Machine Learning Protocols for Water and Environmental Modeling

Machine Learning in Water and Environmental Modeling Workshop
May 2, 2025
Module #1

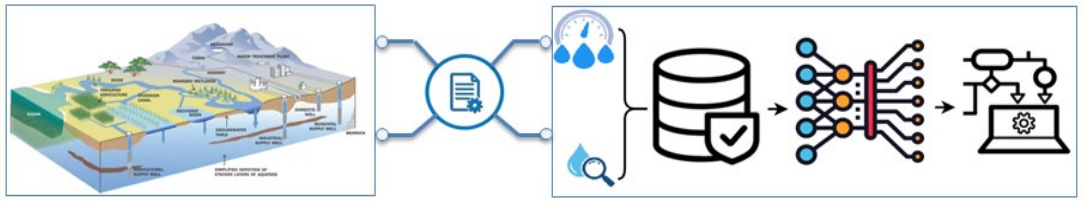
Kevin He
DWR Modeling Support Office

CALIFORNIA DEPARTMENT OF
WATER RESOURCES

Technical Report

Machine Learning Protocols for Water and Environmental Modeling

October 2024



Peer-Reviewed Article

hydrology

IMPACT FACTOR 3.1

CITESCORE 4.9

Review https://doi.org/10.3390/hydrology12030059

Protocols for Water and Environmental Modeling Using Machine Learning in California

Minxue He, Prabhjot Sandhu, Peyman Namadi, Erik Reyes, Kamyar Guivetchi and Francis Chung

Acknowledgements

DWR Modeling Support Office

Bay-Delta Modeling Forum (BDMF)
–Modeling Protocols Committee (2000)

California Water and Environmental Modeling Forum (CWEMF)
–Modeling Protocols Committee (2021)

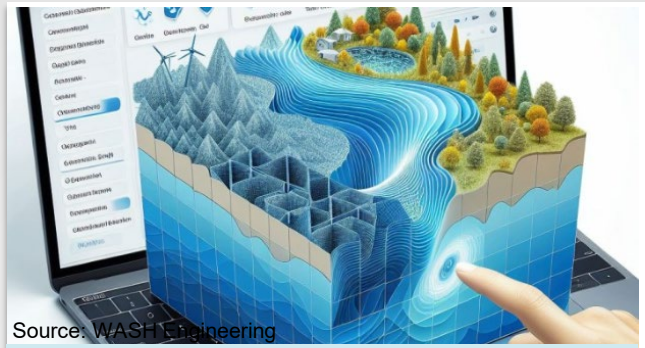
Machine Learning Community

Outline

1. Why machine learning (ML) protocols?

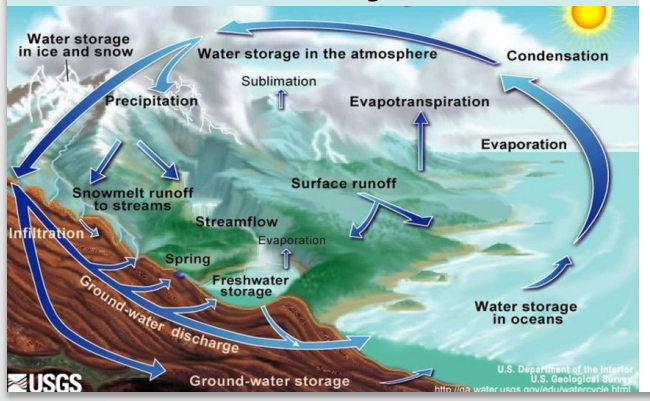
2. Why us?

