MODELING SALINITY IN SUISUN BAY AND THE WESTERN DELTA USING ARTIFICIAL NEURAL NETWORKS

Final Report

Prepared for:

Paul Hutton, Ph.D., P.E.
Metropolitan Water District of Southern California
1121 L Street, Suite 900
Sacramento CA 95814-3974

Prepared by:

Limin Chen, John S. Rath and Sujoy B. Roy
Tetra Tech Inc.
3746 Mt. Diablo Blvd, Suite 300
Lafayette, CA 94549

April 18, 2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ANN	Artificial Neural Network
DSM2	Delta Simulation Model II
CDEC	California Data Exchange Center
CDFG	California Department of Fish and Game
DWR	Department of Water Resources
FFW	Feed Forward Network
IEP	Interagency Ecological Program
MLP	Multi-layer Perceptron
MSL	Mean Sea Level
NARX	Nonlinear Autoregressive Network with Exogenous Inputs
SE	Standard Error
USGS	United States Geological Survey

EXECUTIVE SUMMARY

The goal of this work was the development of an Artificial Neural Network (ANN) model of salinity in Suisun Bay and the western Delta using as input the following variables: flows past Rio Vista on the Sacramento River and past Jersey Point on the San Joaquin River (identified as Qwest), astronomical tide, and the residual between actual and astronomical tide. These specific inputs were selected based on an evaluation of a wider range of input sets in earlier work.

Available salinity data for this work were based on a parallel data compilation and cleaning effort and cover water years 1922–2012. Data from the early period (1922-1971) are mostly grab sample measurements, and from 1965-2012 include continuous measurements as electrical conductivity (EC). Using both data sources, a daily record of salinity at multiple stations was created, albeit with some gaps in the early part of the record.

Two approaches were used for ANN training, either using the observed salinity directly, or fitting a model to the salinity data and training the model parameters using an ANN.

- In **Approach 1**, ANNs were developed using station-level data for the period where daily continuous records of salinity (using on-line conductivity sensors, as opposed to grab samples) were available with minimal data gaps (October 1974 to June 2012, WY 1975-2012).
- In **Approach 2** the Delta Salinity Gradient (DSG) model (Hutton 2013) was used to fit to observed salinity, and the best fit parameters of the DSG model were used for the ANN training. This has two major advantages: it explicitly incorporates the basic conceptual model of salinity transport in the estuary, with a strong west-to-east gradient, and it allows for variations in the presence/absence of data at individual stations.

Separate ANNs were developed for stations along the Sacramento and San Joaquin river reaches.

Using the trained ANNs, the outputs were compared to EC observations at individual stations and to the position of the X2 isohaline. Approach 1 fits were better (typical $R^2 > 0.95$ for the Sacramento River model, and $R^2 > 0.92$ for the San Joaquin River model). Approach 2 fits as R^2 were 0.9 or better for both models. In both cases, some eastern stations were not fit as well, and for a subset of stations (Emmaton and Jersey Point) of importance for current salinity compliance, station-specific ANNs were developed to achieve good fits.

Model evaluation was also performed for X2 interpolated from observed data and ANN-predicted X2. The overall fit of Approach 1 ANN-X2 to interpolated X2, represented as R^2 was 0.94 and 0.91 for the Sacramento and San Joaquin River models. The overall fits of the Approach 2 ANN-X2 to interpolated X2 values, were slightly better (R^2 = 0.95 and 0.93 for the Sacramento and San Joaquin River models).

Additional comparisons of the X2 isohaline location from two other models were also evaluated (the Kimmerer-Monismith equation and the DSG equation with constant parameters) to compare against ANN performance. The interpolated data were fit quite well with the daily DSG model, although the R^2 values were slightly lower than the ANN model. The K-M fit was not as good, and was also limited by the ability of this model to represent negative Delta flows. The daily DSG model fits can be improved with bias correction as an alternative to future applications. For station specific application, the ANN approach was found to be a useful alternative to methods that are focused on the entire gradient. Future application may also consider more than one modeling approach to calculate X2 for specified conditions to obtain more robust estimates of X2 and salinity at specific stations.

