Modeling Delta Salinity Constituents Using Machine Learning (ML)

Machine Learning Brown Bag
April 7, 2023

Peyman Namadi
DWR Modeling Support Office

CALIFORNIA DEPARTMENT OF
WATER RESOURCES

Background

- **Previous Work: Delta Salinity (EC) – Ion Constituent Conversion**

1986

Author: Kamyar Guivetchi, DWR
Variables: Salinity (EC), Total Dissolved Solids (TDS), Chloride (Cl)
Data Period: 1968-1981
Study Area: Delta wide
Model: Linear equations (water year type and site specific)

2002

Author: Bob Suits, DWR
Variables: EC & 9 constituents
Data Period: 1971-1999
Study Area: Delta Wide
Model: Quadratic equations (region specific)

2015

Author: Richard Denton, Consultant
Variables: EC & 10 constituents
Data Period: All Available
Study Area: Delta wide
Model: Quadratic equations (water year type, site/region, season specific)

2021

Author: NCRO WQES, DWR
Variables: EC & 12 constituents
Data Period: 2018-2020
Study Area: 7 South Delta locations
Model: Quadratic equations (performance not ideal for nitrate, potassium, and boron)

2022

Authors: Paul Hutton et al., Tetra Tech
Variables: EC & 9 constituents
Data Period: Up to 2019 (not including NCRO dataset)
Study Area: Delta wide
Model: Quadratic equations (water year type, site/region, month, X2 position (optional))

Current Work

$$Ion = A \times EC^2 + B \times EC + C$$

CALIFORNIA DEPARTMENT OF WATER RESOURCES

Goals and Methodology

Explore alternative models (including ML) that may outperform the simple regression method in mapping EC to ion constituents in the Delta.

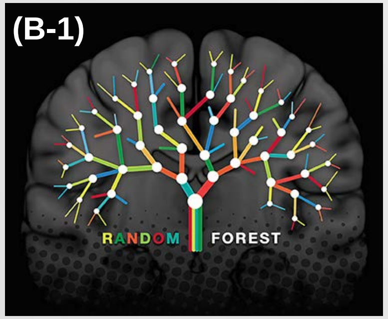
A. Regression Trees (RT)
B. Ensemble method:
  B-1. Random Forest (RF)
  B-2. Gradient Boosting (GB)
C. Artificial Neural Networks (ANN)

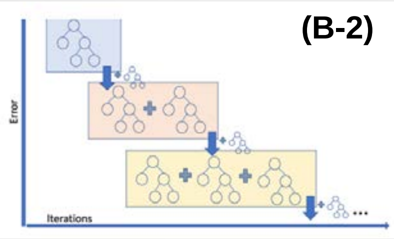
❖ Statistical metrics: R-squared and Mean Absolute Error (MAE)

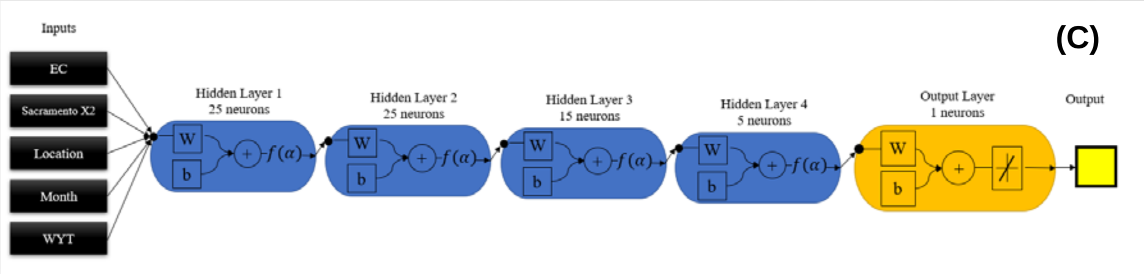
$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \hat{y}_i)^2 + \sum(\hat{y}_i - \bar{y})^2}$$ (Perfect value = 1)

$$MAE = \frac{\sum|\hat{y}_i - y_i|}{n}$$ (Perfect value = 0)

y_i = observed values
$\hat{y}_i$ = predicted values
$\bar{y}$ = mean of observed values
$\hat{\bar{y}}$ = mean of predicted values