3. What protocols?


Source: WASH Engineering

Machine Learning Models

Water Cycle



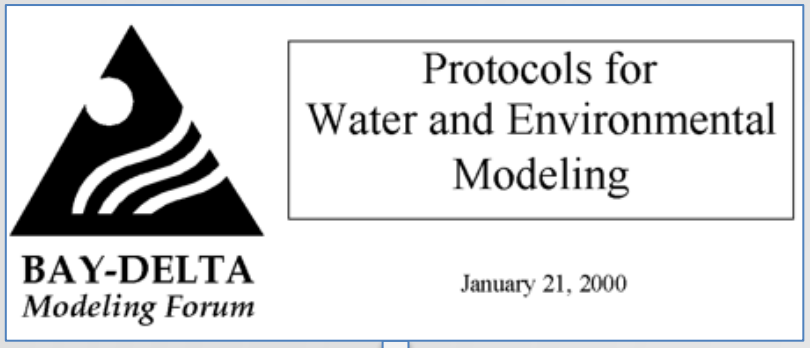
Major Water Projects

Source: California Water Plan

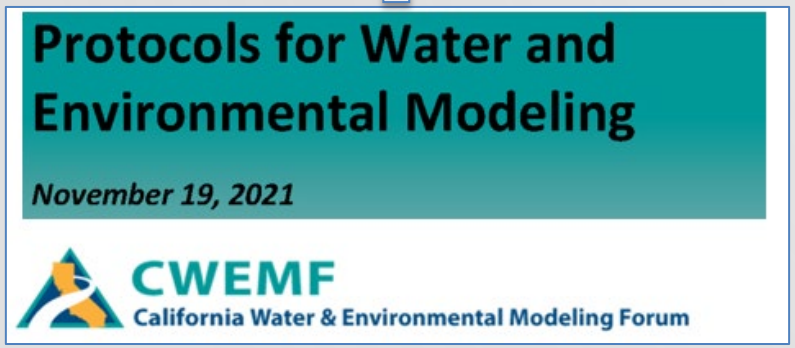
CALIFORNIA DEPARTMENT OF WATER RESOURCES

Why do we need ML modeling protocols?

➢ Existing Protocols



Model Credibility



➢ Conventional Bay-Delta models vs. ML models

	SCHISM* 3-D Model	DSM2* 1-D Model	ML
Approach	Physically-based	Simplified physics	Data-driven
Structure	Rigidly structured	Structured	Flexible
Runtime	Days	Hours	Seconds
Interpretability	"White box"	"Grey box"	"Black box"

Credible/ Trustworthy ML Models ⟵ **Protocols**

***SCHISM**: Semi-implicit Cross-scale Hydroscience Integrated System Model **DSM2**: Delta Simulation Model II

ML experience of the Modeling Support Office

Timeline

DWRSIM–MDO*	CalSim II–G Model*	CalSim II–DSM2*	CalSim 3–DSM2*	
1995	**2001/2002**	**2007**	**2020**	**Post-2020**

* **DWRSIM, CalSim II, CalSim 3**: 1st, 2nd, and current generation of DWR's planning model of the SWP and CVP

* **MDO**: Minimum Delta Outflow **G Model**: Salinity Gradient Model **DSM2**: Delta Simulation Model II

