TETRA TECH

A hybrid empirical-Bayesian artificial neural network model of salinity in the San Francisco Bay-Delta estuary

Christine Lew, Tetra Tech
christine.lew@tetratech.com

Presentation at the
19th Conference on Artificial Intelligence for Environmental Science
American Meteorological Society Annual Conference

January 14, 2020

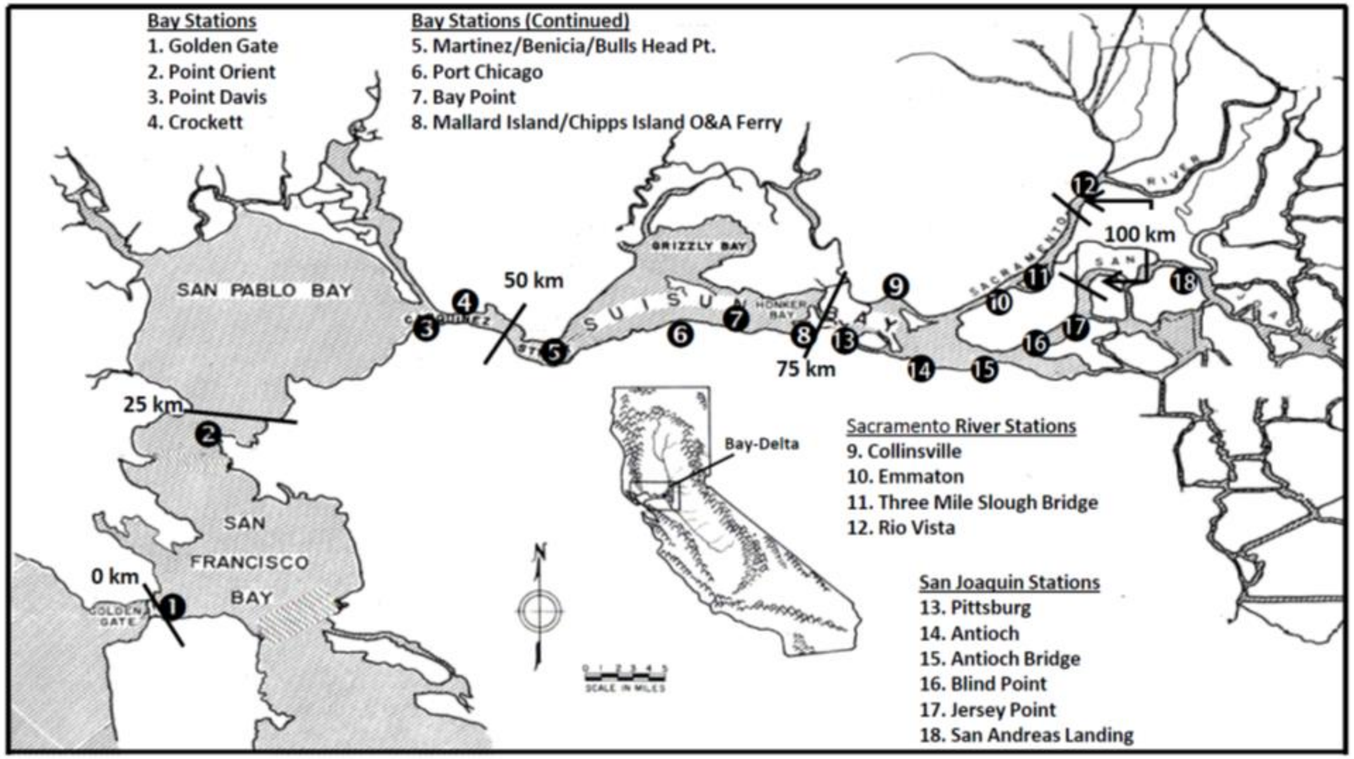
Presentation Overview

- Why develop an ANN model for salinity?
- Data-driven ANN models vs Hybrid ANN model
- Predictive and structural validation
- Input data and model selection, model training approach
- Model results
- Sea level rise case study

Why develop an ANN model for salinity?