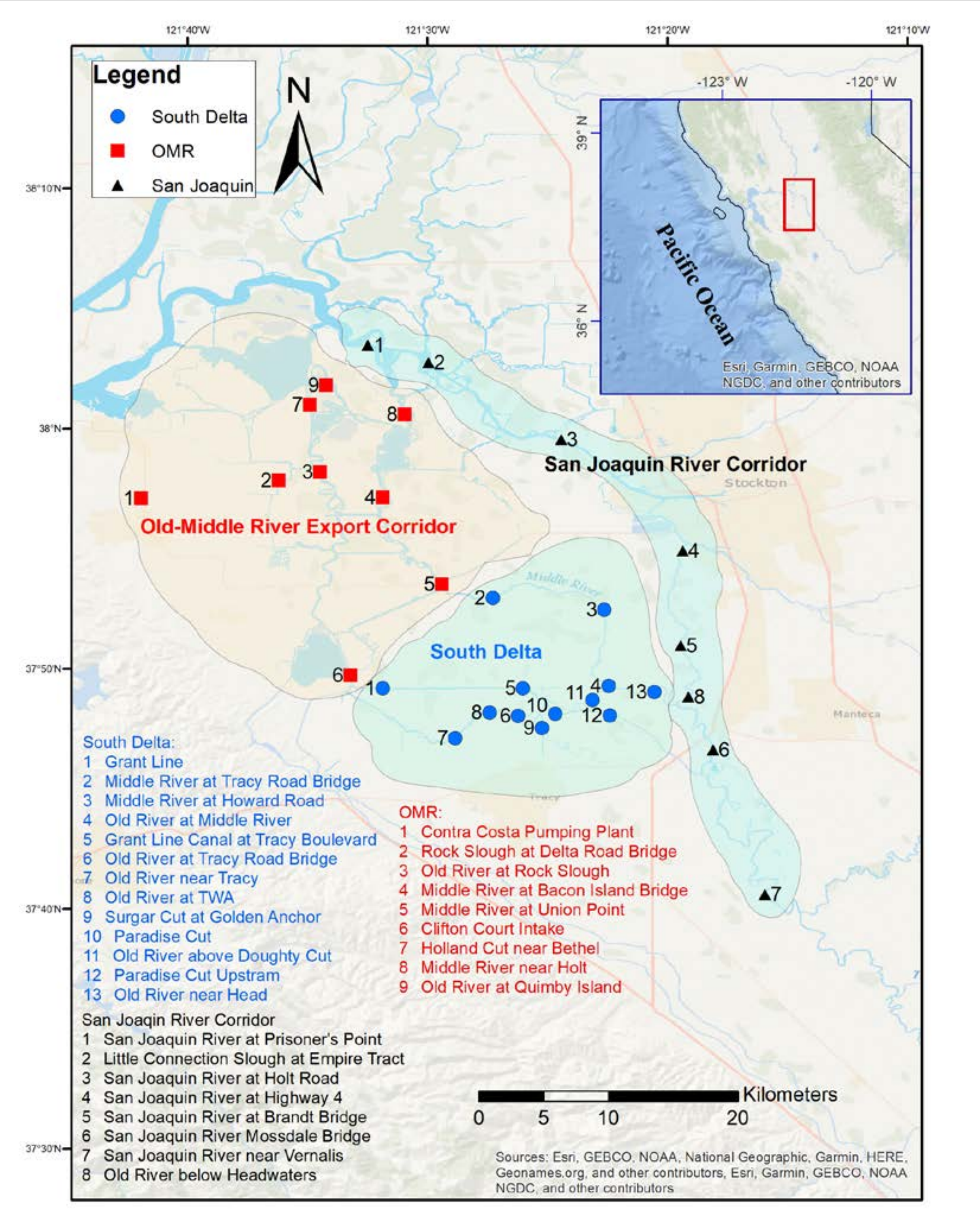
Study Dataset and Area

- Three sources of grab samples:
  1. Hutton et al. (2022): 1959 to 2018 at 19 stations.
  2. Department of Water Resources: 2018 to 2020 at 7 stations.
  3. California Data Exchange Center (CDEC): 2018 to 2022 at 13 stations

- Interior Delta Region further classified into three subregions:
  1. Old-Middle River Export Corridor (OMR)
  2. San Joaquin River Corridor
  3. South Delta

1- EC
2- Sacramento X2
3- Month
4- Water Year Type
5- Location

1- Total Dissolved Solids
2- Bromide
3- Chloride
4- Sulfate
5- Alkalinity
6- Sodium
7- Calcium
8- Magnesium
9- Potassium