2020 - Present

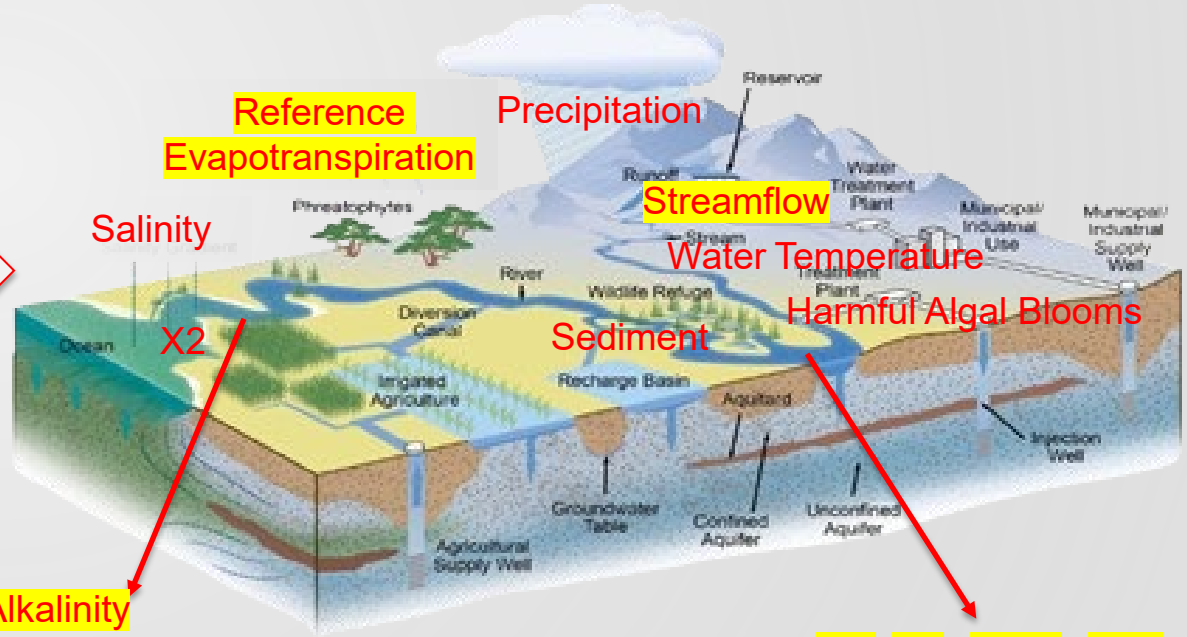
20+ ML Models

20 Variables

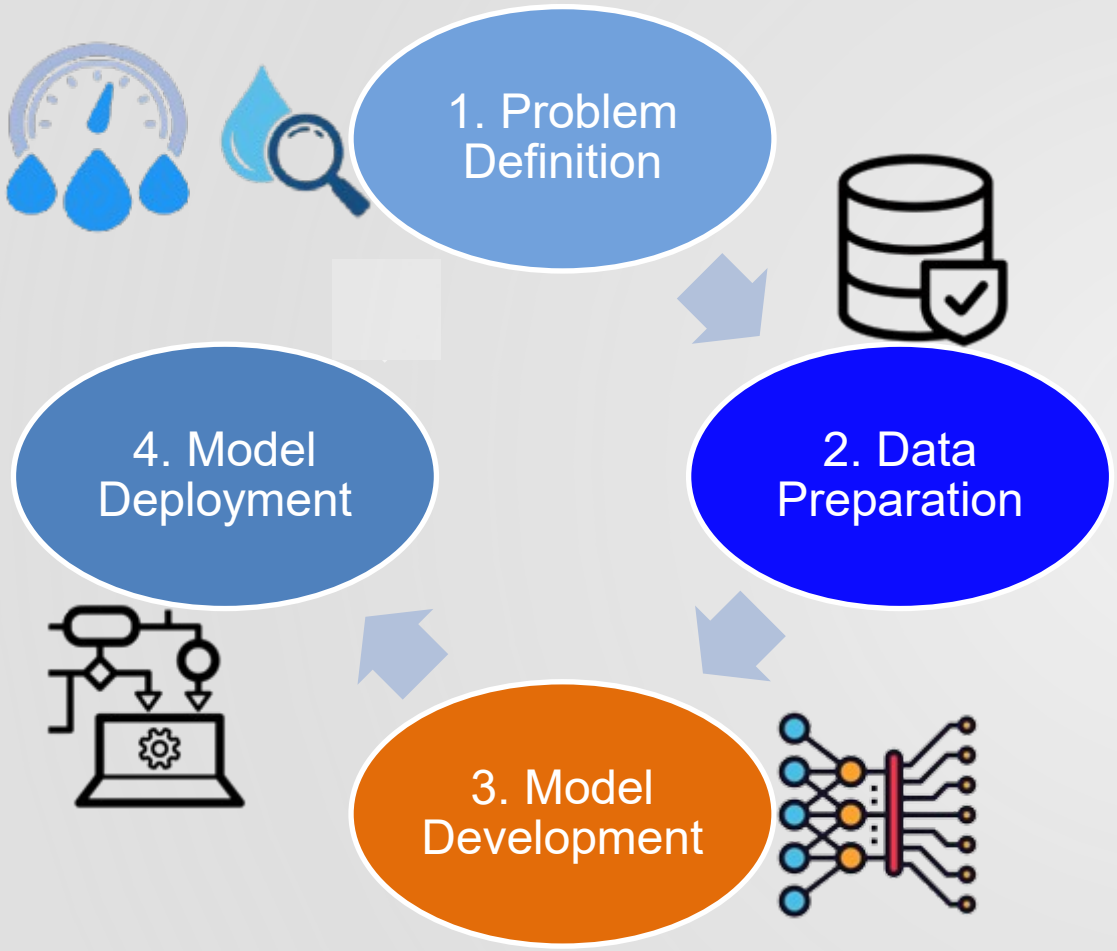
12 Articles

