Planning for Floods & Droughts: Intro to AI-Driven Hydrological Modeling

Kshitij Tayal[1], Arvind Renganathan[2], Siyan Liu[1], Dan Lu[1]
[1]Oak Ridge National Laboratory, [2]University of Minnesota

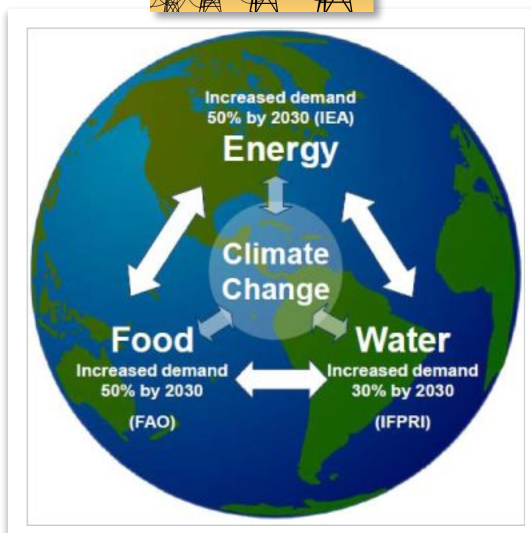
Environmental Grand Challenges of the 21st Century

IPCC Report warns of 'irreversible' impacts of global warming *2/28/2022*

Increasing frequency of natural disasters

U.S. drought one of the worst in 1,200 years
Science News APRIL 16, 2020

North American Drought nearly 50 percent more severe



Extreme flooding to increase as temperatures rise
The Washington Post September 13, 2021

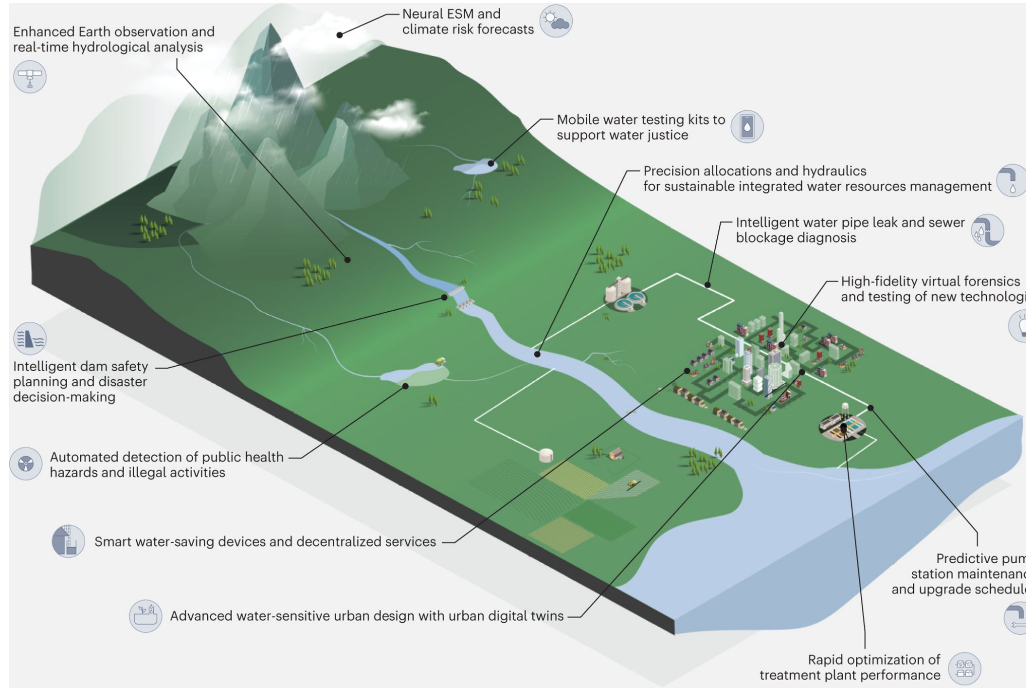
The Ahr River floats past destroyed houses in Insul, Germany