1. INTRODUCTION

The abundance of several biological populations in the eastern reaches of San Francisco Estuary is related to the location of the low salinity zone, which in turn depends on freshwater outflows from the Delta (Jassby et al., 1995). The position of the 2 parts per thousand (ppt) bottom salinity isohaline, termed X2, is a key component of the salinity standard in the estuary (State Water Resources Control Board, 2006). Under current regulations, it is interpolated as an equivalent surface salinity from fixed monitoring stations and reported as a distance from Golden Gate Bridge. Besides the X2 position, which is largely driven by habitat considerations, there are also salinity compliance points further east in the Delta for municipal and agricultural beneficial uses. Salinity behavior in an environment such as the San Francisco Bay Delta estuary is known to be dynamic and dependent on tides as well as current and antecedent freshwater flows (Harder, 1977; Denton, 1993).

In support of inflow management in Suisun Bay and the western Delta and for retrospective evaluation of salinity over changing conditions, there is a need to develop tools that provide information on salinity at specific locations and the X2 position as a function of other inputs that can be predicted or defined. Over the past two decades, various modeling frameworks have been applied to the prediction of the X2 position and of the salinity patterns in the Delta and San Francisco Bay, ranging from simple statistical models to complex three-dimensional hydrodynamic models, as reviewed briefly in the following chapter.

The focus of this work is on artificial neural network (ANN)-based modeling for salinity. The ANN approach contains some of the black-box aspects of all statistical models; however, in the Delta and elsewhere, ANN-based prediction frameworks have shown the ability to represent complex processes well, and may be considered an alternative to conventional statistical methods and mechanistic models. ANNs use simple elements (neurons) and connections between elements using a range of functional forms to represent complex real-world data. The ANN methodology has found broad application in the prediction and control of complex systems, specifically in the water resources domain (Maier et al., 2010; American Society of Civil Engineers, 2000). An ANN can be trained, in a manner similar to calibrating a model, to perform a particular function through adjusting values that form the connections between elements (weights). The ANN approach has been used broadly in the Sacramento–San Joaquin Delta for predicting salinity at various interior locations by the California Department of Water Resources (Finch and Sandhu, 1995; Sandhu et al., 1999) and for predicting salinity and impacts of sea level rise (Seneviratne et al., 2008). The salinity ANNs being developed by the Department of Water Resources are being trained on DSM2 results that may represent historical or future conditions, through taking into account individual flow components and operational parameters as model inputs.

The goal of this work is the development of an ANN model considering a wider range of inputs than used in the most common statistical modeling frameworks for salinity in this region (e.g., Kimmerer and Monismith, 1992). This work builds on a prior effort by us to develop a salinity ANN model (Chen and Roy, 2013). In the preceding work, the salinity data used were daily values collected over 1974-2012. A major focus of the work was the testing of different input combinations to identify suitable models for predicting salinity as a function of distance. Key inputs that were examined included the following: net Delta outflow, flows past Rio Vista on the Sacramento River and past Jersey Point on the San Joaquin River (identified as Qwest in the DAYFLOW model); tidal terms at different locations and astronomical tide at Golden Gate; and channel depth in the Western Delta. For each combination, two models were developed, one for the Sacramento River and one for the San Joaquin River. The outcomes of the training suggested that including both Rio Vista and Qwest flows in the training improved the results versus use of only the net Delta outflow term. Models that used both flows as inputs, or those that used the Rio Vista flow and a residual between the Rio Vista-Qwest flow correlation as inputs performed similarly, and either formulation of the flow inputs was considered acceptable for future application. The comparison of models with respect to different tidal inputs suggested that relatively good agreement between observed and model predicted values could be achieved through using one or two tidal terms either as tidal range, the astronomical tide, or the residual between actual and astronomical tide. Finally, although the ANNs for each river did well at representing the salinity gradient, some eastern stations were not as well represented. Some of these are important compliance stations and were better represented by single-station ANNs. All of these initial findings were used to guide this phase of the ANN training.