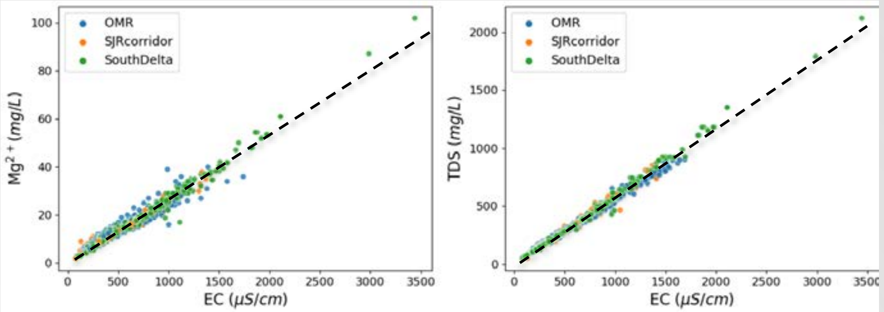
CALIFORNIA DEPARTMENT OF WATER RESOURCES

- Scatterplot matrix showing the relationship between Electrical Conductance and other water quality variables.
- Duration: 1959-2022

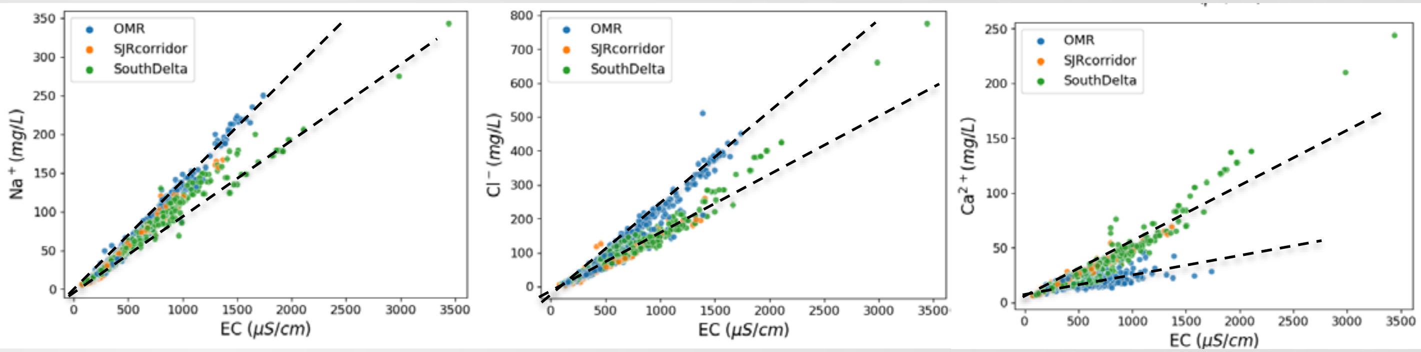
Observations:

- Group 1: TDS and Mg (linear)
- Group 2: Cl, Ca, and Na (Dual linear, Seawater and agriculture drainage)
- Group 3: Alkalinity, K, SO4, and Br (Non-Linear)

