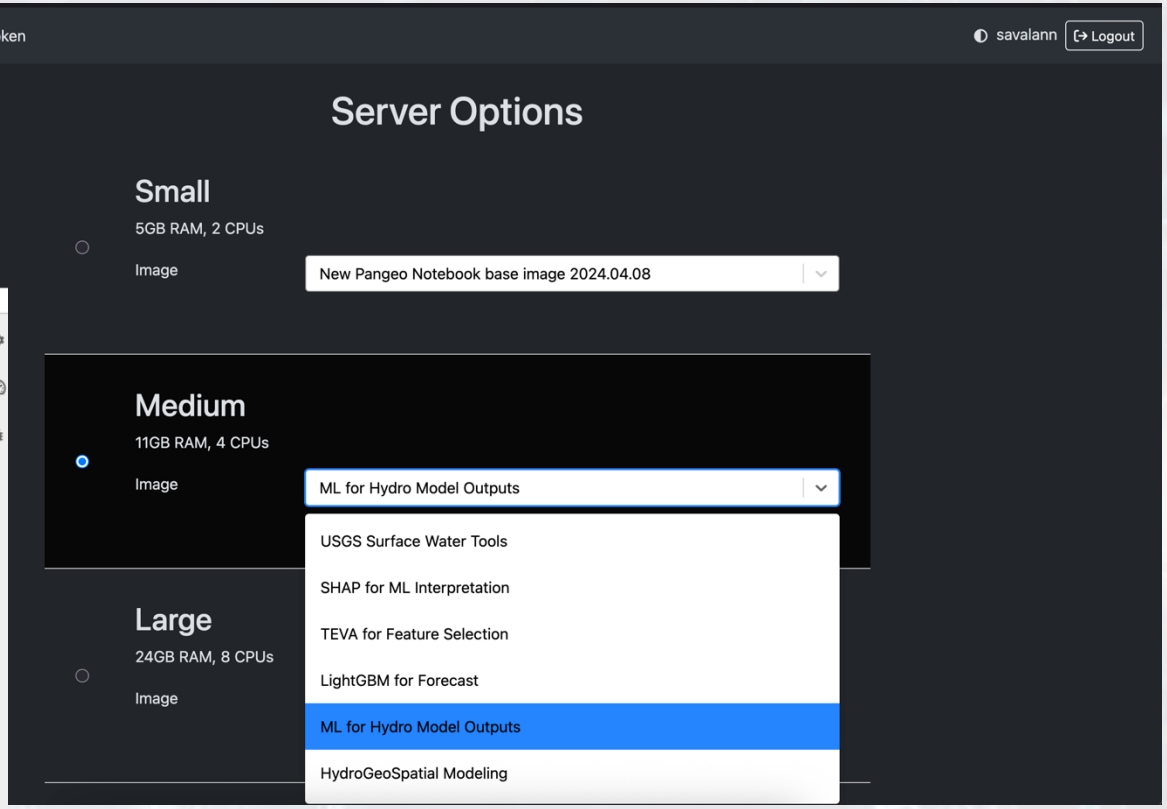
## Evaluation Metrics/Methods

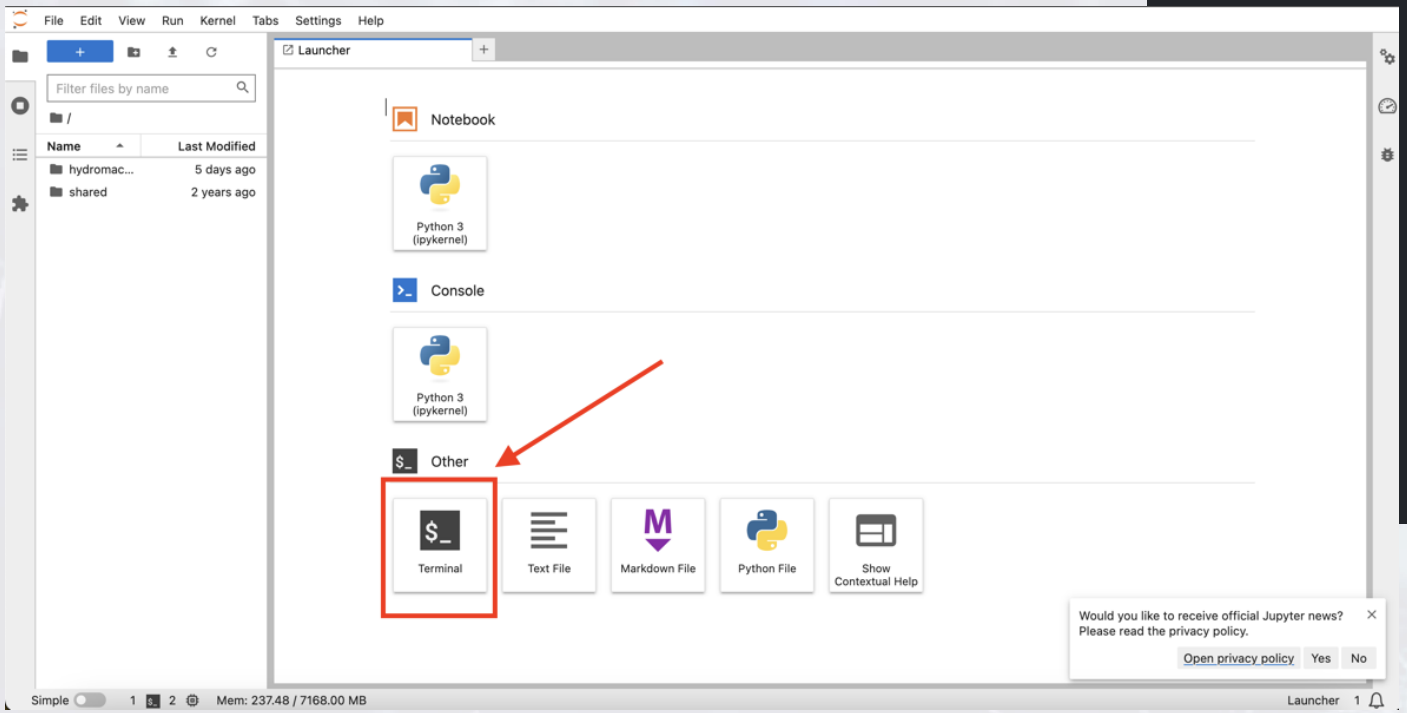
- KGE, PBias, and RMSE
- Streamflow Regimes (Low, High, Normal)
- Hydrological Signatures

## Input Features

- SWE
- Catchment Characteristics
- CFE Streamflow Results
- Upstream Storage
- Precipitation and Temperature
- Seasonality Index

