# Modeling Salinity in the San Francisco Bay-Delta Estuary using Artificial Neural Networks

John S. Rath[1], Paul H. Hutton[2], Limin Chen[1], and Sujoy B. Roy[1]

[1]Tetra Tech Inc. and [2]Metropolitan Water District of Southern California

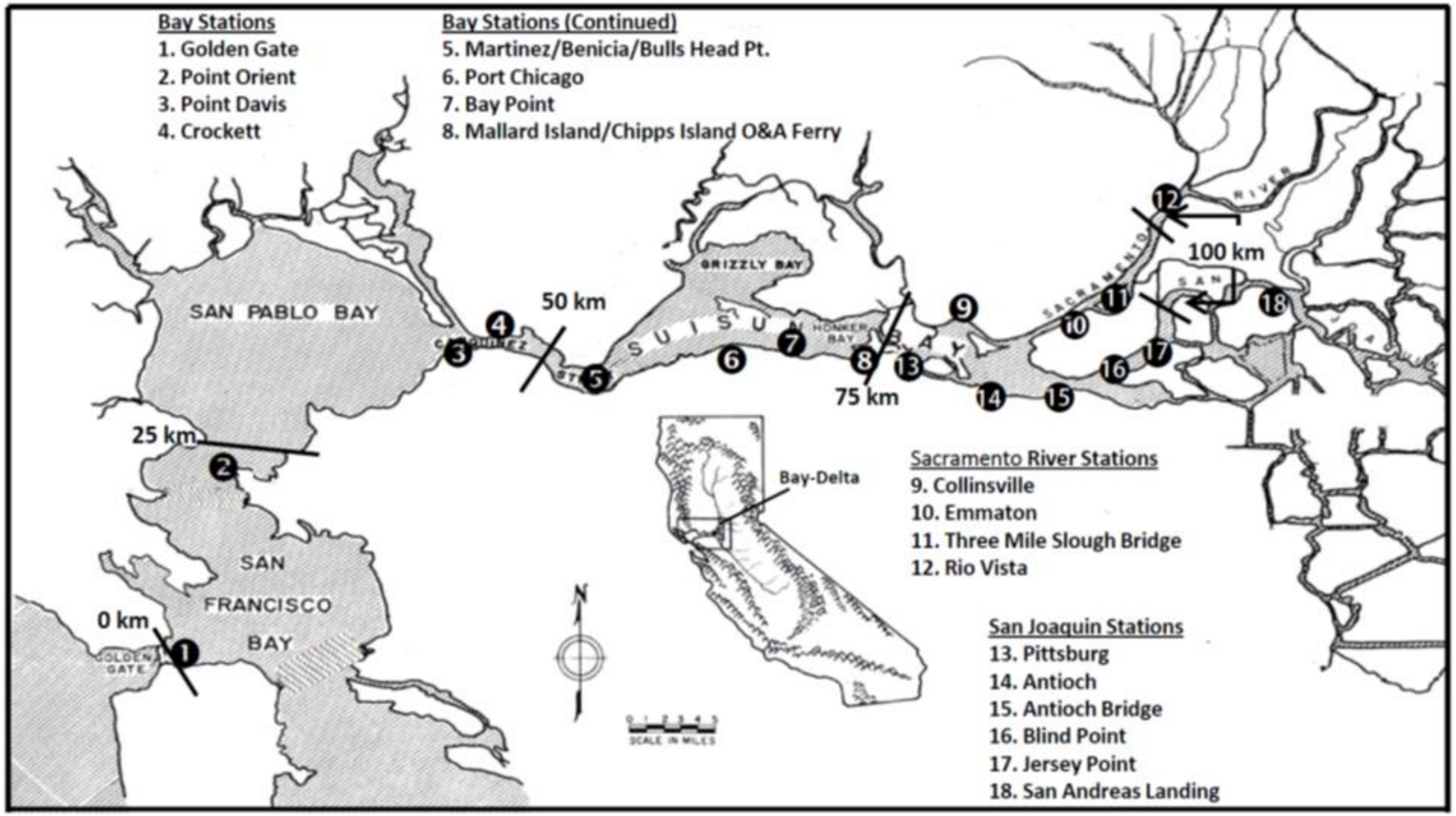
Presentation at the California Water and Environmental Modeling Forum