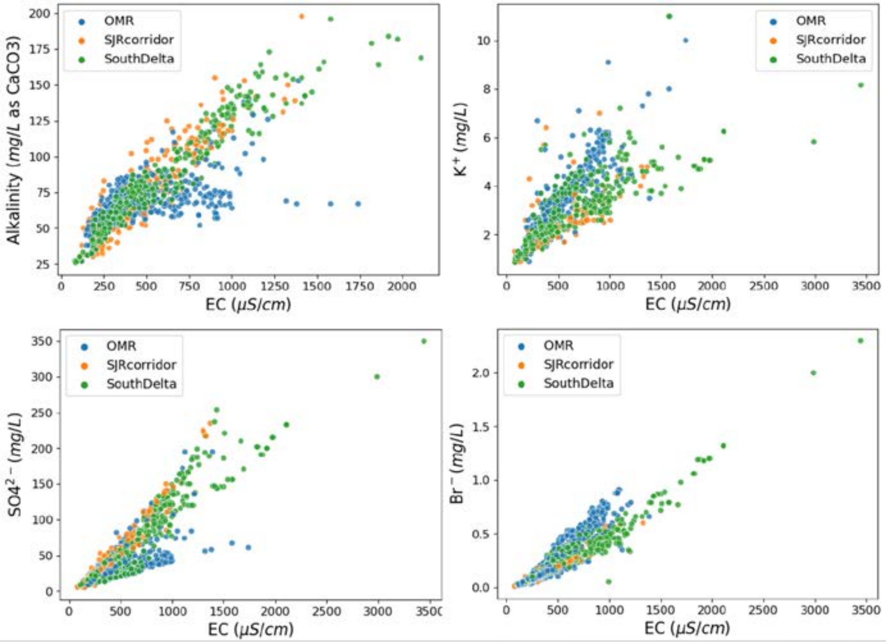
Group 1



Group 2



Group 3



CALIFORNIA DEPARTMENT OF WATER RESOURCES

Tetra Tech Method

Boundary Regions
- Seawater Boundary
- San Joaquin River Boundary
- Blended Boundary

❑ The Seawater Boundary

1. “Low”: 100 ≤ [EC] < 250 µS/cm
2. “Low-Medium” 250 ≤ [EC]
   - When EC < 105 µS/cm, set Br- to 0.03 mg/L

❑ The Blended Boundary

150 ≤ [EC] < 1300 µS/cm

❑ The San Joaquin River

100 ≤ [EC] < 1600 µS/cm

- Develop two unique regression equations.

$$Ion = K_1 + K_2 \times EC^{0.5} + K_3 \times EC + K_4 \times EC^{1.5} + K_5 \times EC^2 + K_6 \times EC^{2.5}$$
$$Ion = A \times EC^2 + B \times EC + C$$

X2 position is known and is ≥ 81 km

WATER YEAR TYPE AND SEASON MATRIX #2

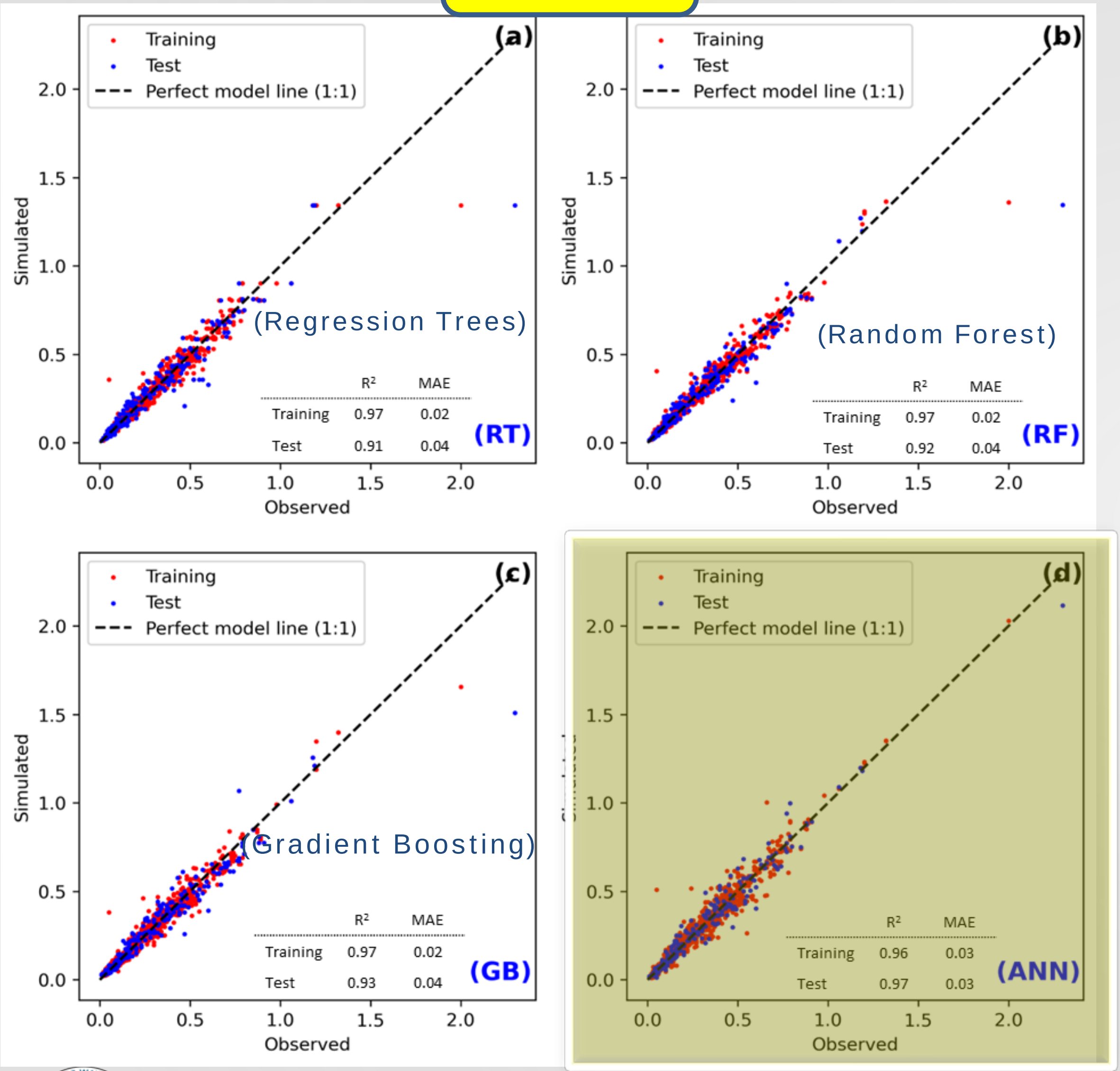
	J	F	M	A	M	J	J	A	S	O	N	D
Old-Middle River Export Corridor Subregion												
W	SEA	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SEA	SEA	SEA	SEA
AN	SEA	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SEA	SEA	SEA	SEA
BN	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA
D	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA
C	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA	SEA
San Joaquin River Corridor Subregion												
W	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR
AN	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR
BN	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR
D	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR
C	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR
South Delta Subregion												
W	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR
AN	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR	SJR
BN	IND	IND	SJR	SJR	SJR	SJR	SJR	IND	IND	IND	IND	IND
D	IND	IND	SJR	SJR	SJR	SJR	SJR	IND	IND	IND	IND	IND
C	IND	IND	SJR	SJR	SJR	IND	IND	IND	IND	IND	IND	IND