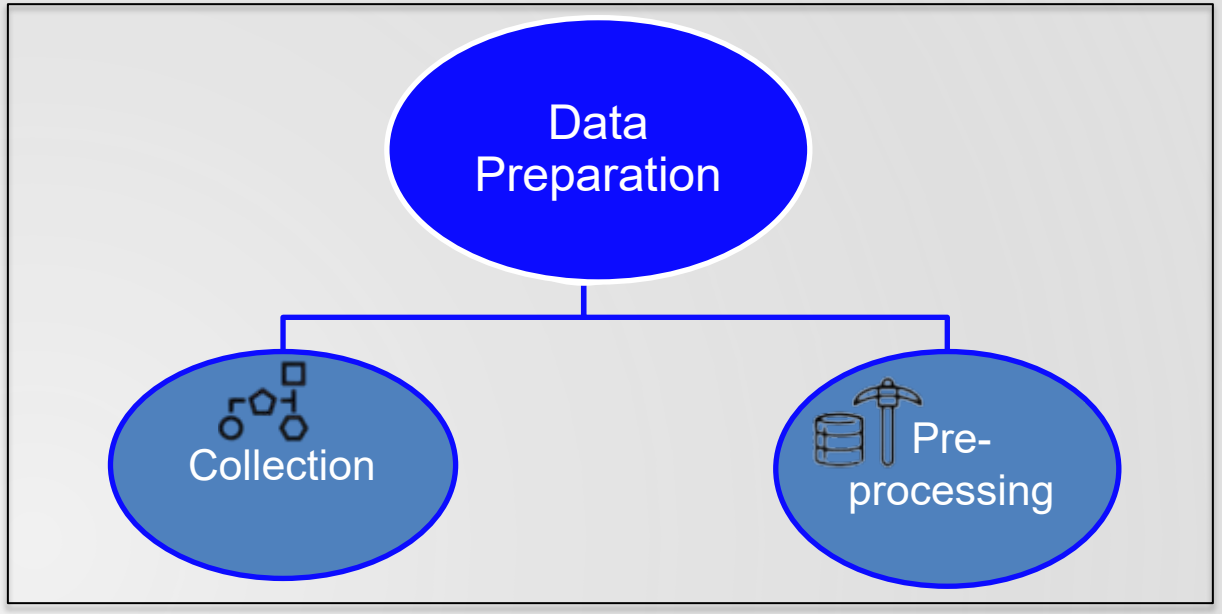
6 Dashboards

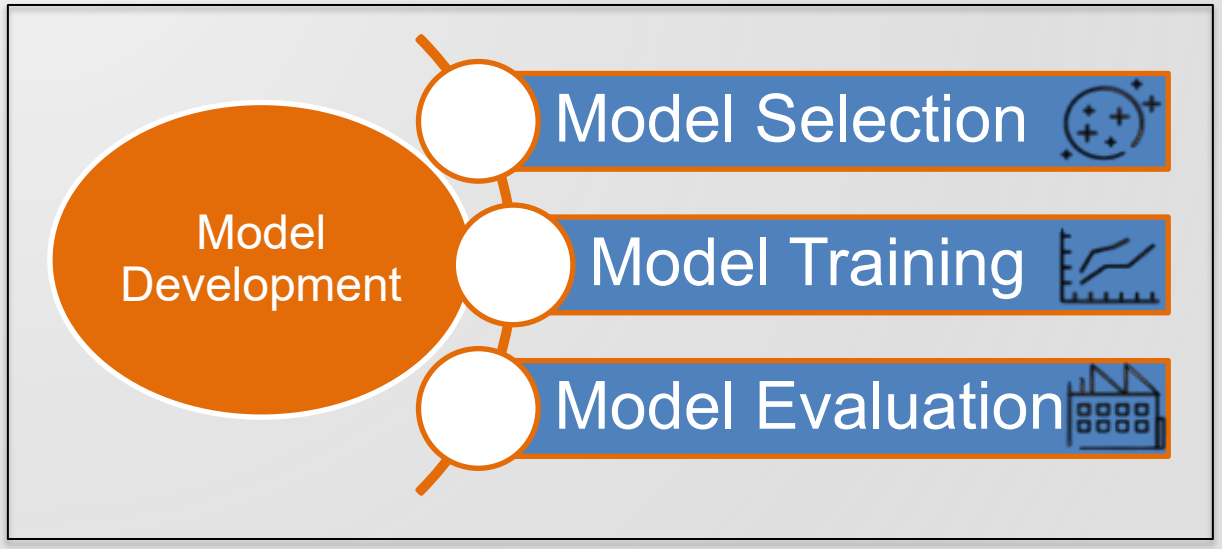
1 Workshop



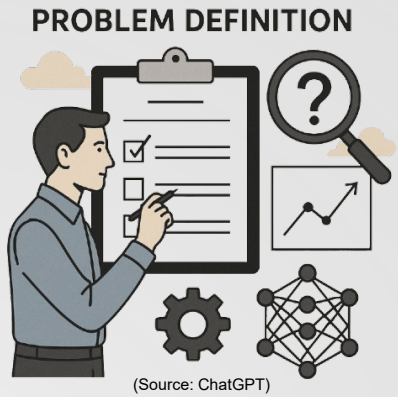
Life-cycle of a typical ML model







Phase 1: Problem Definition



(Source: ChatGPT)

Problem Statement

Modeling Required ?

YES

ML Modeling Required ?

YES

Data Available ?