April 11, 2016

# Presentation Overview

- Why develop an ANN model?
- Data used in ANN development
- Model selection and training approach
- Data-driven ANN models
- Hybrid ANN model
- Predictive and structural validation
- Example applications
- Sea level rise case study

# Salinity in the western Delta and San Francisco Bay



# Why develop an ANN model for salinity?

- Supports the need for making rapid salinity predictions across multiple locations for Delta outflow management
- 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)

# Overview of Approach

- Used feed-forward 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

# Model Inputs: Initial Phase

- Station distance (km) from Golden Gate, estimated independently along estuary center depth
- Flow variables – Rio Vista flow (on the Sacramento River), Qwest flow (on the San Joaquin River past Jersey Point), and net Delta outflow from the DAYFLOW program
- Tides (mean and range) –Golden Gate and other locations
- Astronomical tide and atmospheric pressure
- Many different combinations of inputs possible; after initial screening, 10 sets of inputs were considered

# Model Inputs: Final

- Station distance (km) from Golden Gate
- Flow variable –net Delta outflow from the DAYFLOW program
- Tides –Golden Gate mean sea level and daily tidal range

# Model Outputs used for Training ANNs

- Used a 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
- Summary of data used and trend evaluation reported in Hutton et al. 2015

Hutton, P.H., J.S. Rath, L. Chen, M. J. Ungs, and S. B. Roy (2015). Nine Decades of Salinity Observations in the San Francisco Bay and Delta: Modeling and Trend Evaluation. ASCE Journal of Water Resources Planning and Management. doi: 10.1061/(ASCE)WR.1943-5452.0000617, 04015069.