Group	Ion	Sample size	Data range	SD	R^2	MAE
Group 1	TDS	1466	49-2120	204	0.99	12.7
	Mg^{2+}	1336	2-102	8.6	0.96	1.24
Group 2	Na^+	1575	6-343	44	0.94	4.77
	Ca^{2+}	1335	5.8-244	18	0.87	3.31
	Cl-	1972	4-775	77	0.92	10.26
Group 3	$SO4^{2-}$	1066	5-350	46.5	0.52	14.61
	Br^-	1239	0.01-2.3	0.22	0.9	0.04
	Alkalinity	1039	26-198	27.6	0.79	9.52
	K^+	1148	0.87-11	1.35	0.62	0.51

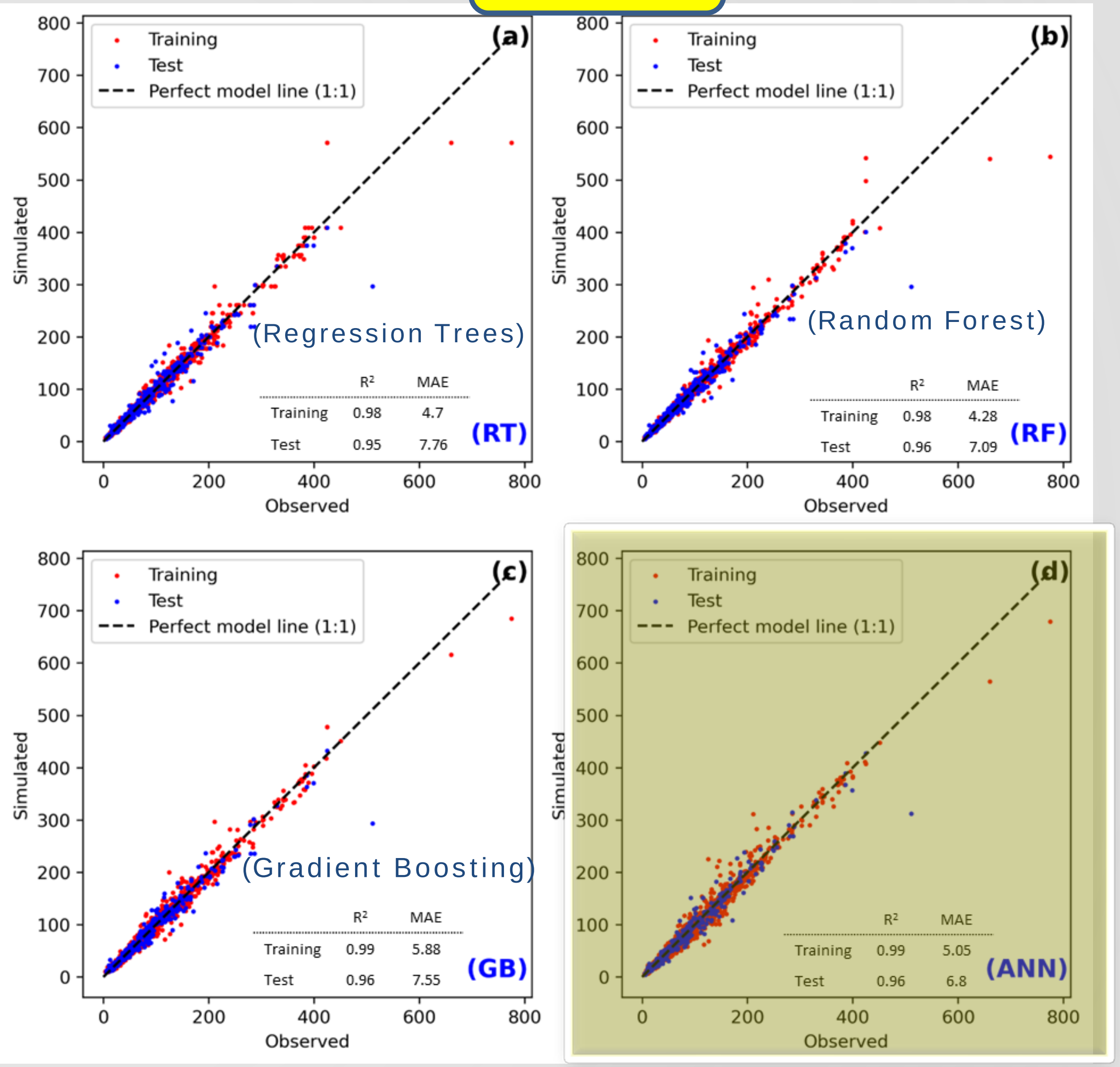
CALIFORNIA DEPARTMENT OF WATER RESOURCES

- Performance of four alternative models. Train for 80% and Test for 20% of samples.