- **Supports the need for making rapid salinity predictions across multiple locations for Delta outflow management**
  - Health of highly diverse ecosystem
  - Management of freshwater withdrawals for consumption, irrigation
- **ANNs are already in use within Delta models; can we improve and supplement the existing tools?**
- **From a process perspective, salinity in the western Delta is a complex function of current and antecedent flows, plus other variables; a key objective of the task was the exploration of these other variables (specifically tidal effects)**

Salinity in the western Delta and San Francisco Bay

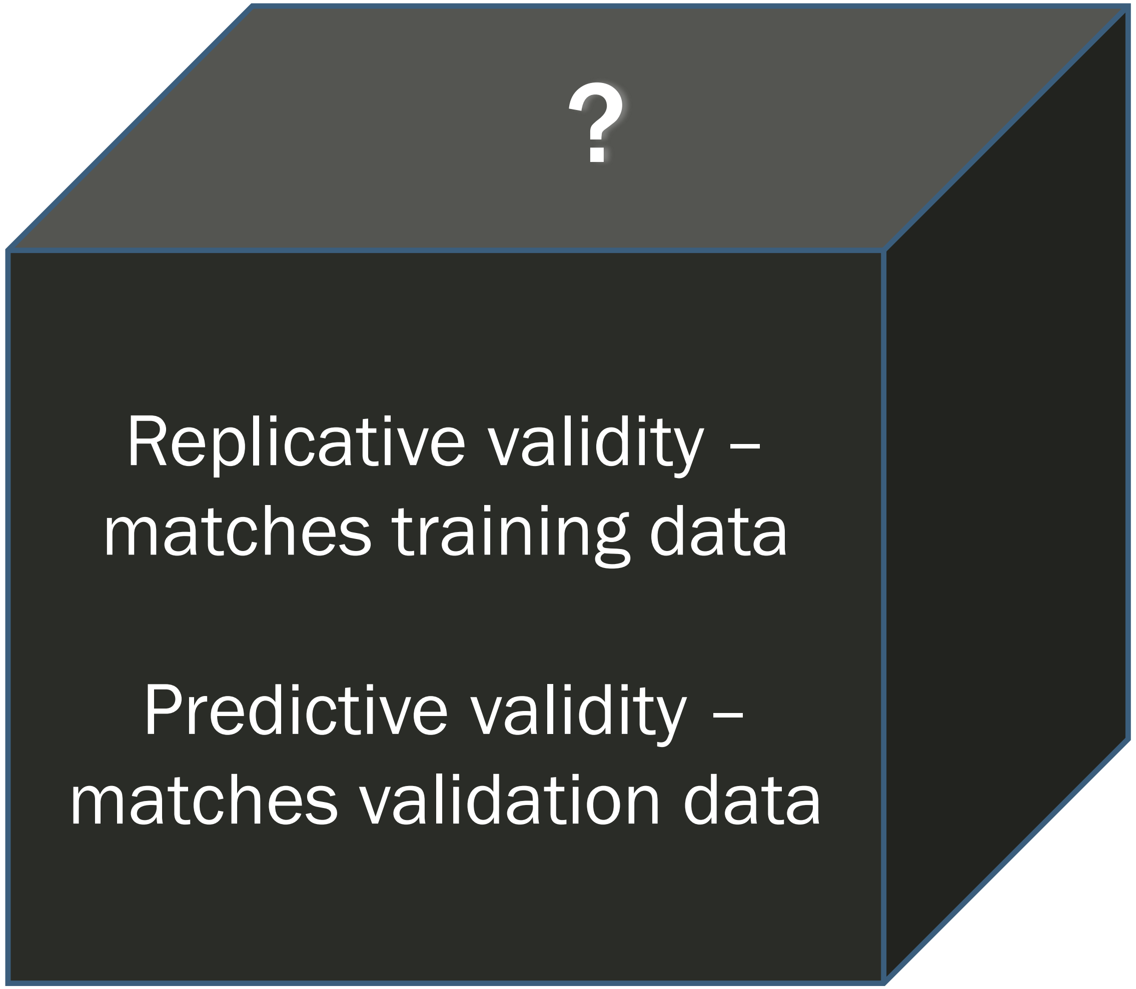


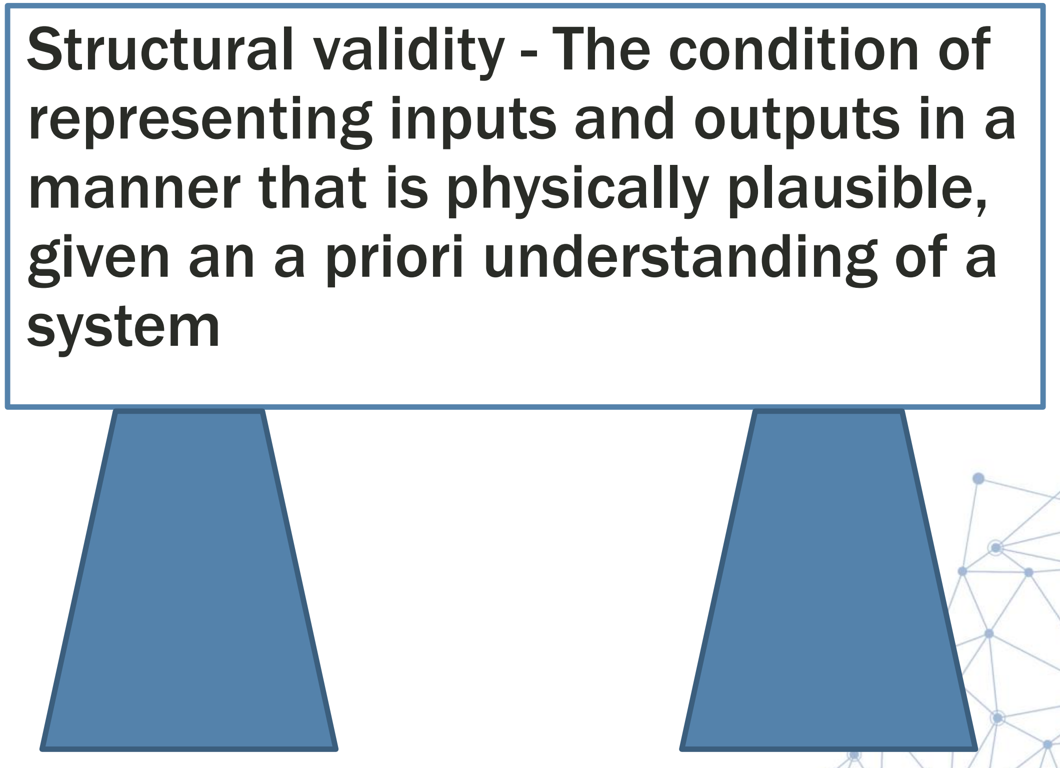
- **X2 position (marks the low salinity zone)**
  - Operationally - position of the 2 parts per thousand bottom salinity isohaline
  - Regulatory - interpolated as an equivalent surface salinity from fixed monitoring stations near the surface, reported as distance from the Golden Gate bridge in km

Overview of Approach

- Used feedforward ANNs, widely used in water resources applications
- Phase 1: **Data-driven ANNs**, uses only input and output data and no pre-defined model structure; data for WY 1974-2012
- Phase 2: **Hybrid ANNs** where an empirical DSG model (Hutton et al., 2015) is used to fit the salinity data and then an ANN is used to correct the empirical model fit; data for WY 1922-2012