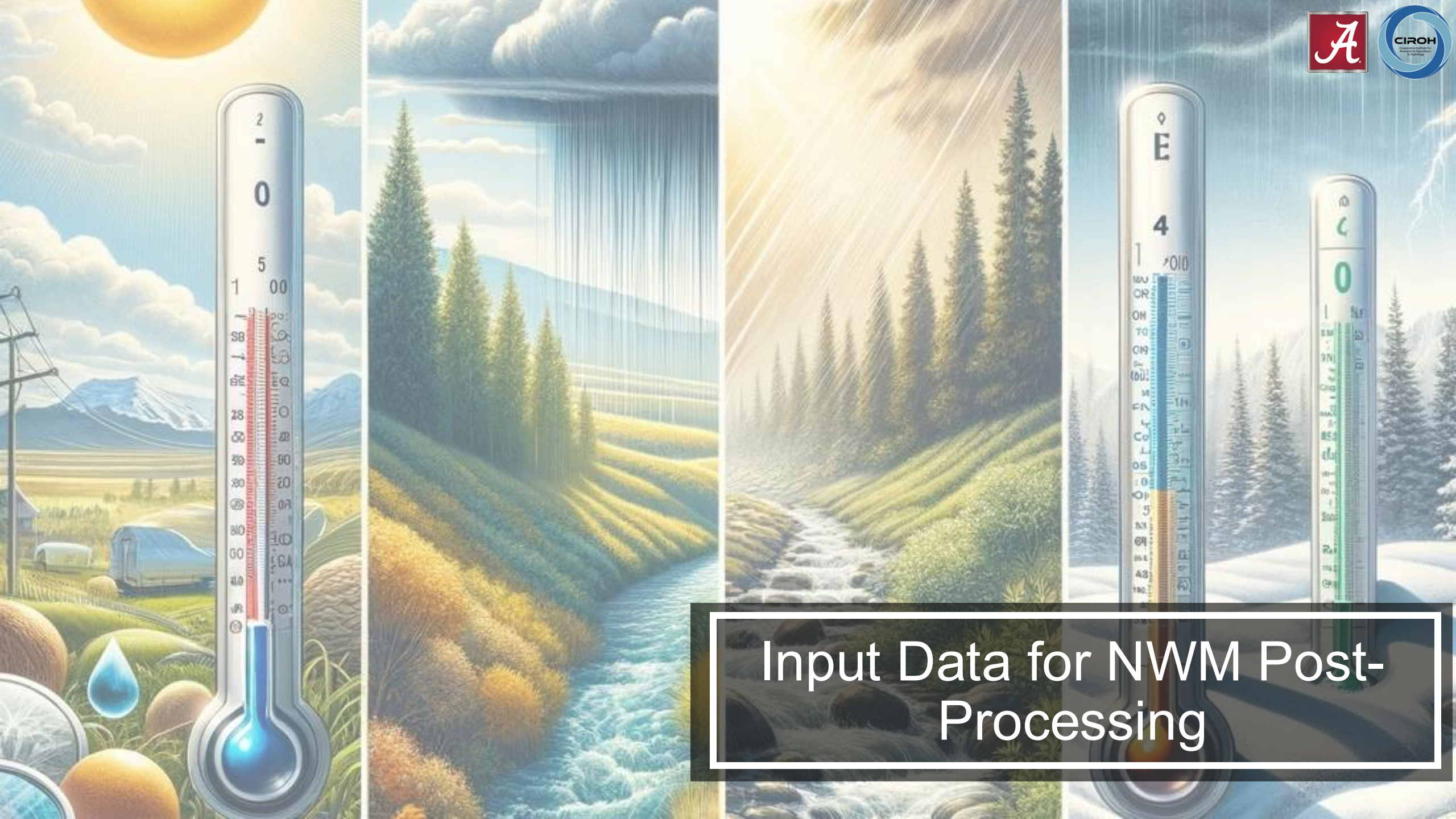
# Let's Start Coding!!!