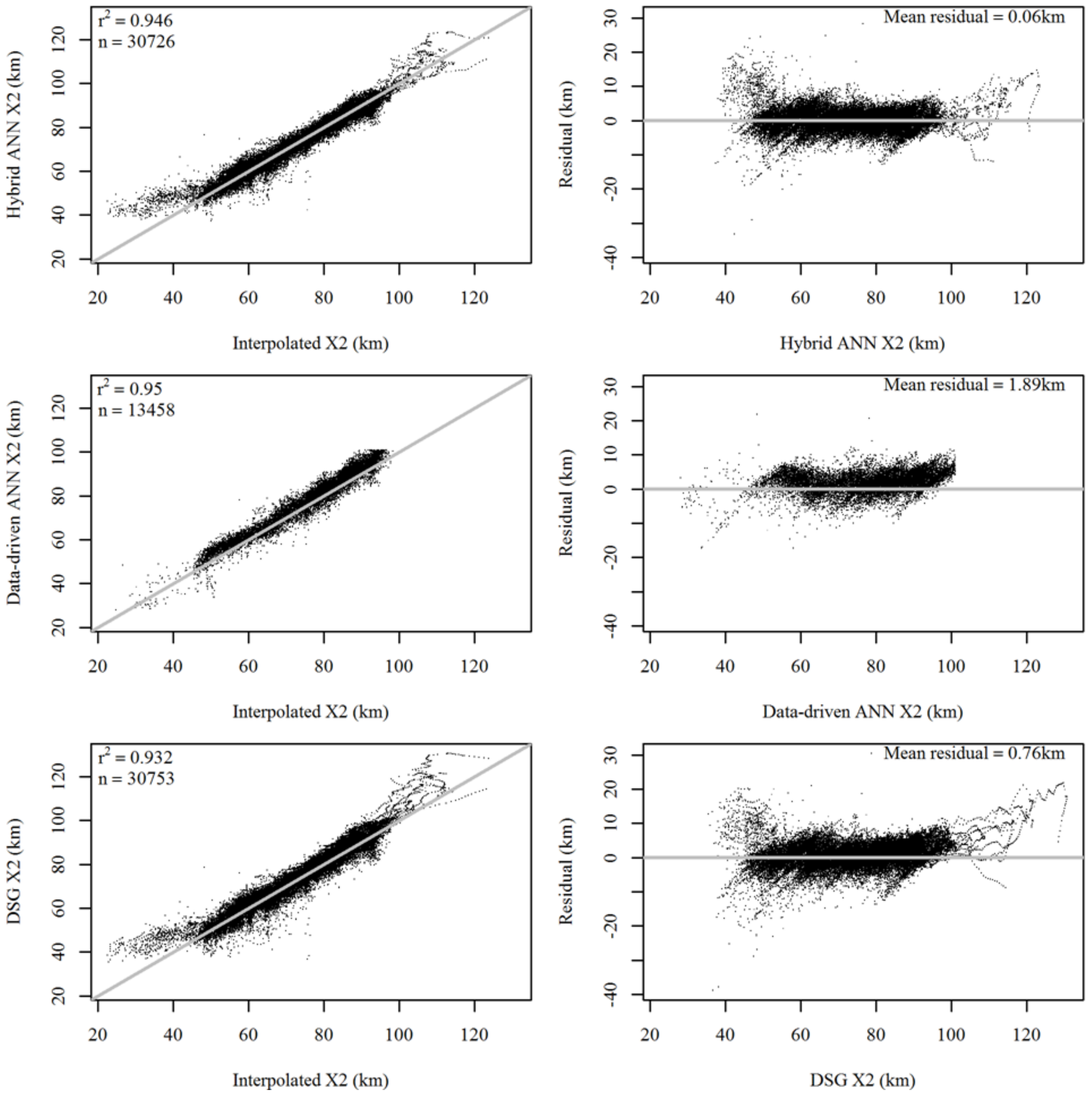
# Model Architecture and Structure

- Data-driven ANNs
  - Feedforward multilayer perceptron
  - Time lag of input variables from 15-120 days
  - 1-10 hidden neurons
- Hybrid ANN
  - Feedforward multilayer perceptron
  - Time lag of 28 days
  - 3 hidden neurons

# Calibration Approach

- **Data-driven ANN models:** Split data into training, validation, and testing subsets (50%, 25% and 25%). Used continuous daily salinity data from WY 1974-2012
- **Hybrid ANN models:** Used a Bayesian approach for training, constraining the weights such that only a positive response to sea level rise was possible; 15% of the data used for training, used data from WY 1922-2012, including periods with gaps in the early part of the record

