Generalized Delta Conservative Constituent Modeling using Artificial Neural Networks: Theoretical Background and Application

Project Team: Limin Chen, Katya Sayenko, and Sujoy B. Roy, Tetra Tech Inc.

Technical Direction: Paul Hutton, Metropolitan Water District of Southern California

Funding: State Water Contractors, Elaine Archibald, Project Manager

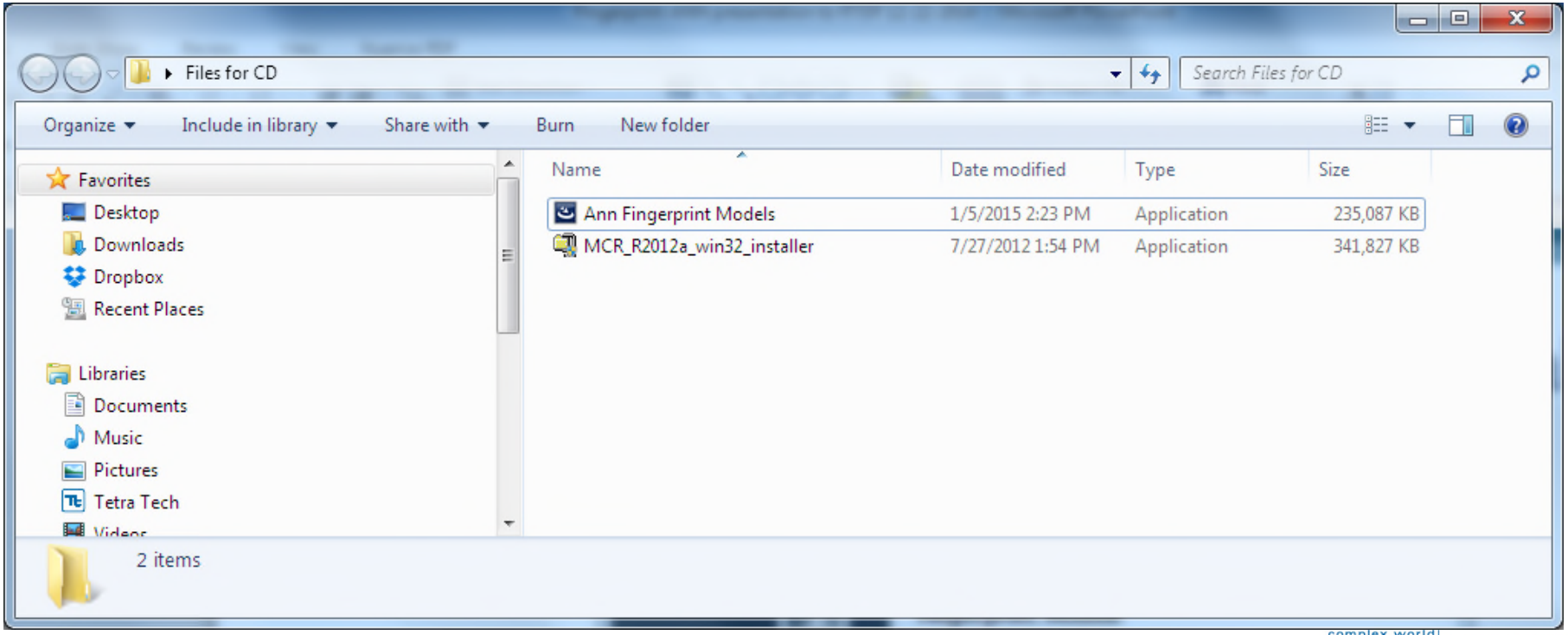
January 6th, 2015

Overview for Today

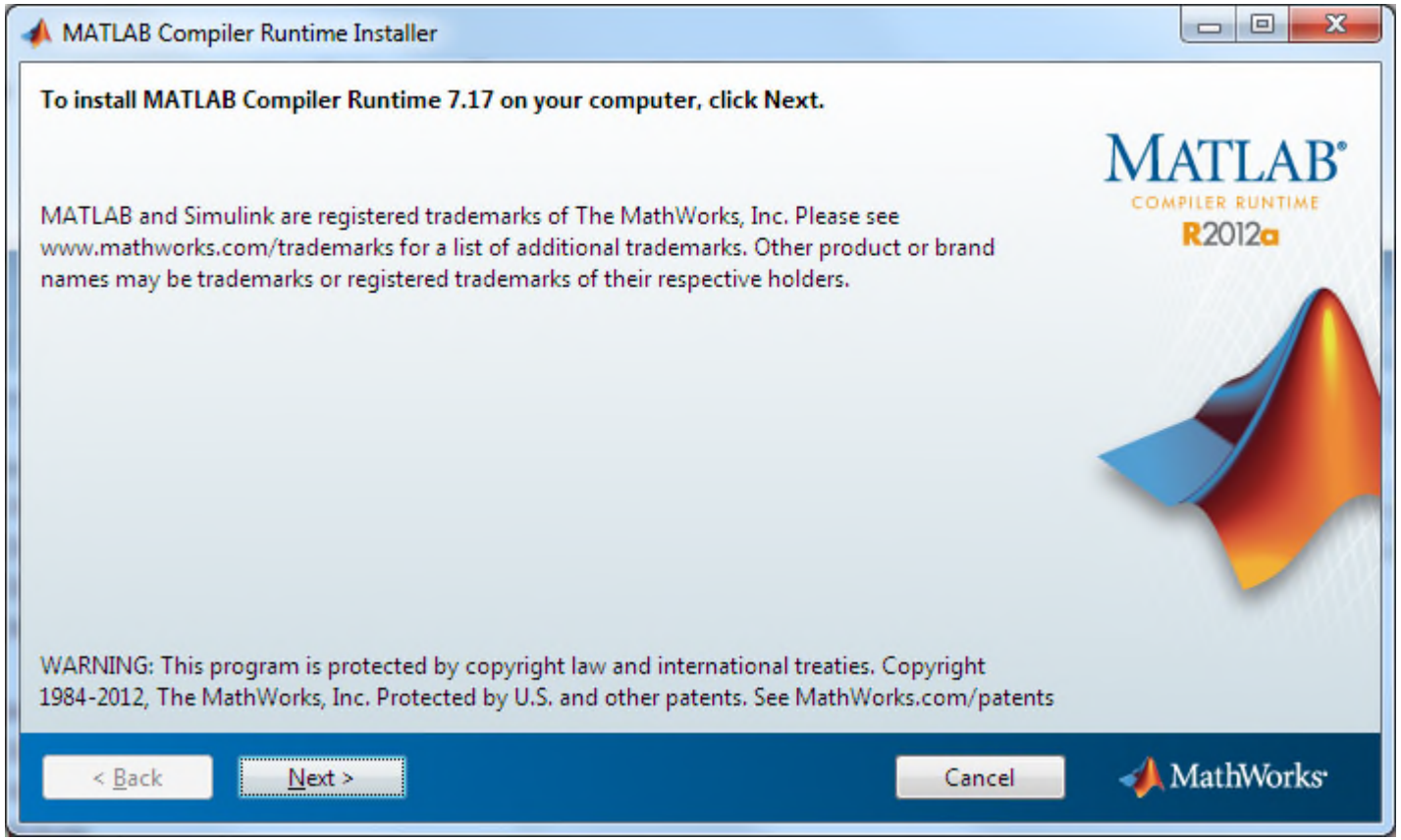
- Installation
- Introduction
- Theory and development
- Example applications
- Future development
- Discussion/Questions

Installation

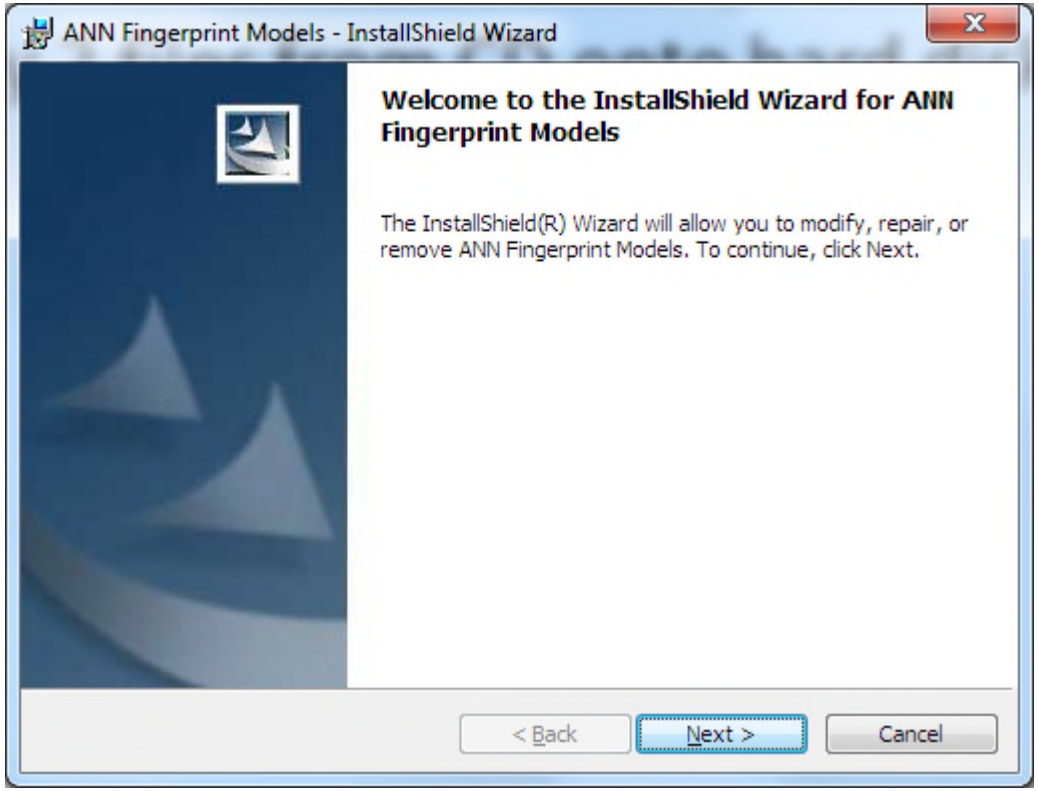
- Copy 2 files from CD onto hard disk
- Important note: Need administrator access to install new software on your computer.



First Install Matlab Runtime



ANN Fingerprint Model Installer



- During the install process, the program will bring up a DOS-like window
- Several DLLs need to be registered
- Answer “Y” when prompted for each DLL

At the end of the installation process:

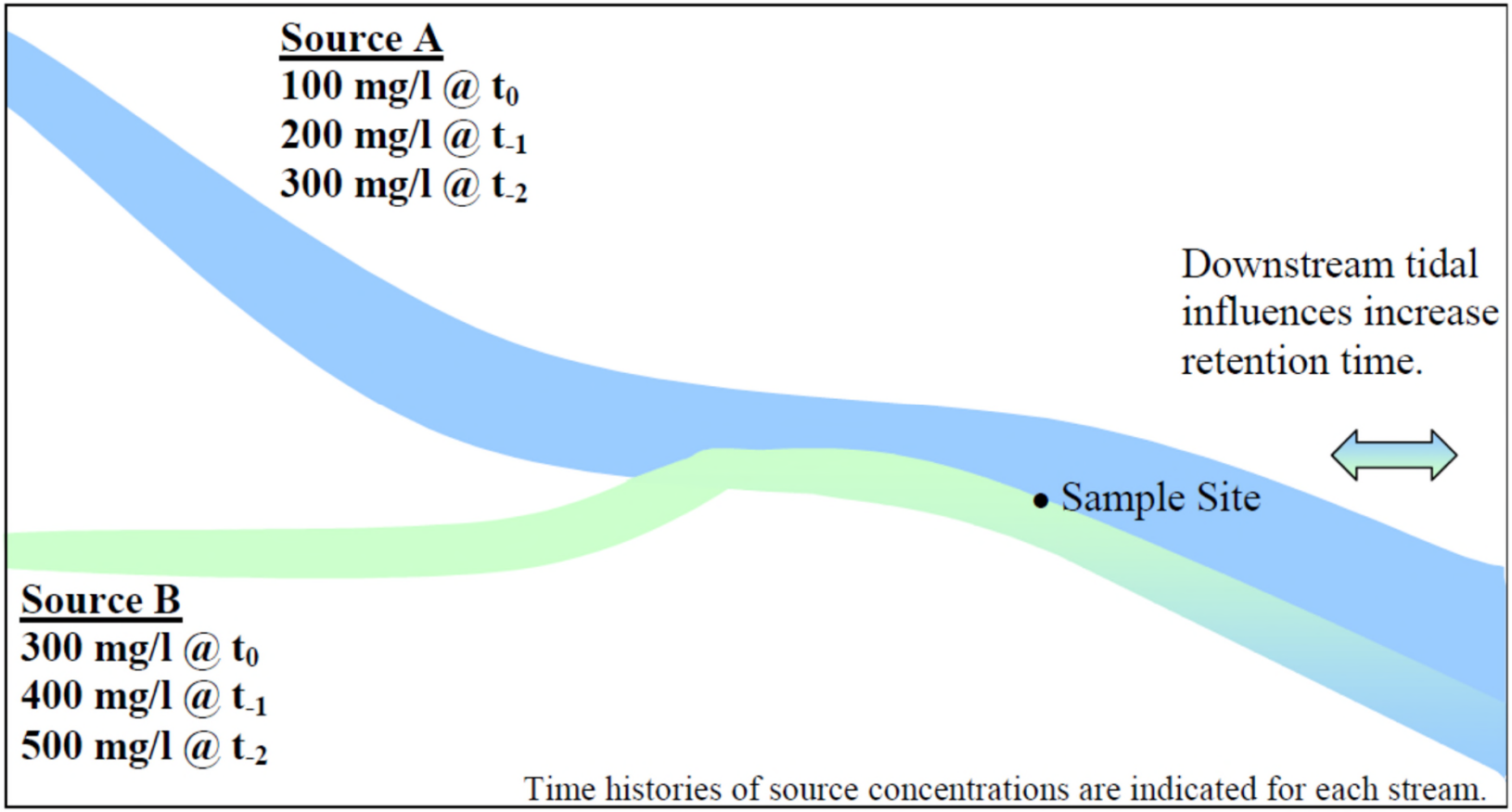
- You will have a run-time version of Matlab. This is largely invisible to the user.
- You will have Excel interface for the ANN fingerprint models
- DLLs for multiple ANNs will be registered to your account. Like Matlab, this is largely invisible to the user.
- The DLLs and Matlab are called through the Excel interface.

Theory and Development

Overview

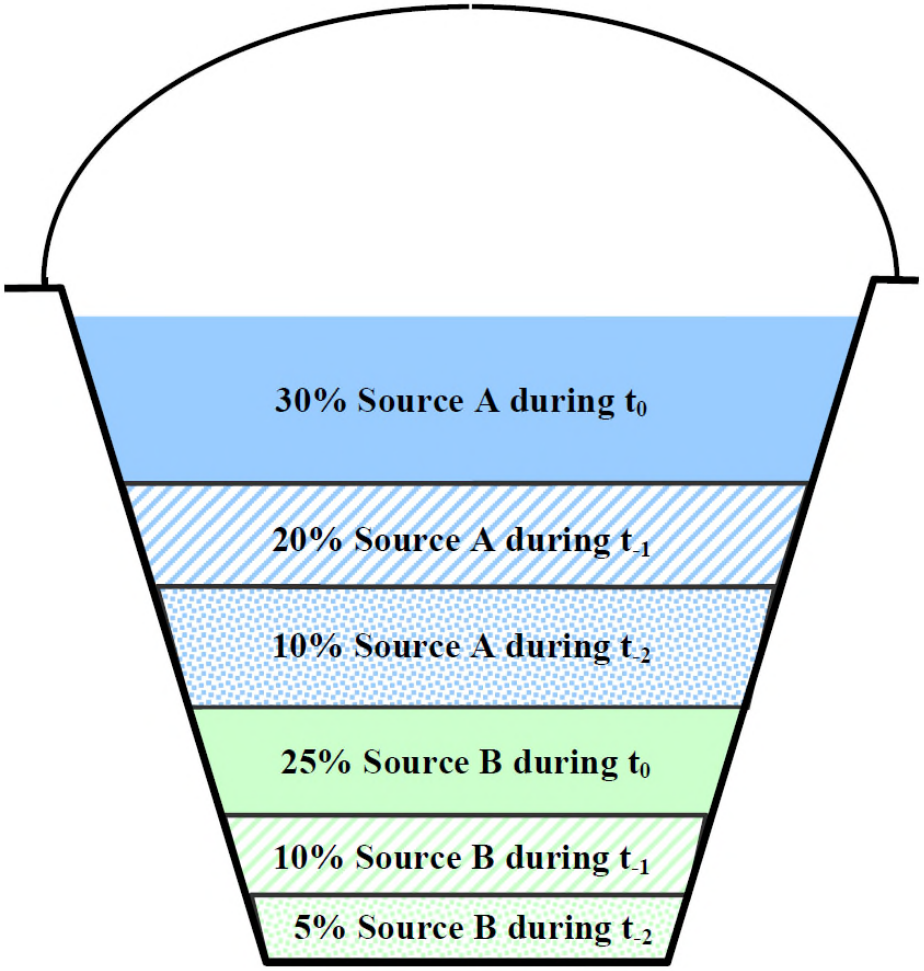
- The DSM2 model was used to simulate volumetric contribution from boundary sources across multiple stations in the Delta
- DSM2 used in the fingerprint mode can be used to relate the sources of water and any conservative constituent at a given location
- In this work we have developed an emulator for DSM2 using an Artificial Neural Network approach
- Goal is to have a simplified tool that performs calculations similar to those performed by DSM2
- The ANN-based model for individual stations was then used to compute the concentrations of three conservative constituents (salinity, bromide, and DOC) given boundary values, thus emulating the DSM2 processes

Fingerprint Model Conceptualization



Source: Jamie Anderson, 2002, DSM2 Fingerprinting Methodology

Concentrations at a Given Location

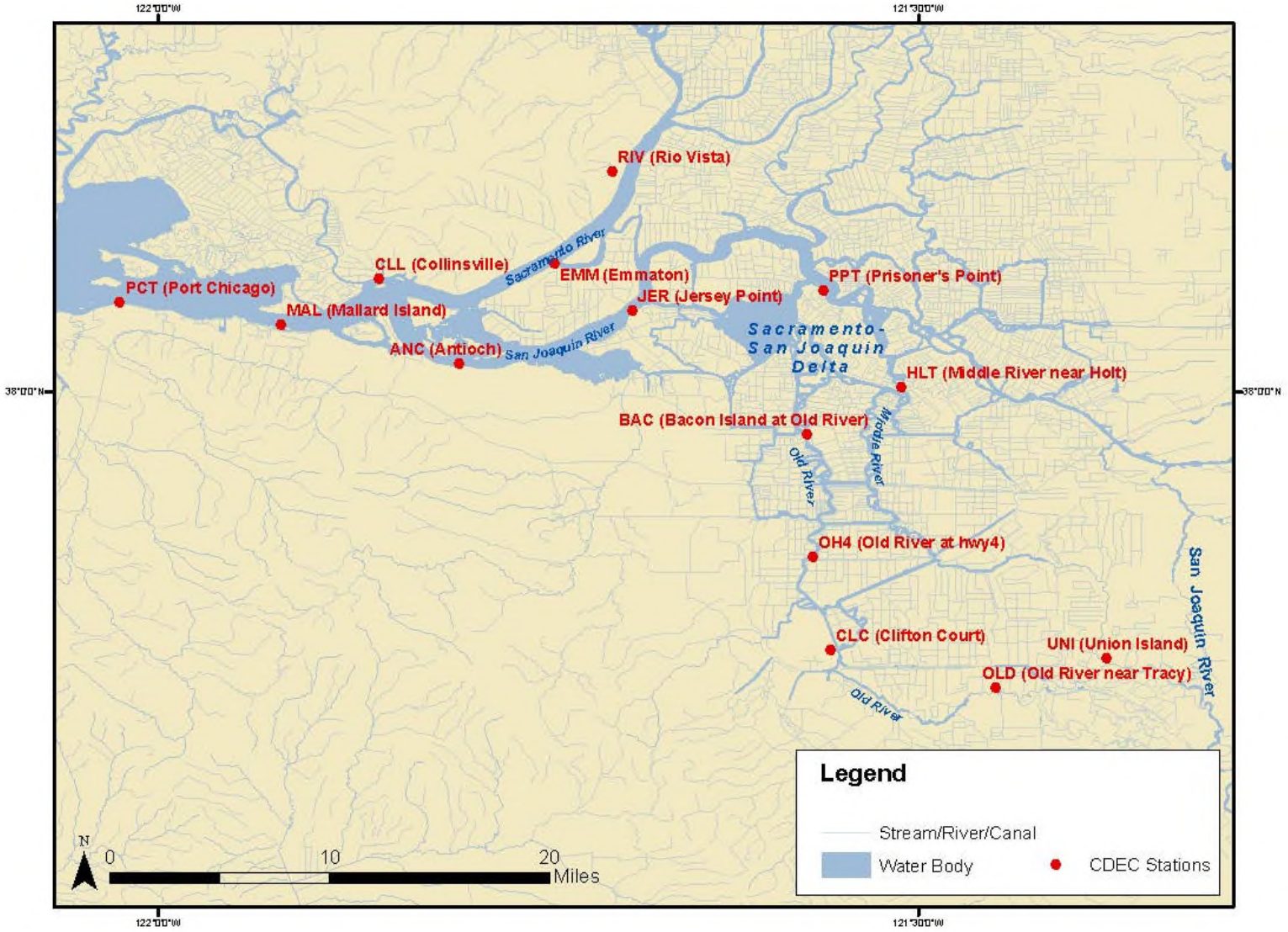


Source: Jamie Anderson, 2002, DSM2 Fingerprinting Methodology

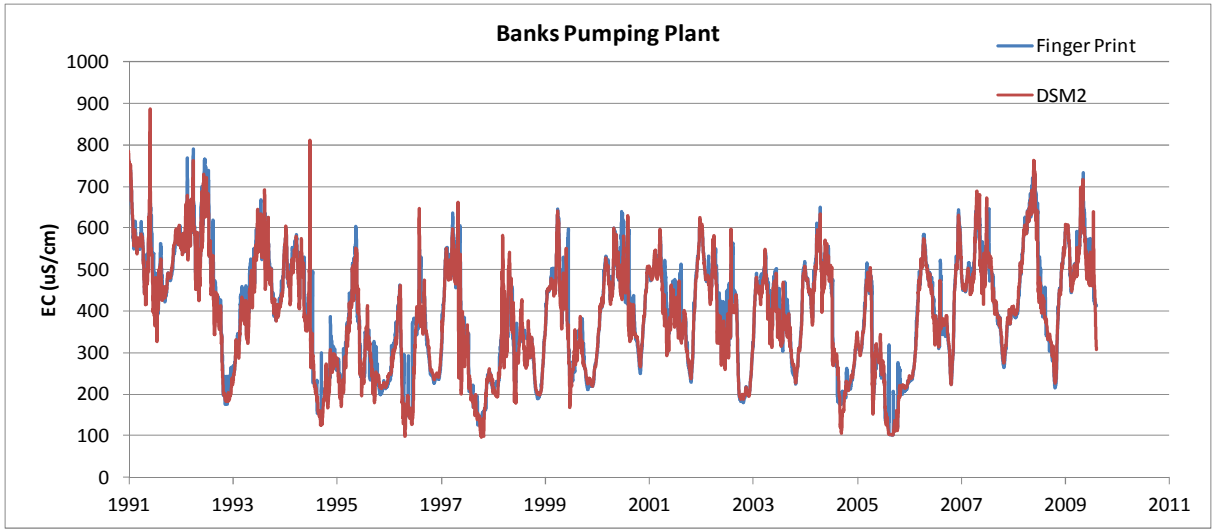
First Step: Does the DSM2 Fingerprinting Model Work as Expected?