YES

Phase 2

NO

STOP

NO

STOP

NO

STOP

Phase 2: Data Preparation

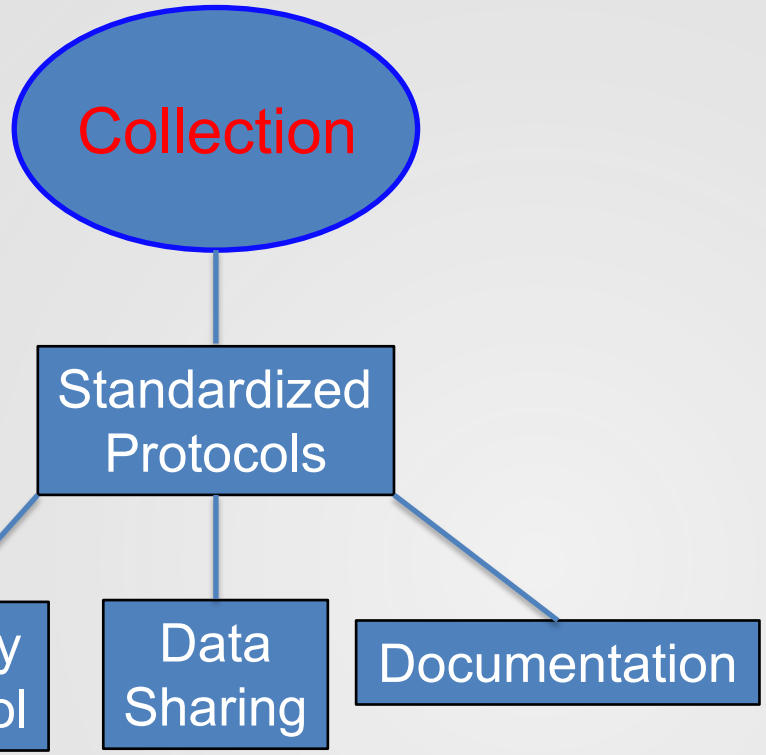
Focus

- Numerical
- Categorical

  Water quantity; water quality; fish abundance and distribution, ……

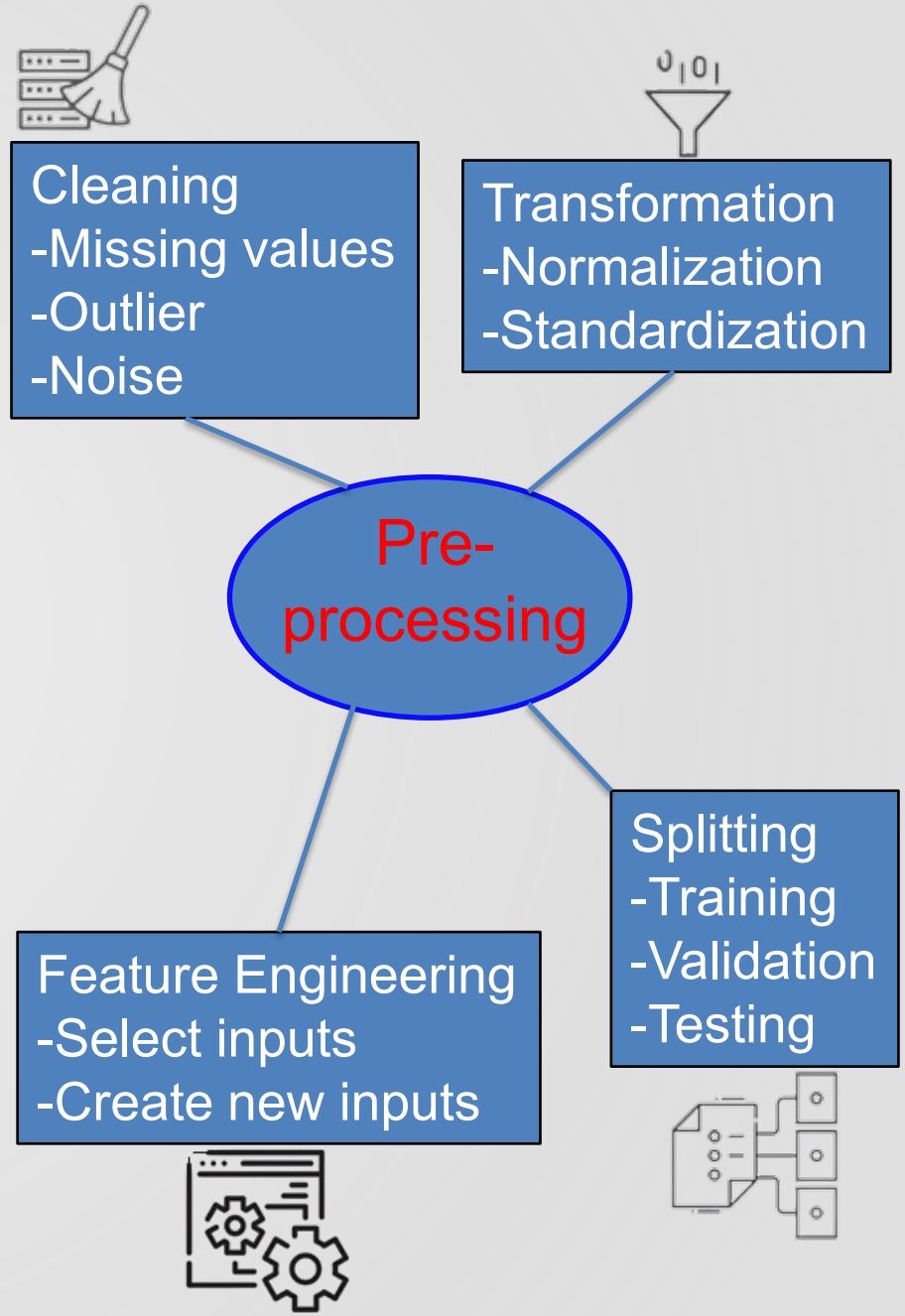
Not Included

- Text
- Image
- Audio
- Video



Example: **Assembly Bill 1755 (AB1755) Open and Transparent Water Data Act**

https://data.cnra.ca.gov/



Phase 3: Model Development

Model Selection

- Data Nature
- Problem Type
- Resources

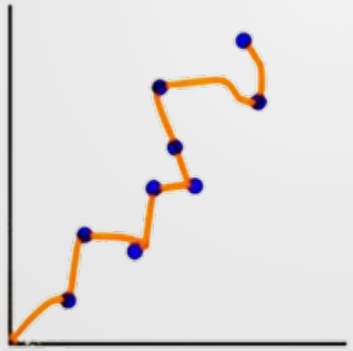
- Start simple

- Start with several models

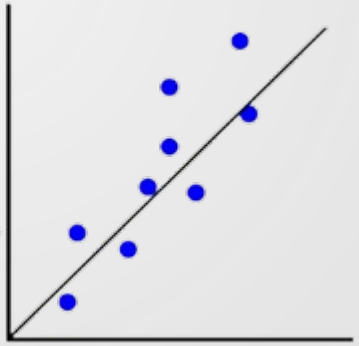
Model Training