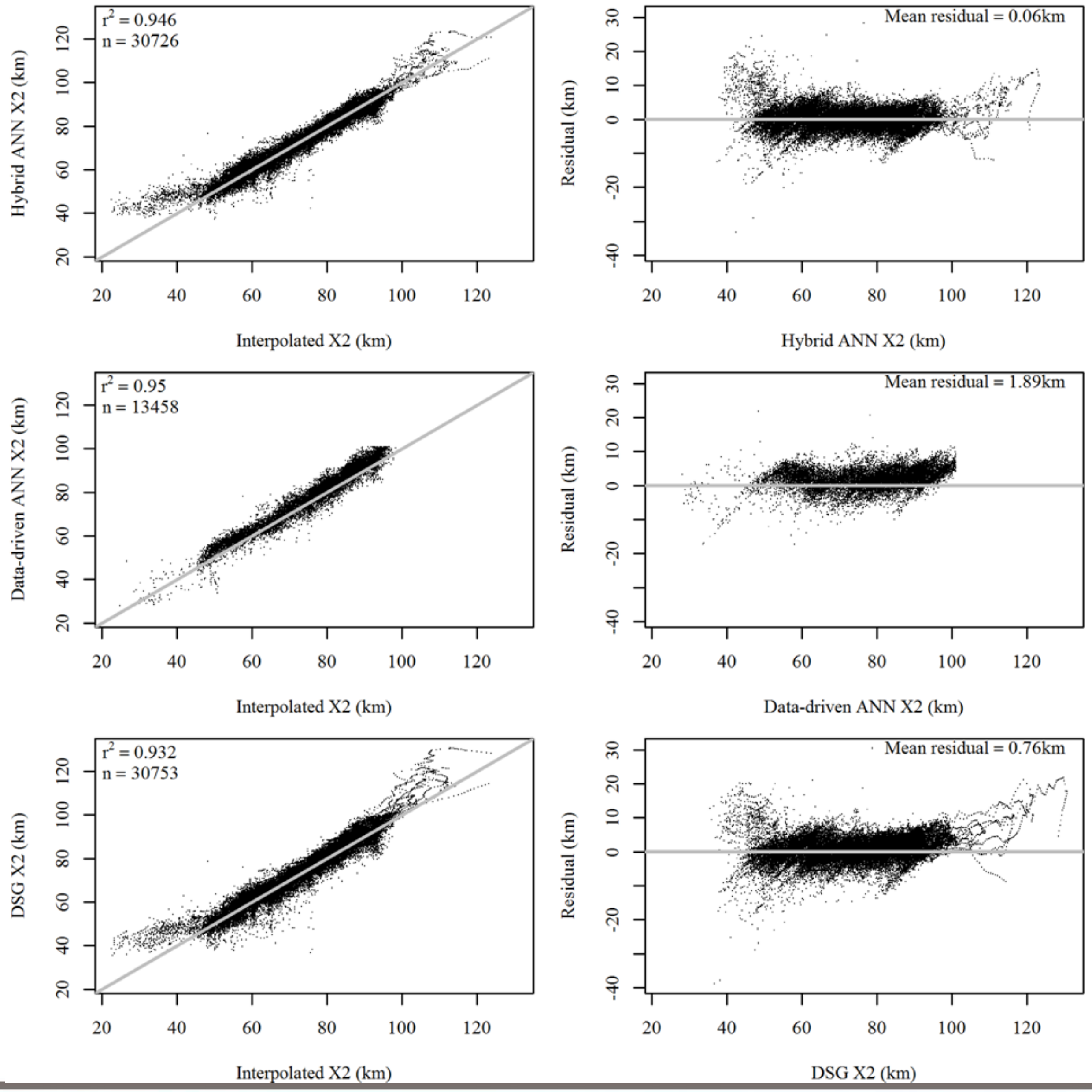
Definition of Terms





TETRA TECH

Fit to X2 Values (Predictive validation)



Hybrid ANN
WY 1922-2012

Data-driven ANN
WY 1974-2012

DSG Model
WY 1922-2012

Structural Validation of Data-Driven Models

- **Evaluate sensitivity of change in the sea level variable**
  - We expect that an increase in sea level should result in a positive effect on salinity at a fixed location, or a positive effect on X2 for the same Delta outflow
  - Data-driven models with more hidden neurons fit the data better, but had an inconsistent response to an imposed sea level change
- **Proposed solution: Vary ANN size to control response to MSL**
- **However, smaller networks had a better structural response, but poorer predictive validation**