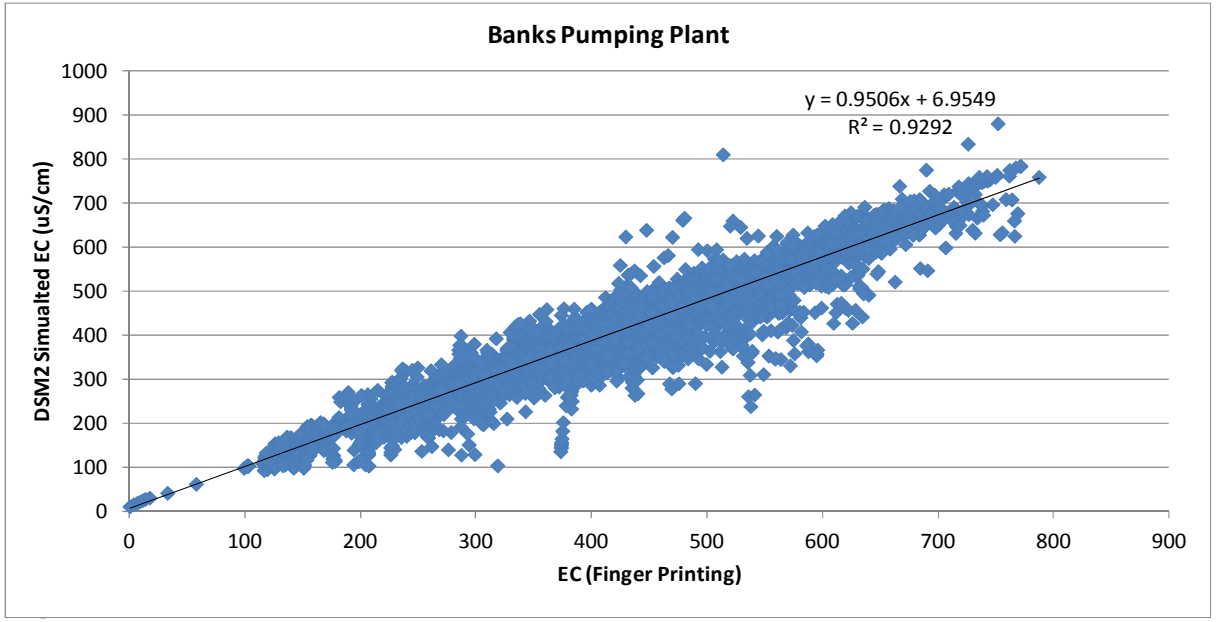
- DSM2 finger printing results, along with EC at boundaries, were used to predict EC at validation locations within Delta
- Predicted EC values from the fingerprint model were compared to DSM2-simulated EC at six validation locations:
  - Jones Pumping Plant
  - Banks Pumping Plant
  - Old River at Bacon Island
  - San Joaquin River at Jersey Point
  - Sacramento River at Mallard Island and
  - Old River at Highway 4

Focus on Delta Stations

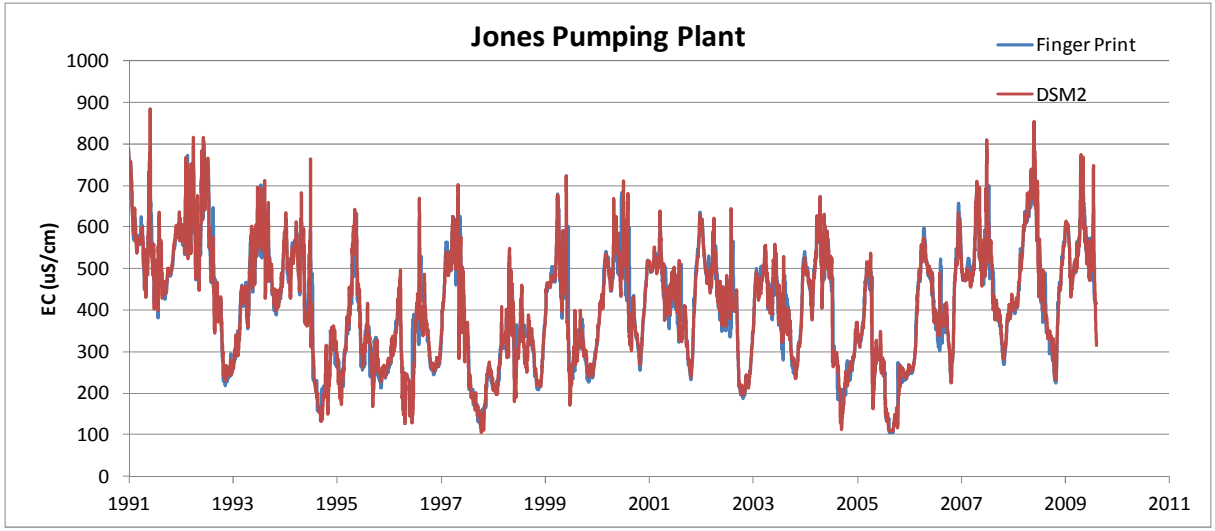


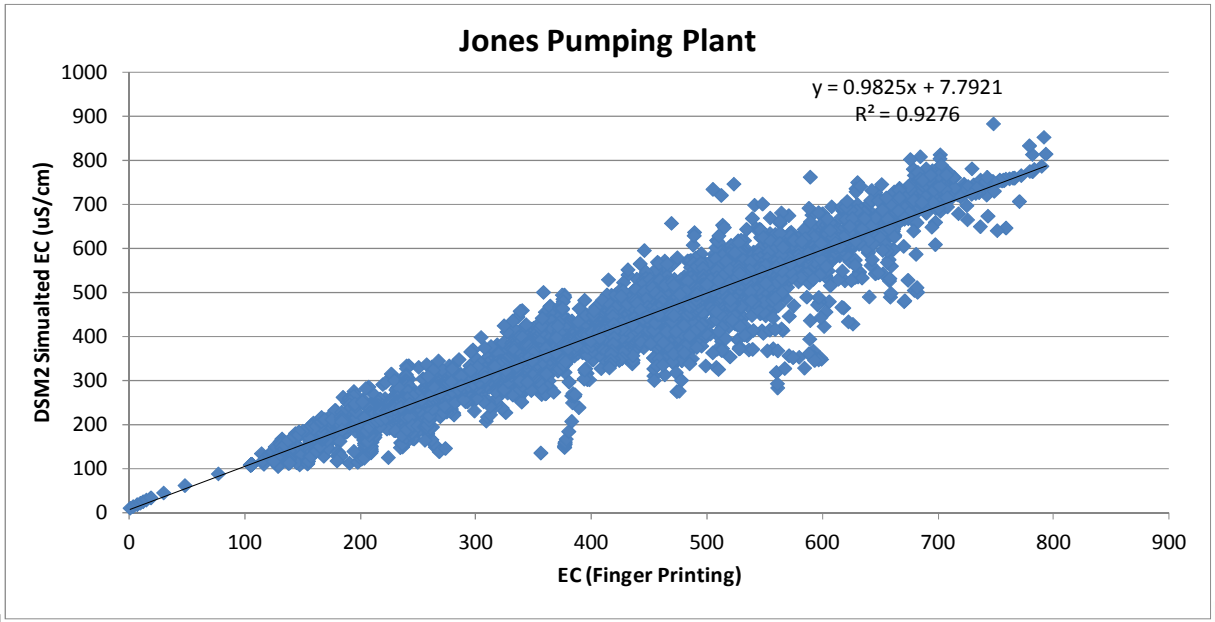
Fingerprint Model Validation: Banks Pumping Plant



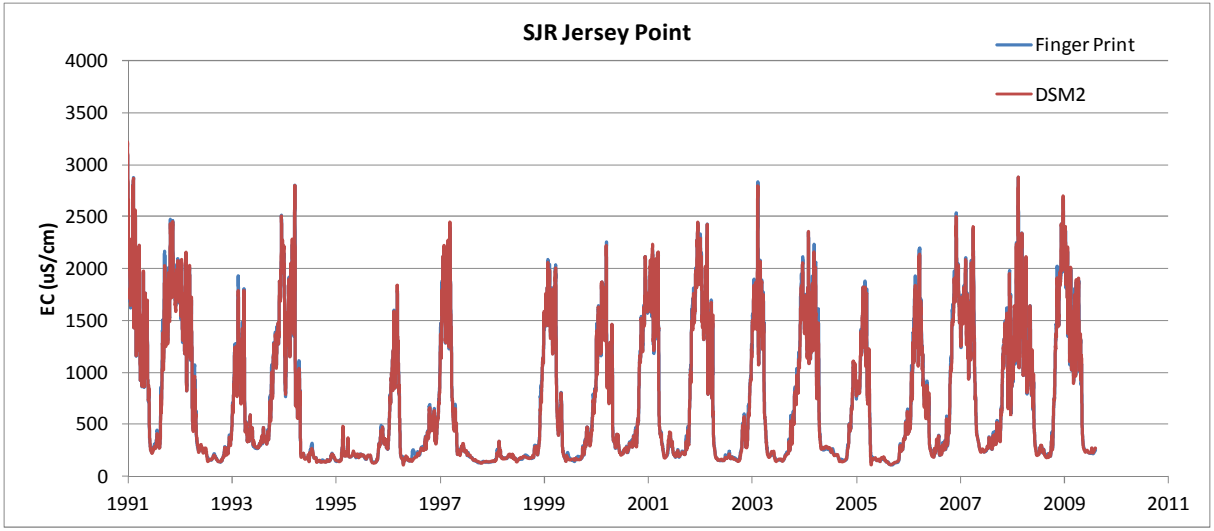


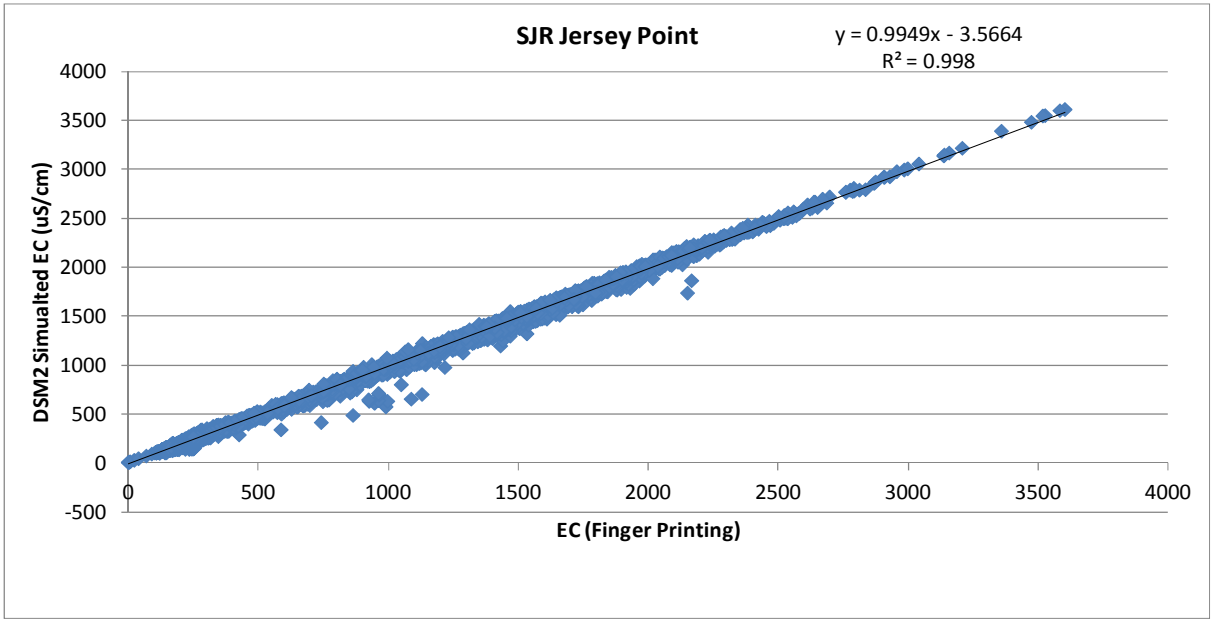
Off Ramp Validation: Jones Pumping





Off Ramp Validation: SJR Jersey Point

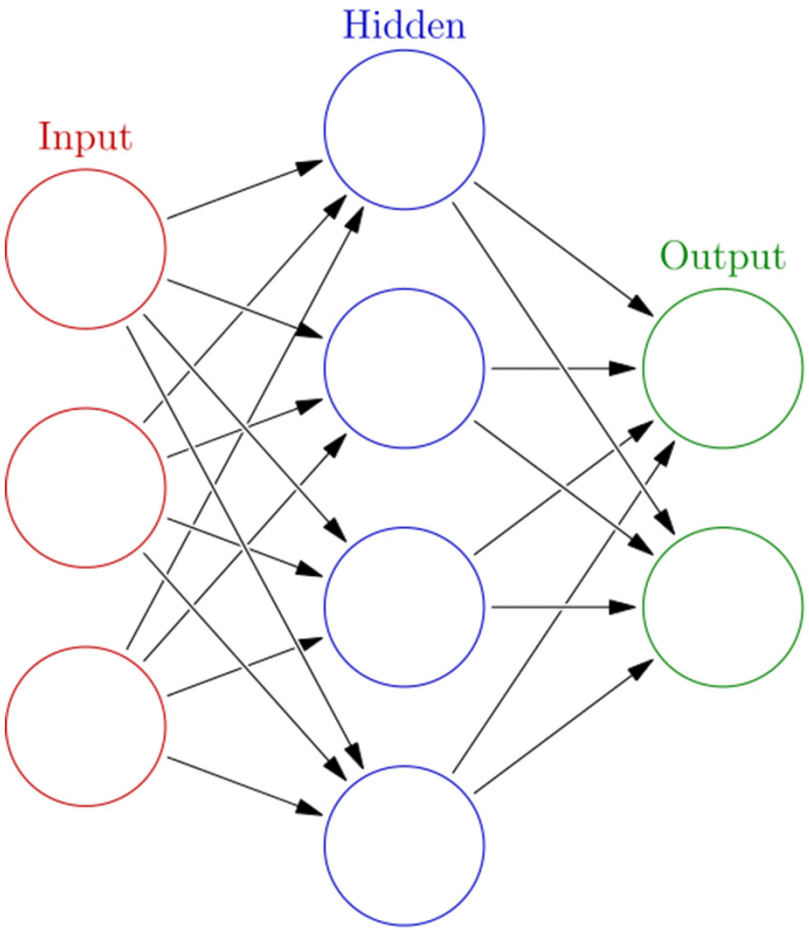




Off-Ramp Summary

- The DSM2 fingerprinting validation was considered adequate to use for ANN development

What are Artificial Neural Networks (ANNs)?

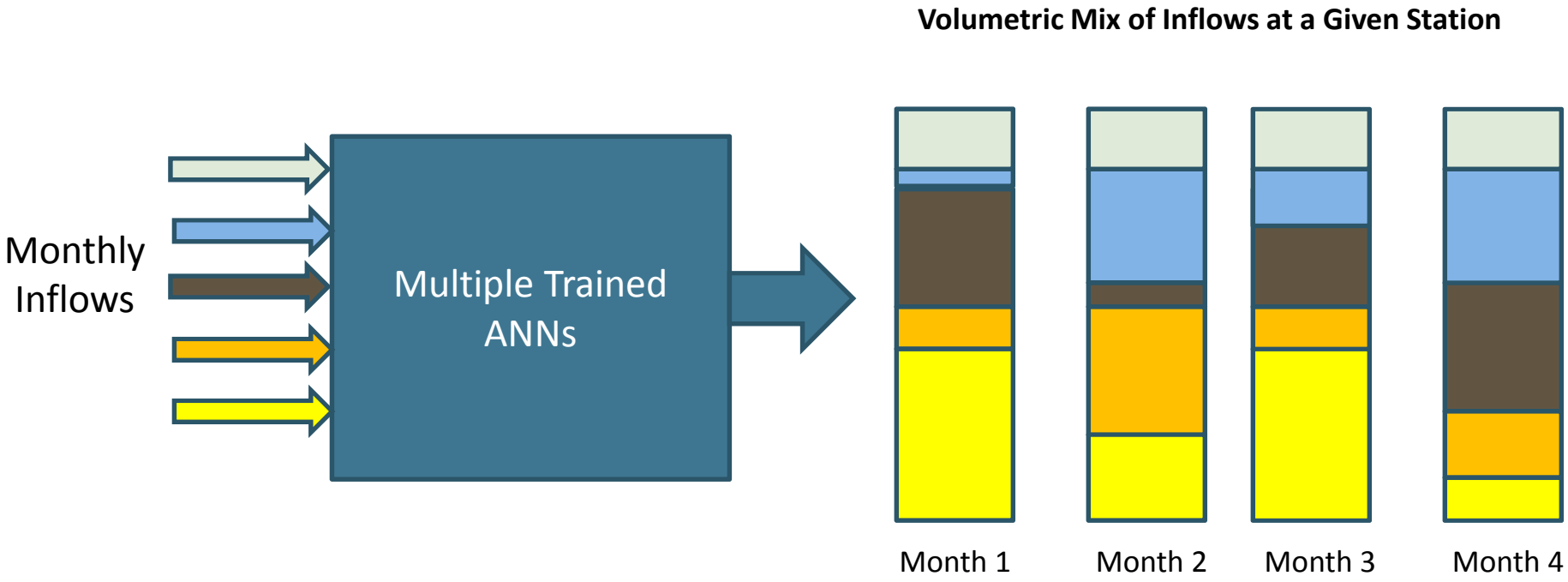


- Mathematical models inspired by biological neural networks
- Similar to natural systems, they can be "trained" using real-world data. In the context of ANNs, this includes varying inputs, hidden layers, and weights associated with individual nodes. This is analogous to model calibration.
- Over the last two decades, ANNs have been used for modeling complex relationships between inputs and outputs or to find patterns in data
- In a mathematical sense, often called universal emulators

Why develop an ANN model to emulate the fingerprint model?

- A generalized tool that can be applied without the expertise needed to run DSM2
- ANNs (developed in Matlab) can be exported to Excel or called as DLLs from other modeling platforms
- Can be embedded within other tools, that seek to optimize some planning objective, especially where the model needs to be called multiple times (say, 100's of calls). This is the current application of the salinity ANN within CALSIM.

What did we implement here?