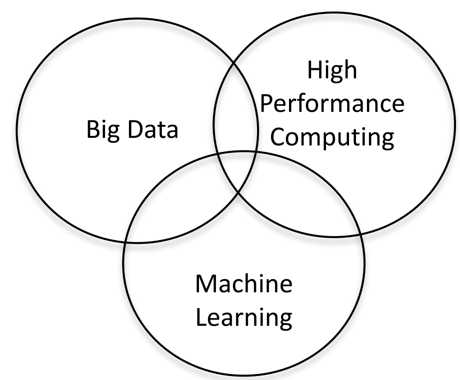
Water under stress

Aral Sea in 1989 Aral Sea in 2014

Harnessing the Data Revolution for Scientific Discovery

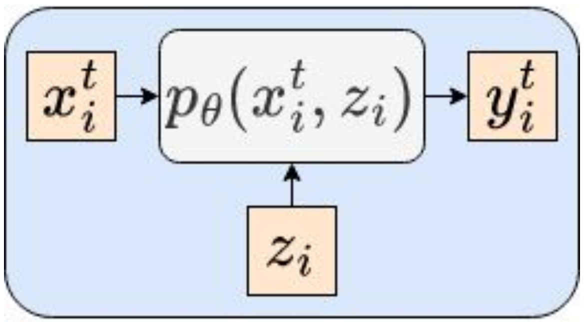
- Advances in ML and high-performance computing fed by big data have revolutionized all aspects of our lives.
- Big data and ML are Increasingly being considered as an alternative to the traditional scientific discovery paradigm.





Richards, Catherine E., et al. "Rewards, risks and responsible deployment of artificial intelligence in water systems." Nature Water 1.5 (2023): 422-432.

Abstract Representation of a Physical System



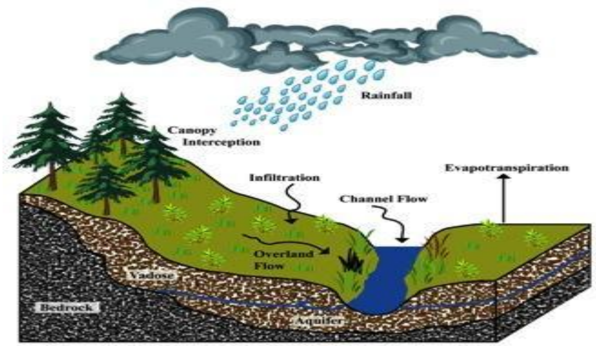
***x**^t* : dynamic inputs at time t

z : set of static characteristics (latent parameters of the system)

***y**^t* : response at time t

***x**^t* and ***y**^t* can have spatial dimensions

Modeling stream flow in a watershed

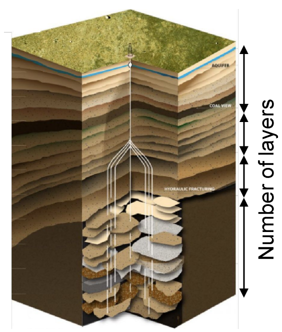


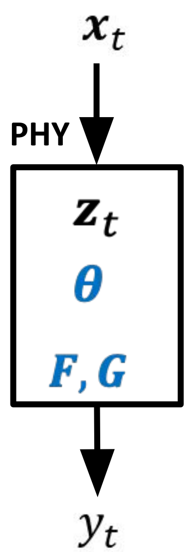
SWAT: physics based model used by hydrological community

Problem Formulation:

- Given input driver x^t and system characteristics z learn to predict response y^t

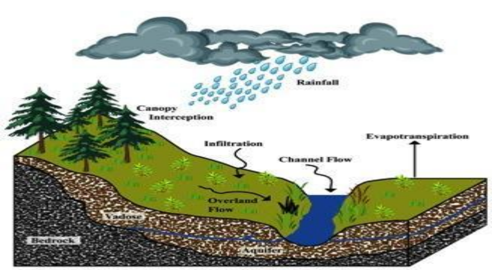
Limitations of Process-based Models





- **Incomplete or missing physics (F,G)**
  - –Physics-based models often use approximate forms to meet "scale-accuracy" trade-off
  - –Results in *inherent model bias*
- **Unknown parameters (θ) need to be "calibrated"**
  - –*Computationally Expensive*
  - –*Easy to overfit*: large number of parameter choices, small number of samples
- **Inefficient use of observations**
  - –Calibration of a Process-based models on a highly observed entity *does not help improve performance* on less observed or un-observed entities
- **ML has the potential to**
  - –provide high predictive power with **sparse observation**
  - –generalize to **unseen scenarios**
  - –produce **physically consistent** results
  - –leverage information from highly observed entities to provide high quality prediction in **unobserved** and **sparsely observed** entities

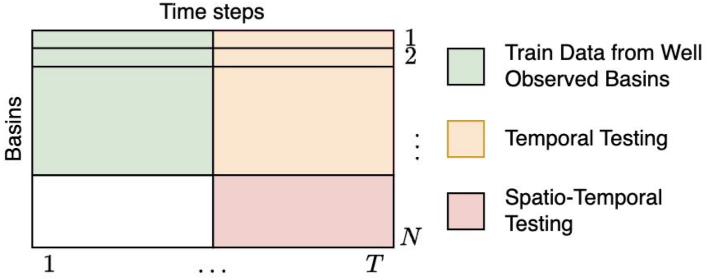
Modeling stream flow in a watershed



SWAT: physics based model used by hydrological community

BUILDING ML MODELS: LOCAL MODELS

- **Temporal Testing:** The training and testing data are from the same entities but time period of training and testing are different.