The data used in this study span a much broader range (1921-2012) than used in our preceding study and are based on a parallel data collection effort (Roy et al., 2014) that is summarized briefly here. Daily salinity data were compiled for the longest record available in Suisun Bay and the western Delta, combining historical grab samples and modern electrical conductivity samples. Grab sample data from historical documents data were based on a compilation from documents from October 1921 to June 1971 from the California Department of Public Works (DPW) and its successor agency, the Department of Water Resources (DWR). Data from scanned paper copies of these bulletins were used to develop an electronic database of salinity throughout the Delta and portions of San Francisco Bay. In addition, modern databases were queried for data in Suisun Bay and the western Delta, reported as continuous measurements of electrical conductivity, including: 1) California Data Exchange Center (CDEC); 2) the Interagency Ecological Program (IEP) water quality data; and 3) USEPA's STORET dataset. The modern data were further supplemented by U.S. Geological Survey data for stations in San Francisco Bay to account for situations where the low salinity zone extends into the Bay, typically under high flows. The combined data gathering effort resulted in a master database containing salinity data from October 1921–September 2012, i.e. water years 1922–2012. Going forward, where differentiation is needed, the grab sample data are referred to as Bulletin 23 data and the continuous EC data are referred to as CDEC data, although it is recognized that both groups contain data from other sources as well.

A significant effort was expended to "clean" the data to remove values that appeared to be clearly inconsistent with other values. Following this cleaning, data gaps stations were filled by interpolation to create the longest possible salinity record without data gaps. Consideration of the long data series is an important factor, because ANNs, being data-driven constructs, perform best within the training range, with poor or undefined behavior when extrapolated beyond the training range. The long data set developed here has the advantage of providing the ANN with a wide variety of hydrologic conditions for training.

To fully evaluate the benefits of the ANN modeling approach, we also compared the results of this work with an analytical model of salinity that is an alternative tool for representing salinity behavior in Suisun Bay and the western Delta (Hutton, 2013) and with a daily version of the Kimmerer-Monismith (K-M) autoregressive model (Kimmerer and Monismith, 1992).

The remaining sections of this report describe previous work on salinity modeling in San Francisco Bay and the Delta (Chapter 2), the ANN modeling approach used (Chapter 3); results from the ANN models, comparison against existing tools and exploration of sensitivity of specific inputs such as flows and sea level (Chapter 4); and a summary of key findings and recommendations on the use of selected models in future applications (Chapter 5).

2. PRIOR MODELING OF SAN FRANCISCO BAY AND DELTA SALINITY

Given the importance of salinity in the San Francisco Bay Delta estuary to habitat, municipal, and agricultural beneficial uses, a variety of quantitative analyses have been performed to describe salinity behavior in the region for different applications. A brief summary of the commonly used tools is presented below.

2.1 STATISTICAL MODELING

A widely used tool is the autoregressive equation between Delta outflow and X2 position, termed the K-M model (Kimmerer and Monismith, 1992; Jassby et al., 1995). This equation was calibrated using salinity data in the Bay and Delta from October 1967 to November 1991, the most complete data set available at the time of publication. The monthly flow-X2 relationship (Kimmerer and Monismith, 1992) has been expressed as[1]:

$$X2(t) = 122.2 + 0.328X2(t-1) - 17.6 \log(Q_{out}(t)) \quad \text{Eq 2.1}$$

where Q_{out} is the mean monthly Delta outflow in terms of cubic feet per second (cfs) and X2(t-1) is the previous month isohaline position expressed as km from Golden Gate. As a general tool for estimating X2 under different flow conditions, the above equation is used widely (referred to as the K-M equation). This equation has also been proposed using an exponent form of the Q_{out} term, rather than the logarithm, albeit using the same surface salinity dataset as in the original analysis (Monismith et al., 2002), although this is in less common use than the K-M equation.

2.2 ONE-DIMENSIONAL EMPIRICAL MODELING

An empirical model of salinity was also developed by Denton (1993) (updated Denton, 1994), utilizing boundary salinity values representative of the downstream ocean and upstream riverine environments, and a concept called antecedent outflow, representing flow time-history in the Delta. The equation can be represented as:

$$S = (S_o - S_b) * \exp[-\alpha * G(t)] + S_b \quad \text{Eq. 2.2}$$

where S is the salinity at a given location, S_o and S_b are the ocean and river boundary salinities, and G(t) is the term representing the flow history, and α is an empirically-determined constant, computed for selected Delta locations based on field data. The G-model estimates salinity at individual locations, rather than the X2 position estimated using the K-M equation. This is important because salinity standards are described in terms of electrical conductivity at individual stations in the current Water Quality Control

[1] A slightly different intercept for this equation has also been reported for flow in m^3/second:
X2(t) = 95 + 0.33X2(t-1) -17.6 log(Qout(t)) (Jassby et al., 1995)

Plan, e.g., Emmaton on the Sacramento River and Jersey Point on San Joaquin River (State Water Resources Control Board, 2006).