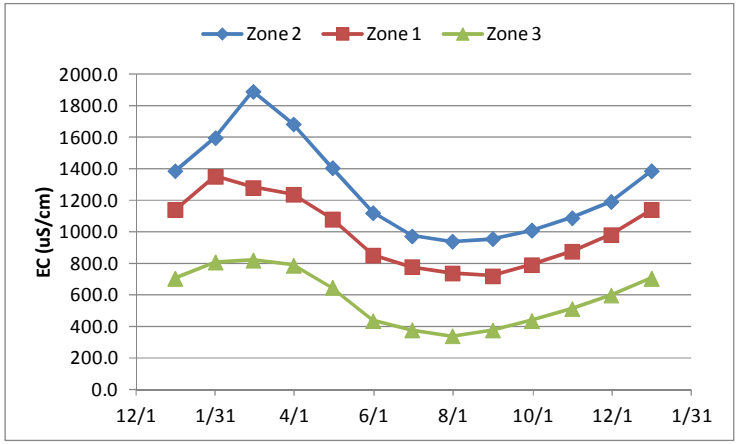


Boundary Inputs Used in ANN Development

- Sacramento River at Freeport (flow)
- San Joaquin River at Vernalis (flow)
- Mokelumne River at Benson's Ferry (flow)
- Calaveras River (flow)
- Yolo Bypass (flow)
- Martinez (flow)
- Delta agricultural return flows (divided into 9 subregions)
- Fifteen (15) possible inflows, we term the first 6 the primary flows, and the rest the DICU flows
- DCC gate status (0: fully closed; 1: one gate is fully open; 2: two gates are fully open) for selected stations

DICU Region Definition

- 3 EC levels for delta islands in DSM2 (3 EC DICU regions)
- 3 DOC levels for delta islands in DSM2 (3 DOC DICU regions)
- The combination of 3 EC regions and 3 DOC regions became 9 regions



Composite region	DOC region	EC region
1	1	1
2	2	1
3	3	1
4	1	2
5	2	2
6	3	2
7	1	3
8	2	3
9	3	3

Output Locations

	Model Output Location	Key Boundary Influence(s) Anticipated
1	Port Chicago	Martinez
2	Mallard/Chipps Island	Martinez
3	Collinsville	Martinez
4	Antioch	Martinez
5	Emmaton	Martinez; Freeport; Yolo
6	Rio Vista	Freeport; Yolo
7	Jersey Point	Martinez; Freeport; Vernalis; Ag Returns
8	Old River @ Bacon Island	Martinez; Freeport; Vernalis; Ag Returns
9	Old River @ Highway 4	Martinez; Freeport; Vernalis; Ag Returns
10	CCF Intake	Martinez; Freeport; Vernalis; Ag Returns
11	Jones Pumping Plant	Martinez; Freeport; Vernalis; Ag Returns
12	SJR @ Prisoner's Point	Martinez; Freeport; Vernalis; Mokelumne; Calaveras; Ag Returns
13	Middle River @ Holt	Martinez; Freeport; Vernalis; Mokelumne; Calaveras; Ag Returns
14	Middle River @ Victoria	Martinez; Freeport; Vernalis; Mokelumne; Ag Returns
15	Old River @ Tracy Rd Bridge	Vernalis; Ag Returns
16	Middle River @ Union Island	Vernalis; Ag Returns
17	SJR @ Hwy 4	Vernalis; Ag Returns; Calaveras

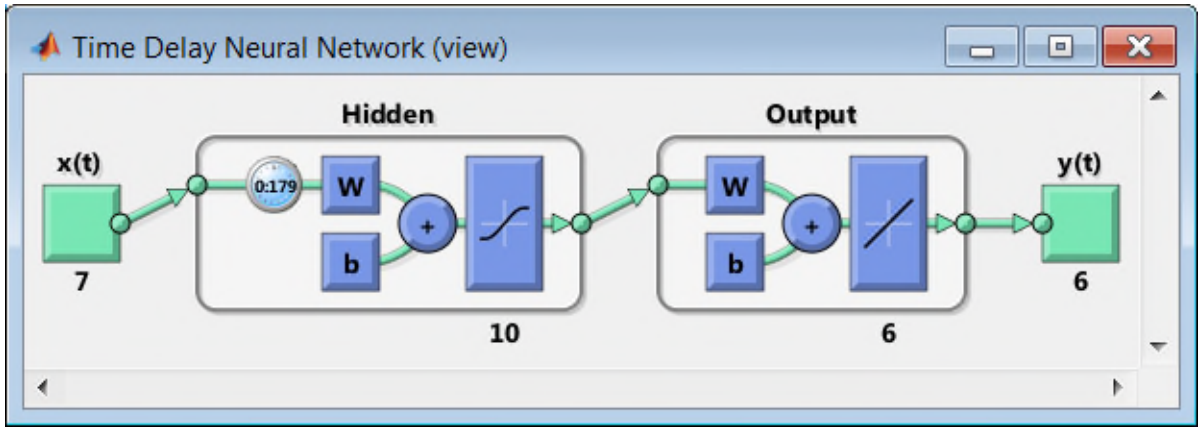
Multiple DSM2 Simulation Scenarios Were Considered

Run #	Ag Barriers	DCC	S. Delta Exports
1	Historical	Historical	Historical
2	In	Open	Historical
3	Out	Open	Historical
4	In	Closed	Historical
5	Out	Closed	Historical
6	Historical	Historical	None
7	In	Open	None
8	Out	Open	None
9	In	Closed	None
10	Out	Closed	None

ANN Modeling Goal

- Using time history of flow, for each of the flows, estimate volumetric contribution at selected output locations
- Thus, a user would provide an estimate of the last six months of the flow inputs, and the ANN would estimate the relative contribution of each of the input volumes (% from each source by month)
- For conservative constituents, the % volumetric contribution multiplied by inflow concentration provides estimates of concentrations at output stations.
- DSM2 model time frame provided the data for training: 1990-2010

ANN Model Structures



- 10 hidden neurons
- 180 days time delay
- Different input structures examined
- All possible inflows were considered for each output location, letting the ANN identify the more important flow contributions
- One ANN was developed for each station for five primary inflows (San Joaquin River at Vernalis, Mokelumne River, Calaveras River, Yolo Bypass, and Martinez); additional ANNs were developed for the Sacramento River for each month; and other ANNs for each DICU inflow.
- 187 ANNs for primary inflows (17 outputs X 11 flows)
- 153 ANNs for DICU inflows (17 outputs X 9 flows)

ANN Training Approach

- With the large dataset used in the training, the training was conducted separately for different stations and sources
- For a typical training, the inputs are time series data of flow from seven sources and the outputs are volumetric contribution from one source to one location at six time steps
- The Scaled Conjugated Gradient (SCG) approach was used for error minimization during training

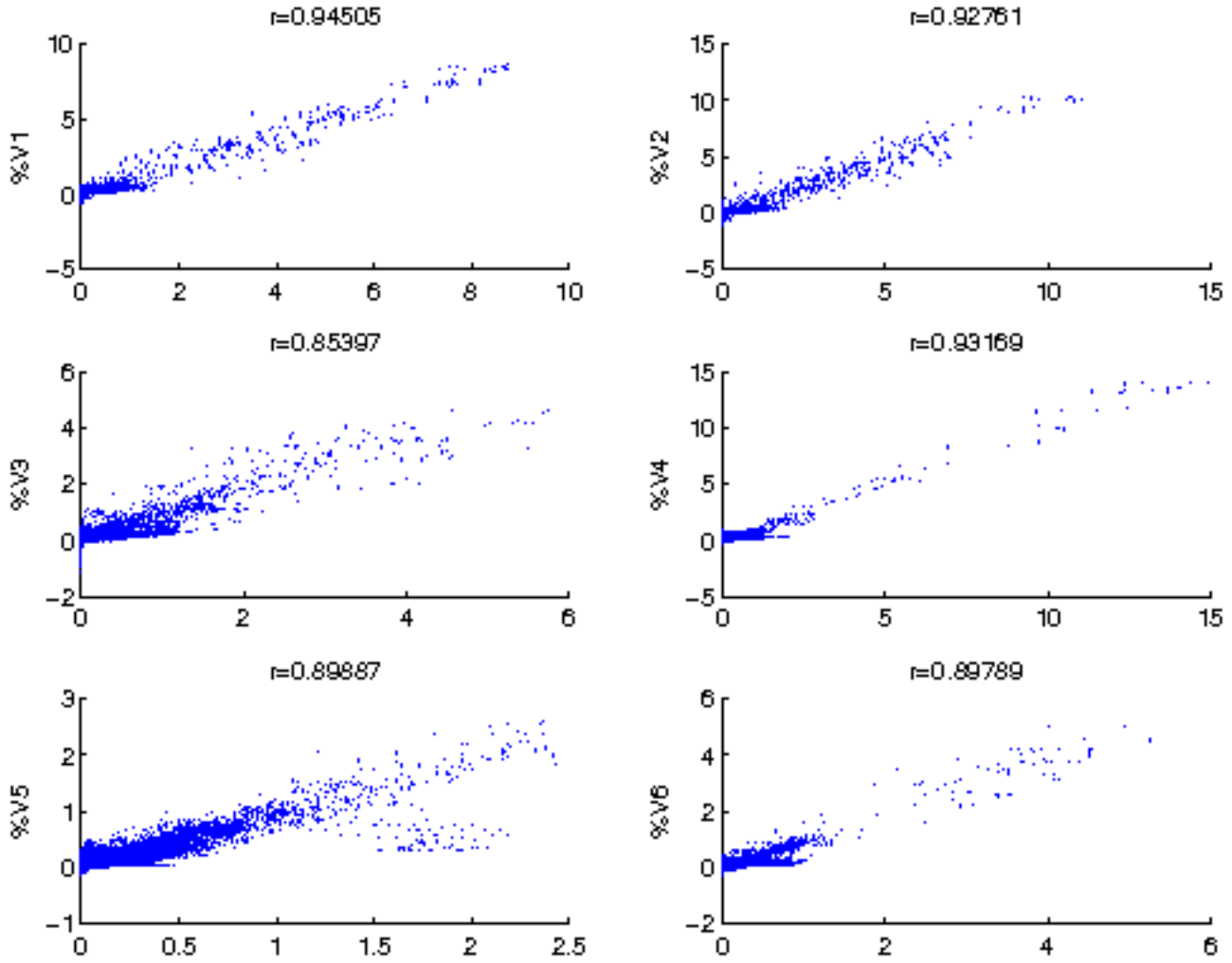
ANN Model Interface

- There were 340 ANNs developed for the 17 stations (20 ANNs each station, 1 ANN for each flow component except for SAC)
- The ANNs were developed using Matlab 2011b.
- The ANNs were wrapped into 17 DLLs (1 DLL for each station)
- A user interface was created to run these ANN models through Excel, through specifying flow and concentration inputs at boundary
- The Excel interface calls the corresponding DLL to calculate flow and concentration outputs at the user selected locations

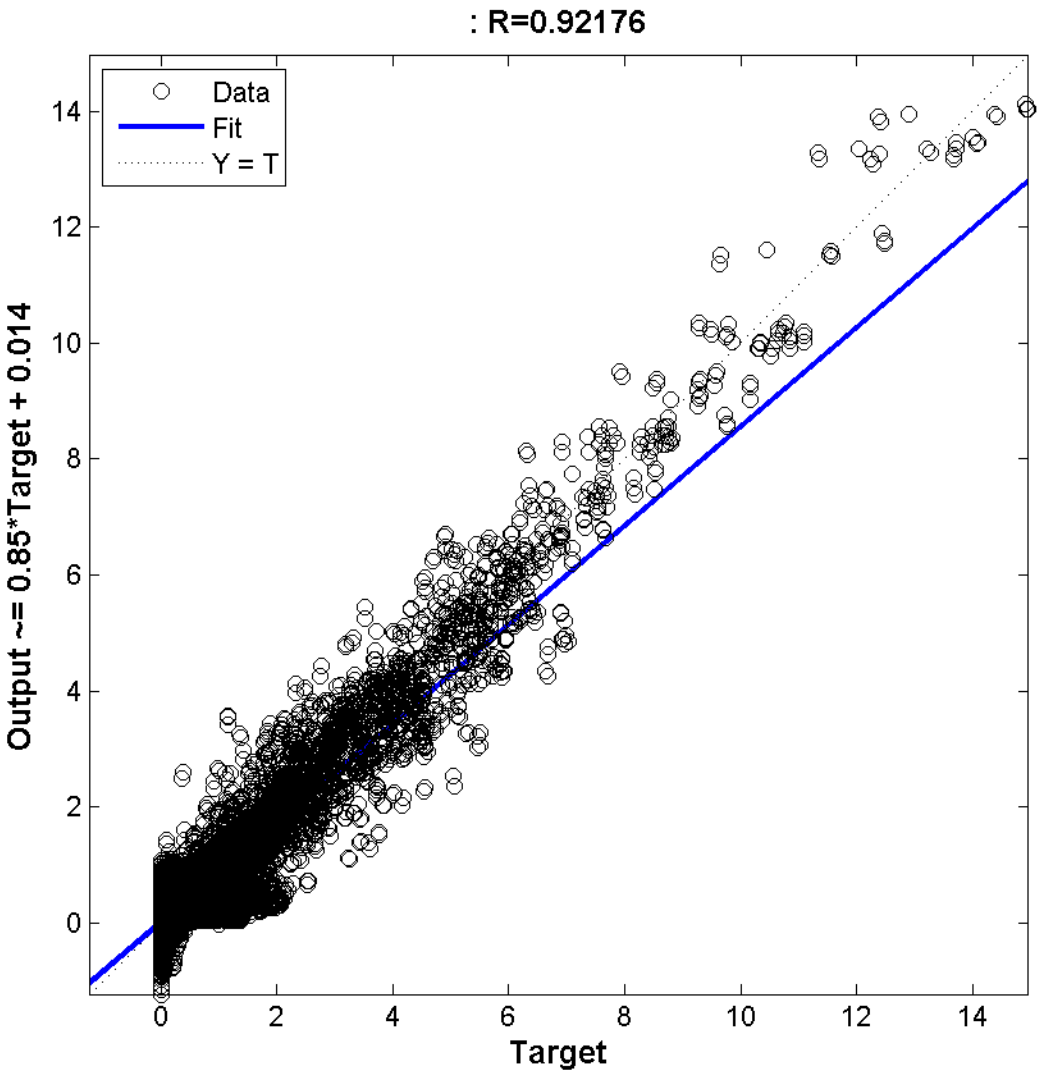
Representative Examples

- We show the training at one station: Antioch
- Plots are shown for the major tributaries, comparing the DSM2-calculated versus the ANN-calculated flow contributions
- For each inflow, there are six possible contributions, from 1 to 6 months prior to the current time
- The model application is shown for individual conservative constituents (DOC, Br, EC) at different stations throughout the Delta

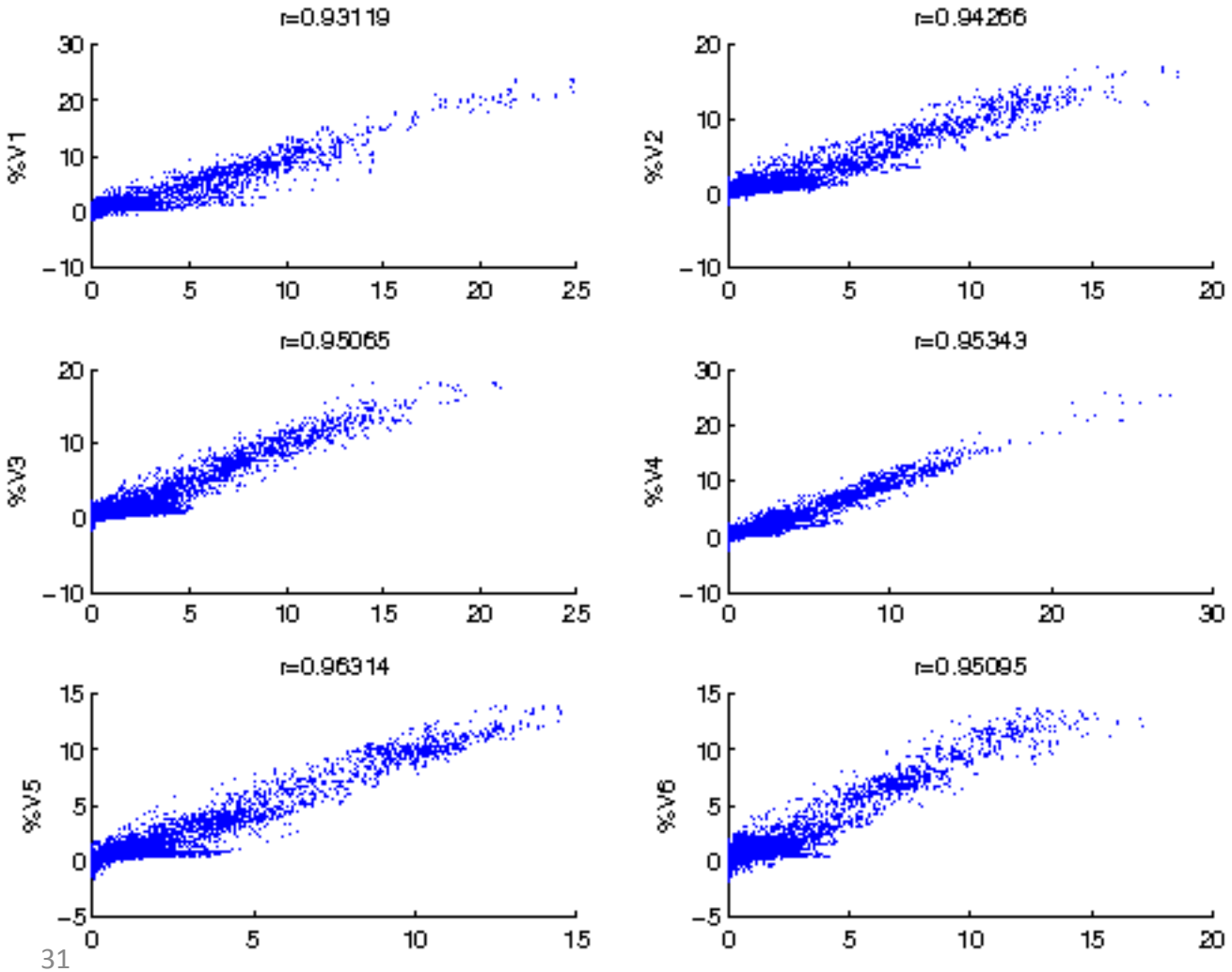
Results at Antioch: Contribution from Calaveras River (Each plot represents one month)



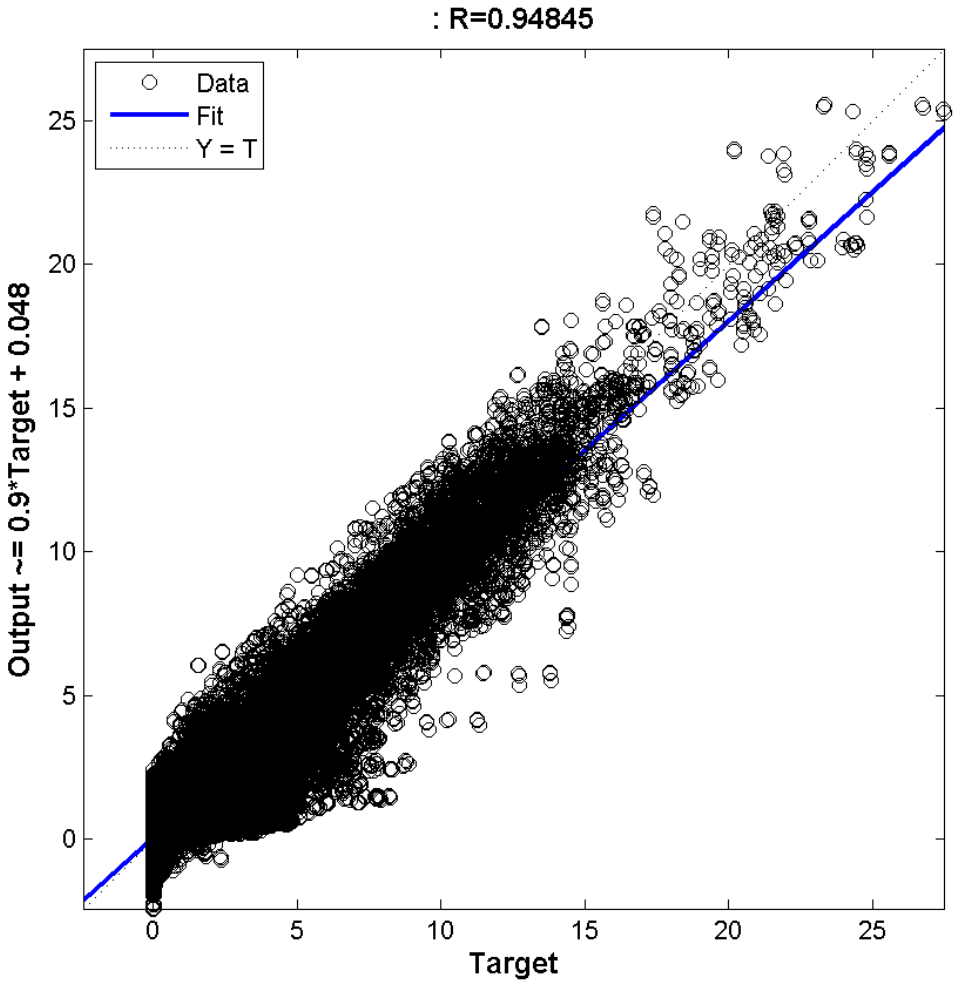
Results at Antioch: Contribution from Calaveras River (All months combined)



Results at Antioch: Contribution from Mokelumne River (Each plot represents one month)



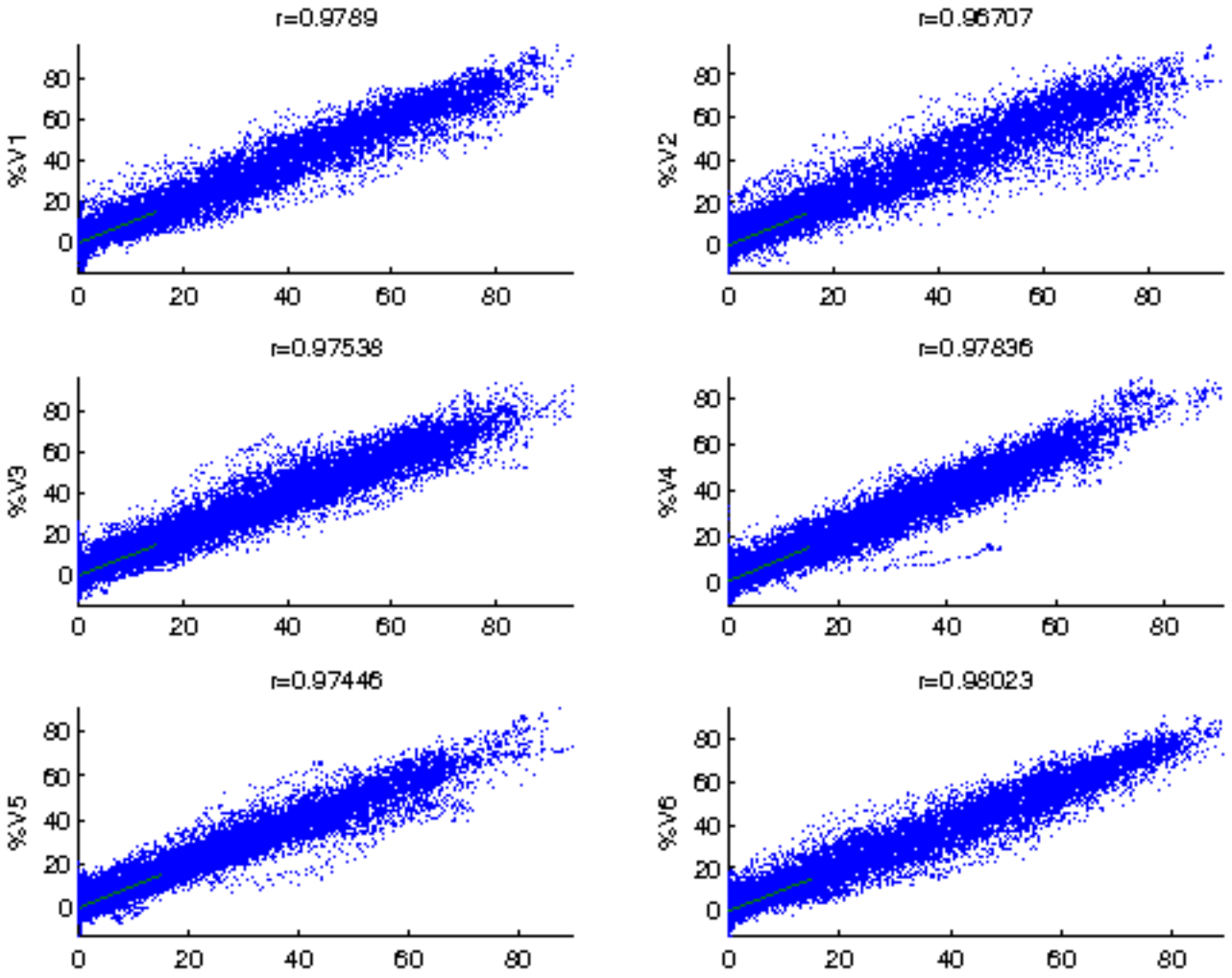
Results at Antioch: Contribution from Mokelumne River (All months combined)