Alternative Approach: Hybrid ANN

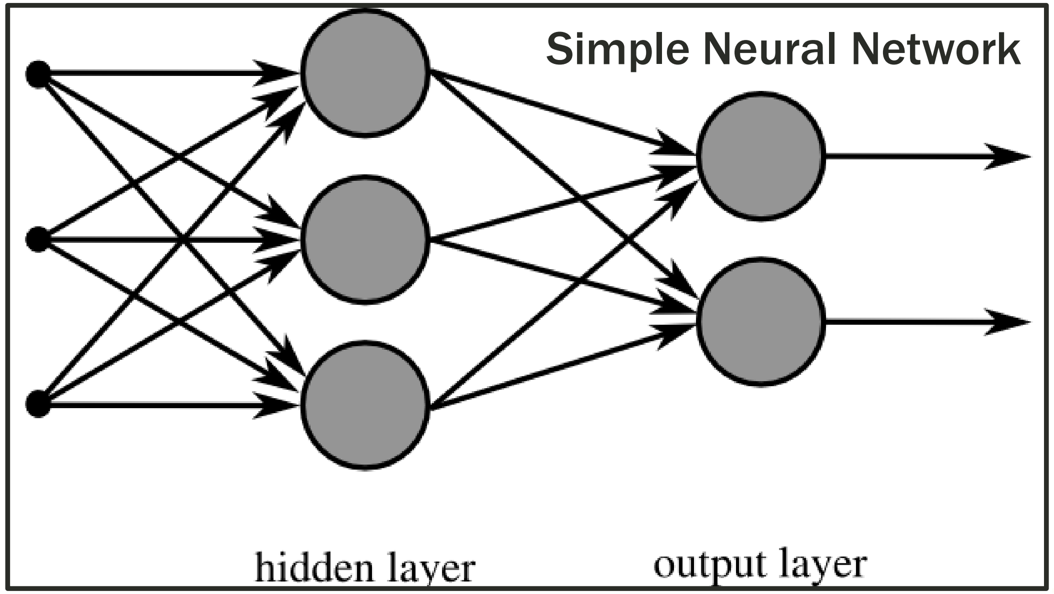
- Takes advantage of knowledge of the system
- Fit salinity data using empirical model, the Delta Salinity Gradient (DSG) model
- Remaining error fit with ANN
- The role of the ANN is thus more limited and only focused on improving the DSG model fit
- Additional constraint:
  - Constrain weights such that MSL increase can only result in a positive effect on X2
  - Only the sign of the change was constrained, not the magnitude

Model Inputs and Outputs

- **Explored various combinations of inputs with regard to ability to fit observed salinity data at key locations along the salinity gradient**
- **Final Inputs**
  - Station distance (km) from Golden Gate
  - Daily freshwater flow to estuary
  - Tides – daily mean coastal water level and daily tidal range (not in original DSG model)
- **Model Outputs used for Training ANNs**
  - Used data set of salinity compiled from grab sample and continuous data, spanning WY 1922-2012
  - Data cleaned to remove erroneous values, and filled where short gaps exist
  - X2 computed based on log-linear interpolation of salinity values
  - Separate sets of X2 developed for the Sacramento and San Joaquin Rivers