Br



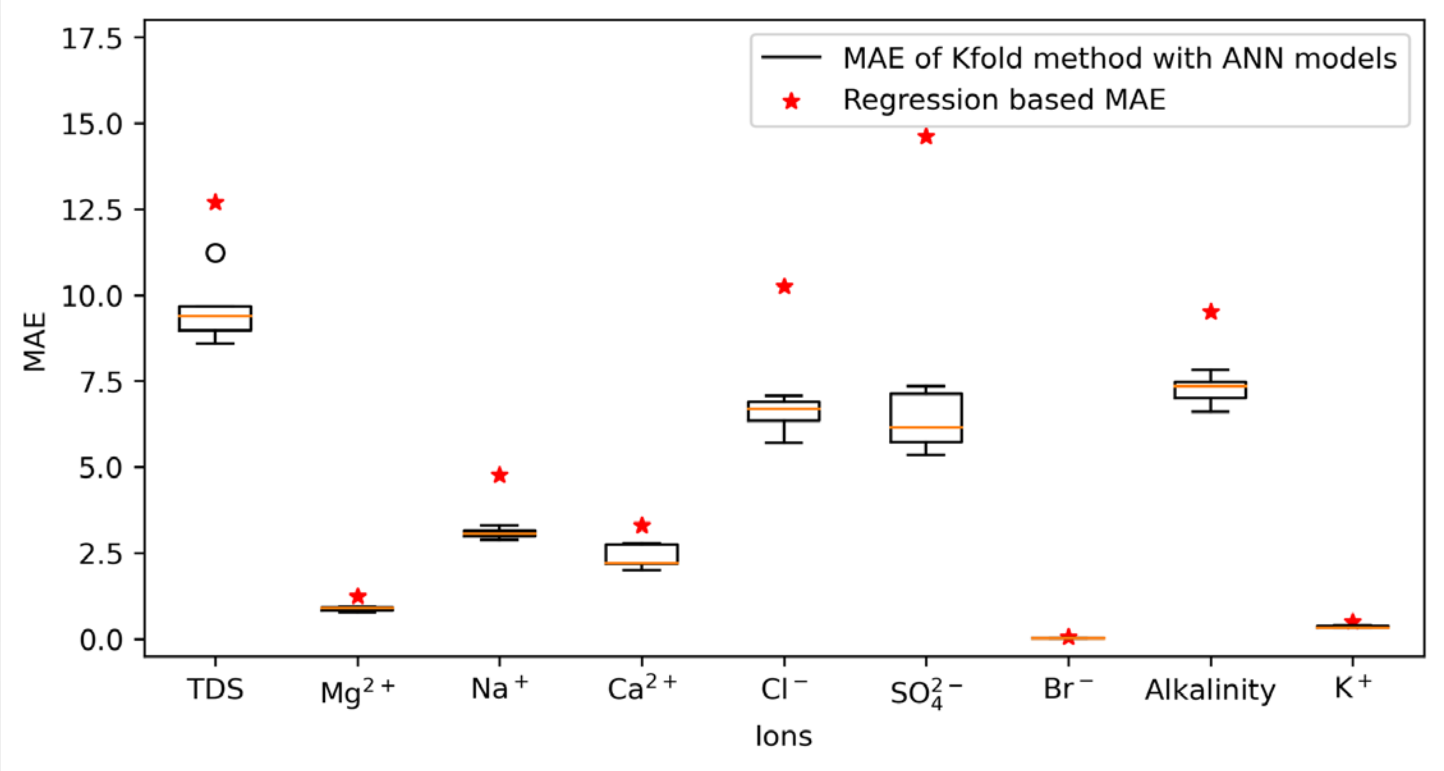
Cl

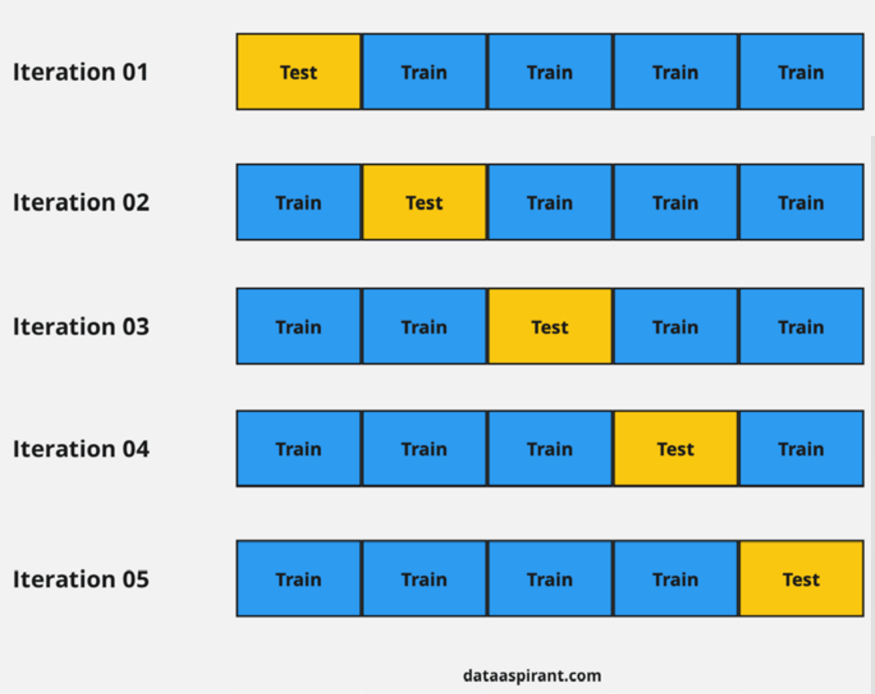


CALIFORNIA DEPARTMENT OF
WATER RESOURCES

K-fold Analysis

K-fold analysis is a technique commonly used in machine learning and statistics for model evaluation and selection.





CALIFORNIA DEPARTMENT OF
WATER RESOURCES

❑ Percentage improvement of K-fold ANN model vs. Regression based model

Observation:

- For the same dataset (60-year data), ANN outperforms Regression based model.

Dashboard

dwrdashion.azurewebsites.net

CALIFORNIA DEPARTMENT OF
WATER RESOURCES

Ion Simulator Dashboard

This dashboard allows you to simulate ion concentrations based on various input parameters. Use the sliders and dropdown menus to select the desired values for EC, Sacramento_X2, Ion, WYT, Location, and Month. Then click the 'Compute' button to generate a bar chart of the predicted ion concentrations.