Volumetric Contribution from ANN (%)

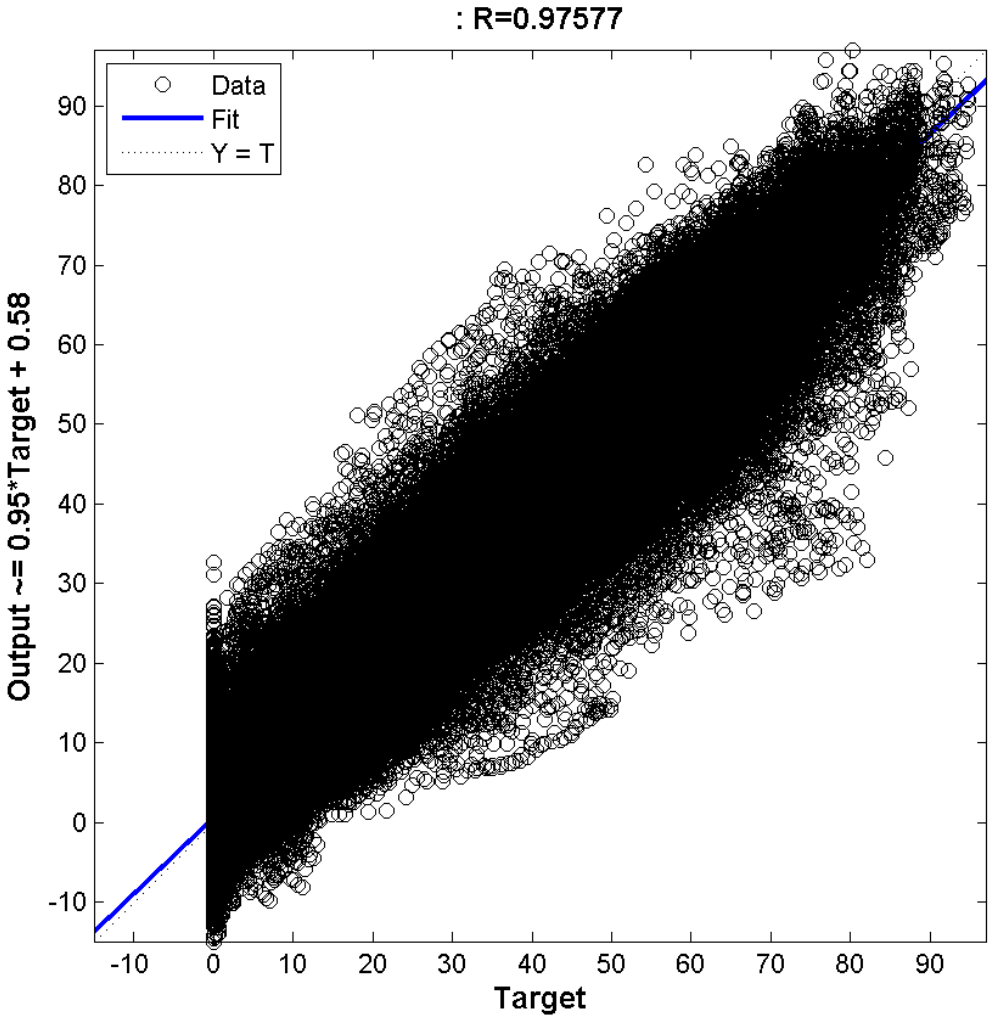
Volumetric Contribution from DSM2 (%)

Results at Antioch: Contribution from Sacramento River (Each plot represents one month)*



*with DCC gate inputs

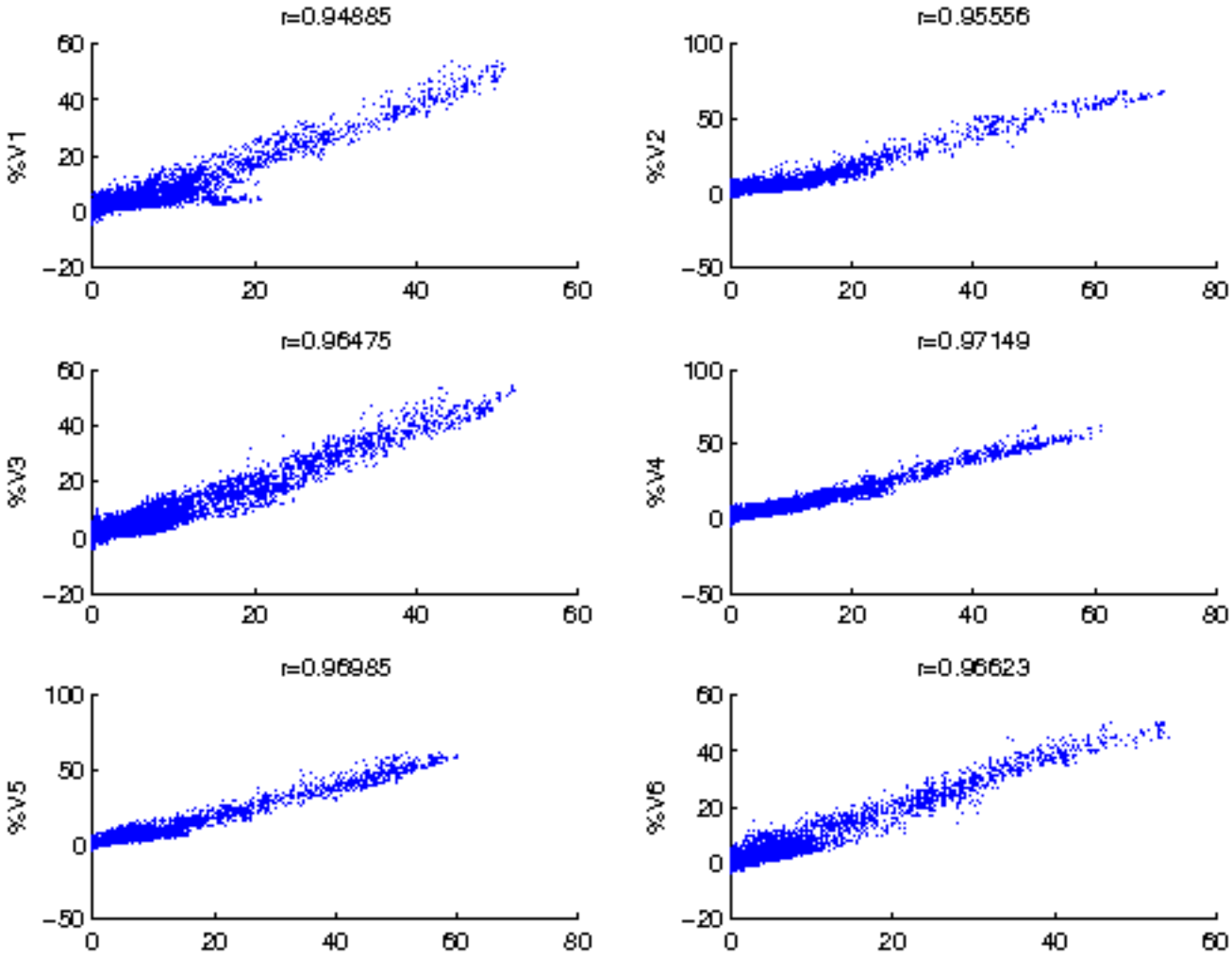
Results at Antioch: Contribution from Sacramento River (All months combined)



Volumetric Contribution from ANN (%)

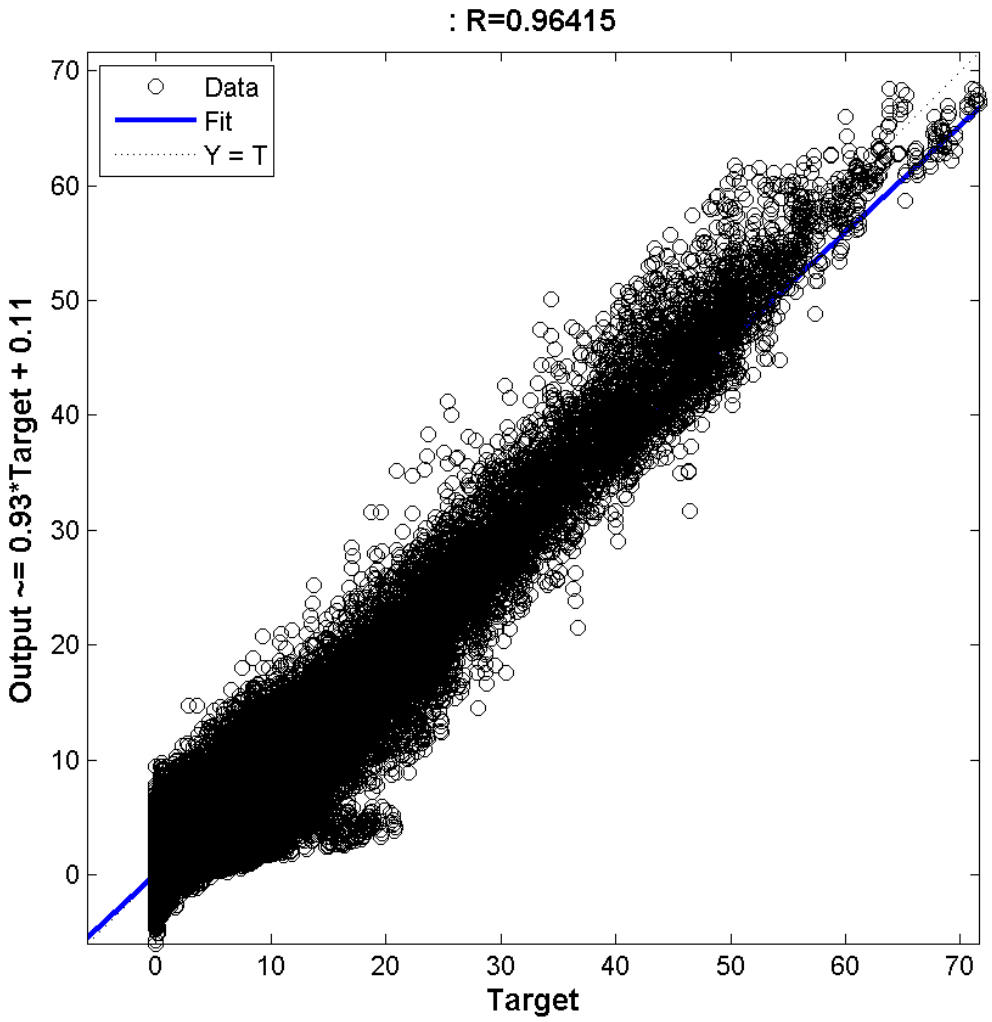
Volumetric Contribution from DSM2 (%)

Results at Antioch: Contribution from San Joaquin River (Each plot represents one month)*



*with DCC gate inputs

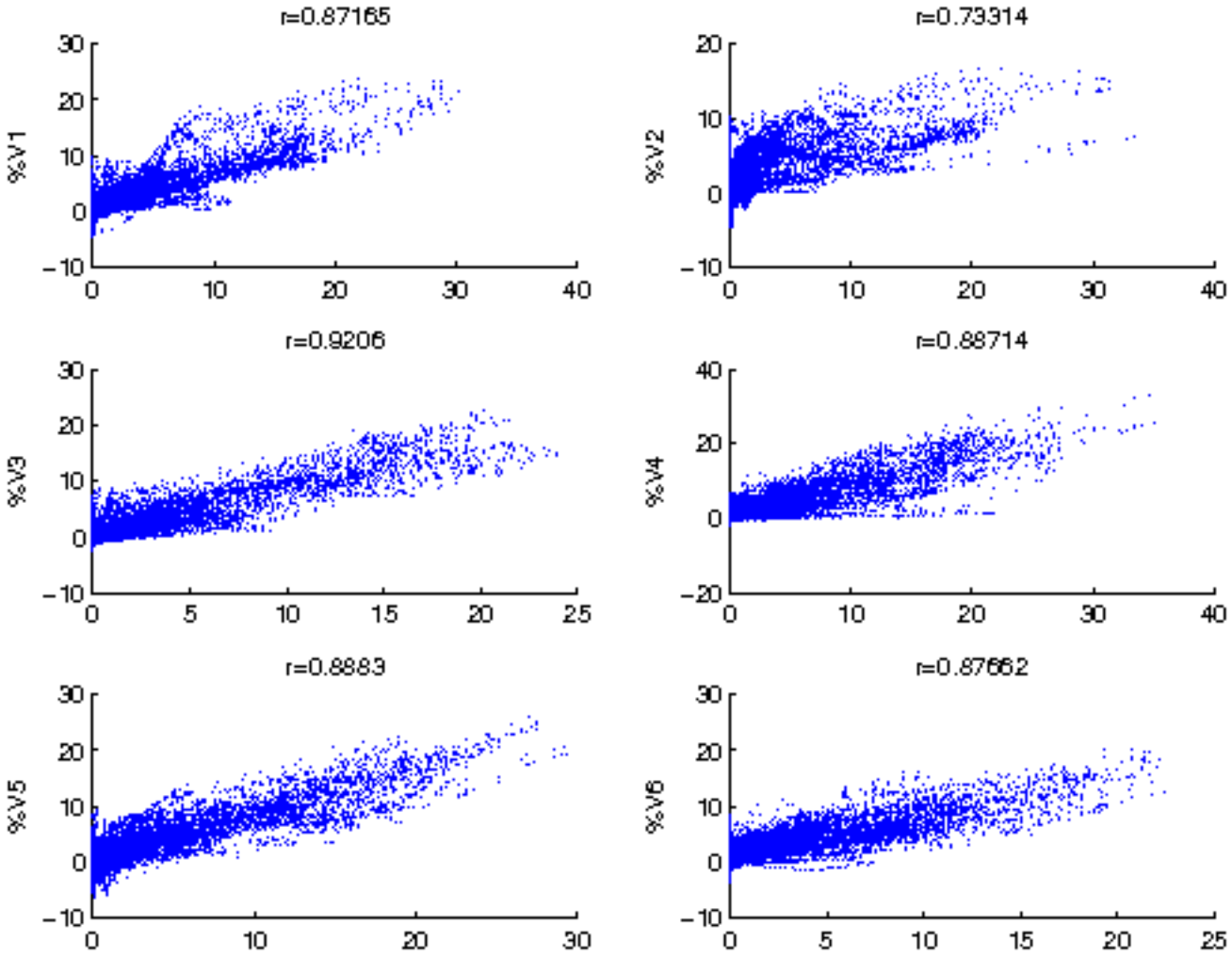
Results at Antioch: Contribution from San Joaquin River (All months combined)



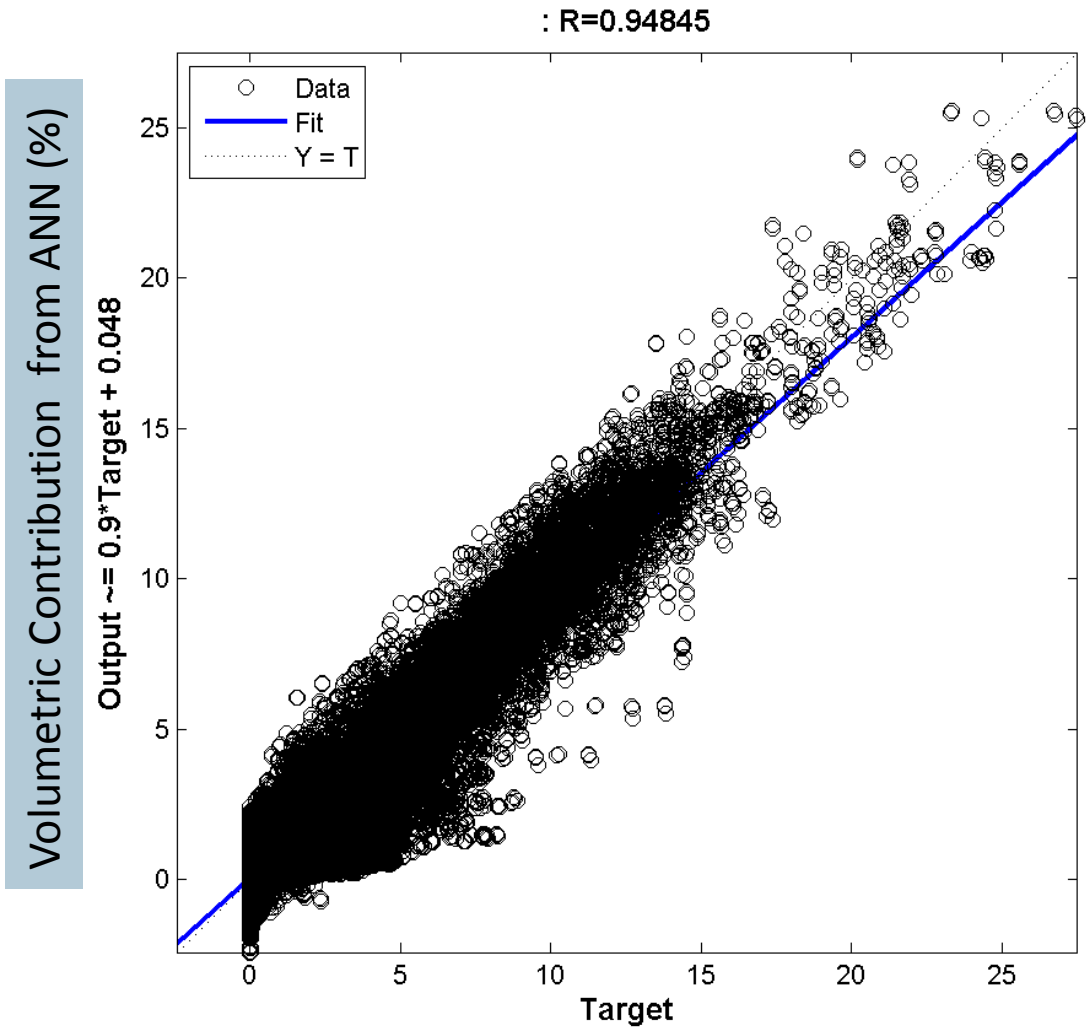
Volumetric Contribution from ANN (%)

Volumetric Contribution from DSM2 (%)

Results at Antioch: Contribution from Martinez (Each plot represents one month)



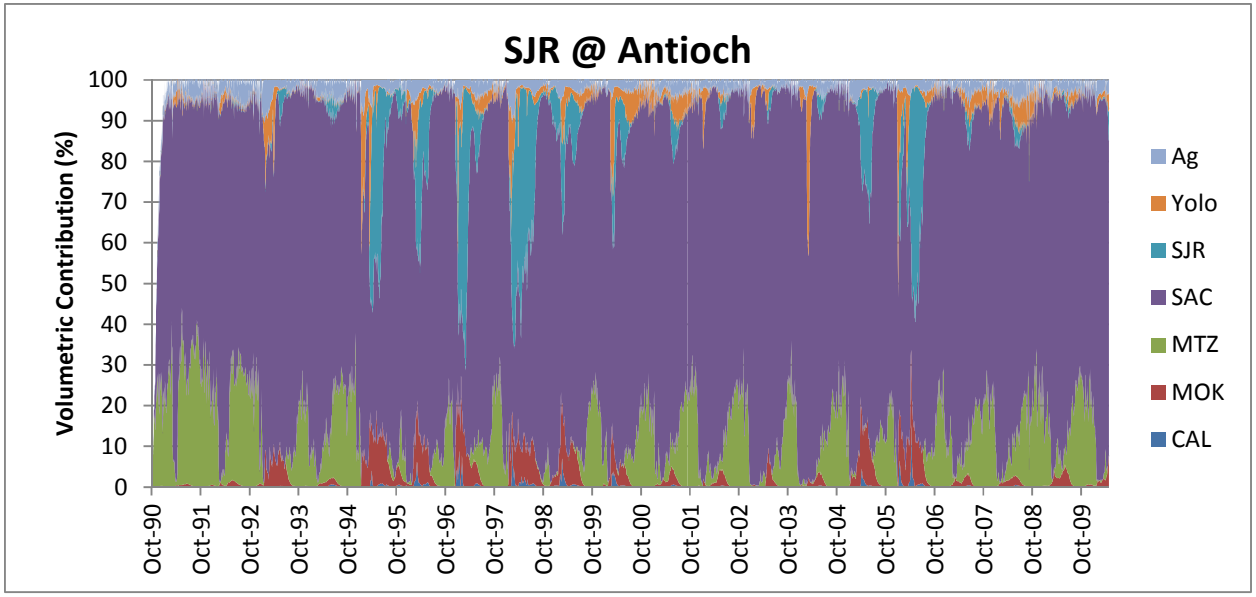
Results at Antioch: Contribution from Martinez (All months combined)

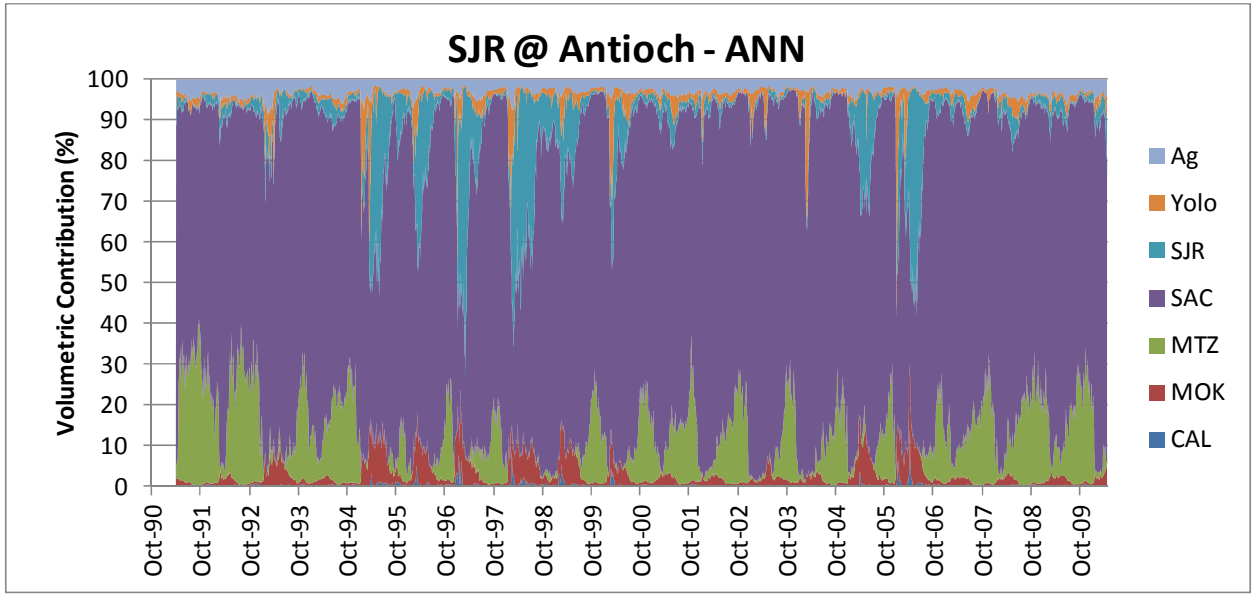


Volumetric Contribution from ANN (%)