# Fit to X2 Values (Predictive validation)




Hybrid ANN
WY 1922-2012

Data-driven ANN
WY 1974-2012

DSG Model
WY 1922-2012

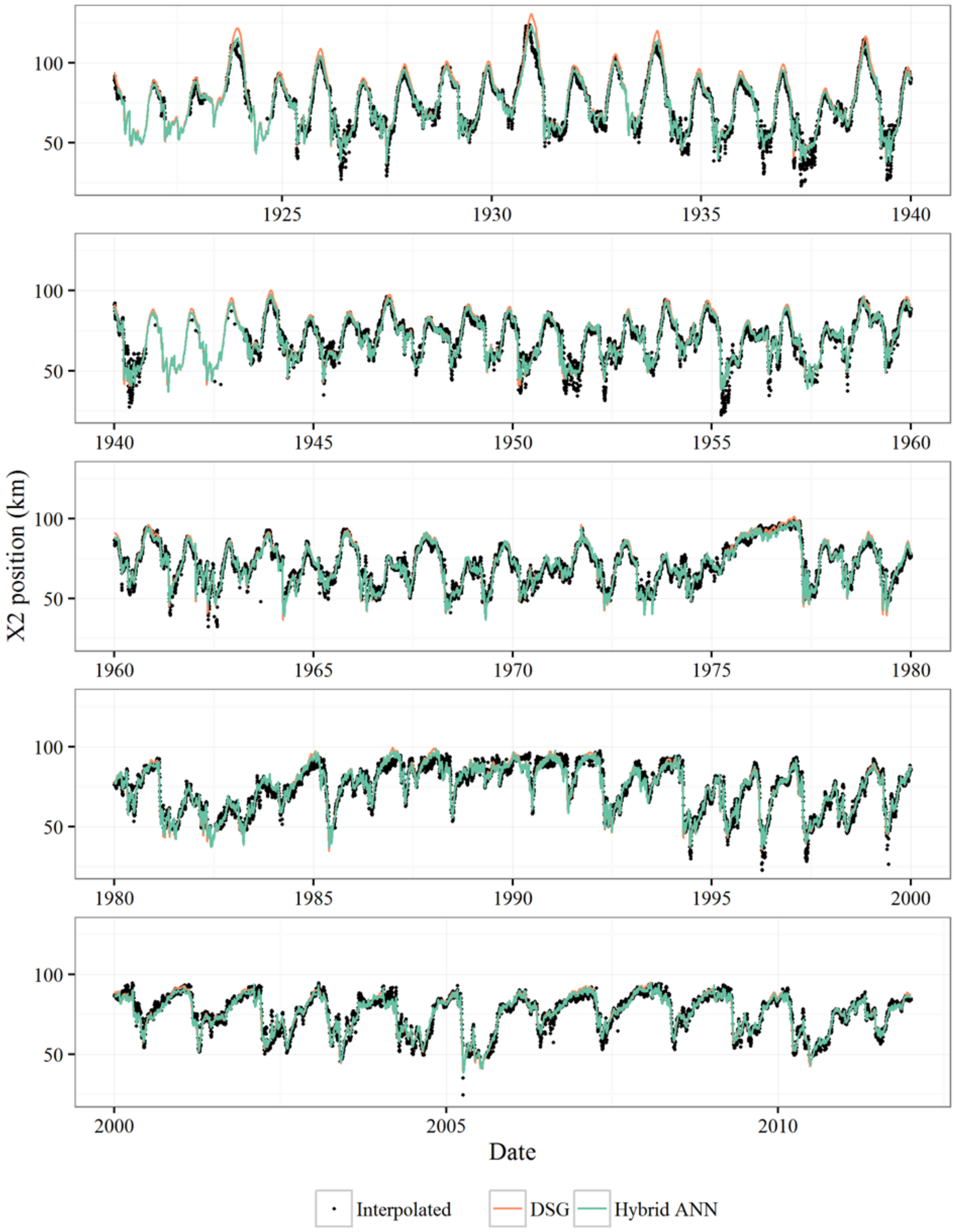
# Structural Validation of Data-Driven Models

- 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 an 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