git clone https://github.com/savalann/hydromachine-tutorials.git

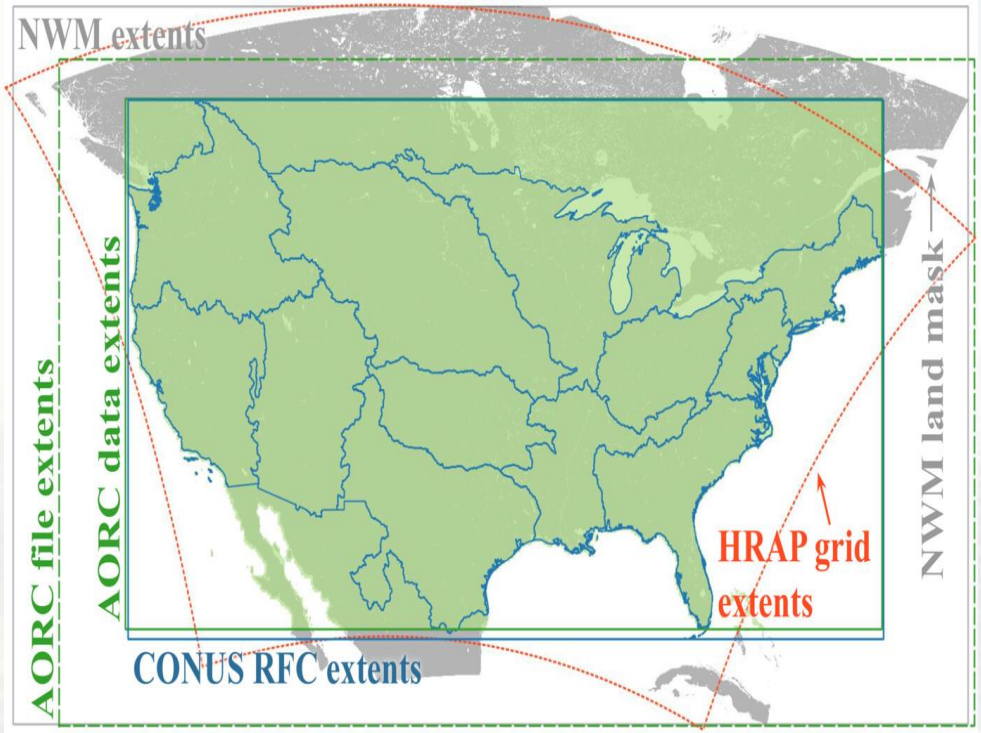
Hydromachine-tutorials →devcon_2025 → 01.script



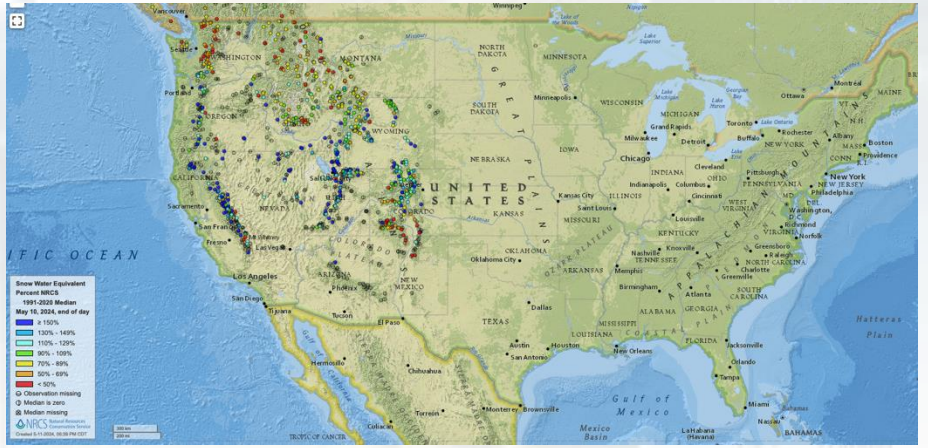
# Input Data for NWM Post-Processing

# Analysis of Record for Calibration (AORC)

- It was developed for operational hydrological modeling.
- It has CONUS-wide gridded historical forcing data from 1979 to now, with near-real-time updates lagging by nine days.
- Spatial resolution is 800 m, and temporal resolution is one hour.
- AORC uses NLDAS-2, LIV16, NEXRAD Stage IV, URMA, and PRISM/NCEI/OWP as data