Volumetric Contribution from DSM2 (%)

Model Comparison ANN vs. DSM2





Lessons Learned

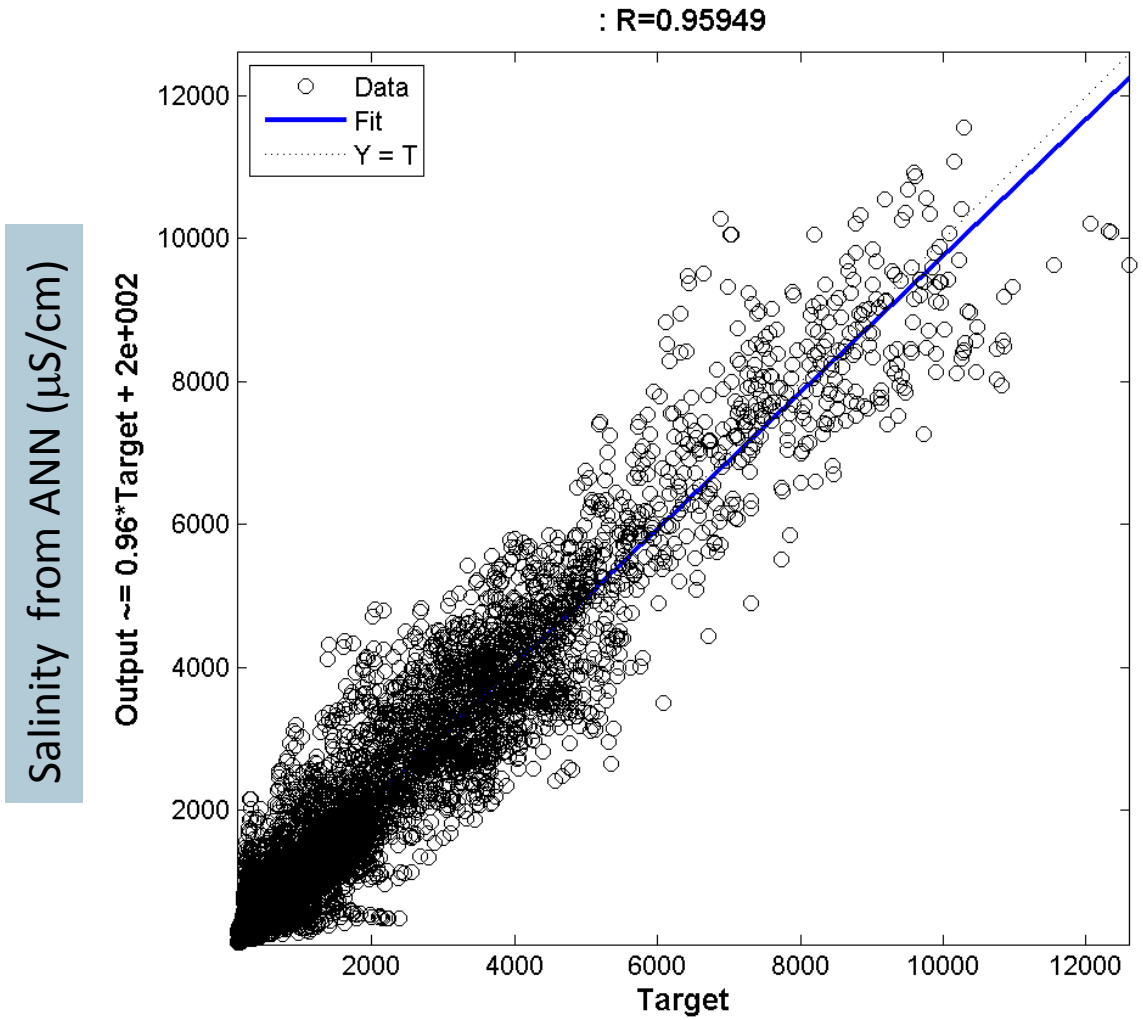
- The fingerprint process is complex, given the number of inflows and the time history for each flow
- Multiple ANNs were required to best capture the relationships; 340 ANNs being used
- Correlations for flow were reasonably good across most stations on the San Joaquin and Sacramento Rivers (>0.9), but not as good for Old River and Middle River stations (i.e., were rarely >0.95, good fits were considered to be in the >0.85 range)
- The consideration of the DCC gate status improved fits at several stations and was made part of the input structure
- The ANN application to estimate concentrations for three constituents showed very good results across 16 stations that were targeted in this work; correlations were better than for flow, and ranged from 0.90 to 0.99
- At least one station was not adequately fit using this approach (Middle River at Union Island)

Application Examples

ANN Application to Compute Concentrations

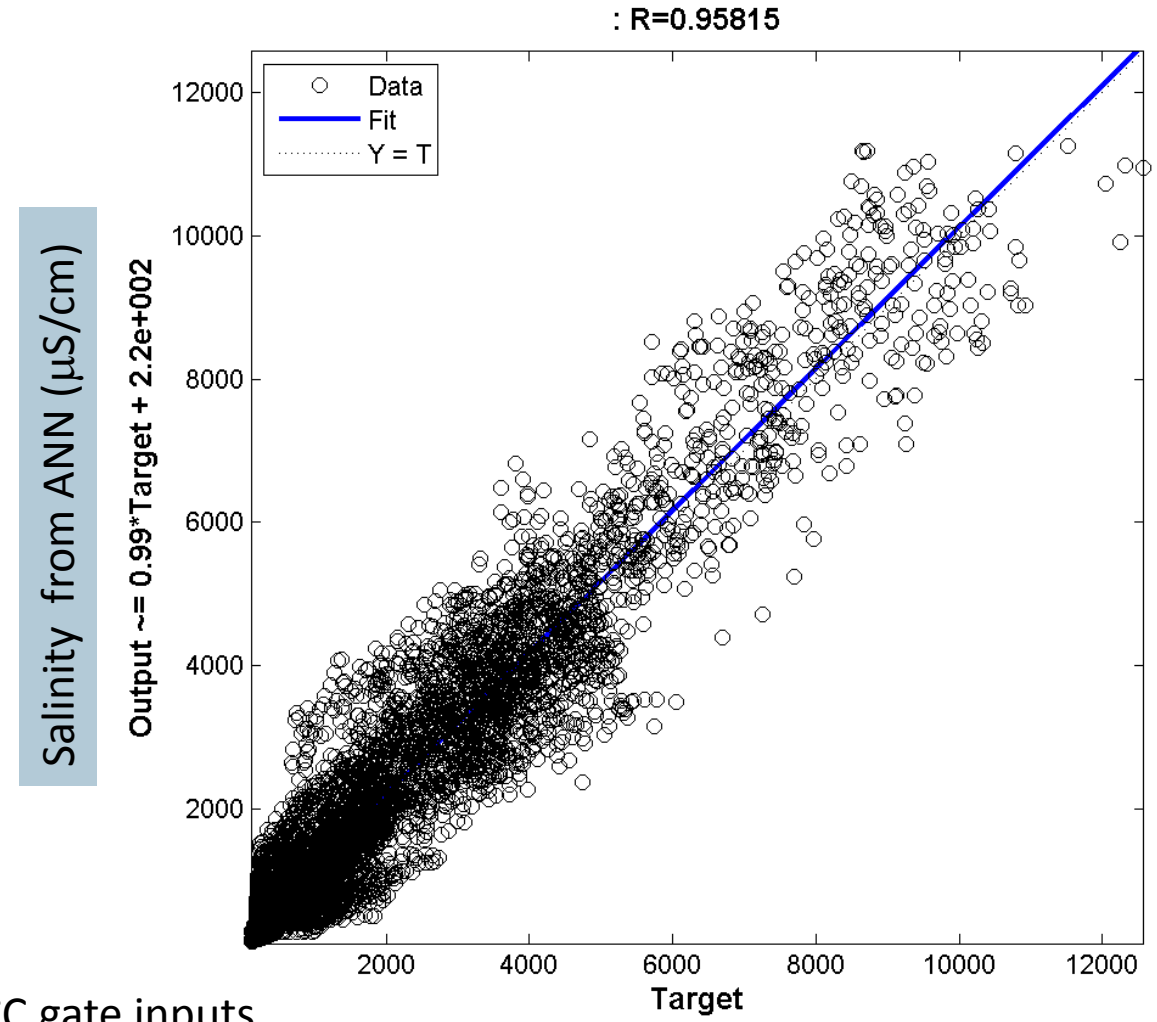
- The ANN-simulated volumetric contribution from different sources, along with conservative constituent concentrations at boundaries, was used to estimate concentrations at different locations within the Delta
- Three conservative constituents were considered:
  - Electrical conductivity (EC)
  - Bromide (Br)
  - Organic carbon (OC)
- The estimated concentrations from the fingerprint ANN were compared to values estimated from the DSM2-simulated volumetric contribution
- The results suggest strong agreement between the two modeling approaches
- Three stations are shown for illustration: Antioch, San Joaquin River at Hwy 4, and Port Chicago

ANN Application to Predict EC at Antioch*



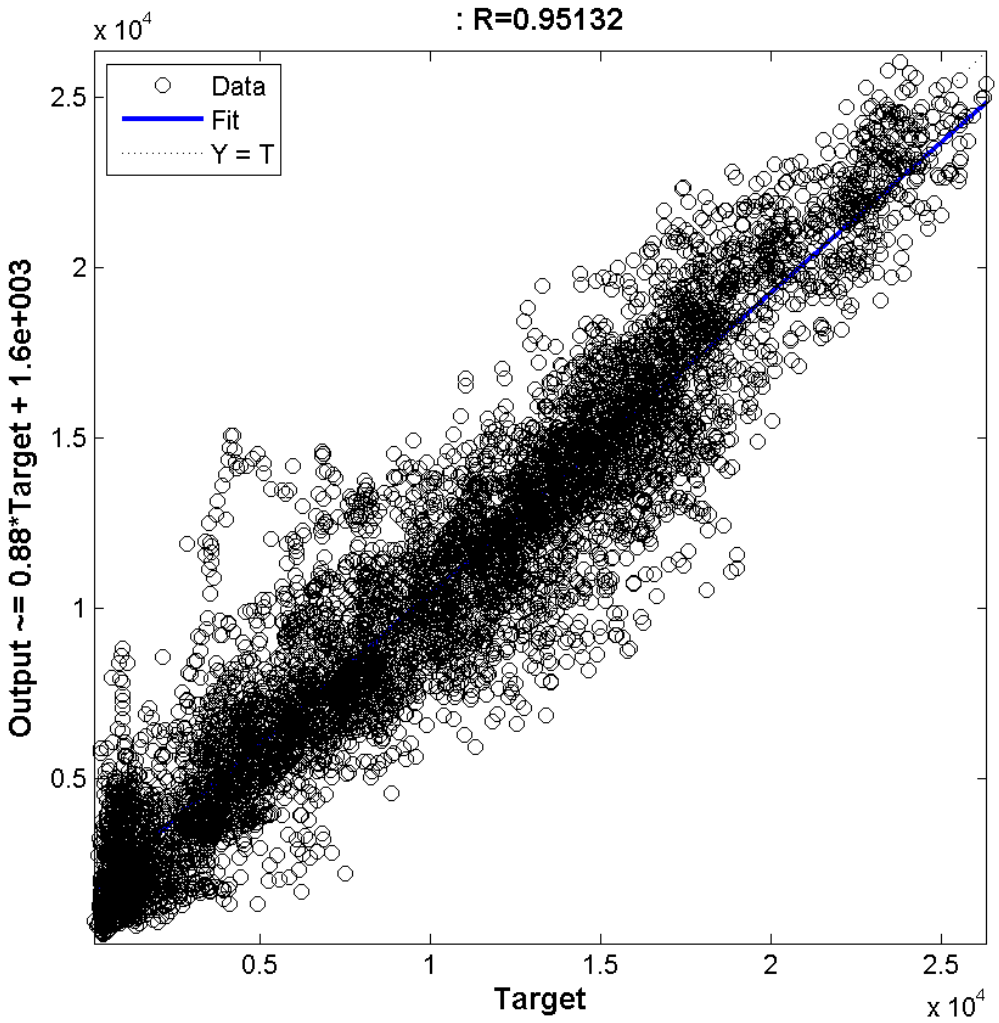
*with DCC gate inputs

ANN Application to Predict EC at SJR @ HWY4*



*with DCC gate inputs

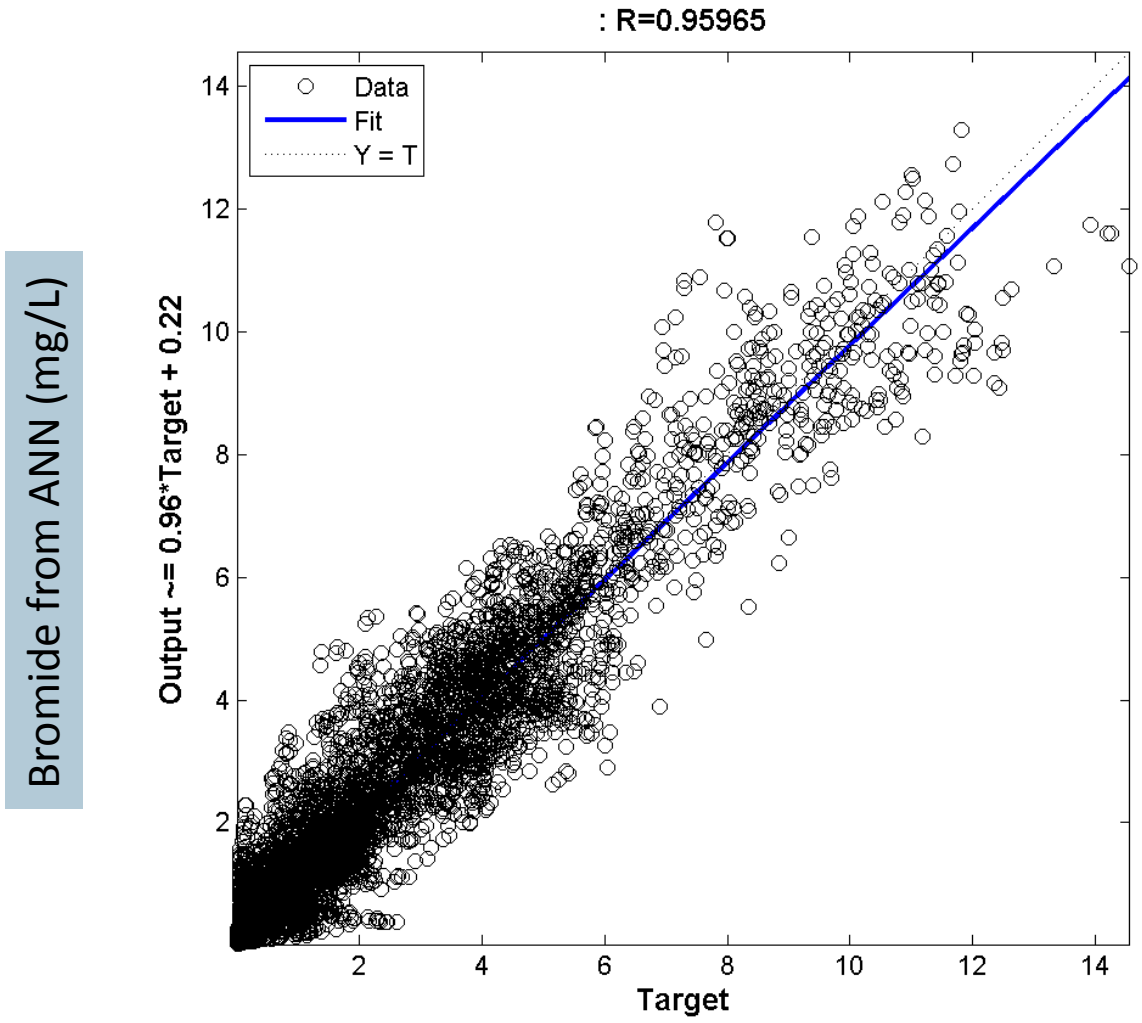
ANN Application to Predict EC at Port Chicago



Salinity from ANN (µS/cm)

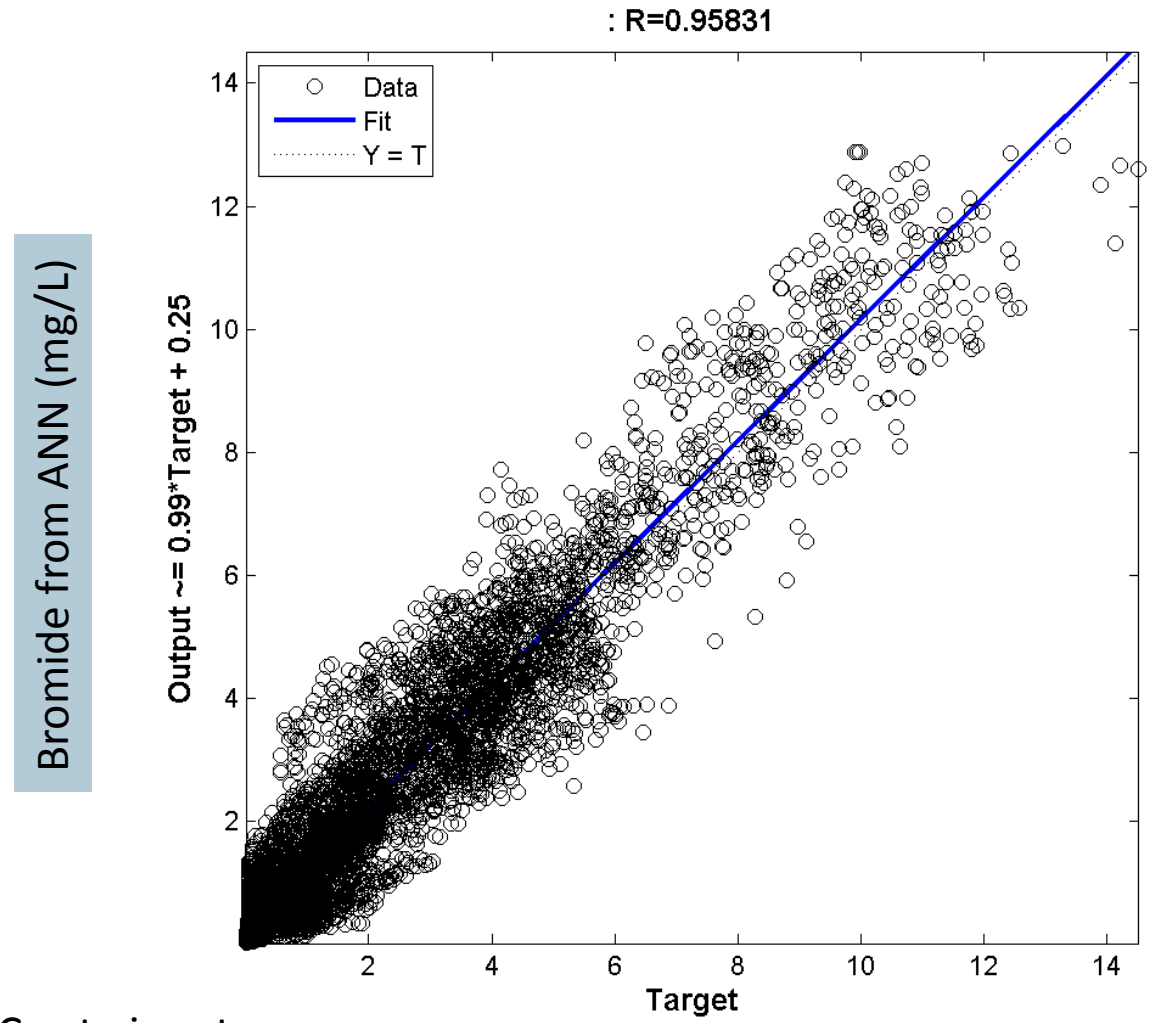
Salinity from DSM2 (µS/cm)