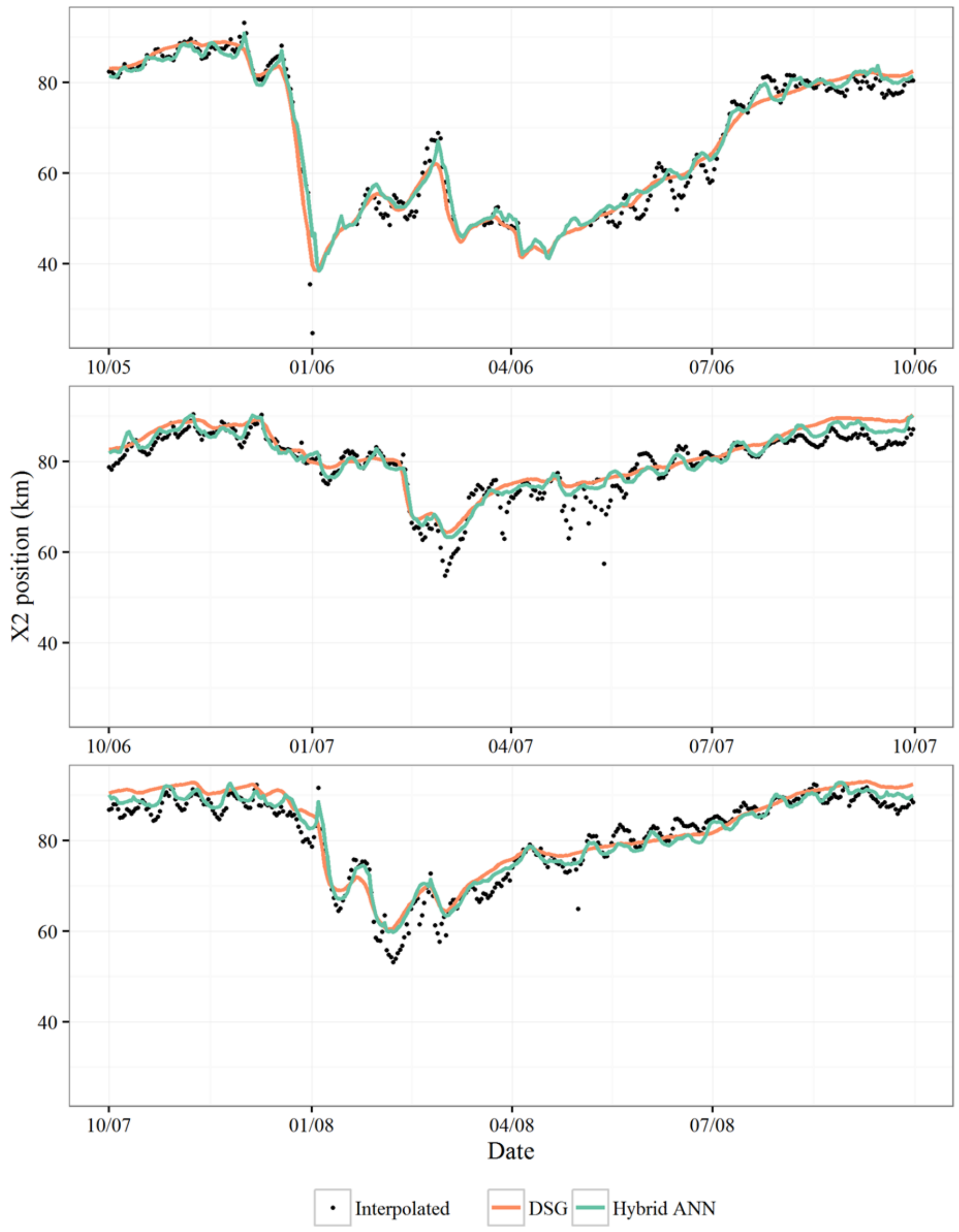
# Observed X2 Fits



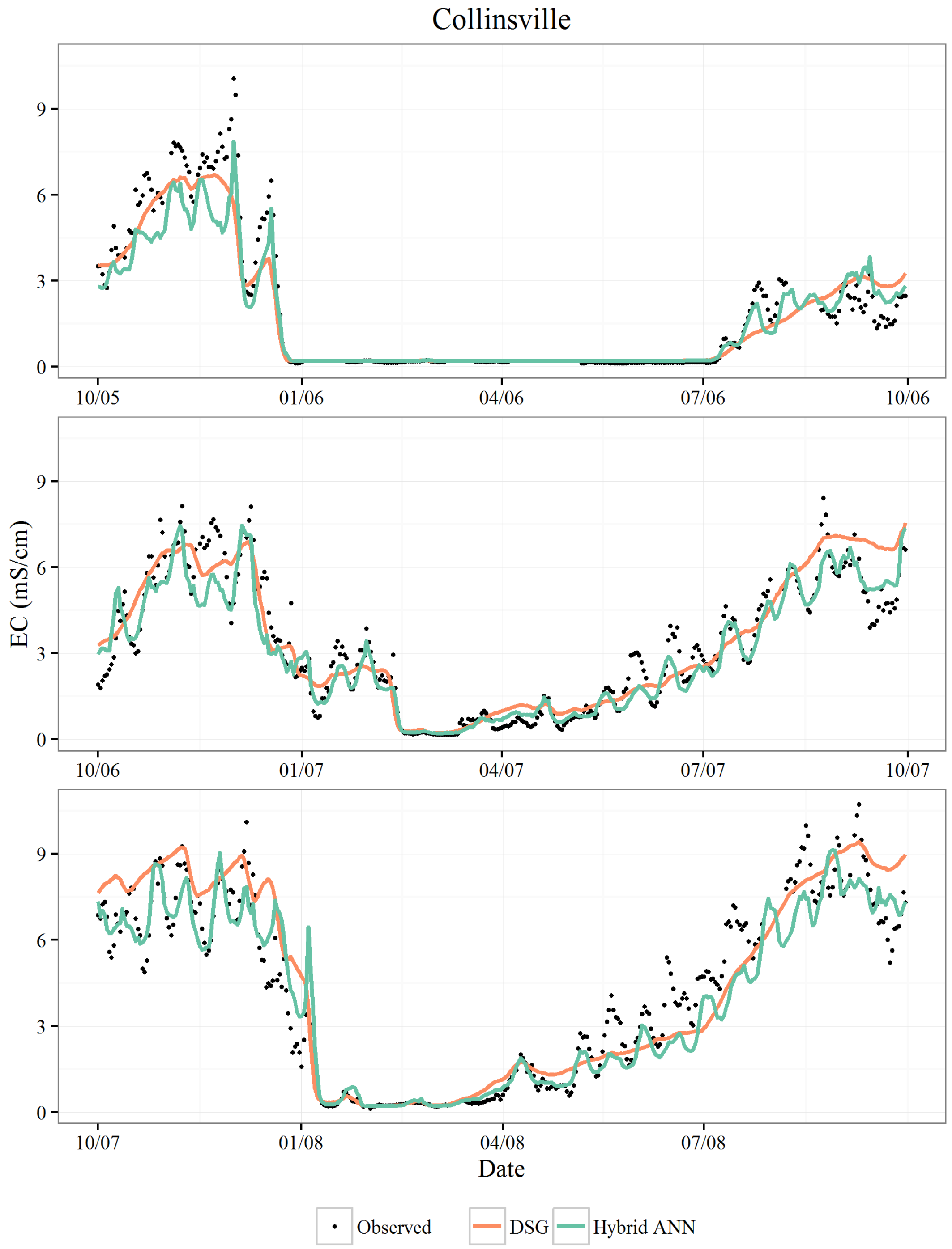
# Statistics Pertaining to X2 Predictions

| River Branch | Count | Mean Residual (km) | | Coefficient of Determination ($r^2$) | | Standard Error (km) | |
|---|---|---|---|---|---|---|---|
| | | DSG | Hybrid ANN | DSG | Hybrid ANN | DSG | Hybrid ANN |
| **Sacramento** | 30,753 | 0.76 | <0.01 | 0.93 | 0.95 | 3.63 | 3.22 |
| **San Joaquin** | 30,224 | -0.31 | -0.07 | 0.93 | 0.94 | 4.03 | 3.73 |