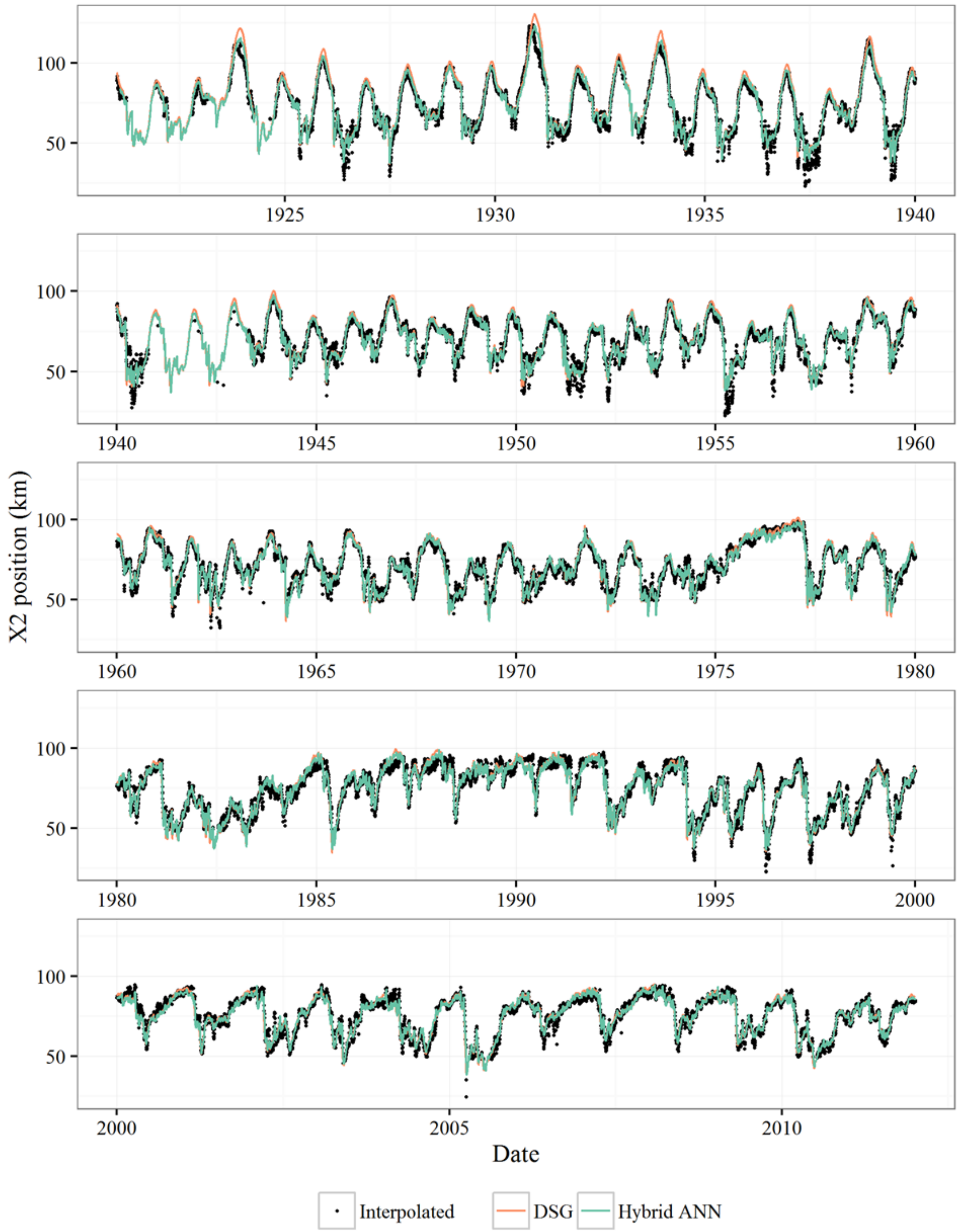
ANN Model Details

- **Architecture and Structure**
  - Feedforward multilayer perceptron
  - Time lag of 28 days
  - 3 hidden neurons
- **Calibration Approach**
  - Used a Bayesian approach for training to constrain the weights such that only a positive response to sea level rise was possible
  - 15% of the data used for training