ANN Application to Predict Br at Antioch*



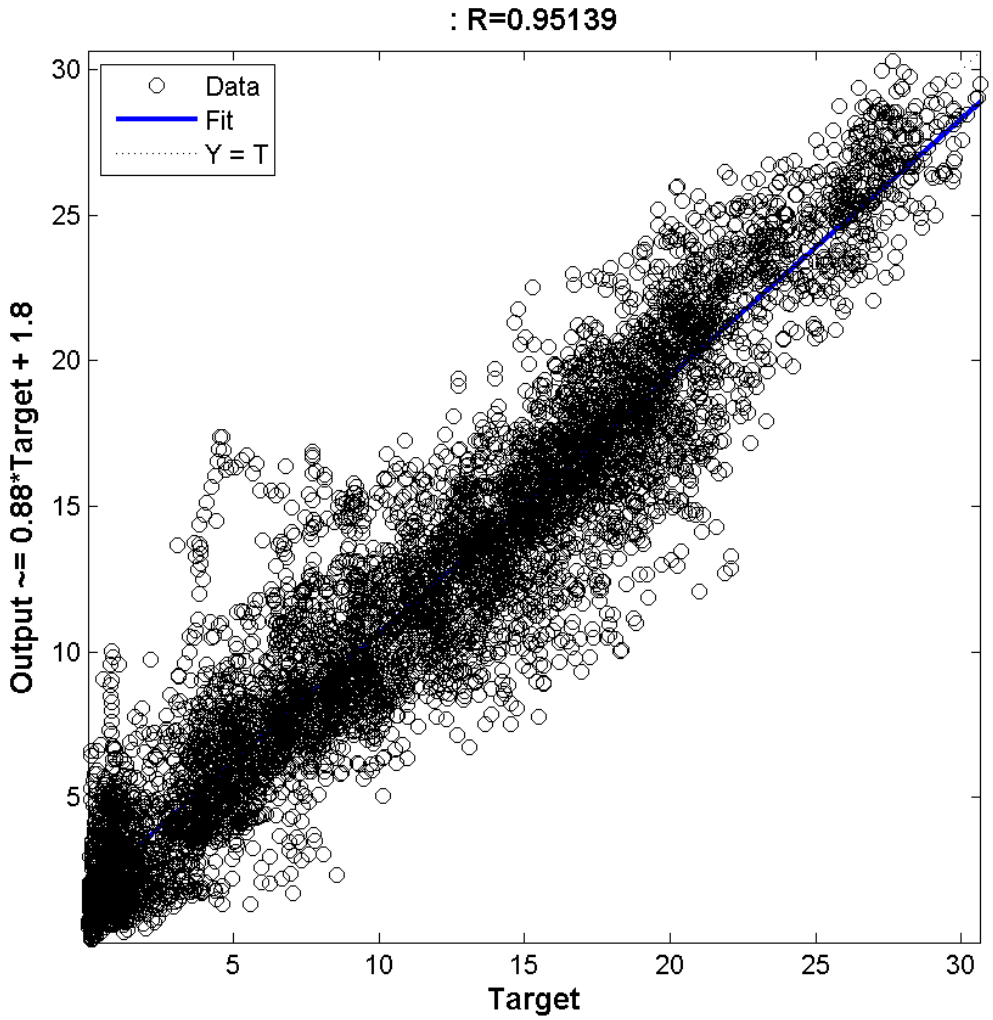
*with DCC gate inputs

ANN Application to Predict Br at SJR @ HWY4*



*with DCC gate inputs

ANN Application to Predict Br at Port Chicago

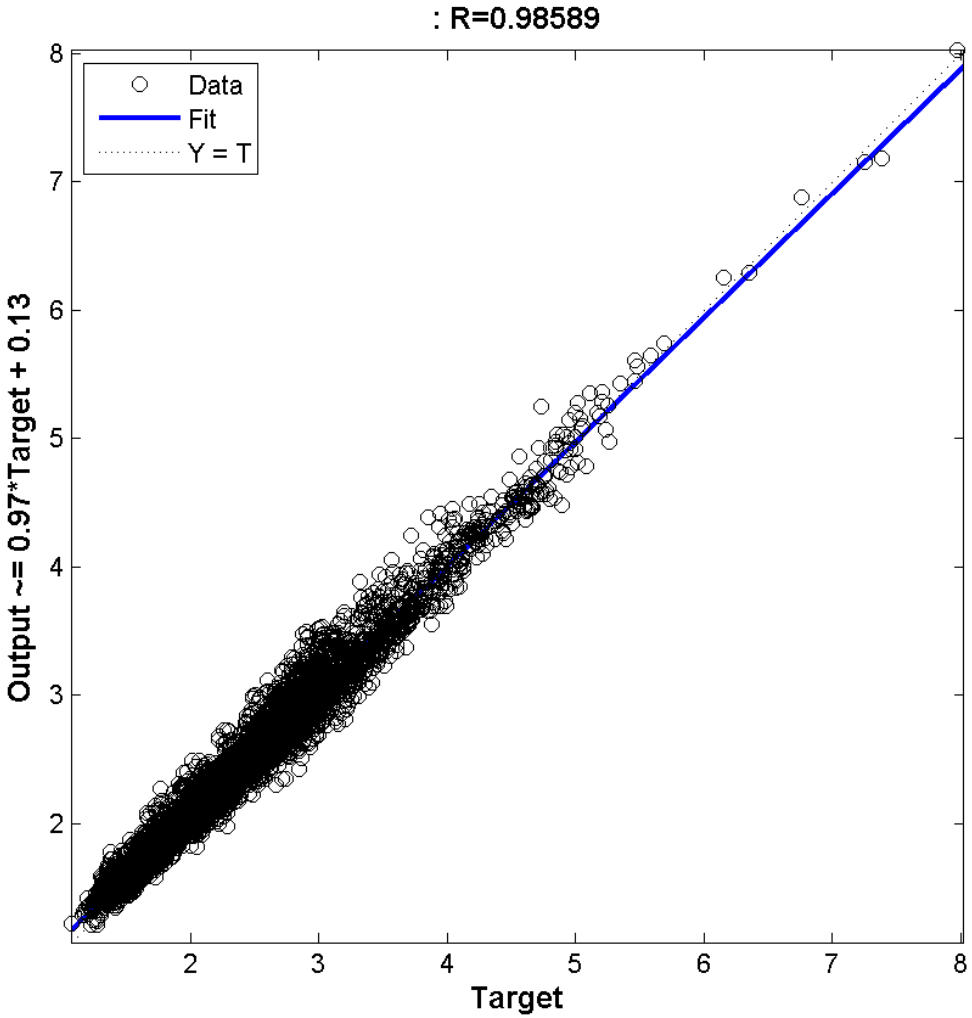


Bromide from ANN (mg/L)

Bromide from DSM2 (mg/L)

ANN Application to Predict DOC at Antioch*

DOC from ANN (mg/L)

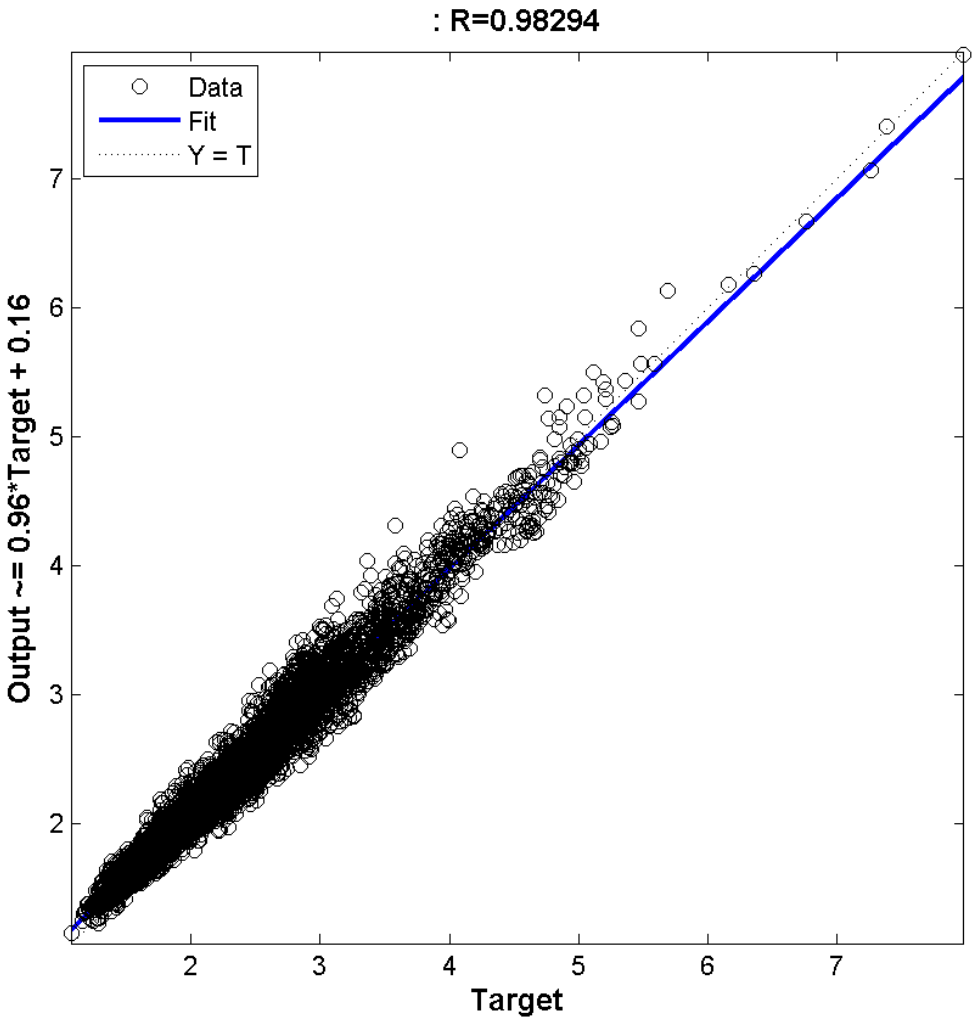


DOC from DSM2 (mg/L)

*with DCC gate inputs

ANN Application to Predict DOC at SJR @ HWY4*

DOC from ANN (mg/L)

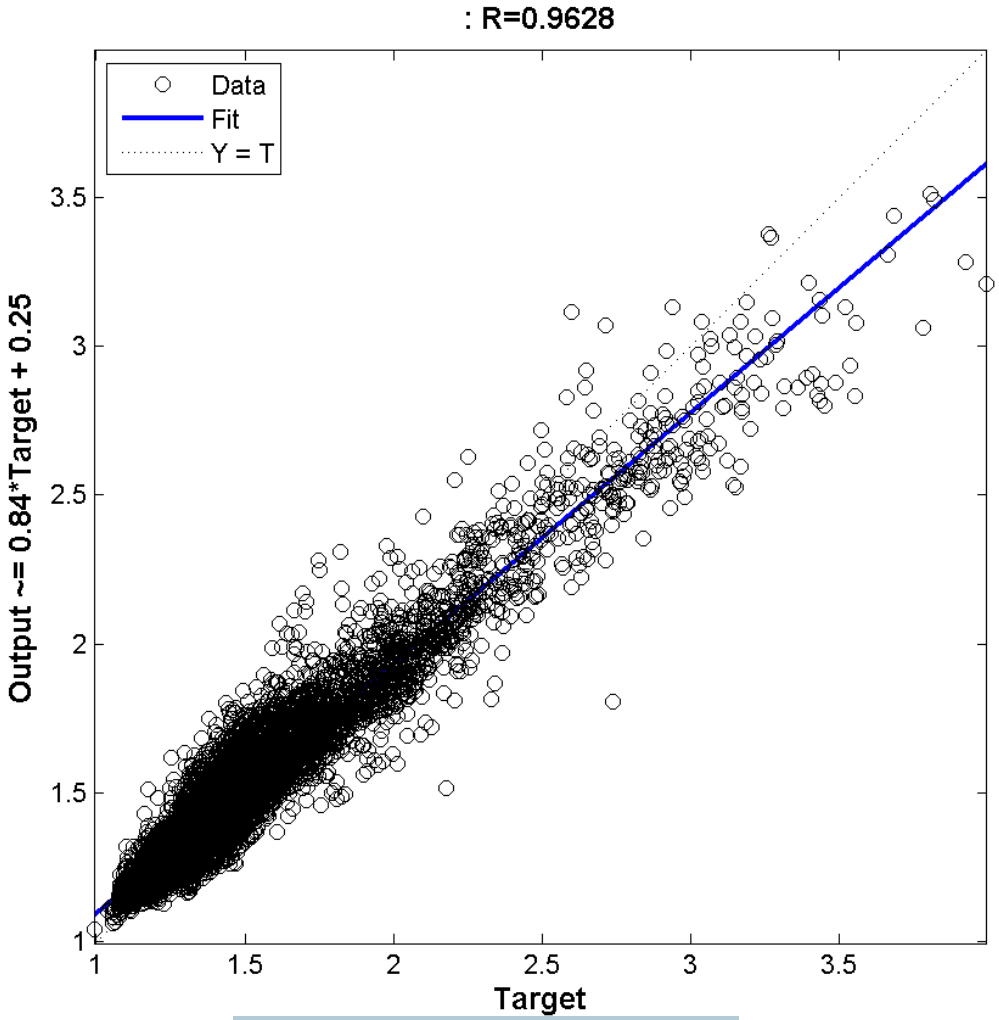


*with DCC gate inputs

DOC from DSM2 (mg/L)

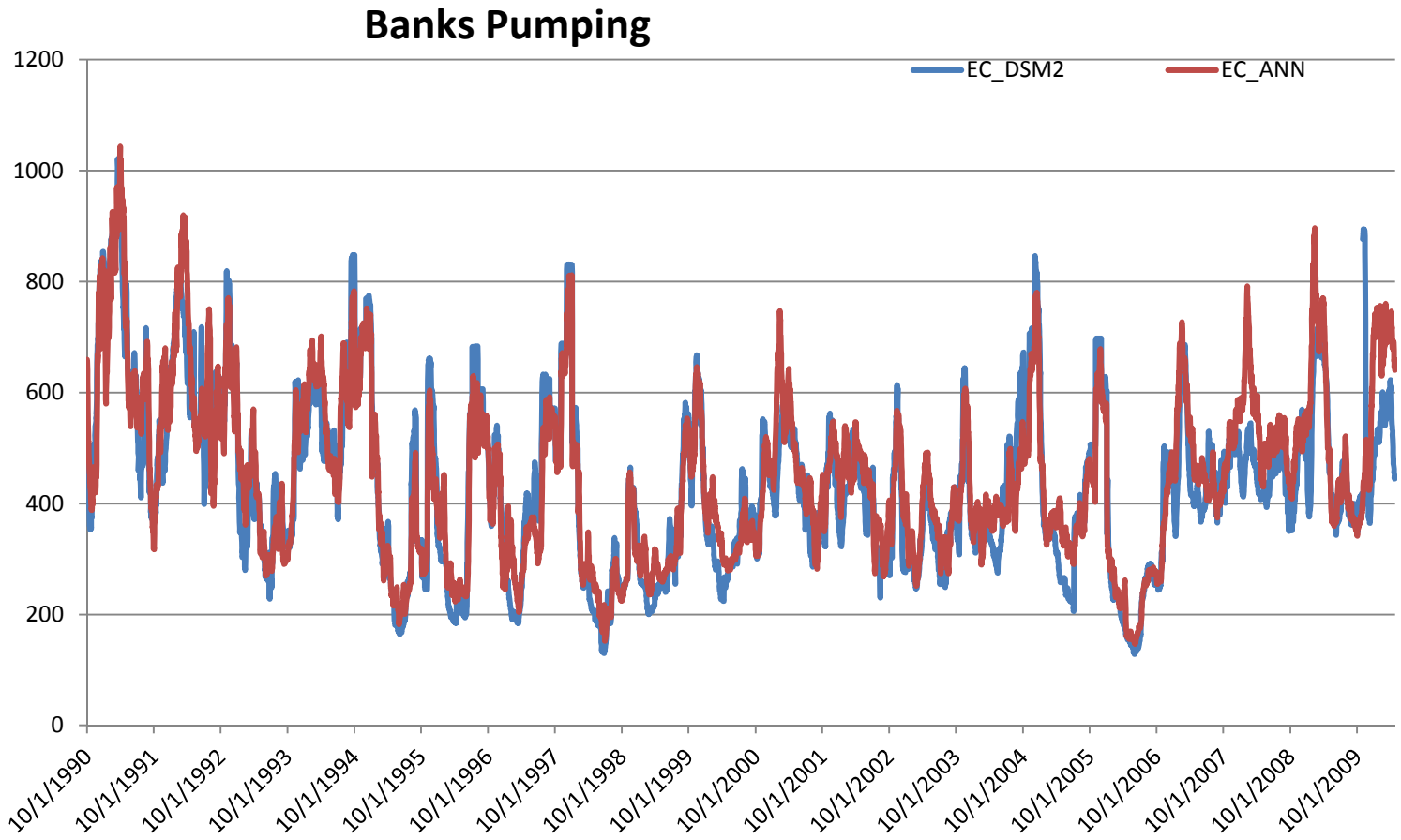
ANN Application to Predict DOC at Port Chicago

DOC from ANN (mg/L)

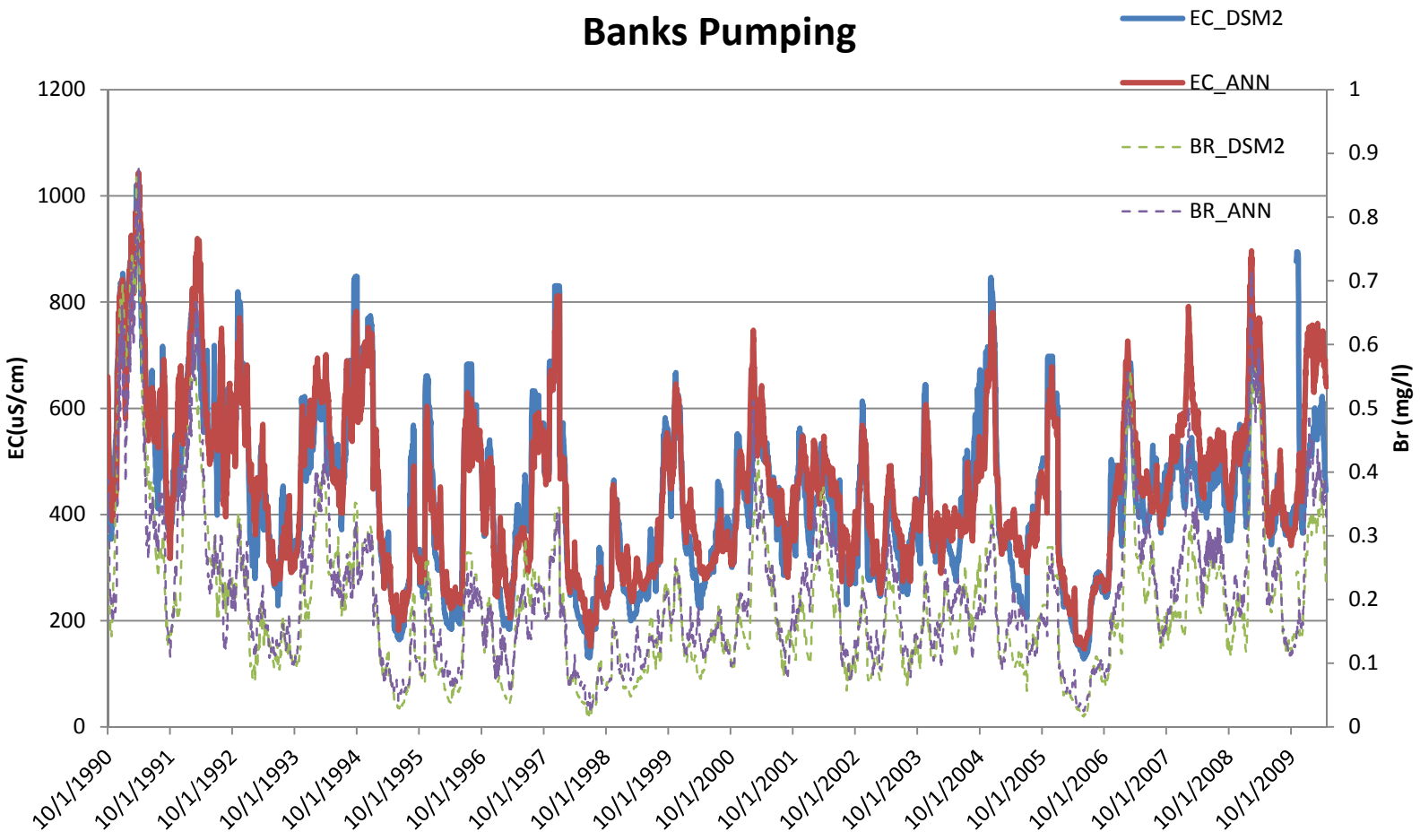


DOC from DSM2 (mg/L)

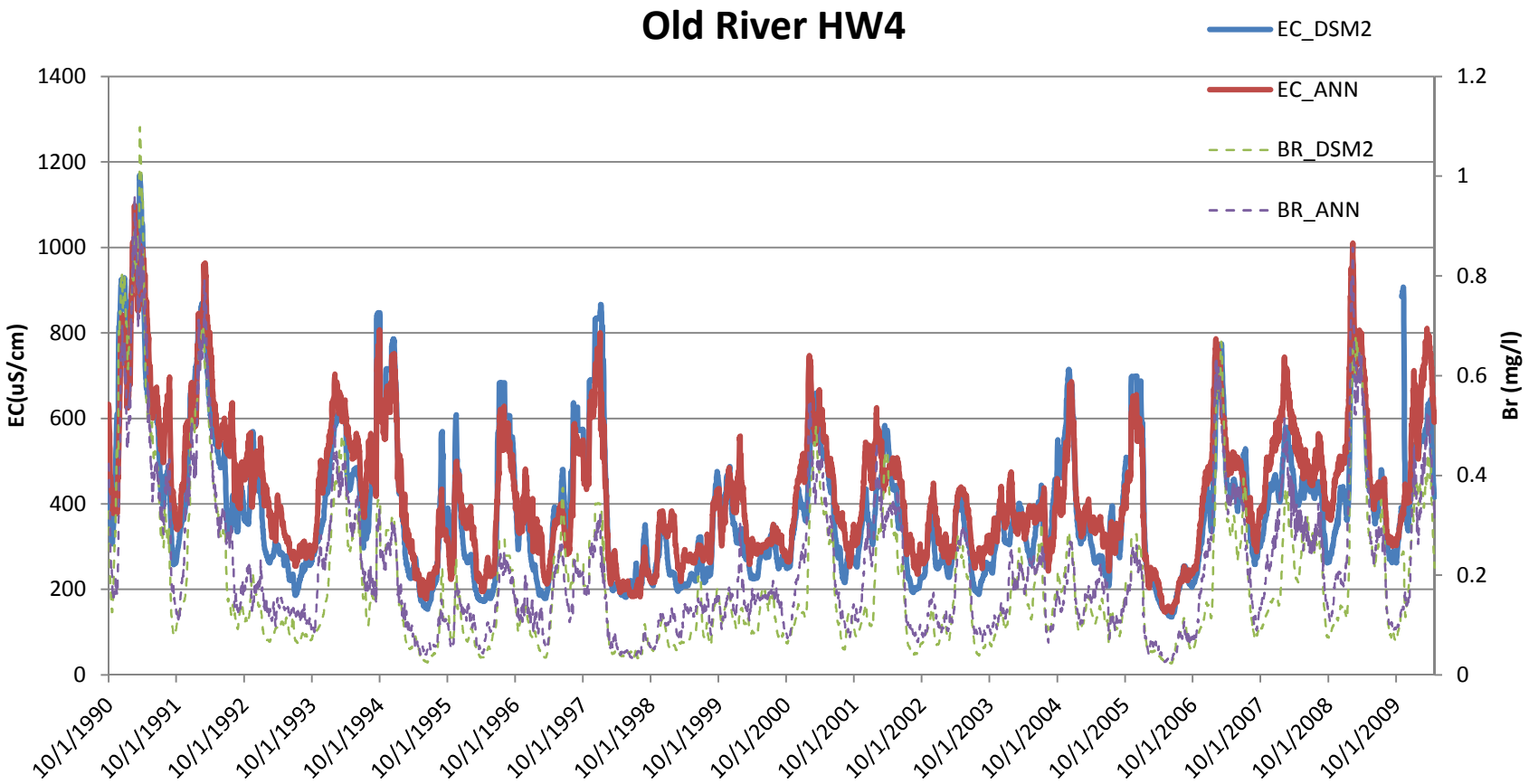
Time Series Example: Banks Pumping Plant EC