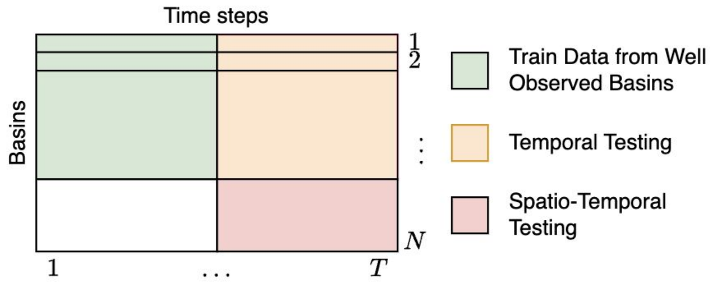


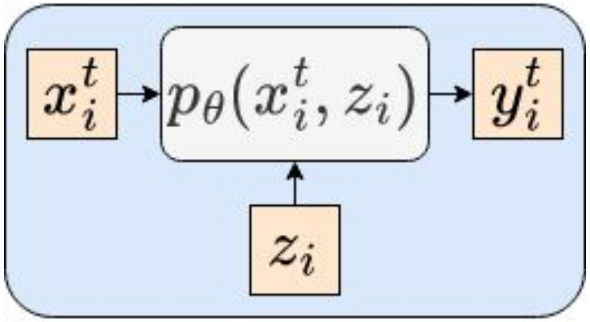
- Build **Local model** for each entity (rows in the top right image)
  - Needs lot of labelled data for each entity

Global ML model with task characteristics outperformed individually calibrated physical models

BUILDING ML MODELS: GLOBAL MODELS

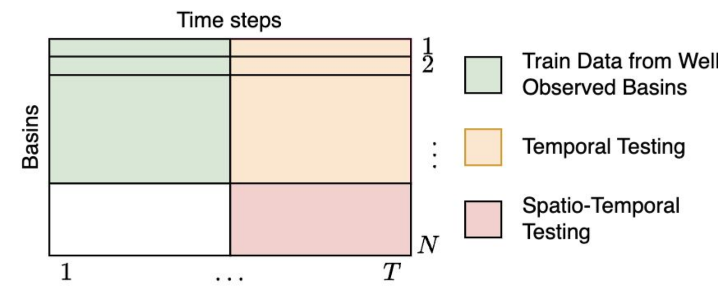
- Build a **global model** using all of the entities together
  - ML can leverage data from diverse cross section of basins
  - Static characteristics (z) introduce heterogeneity in driver- response relationship.
  - Trivial merging would lead to sub-optimal personalized predictions **(Global Model Without static)**



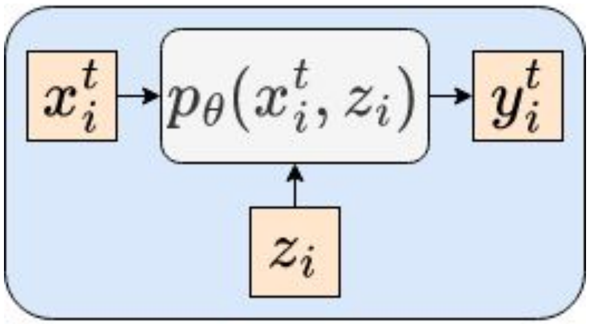


GLOBAL MODEL WITH STATIC CHARACTERISTICS

- Build a **global model** using all of the entities together
  - ML can leverage data from diverse cross section of basins
  - Static characteristics (z) introduce heterogeneity in driver- response relationship.
  - Incorporate static characteristics during training **(Global Model with static characteristics)**
  - CT-LSTM is on way of Incorporating static characteristics