A hybrid of the K-M equation and G-model, proposed by Hutton (2013), is called the Delta Salinity Gradient (DSG) model. In this model, by assuming the modified form of the X2 equation (Monismith et. al. 2002) and steady-state conditions, X2 is related to antecedent outflow as follows:

$$X2(t) = \Phi_1 * G(t)^{\Phi_2} \quad \text{Eq. 2.3}$$

where Φ_1 and Φ_2 are empirically determined coefficients. Salinity is then estimated at individual locations through the following relationship:

$$S = (S_o - S_b) * \exp[\tau * (X/X2)^{-1/\Phi 2}] + S_b \quad \text{Eq. 2.4}$$

where S is the salinity at a given location in mS/cm, S_o and S_b are representative downstream ocean and upstream riverine boundary salinities, and $\tau = \ln[(2.64 - S_b)/(S_o - S_b)]$. This equation can be used to determine salinity at any longitudinal distance from Golden Gate (X) given X2 and Φ_2 and assuming reasonable values for S_o and S_b.

2.3 NUMERICAL MODELING

Numerical models of hydrodynamics and salinity, albeit more complex and demanding of computer time and user expertise, have also been considered for different applications.

The one-dimensional link-node modeling of hydrodynamics and salinity of the Delta using the California Department of Water Resources' (DWR) Delta Simulation Model (DSM2) is used widely to represent salinity under different hydrologic conditions, and for changes from background conditions, such as for increases in sea level at the western boundary, or for consideration of changes to the Delta studies as part of the Bay Delta Conservation Plan (BDCP) (http://baydeltaconservationplan.com/). The DSM2 model is also used extensively for DWR's annual reporting to the State Water Resources Control Board.

Resource Management Associates (RMA) has developed a two-dimensional model of salinity, the RMA-Bay Delta model. Used extensively, this model has recently been used to examine the effects of sea level rise as part of the BDCP effort (Administrative Draft, March 2013, Appendix 29A).

Finally, three-dimensional modeling for salinity and flow in the entire bay and estuary has been performed for evaluating specific projects and for understanding the mechanistic processes of salinity intrusion under different flow and tidal conditions (Gross et al., 2007, 2010).

Although theoretically rigorous, the computational demands of the two- and three-dimensional models limit application within planning simulations that run over decades and consider multiple scenarios.

2.4 ANN-BASED MODELING

ANNs, the specific focus of this study, have also been used to represent flow and salinity in the Delta (Finch and Sandhu, 1995; Wilbur and Munevar, 2001; Mierzwa, 2002; Seneviratne et al., 2008). The ANN approach has also been used extensively by DWR to represent salinity at different locations in the Delta, with the ANNs being trained on synthetic data generated from DSM2, including scenarios that are different from current/historical conditions and employ changes in sea level and tidal amplitude. Because ANNs run significantly faster than the mechanistic models they are trained on, they can be employed within planning models, such as the CALSIM model, where there is a need to return results rapidly. When used in this manner, ANNs are emulating DSM2 behavior, and not serving as an independent model of salinity.

3. ARTIFICIAL NEURAL NETWORK MODELING APPROACH

This section provides an overview of the ANN modeling approach and the data used. As presented in Chapter 1, this work builds on the initial ANN modeling effort presented in Chen and Roy (2013) in defining the model structure and inputs that are used. With respect to outputs, in the original work we used salinity data at specific stations for training. Thus, given a set of inputs of flows, mean sea level, one or more tidal terms, and distance from Golden Gate, the ANN was trained for salinity at different locations. This approach was suitable for the original application because a complete daily time series of salinity across a consistent set of stations was available after a filling procedure was followed. In principle, the same method can be applied to the significantly enhanced data set from Water Year (WY) 1922-2012. However, because of the nature of the grab sampling, there are substantial gaps in the daily salinity even with our best efforts at filling. Therefore, two approaches were used, limiting the station-based approach to the period where continuous EC data are available, and applying a salinity model to the data before fitting:

- **Approach 1**, to retain the good quality of the fit observed in the previous effort, ANNs were developed using station-level data for the period where daily continuous records of salinity (using on-line conductivity sensors, as opposed to grab samples) were available with minimal data gaps (October 1974 to June 2012, WY 1975-2012).
- **Approach 2** integrates the Delta Salinity Gradient model (Hutton 2013) with the ANN approach. In this method, the DSG model was used to fit to observed salinity, and the best fit parameters of the DSG model were used for the ANN training. This has two major advantages: it explicitly incorporates the basic conceptual model of salinity transport in the estuary, with a strong west-to-east gradient, and it allows for variations in the presence/absence of data at individual stations.