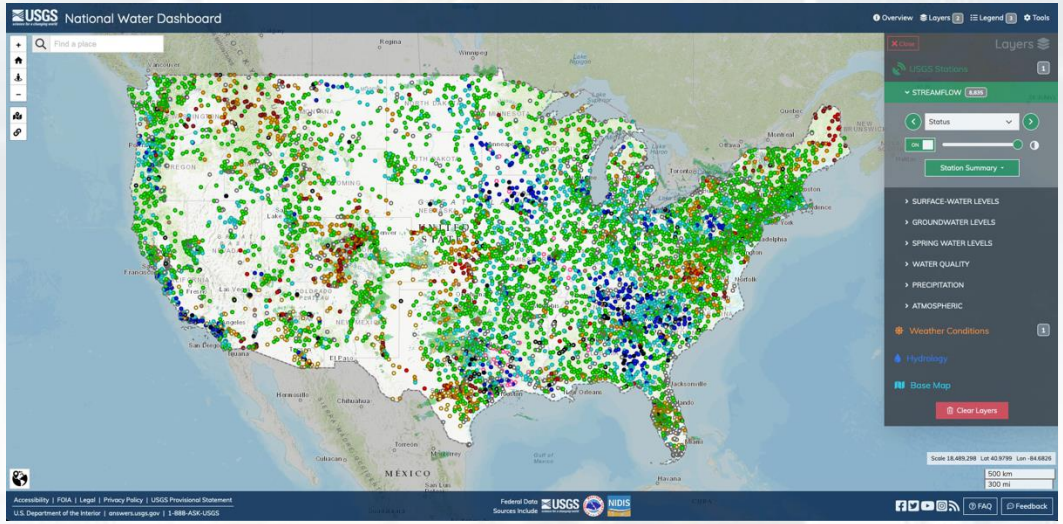


# Snow Telemetry (SNOTEL) Network



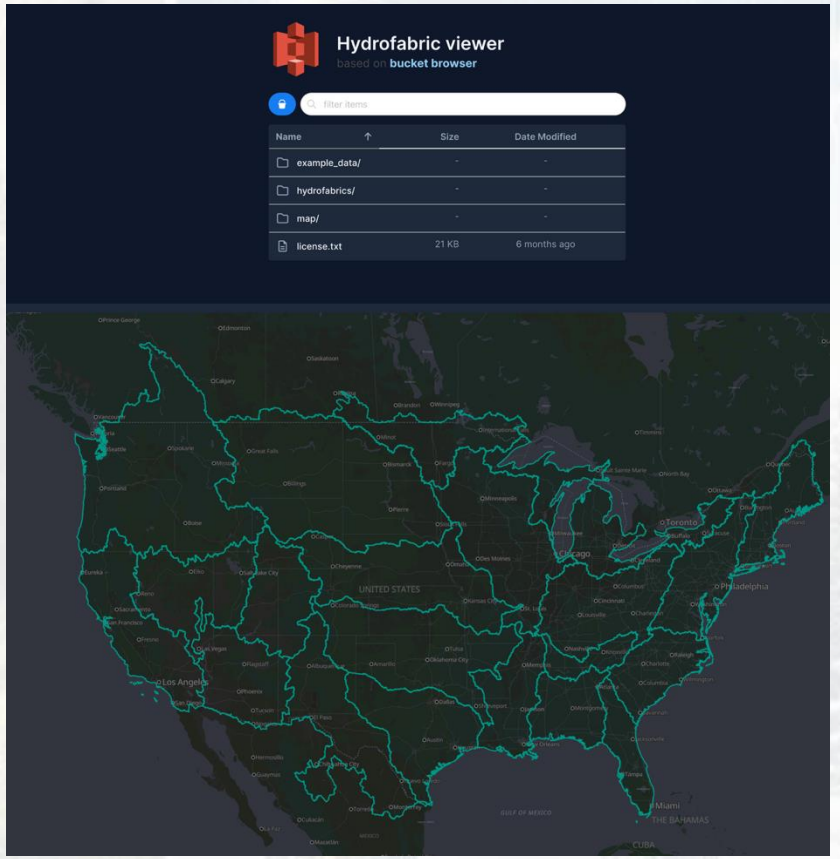

- A part of the Snow Survey and Water Supply Forecasting (SSWSF) Program.
- The SNOTEL network comprises over 900 automated data collection sites in remote, high-elevation mountain watersheds in the western U.S.
- SNOTEL sites are designed to operate unattended without maintenance for a year or more.
- They collect other data, such as soil moisture and temperature measurements at various depths, solar radiation, wind speed, and relative humidity.