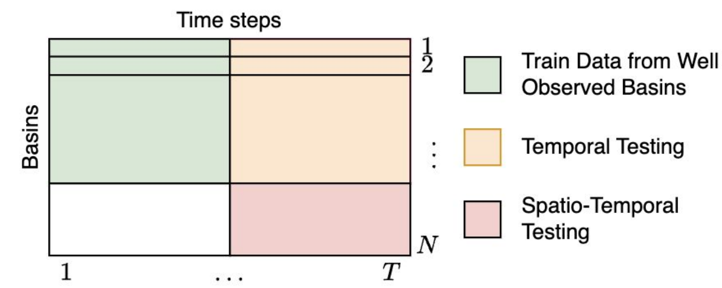


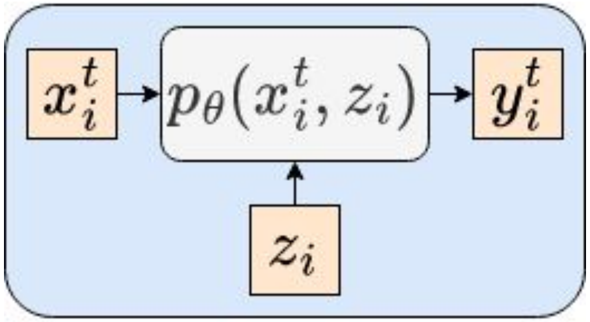
Global ML model with task characteristics outperformed individually calibrated physical models



GLOBAL MODEL WITH STATIC CHARACTERISTICS

- **Spatio-Temporal Testing:** The training and testing data are from different entities and different time period.
- **No observation is available [Ungauged Prediction]**
  - Build a global model using basin characteristics and training data from well observed basins
  - Use static characteristics to transfer knowledge from a learned global model to a new basin in **zero-shot** fashion



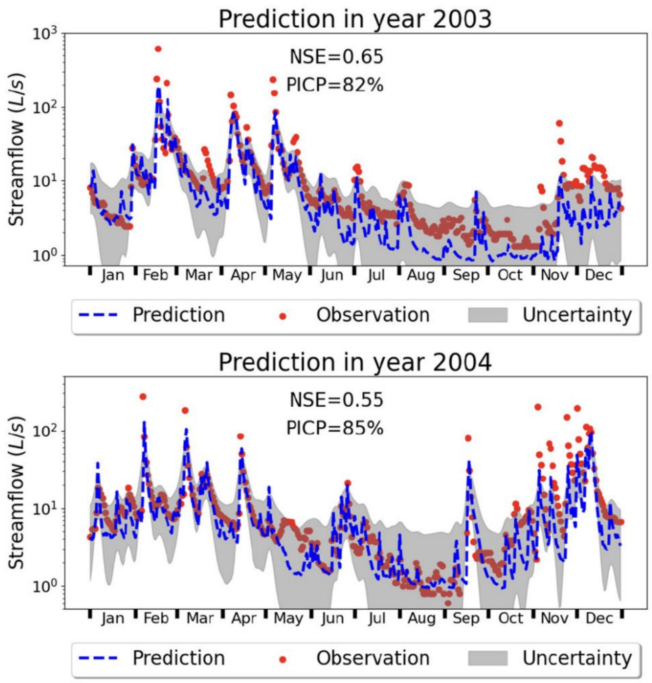


UNCERTAINTY ANALYSIS

- **Importance:**
  - Improving the reliability and robustness of models

- **Sources of uncertainty in hydrology**
  - Input data uncertainty (e.g., precipitation, temperature, land use)
  - Parameter uncertainty (e.g., soil properties, vegetation characteristics)
  - Model structure uncertainty (e.g., simplifications, assumptions)
  - Natural variability and climate change

- **Uncertainty quantification methods**
  - Prediction Interval methods [shown in notebook]
  - Monte Carlo dropout
  - Bayesian inference
  - Generalized likelihood uncertainty estimation (GLUE)
  - Markov Chain Monte Carlo (MCMC) methods



Liu, Siyan, et al. "Uncertainty quantification of machine learning models to improve streamflow prediction under changing climate and environmental conditions." *Frontiers in Water* 5 (2023)

Concluding Remarks

- Big data and machine learning offers great opportunity to increase our understanding of the Earth's climate and environment.

- In the presentation, we provide a high-level picture of the topic. For a more detailed understanding, there is a more comprehensive tutorial available at google colab, which covers the basics. Do check it out for a deeper dive into the subject matter.

- Methods discussed above have wide applicability across diverse domains:
  - Agriculture: Optimizing irrigation systems and managing soil moisture
  - Urban planning: Designing effective drainage and stormwater management infrastructure
  - Climate science: Modeling the water cycle and its interactions with the atmosphere and land surface
  - Energy: Assessing water resources for hydroelectric power generation
  - Disaster management: Predicting and mitigating floods, droughts, and other water-related hazards