As in the previous work, separate ANNs were developed for stations along the Sacramento and San Joaquin river reaches. Details of the ANN modeling approach are presented in this chapter. Analyses were also performed using the station-based approach for WY 1930-2012, but were generally not as good as for Approach 1 and 2, as defined above, and are therefore not presented in this report.

3.1 ARTIFICIAL NEURAL NETWORK (ANN) MODEL STRUCTURE

3.1.1 Model Inputs

Although two alternative methods for salinity modeling were considered, the inputs to the ANNs remained the same, and were directly based on performance of multiple combinations of inputs that were evaluated in Chen and Roy (2013). An overview of these model structures, presented in Table 3-1, demonstrates various combinations of flow and tidal inputs. Ten different model structures (labeled Model 1 through Model 10)

were evaluated based on the quality of the fit and the parsimony of the input variables. Based on both these conditions, Model 2 was identified for further use and was considered as the primary structure in the present work. The Model 2 inputs used in the training included:

- Station distance (km) from Golden Gate
- Flow variables – Rio Vista flow (on the Sacramento River) and Qwest flow (on the San Joaquin River downstream of Jersey Point). Daily values for these flow variables were obtained from the DAYFLOW model from WY 1930 to the present. Additional daily flow data covering water years 1922 through 1929 were developed from data reported in the Bulletin 23 series (Paul Hutton, personal communication). With these additional data, a complete flow record corresponding to the salinity data was available.
- Ocean boundary – astronomical tide and a residual between astronomical tide and actual tide. The residual was correlated with the atmospheric pressure and when used in a predictive mode, the atmospheric pressure can be used to estimate the tidal residual.
- A time lag of 30-120 days was considered for the various inputs in the present study, i.e., training the ANN with input values of the last n days (where n = 30, 60, or 120). To simplify the ANN input requirements, a time lag of 30 days was generally used, when no improvement with longer time lags was indicated. However, a longer time lag was retained for station-specific ANNs developed for Emmaton and Jersey Point (60-day delays for both station-specific ANNs were used). For these ANNs there was an improvement over the 30-day ANNs.

3.1.2 ANN Model Structure

The dynamic nature of flow and salinity in Suisun Bay and the western Delta requires a network structure that takes into account the time-series of inputs. Although other network structures were used in different applications, the multi-layer perceptrons (MLPs) are by far the most popular network structures used in water resource applications to date, representing more than 90% of the peer-reviewed applications in the water resources field (Maier et al. 2010). For this reason, the feedforward MLP network was selected for this application.

For much of this analysis, ANNs were developed by accounting for station distance (integrating all stations along a river). However, based on the results of these salinity-distance ANNs, and poorer performance at some eastern locations, station-specific ANNs were developed for selected stations.

3.1.3 Outputs for Training

For each input set, two separate ANN models were developed for the lower Sacramento River and lower San Joaquin River stations. The training was based on salinity at fixed stations on each river. Training stations for the Sacramento River ANN model were at a number of locations along the lower Sacramento River and several stations in the Bay. Training stations for the San Joaquin River ANN model were at a number of locations along the lower San Joaquin River and several stations in the Bay.

Both the Bulletin 23 dataset (WY 1922-1971) and the continuous EC CDEC dataset (WY 1965-2012) were used in the training. As discussed above, two approaches were used in the ANN training:

1. Approach 1, training using the salinity dataset at specific stations, without preprocessing with a model, but limited to the period with continuous salinity data as EC, October 1974 to June 2012;
2. Approach 2, training by fitting the daily data using the Delta Salinity Gradient (DSG) model (Hutton, 2013), Equations 2.3 and 2.4. X2 in the equations is obtained by fitting a second ANN to the interpolated data.