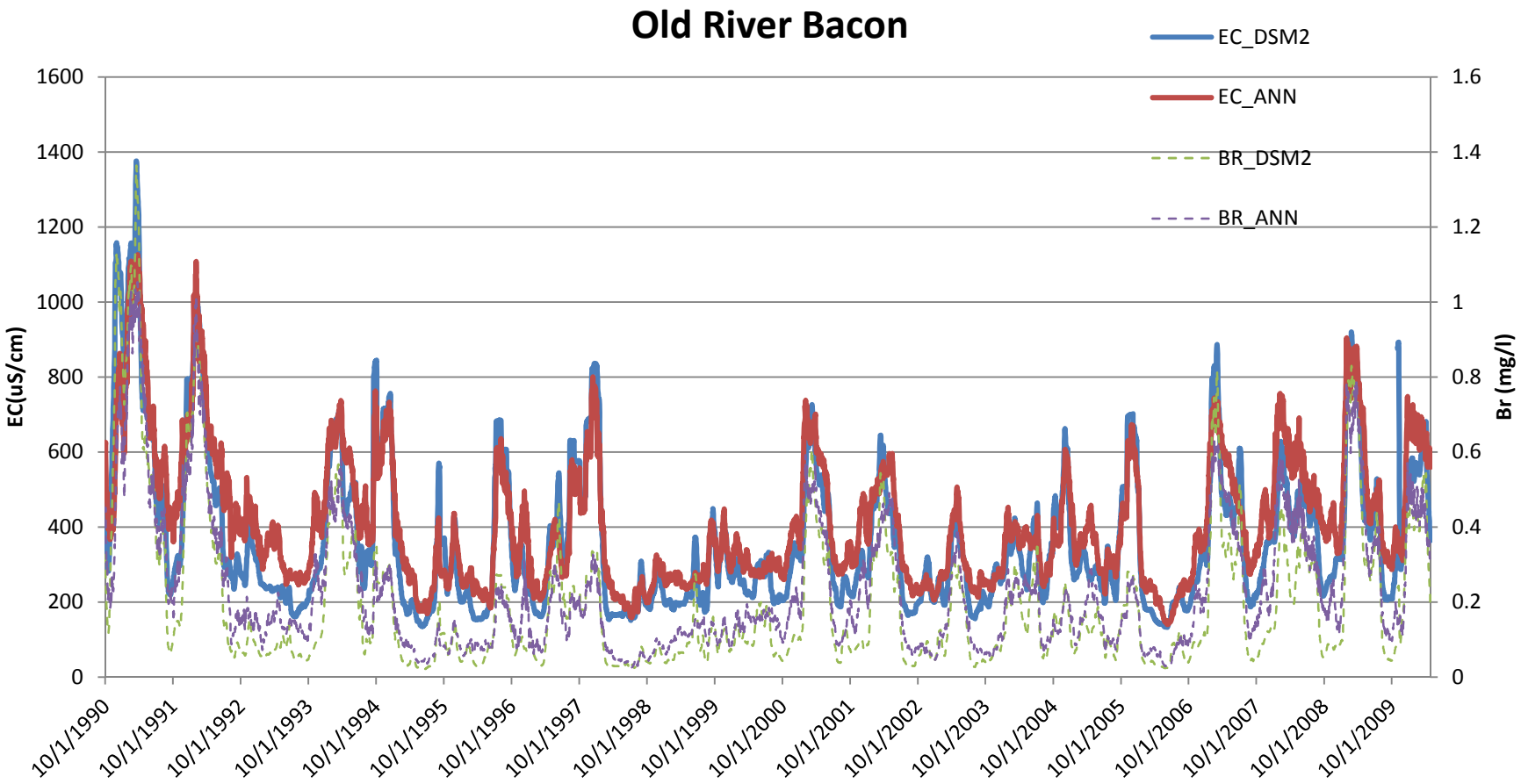
Time Series Example: Banks Pumping Plant EC and Br



Time Series Example: Old River at Hwy 4 EC and Br



Time Series Example: Old River at Hwy 4 EC and Br



Interface Operation

Excel Interface Application Steps

- Step1: Populate flow inputs from boundaries and DCC gate:
  - Calaveras
  - Mokelumne
  - Martinez
  - Sacramento
  - San Joaquin
  - Yolo
  - DICU (total)
  - DCC gates positions

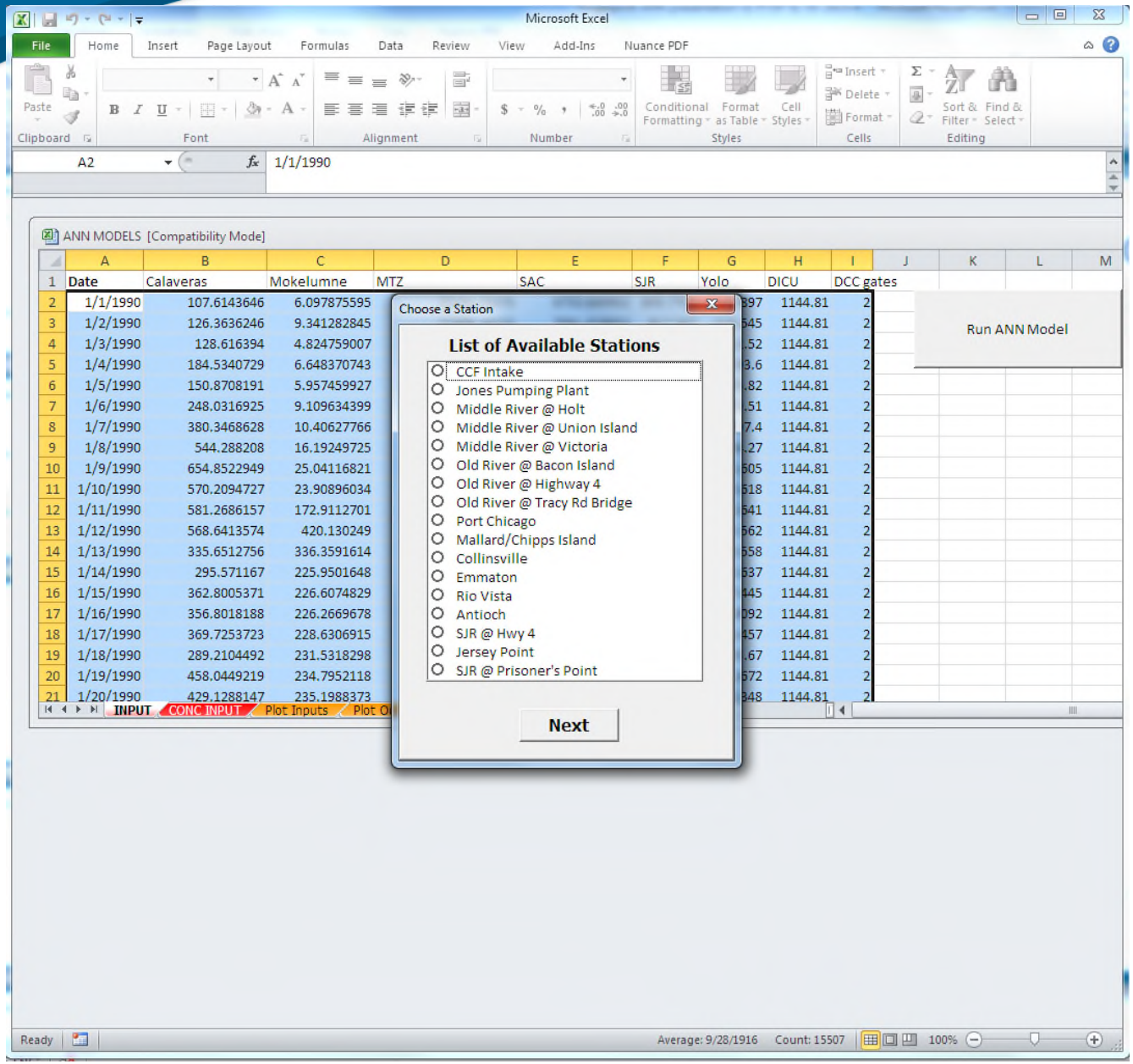
Excel Interface Application Steps

- Step2: Populate concentration inputs from boundaries:
  - Calaveras
  - Mokelumne
  - Martinez
  - Sacramento
  - San Joaquin
  - Yolo
  - DICU concentrations (region 1, region 2, region 3)

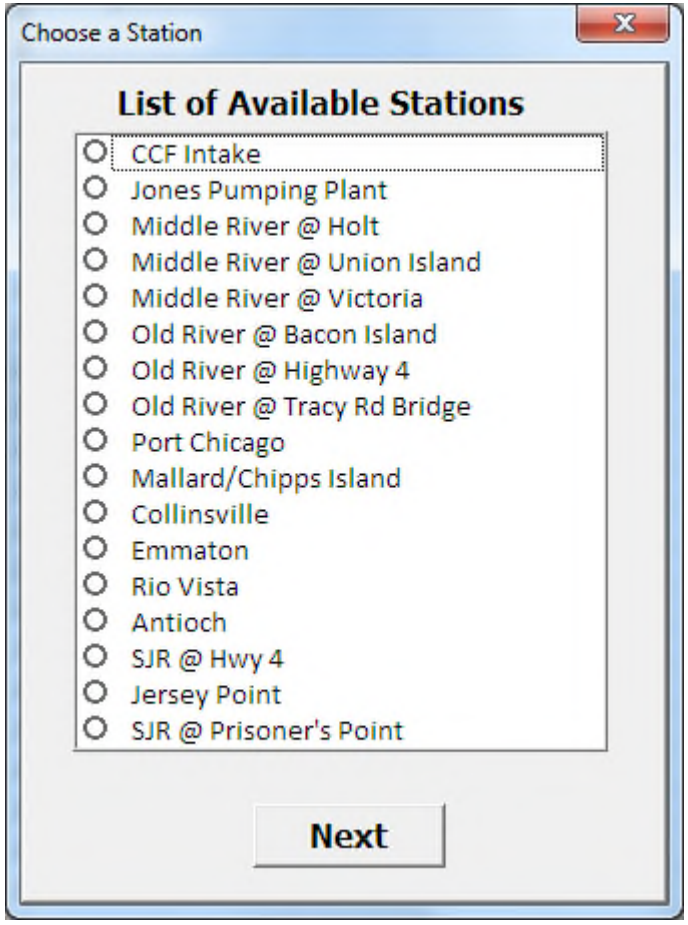
Excel Interface Application Steps

- Step 3: Select constituent for calculation
  - 1: EC
  - 0: DOC
- Step 4: Select stations for calculation
  - Clifton Court Forebay - Mallard/Chipps Island
  - Jones Pumping - Collinsville
  - Middle River @ Holt - Emmaton
  - Middle River @ Union Island - Rio Vista
  - Middle River @ Victoria - Antioch
  - Old River @ Bacon Island - SJR @ Hwy 4
  - Old River @ Highway 4 - Jersey Point
  - Old River @ Tracy Rd. Bridge - SJR @ Prisoner's Point
  - Port Chicago
- Step 5: Run Model

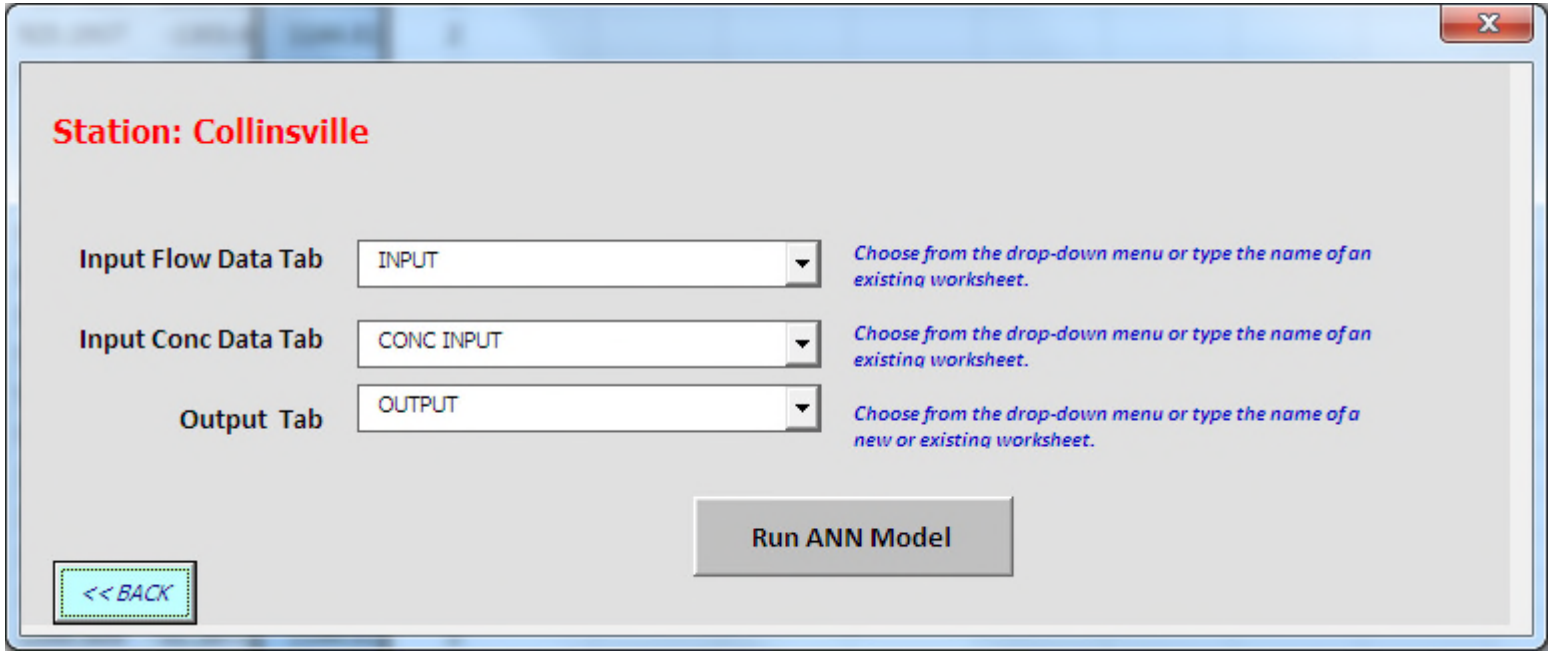
Opening Screen



Select a Station

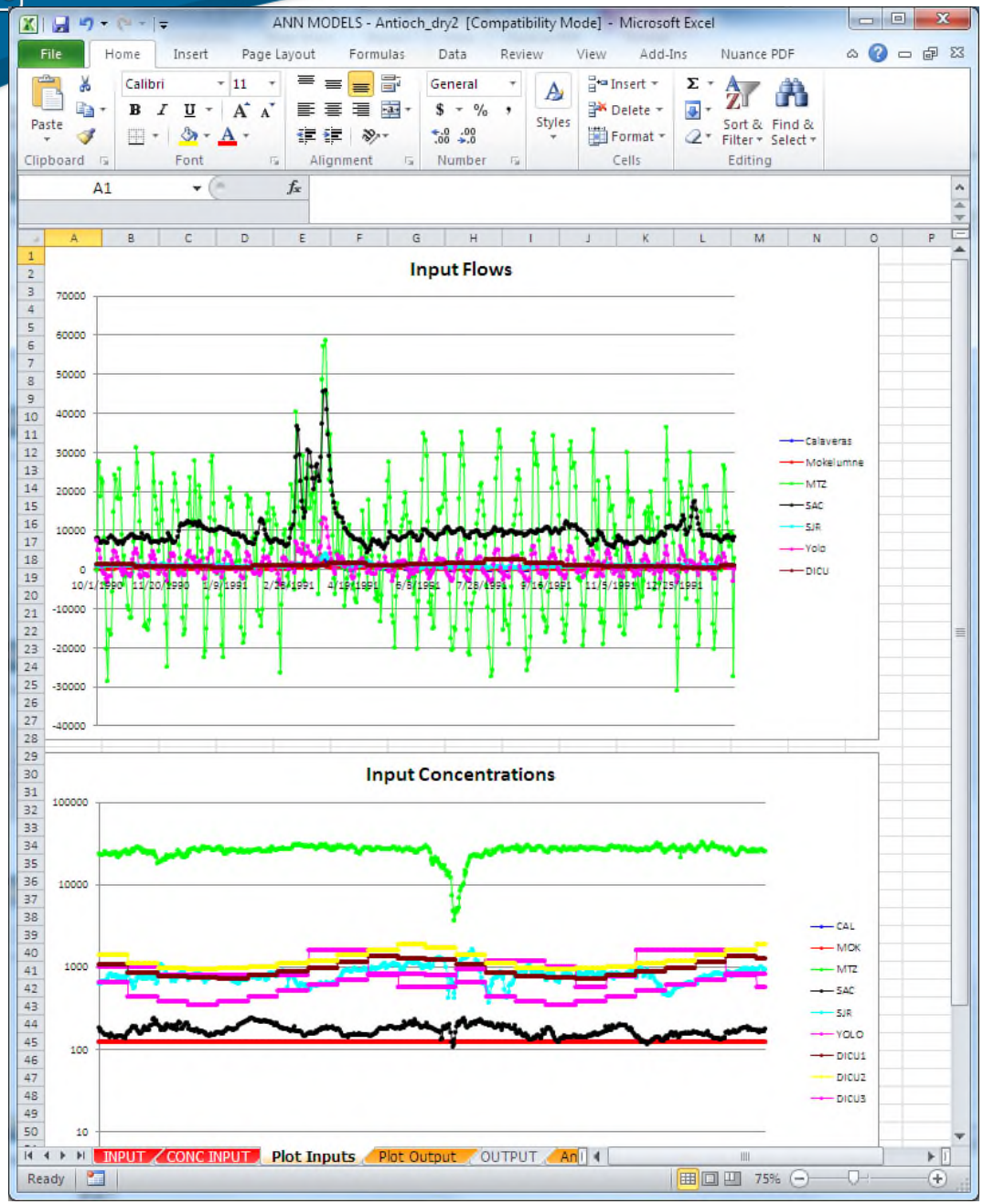


Identify Output Worksheets



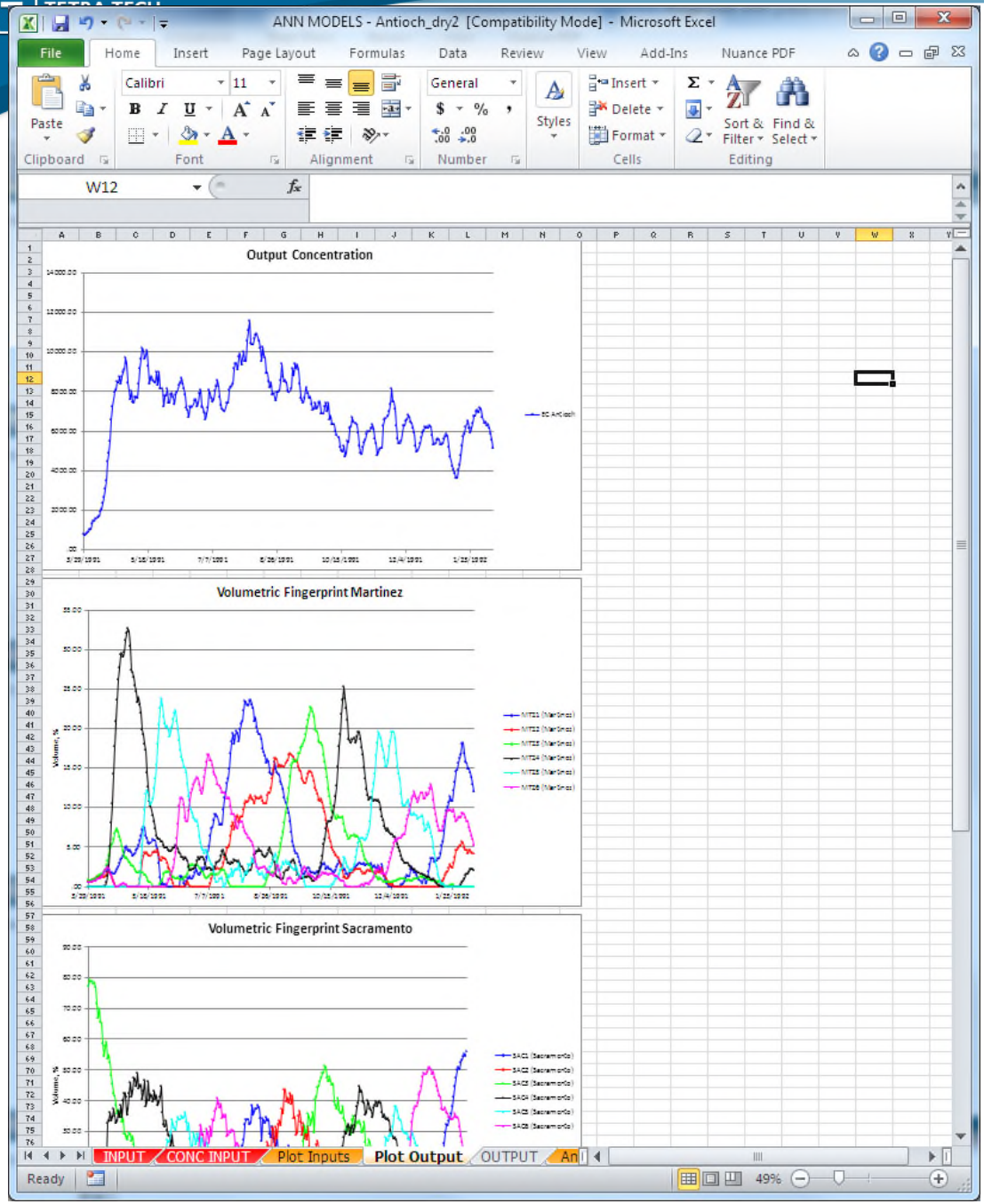
The names of existing worksheets can be edited to save results.