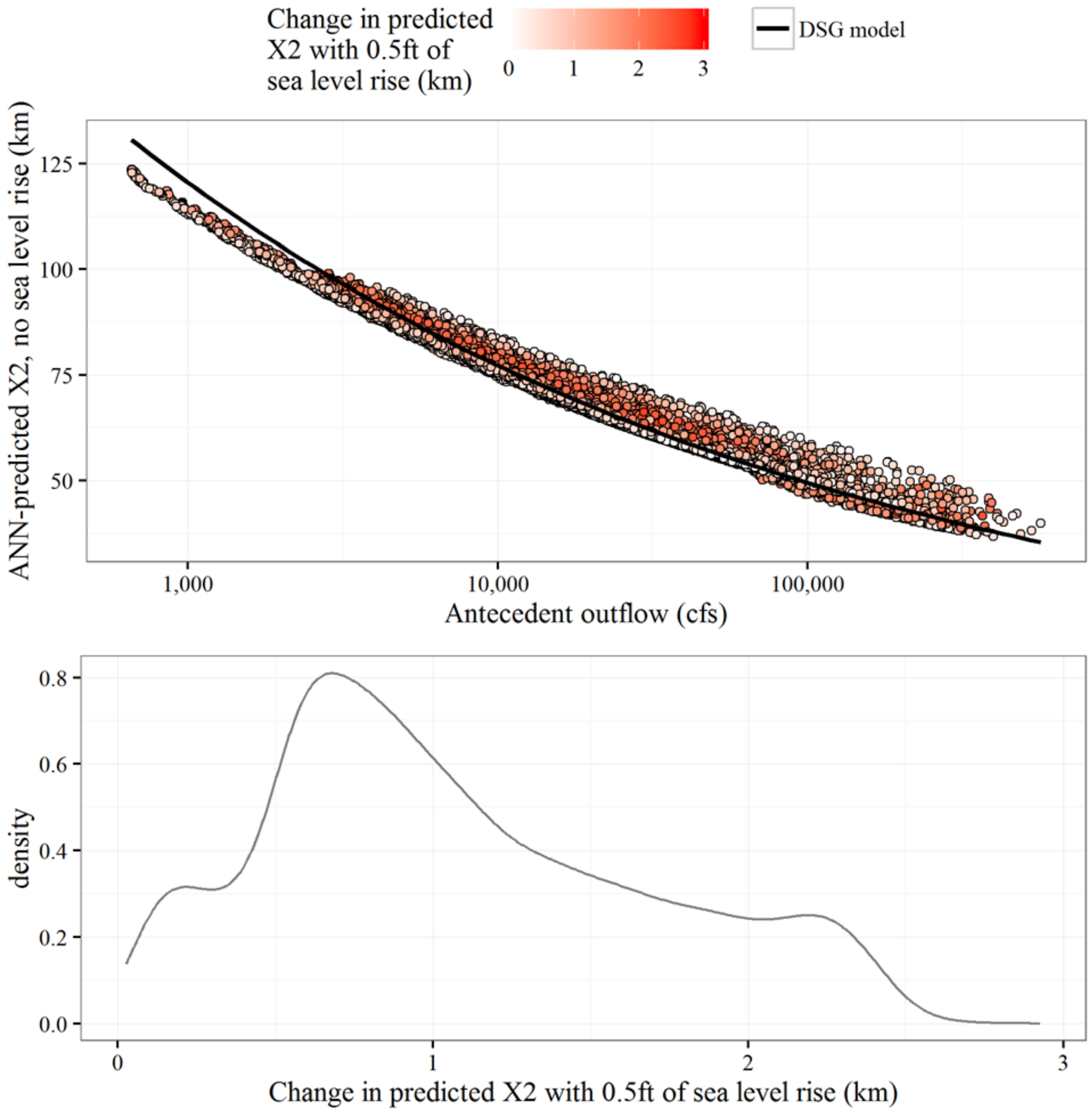
# X2 over a shorter period



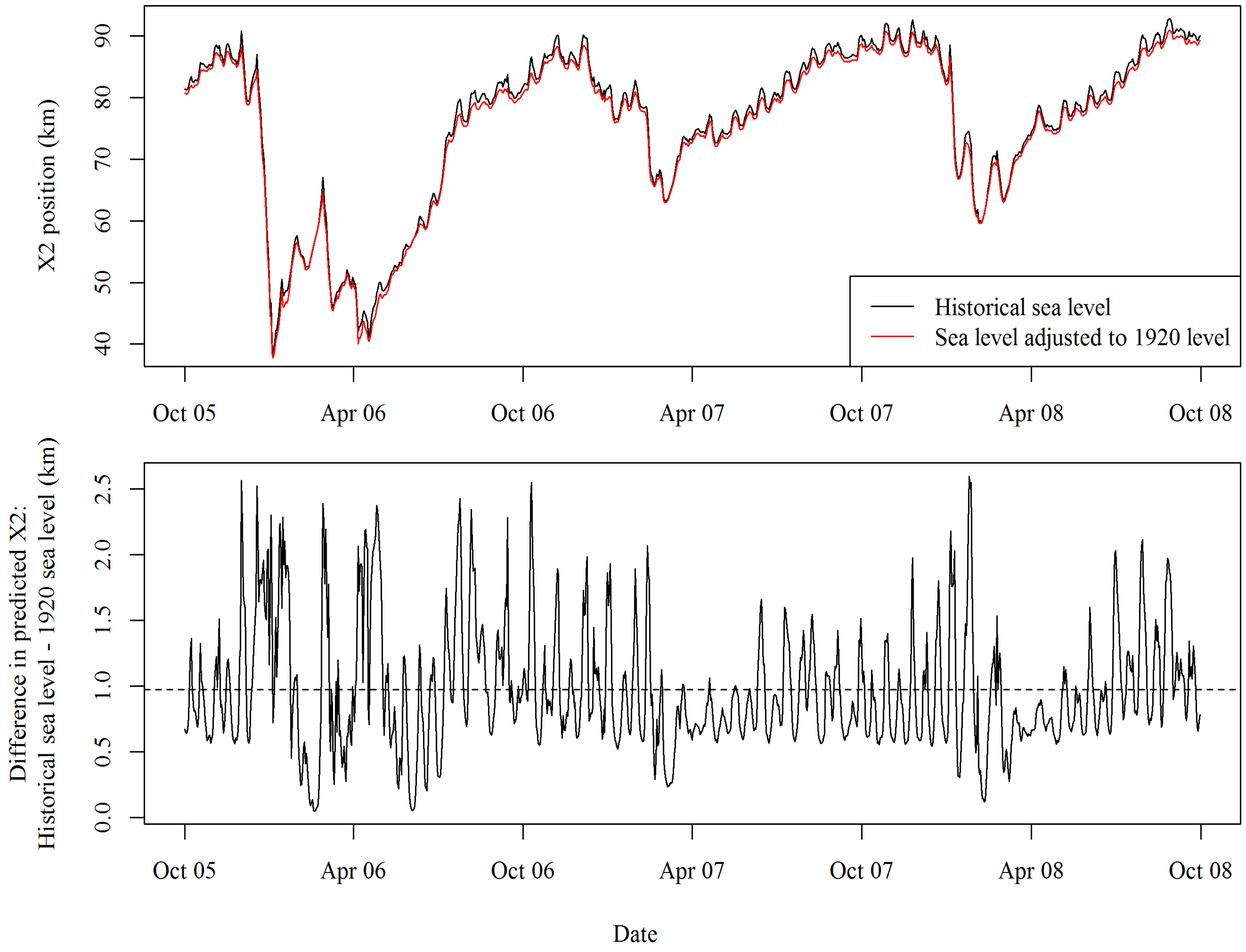
# Salinity at Collinsville



# Effect of Sea Level Rise of 0.5′ on X2



# Sea Level Rise Effects: 1920 to current levels (0.6 feet)



# Key Findings

- Data-driven networks had strong predictive validity; however, even with the large database available, structural validation was a challenge
- One solution was to reduce the size of the data-driven networks, but this adversely affected the quality of the fits
- Therefore, we turned to hybrid models, 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
- Not all error could be explained, and may be related to salinity measurement error, X2 interpolation error, changes in the estuary bathymetry and hydrodynamics