Instructions:

1. Adjust the sliders and drop-down menus to select the desired input values.
2. Click the **Compute** button to run the simulation.
3. The bar chart will display the predicted ion concentrations for different machine learning models.

Notes:

- Electrical conductivity (EC) is measured in microsiemens per centimeter (µS/cm).
- Sacramento_X2 is the percentage of Sacramento River flow that is estimated to reach the Delta. The exact location of the Sacramento X2 point is determined by the California Department of Water Resources (DWR) based on the specific hydraulic conditions and water flows in the Sacramento River. The DWR uses a combination of hydrological models, flow measurements, and other data to determine the location of the Sacramento X2 point.
- The Water Year Type (WYT) is a classification of the water year based on its hydrological characteristics. Water Year Type that includes the following categories: 1- Wet (W), 2- Critical (C), 3- Dry (D), 4- Above-Normal (AN), 5- Below-Normal (BN)
- Location refers to monitoring regions that includes: 1- Old-Middle River (OMR), 2- San Joaquin River Corridor (SJRcorridor), and 3- South Delta (SouthDelta).
- Month refers to the month of the year.

EC: **200**

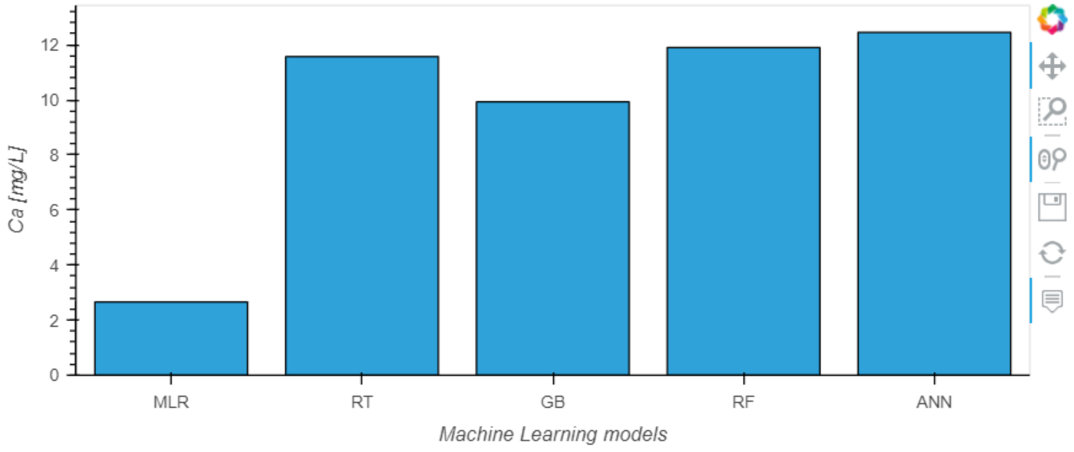
Sacramento_X2: **90**

Ion: Ca

WYT: AN

Location: OMR

Month: September



Compute

Summary and future work

Summary:

- Out of four alternative models investigated, the Artificial Neural Networks (ANN) model performs the best.
- The ANN model outperforms quadratic equations (Tetra Tech) in EC ~ constituent conversion, particularly for constituents with strong non-linear relationships with EC (e.g., group 3 constituents).
- The user-friendly dashboard is accessible for users with our without programming knowledge and facilitates ion level simulation in the Delta.

Future work:

- Model Enhancements
  - More inputs: volumetric fingerprint simulations, operations.
  - More training data: run hydrodynamic and water quality models under different management and (extreme) hydrology scenarios.
- Products
  - Interactive dashboard
  - Data and Source Code
  - Technical reports & journal articles

Acknowledgements

DWR

Kevin He, Nicky Sandhu Bob Suits, Kijin Nam, Hari Rajbhandari, Bradley Tom (Modeling Support Office)

Jared Frantzich, David Colvin (NCRO)

Shaun Philippart (DISE)

Francis Chung (Executive)

Tetra Tech:
Paul Hutton, Sujoy Roy

Palencia Consulting Engineers

Leslie Palencia

CALIFORNIA DEPARTMENT OF
WATER RESOURCES

Contact: kevin.he@water.ca.gov


QUESTIONS