Input Plots


ANN MODELS - Antioch_dry2 [Compatibility Mode] - Microsoft Excel
Input Flows
Calaveras
Mokelumne
MTZ
SAC
SJR
Yolo
DICU
Input Concentrations
CAL
MOK
MTZ
SAC
SJR
YOLO
DICU1
DICU2
DICU3
INPUT
CONC INPUT
Plot Inputs
Plot Output
OUTPUT
Ready

Output Plots

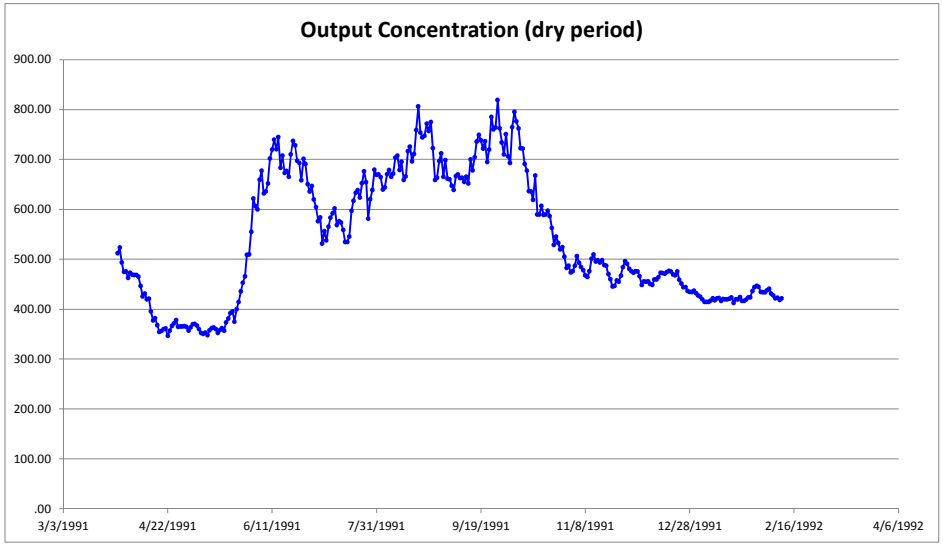
All plots can be copied and edited

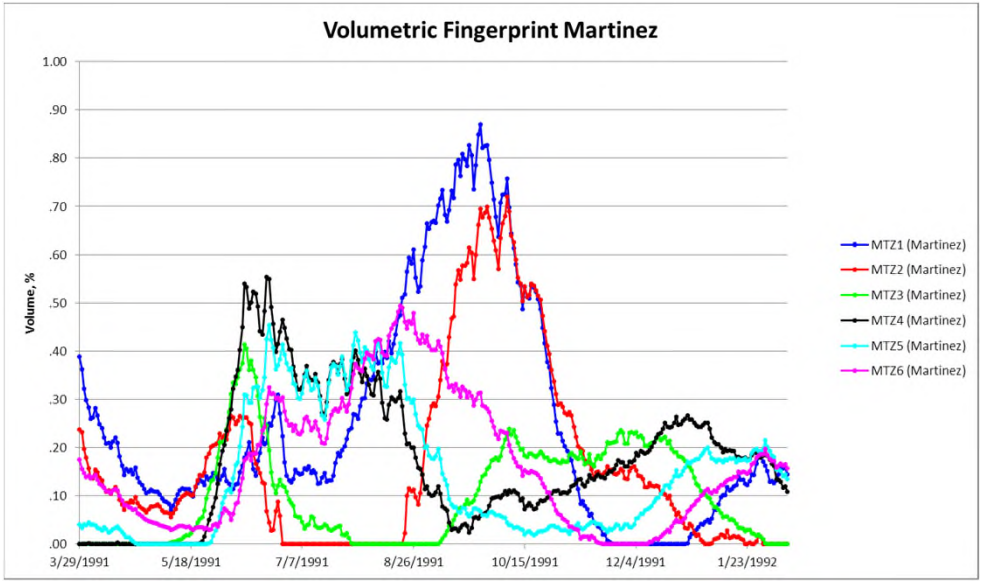


Example Runs

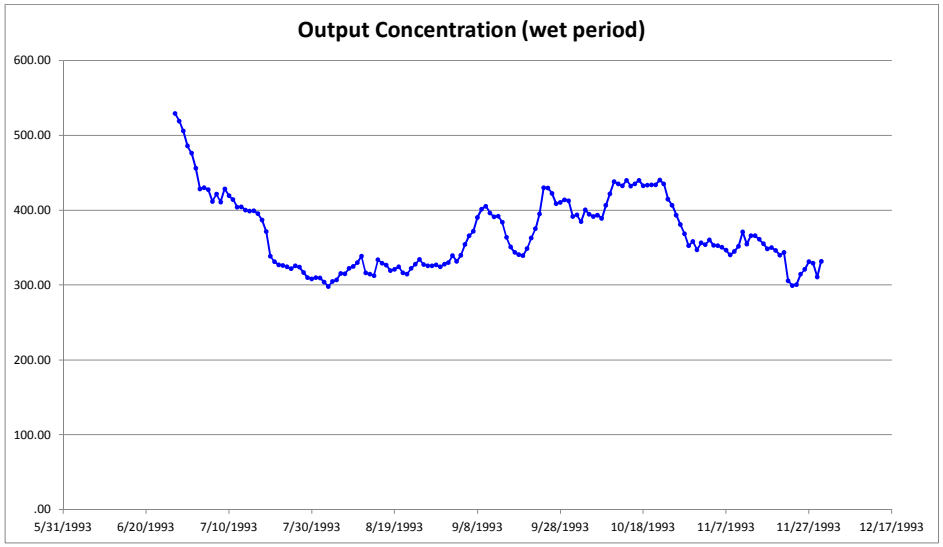
- Example 1:
  - Mid River @ Holt
    - Wet Period (4/30/1991-2/10/1992)
    - Dry Period (12/30/1992- 11/30/1993)
- Example 2:
  - Old River @ Bacon
    - Wet Period (4/30/1991-2/10/1992)
    - Dry Period (12/30/1992- 11/30/1993)
- Example 3:
  - Antioch
    - Wet Period (4/30/1991-2/10/1992)
    - Dry Period (12/30/1992- 11/30/1993)

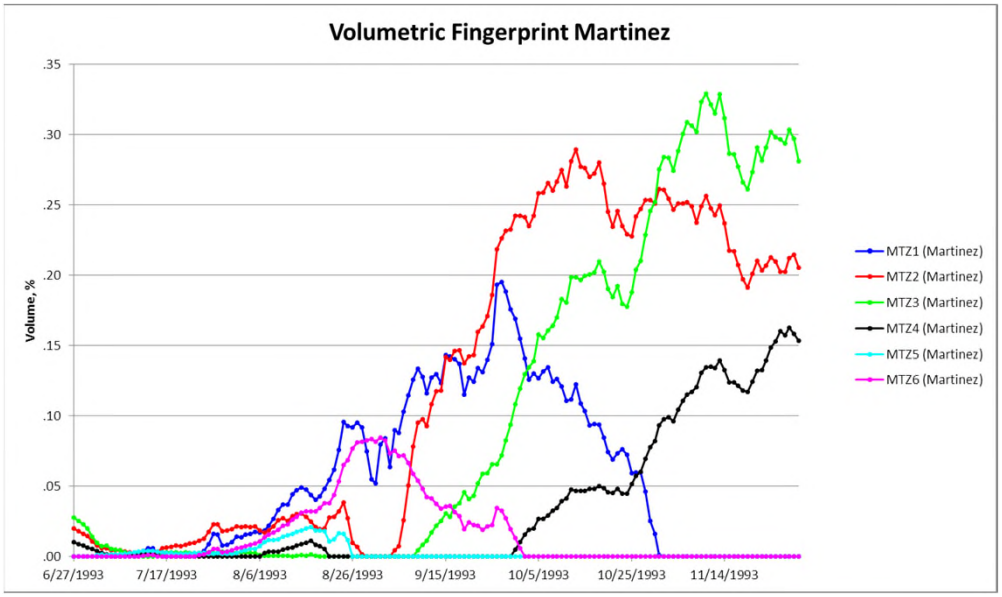
Mid River @ Holt (dry period)



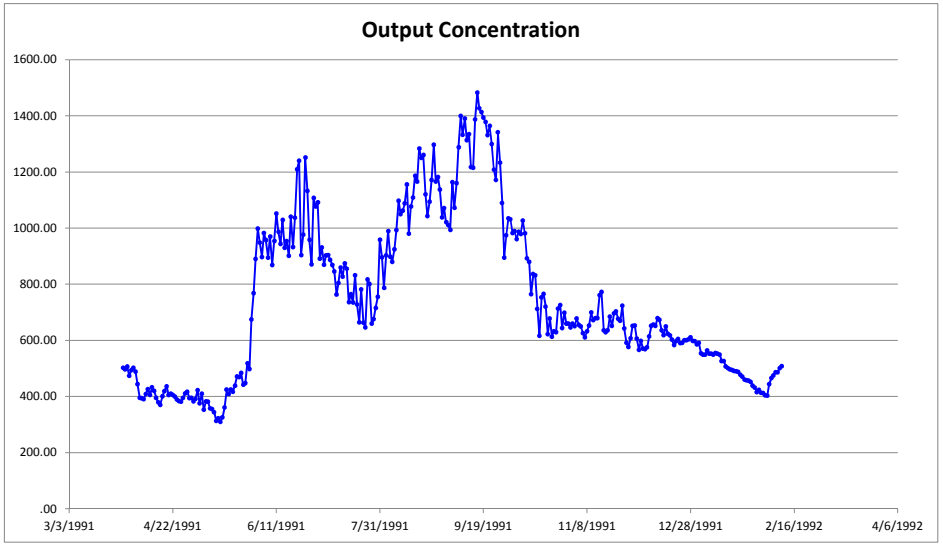


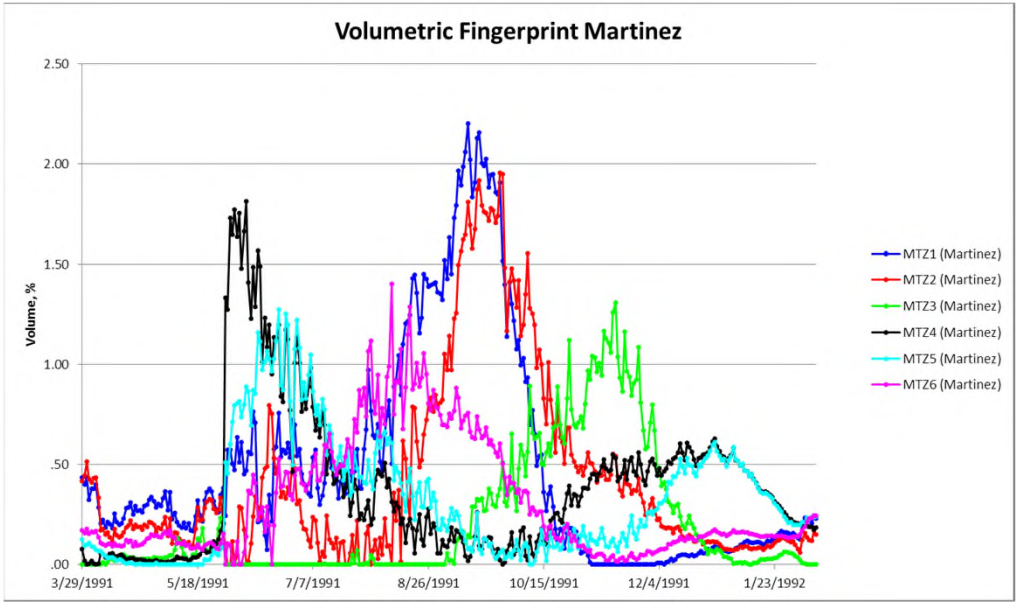
Mid River @ Holt (wet period)



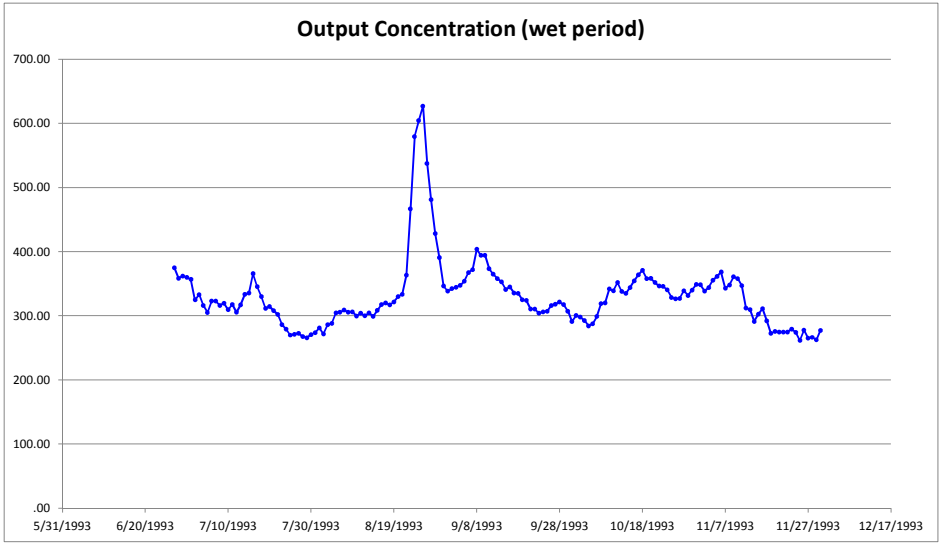


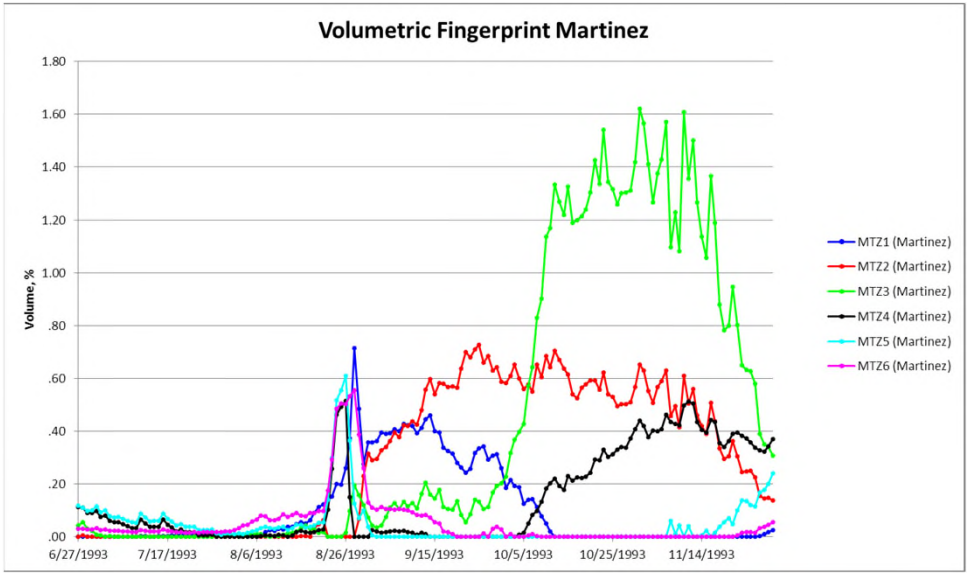
Old River @ Bacon (dry period)



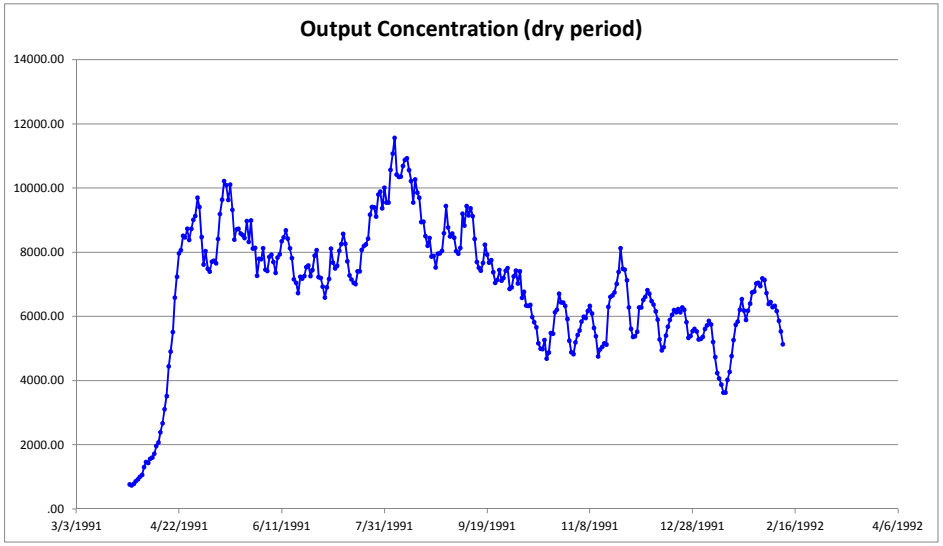


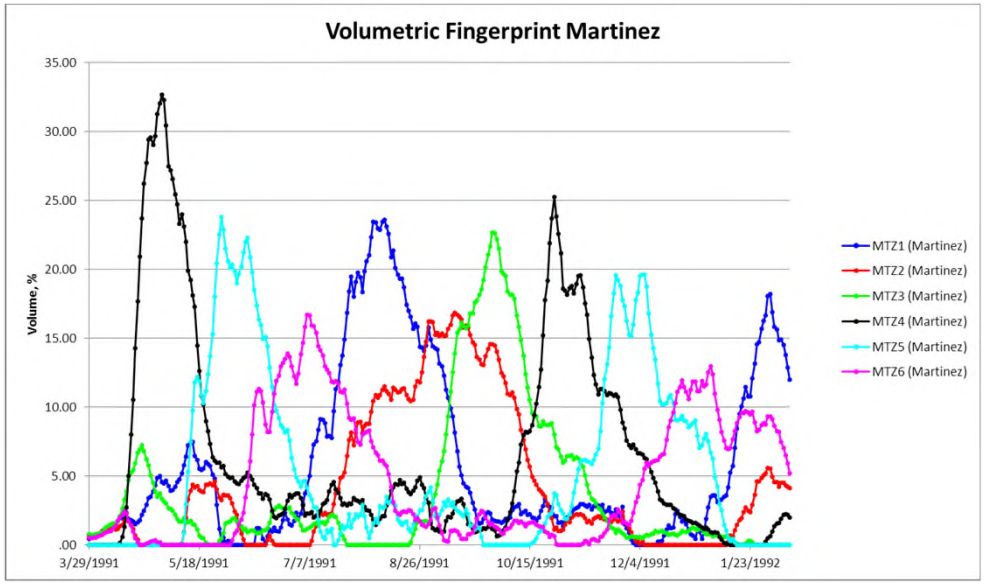
Old River @ Bacon (wet period)



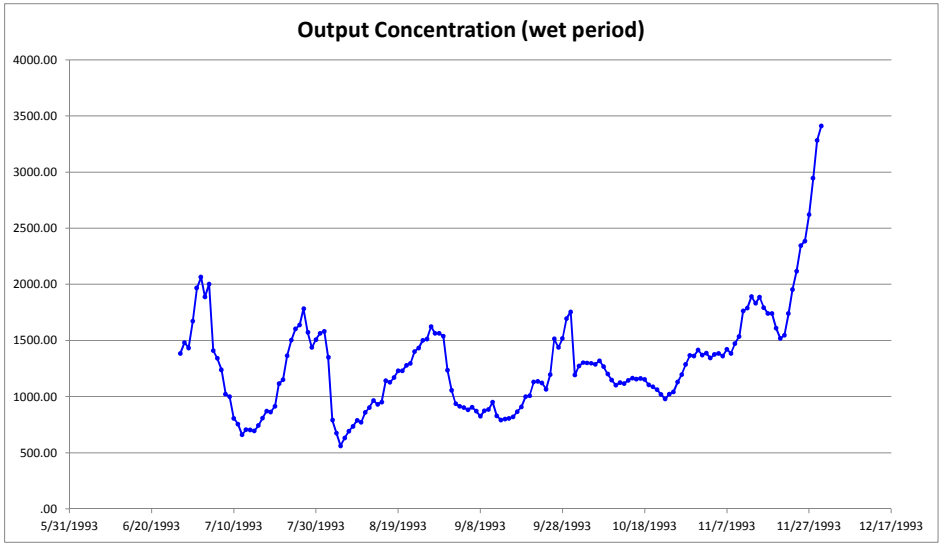


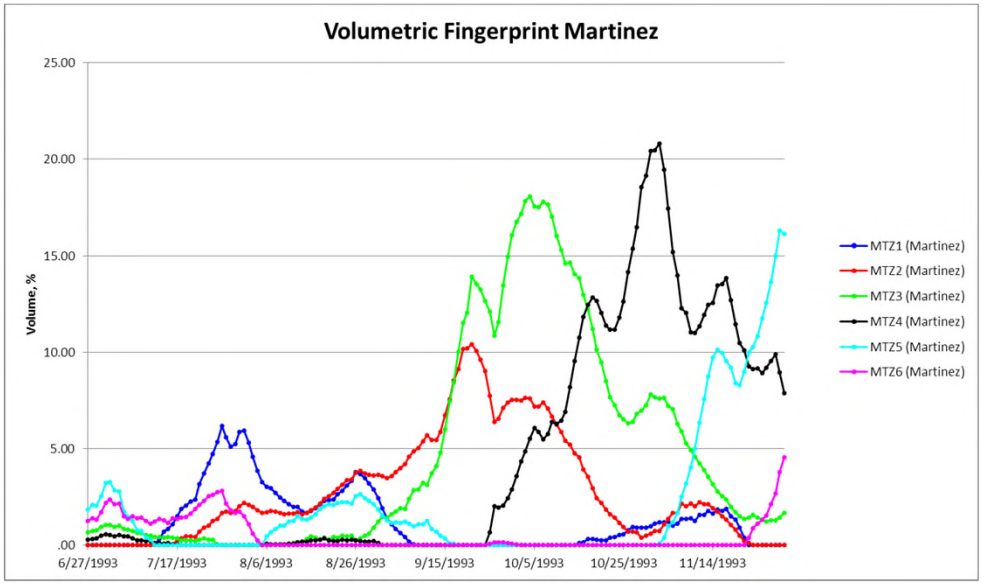
Antioch (dry period)





Antioch (wet period)





Questions and Discussion