- Use ML Frameworks
- Prevent overfitting

Overfitting

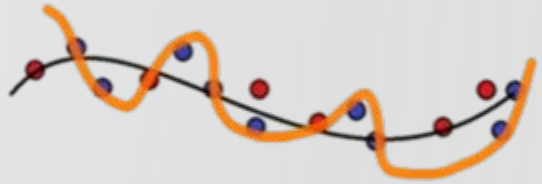


No Overfitting



Model Evaluation

- Evaluation metrics
- Visual inspection
- Generalization



Training Data New Data

Orange Model < Black Model

Domain Knowledge

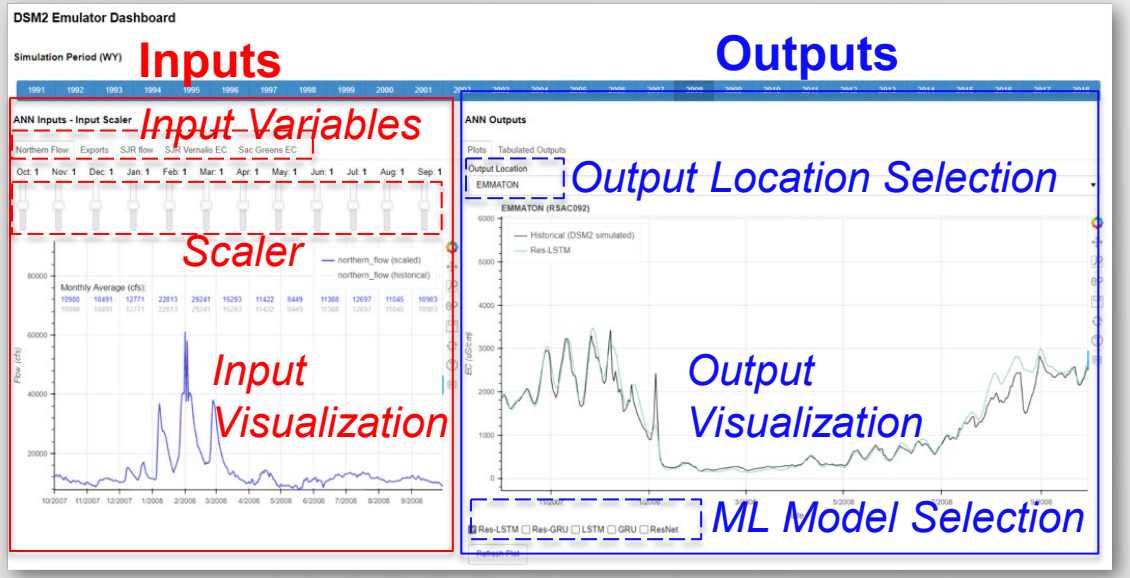
Phase 4: Model Deployment

1.Product

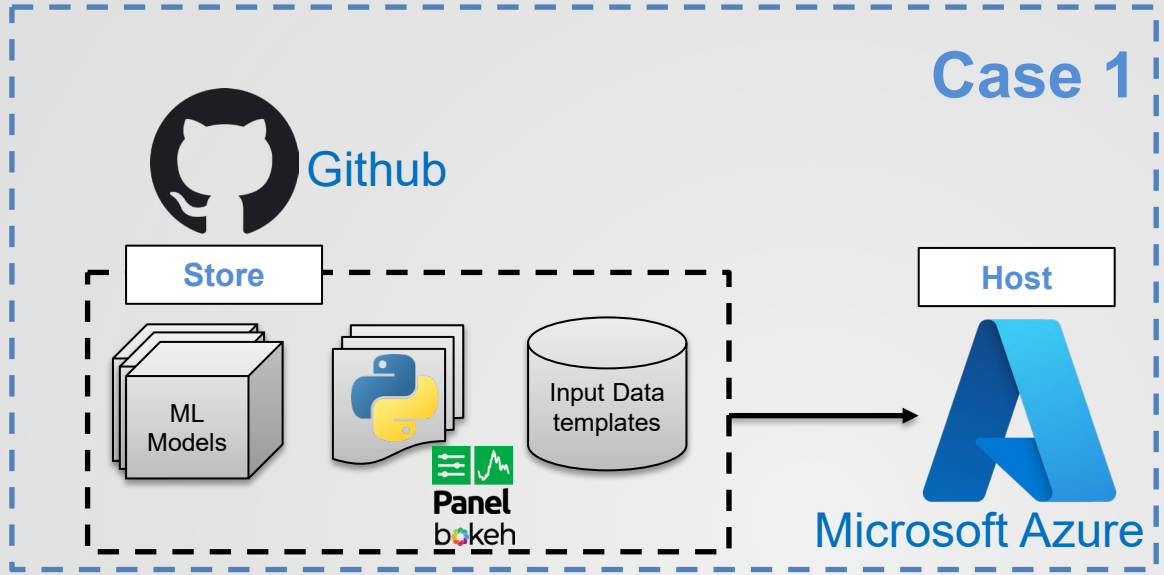
- Users
- Accessibility

Example: Salinity Dashboard (**Case 1 product**)

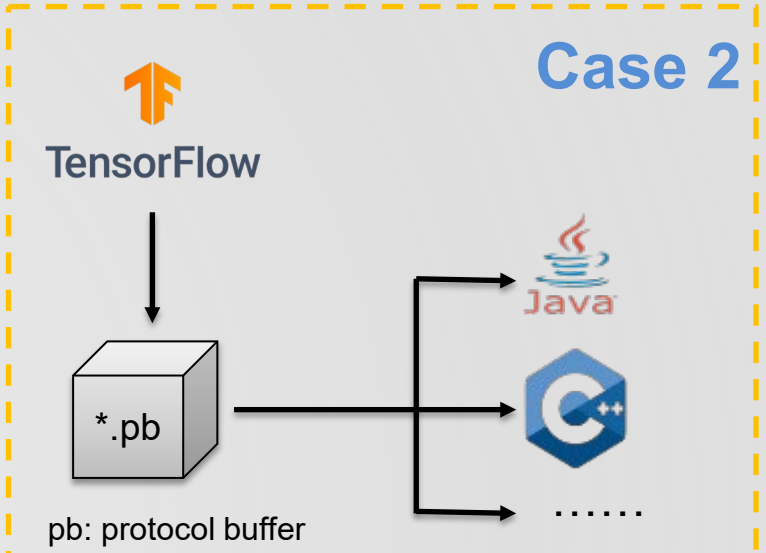
https://dwrbdodash.azurewebsites.net/dashboard



Case 1



Case 2



Case 2 product: .pb file

2.Documentation

- Transparency
- Peer-review

Docs

Code & Data

3.Training

- Workshop
- Hands-on

Recording

4.Maintain

- Monitor
- Improve

Take-Home Messages

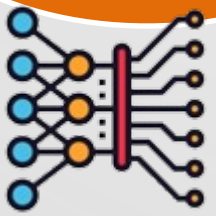
ML→ powerful, fast, but not a replacement

1. Problem Definition

2. Data Preparation

- Collection — Data is the new oil
- Pre-processing — Don't throw raw data into ML

3. Model Development



- Selection — Don't use a sledgehammer to crack a nut
- Training — Overfitting is your enemy
- Evaluation — Generalization is key

Don't let it rust

4. Model Deployment

Technical Report

Peer-Reviewed Journal Article

Questions?

CALIFORNIA DEPARTMENT OF
WATER RESOURCES

Contact: Kevin He Kevin.He@water.ca.gov