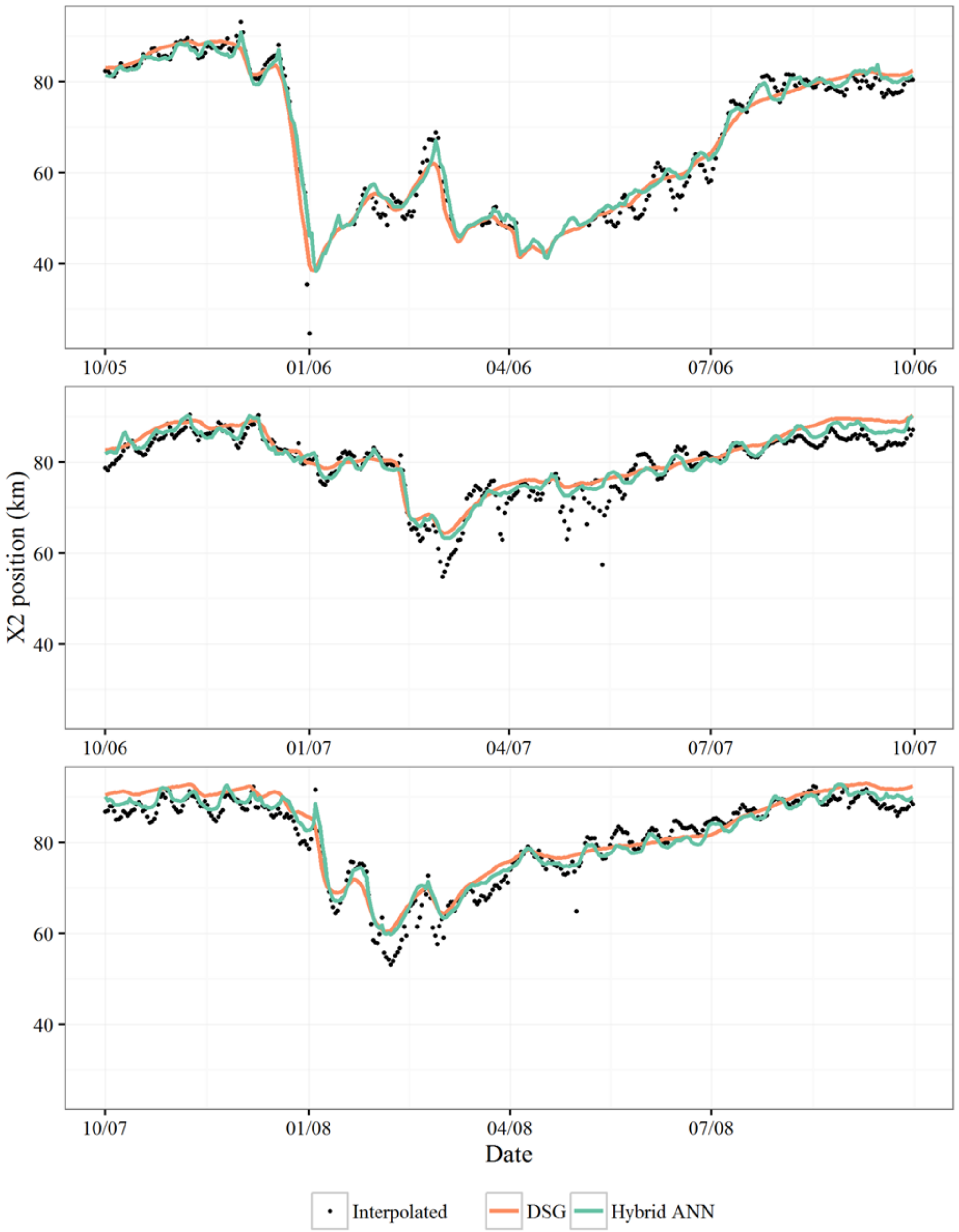


Results: Observed X2 Fits

		Sacramento	San Joaquin
Count		30,753	30,224
Mean Residual (km)	DSG	0.76	-0.31
	Hybrid ANN	<0.01	-0.07
Coefficient of Determination (r2)	DSG	0.93	0.93
	Hybrid ANN	0.95	0.94
Standard Error (km)	DSG	3.63	4.03
	Hybrid ANN	3.22	3.73