# USGS



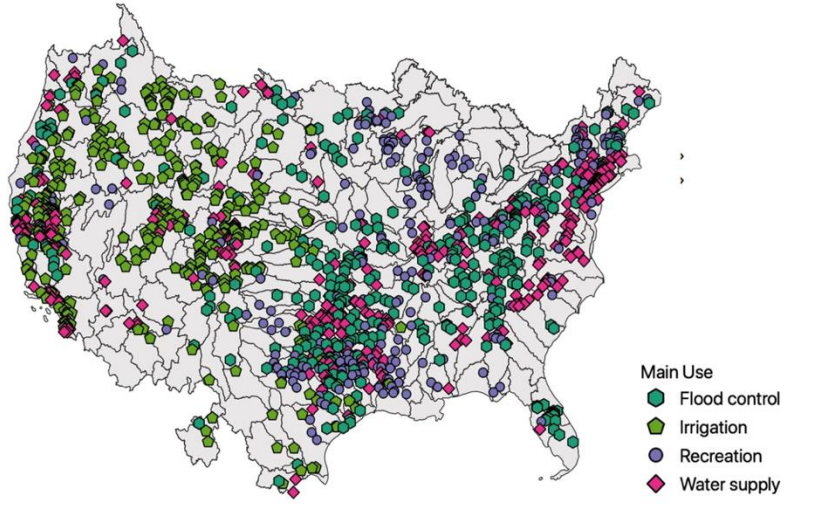

- The United States Geological Survey (USGS) has collected data on water resources at approximately 1.5 million sites in all 50 states.
- The data types collected are of surface water and groundwater.
- Surface-water data include gage height (stage) and streamflow (discharge).
- Groundwater data, such as water level, are collected at wells and springs.

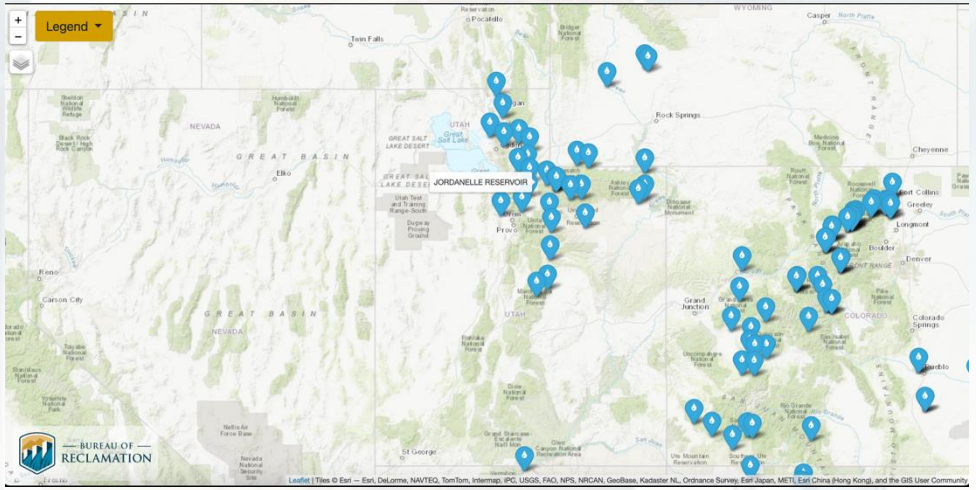
# Catchment Characteristics




- A hydrofabric is a geospatial dataset representing hydrological features, such as rivers, lakes, and catchments, and their interconnections.
- Using the NGIAB data preprocessing tool, we identified the hydrofabric watershed corresponding to each USGS station, including all upstream hydrofabric basins contributing to its flow.

# ResOps Dataset




- The Bureau of Reclamation (BOR) manages most of the reservoirs in the Western US and provides data on stream stations and water diversions.
- ResOps is a dataset that provides historical reservoir operations data, including inflows, outflows, and storage levels, for major U.S. reservoirs, enabling analysis of water management practices and hydrologic modeling.