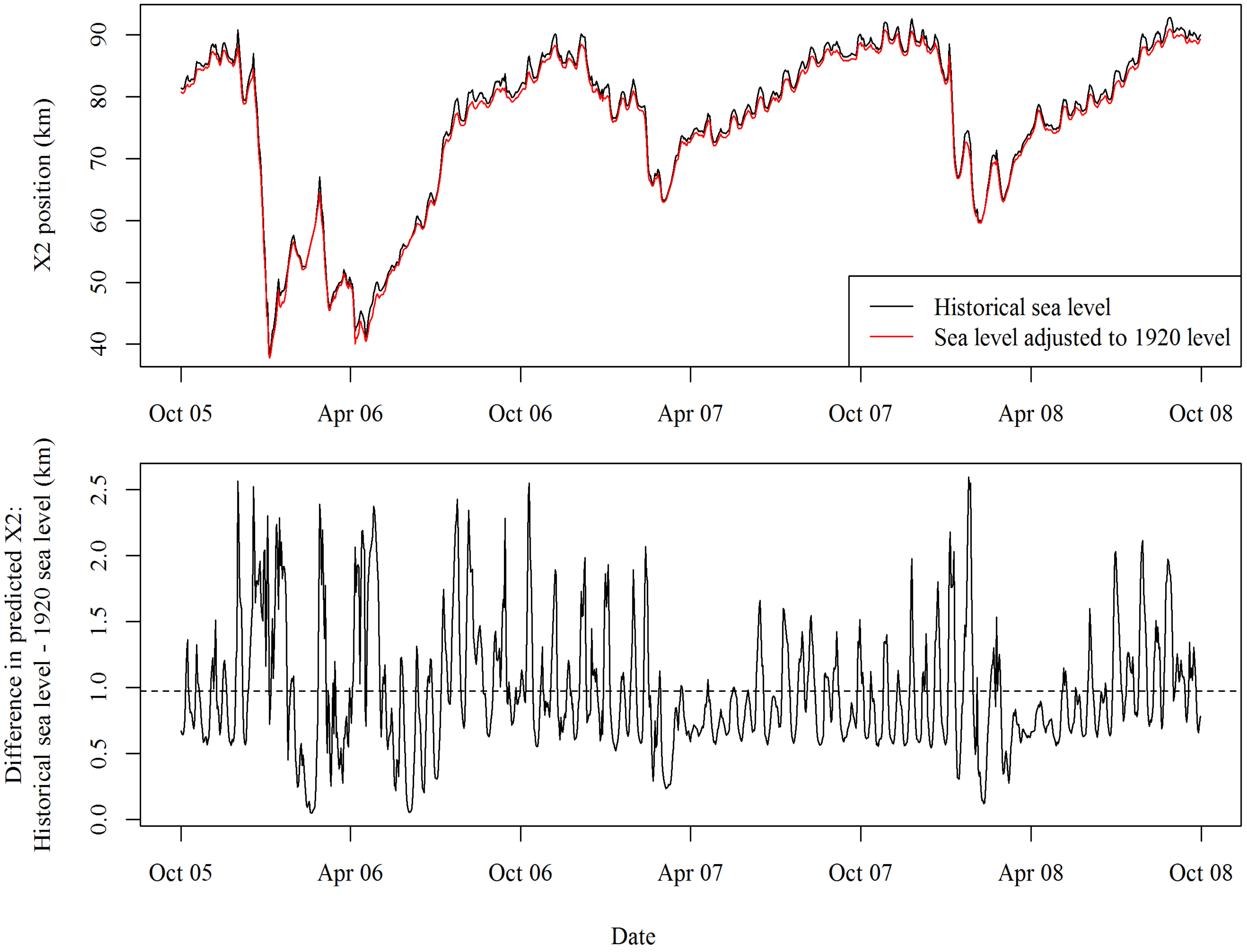
X2 over a shorter period



TETRA TECH

Case Study: Isolate Historical Sea Level Rise Effects

- **Setup**
  - De-trended historical Golden Gate mean water level time series using linear regression to derive a 1920-level input time series
  - Maintained other model inputs at baseline conditions (with a water level difference of +18.3 cm between the two scenarios in 2012)
- **Results**
  - Historical sea level rise has resulted in slightly greater salinity intrusion into the estuary (X2 shifted ~1km upstream)



TETRA TECH

Key Findings

- **Data-driven networks had strong predictive validity; however, even with the large database available, structural validation was a challenge**
- **Used hybrid approach, that combined an empirical model (DSG model) with an ANN**
  - This approach, in conjunction with a constraint on the weights in relationship to sea level rise, resulted in hybrid ANN models that were predictively and structurally valid—and provided better fits than existing models over a 91-year period of record
- **Hybrid approach can be considered a maturation of ANN methodology in water resources, where the strengths are best used in conjunction with existing modeling approaches**
  - Use of model to fit data a priori allows for use of longer data set, exposing model to wider range of conditions
  - Allows rapid, more robust modeling predictions
  - Addresses skepticism of "black box" ANN models

Publication

- This work is published in *Environmental Modeling and Software (open access)*

Rath, J.S., P.H. Hutton, L. Chen, and S.B. Roy. A hybrid empirical-Bayesian artificial neural network model of salinity in the San Francisco Bay-Delta estuary. Environmental Modeling & Software 93 (2017) 193-208.