The training for the station-based approach in Approach 1 was also fine-tuned through applying a "nudging" technique on distance. This step acknowledges that the reported distance for a given station may not be exact or that a station behaves in a certain manner because of the particular hydrodynamics of estuarine flow at the location. The purpose of the nudging was to vary the distance within a certain range (here ± 3 km at 0.1 km intervals), to examine whether the prediction for EC could be improved. The nudging evaluation was performed using data from 1965-2012.

The results of the training are individual ANNs that predict salinity as a function of distance for the Sacramento and San Joaquin Rivers, based on inputs of flow, mean sea level, and tide.

When using Approach 2, the results of the training also included intermediate ANN models that predicted the three parameters in Equation 1, which were then combined with a model for X2 to predict salinity at any arbitrary distance.

In addition to ANNs that were used to represent the entire salinity gradient, ANNs were also developed for two stations separately because of their importance in the existing salinity compliance regulations in the 2006 Water Quality Control Plan for the San Francisco Bay/Sacramento-San Joaquin Delta Estuary: Emmaton on the Sacramento River, and Jersey Point on the San Joaquin River. These stations were fit using data over the period used for Approach 1 (1974-2012).

3.1.4 ANN Training Method

In this work, the data were divided in the following manner: 70%, 15%, and 15% for training, validation and testing, respectively. The training and validation data were used together in calculating the biases and weights that form the ANN, and the test data set were completely independent for additional evaluation of model performance. The dates for training, validation and testing were randomly selected from the entire dataset for each training cycle.

The ANN training used the back-propagation (Levenberg-Marquardt back-propagation) method for error minimization. For each model structure, the training was repeated until a correlation of >0.98 was obtained. The ANN training was performed the Neural Network Toolkit within the Matlab programming environment.

3.2 MODEL INPUT DATA

3.2.1 Flow

Flow data used in the ANN models were obtained from the DAYFLOW program for WYs 1930-2012, with additional data provided by Paul Hutton (personal communication) for WYs 1922-1929. The role of freshwater flow in regulating salinity in the Delta was evaluated by using Rio Vista and Qwest flow as two separate terms in the training.

3.2.2 Tide

The astronomical tide and the actual tide at Golden Gate used in the training were obtained from the National Oceanic and Atmospheric Administration (NOAA) (mean seal level, MSL at hourly time steps). When using the astronomical tide, the tides are expressed as the astronomical tides and residuals between the actual and the astronomical tides. The residuals between actual tide and the astronomical tide were found to be a function of atmospheric pressure (Figure 3-1). Residuals between the actual and astronomical tides were not found to be correlated to other meteorological variables. The correlation between tidal residuals and atmospheric pressure allows use of the model in a predictive mode, given knowledge of astronomical tide and forecasts of pressure.

3.2.3 Salinity

3.2.3.1 CDEC Data

The salinity data (in terms of electrical conductivity, EC, and reported in units of µS/cm) used in the training were obtained from CDEC, IEP, and STORET for the period of 1964-2012, at a number of stations, which were then cleaned and filled. The data cleaning was done based on the expected relationships between EC and flow at different locations, and expected correlations between the adjacent stations. These expected functions were used to identify potential data errors in the dataset that were outside a certain range of the expected functions (e.g., two standard errors). The data cleaning procedures are described in Roy et al. (2014). The data filling was done using linear interpolation for data gaps less than 8 days. For data gaps that are more than 8 days, correlations with nearby stations were used to fill the gaps.

3.2.3.2 US Geological Survey (USGS) Data

The data obtained from the USGS for stations in the Bay were for salinity in practical salinity units (psu). To be consistent with the EC data in µS/cm, the salinity data from USGS were converted to EC using the approach outlined by Schemel (2001).

$$X_{25,S} = \left(\frac{S}{35}\right) \times (53087) + S(S-35) \times [J_1 + \left(J_2 \times S^{\frac{1}{2}}\right) + (J_3 \times S) + \left(J_4 \times S^{\frac{3}{2}}\right)]$$

Eq. 3.1

Where,

$X_{25,S}$ = EC at 25 ^{0}C, J_1= -16.072, J_2 = 4.1495, J_3 = -0.5345, J_4 = 0.0261.

Similar to the CDEC data, correlations between adjacent stations were used to fill larger data gaps (> 8 days). The salinity data obtained from the USGS for stations in the Bay included Point San Pablo (PSP) at near-surface and Carquinez (CAR) at mid-depth. The CAR station did not have measurements at near-surface depths. Previous studies have

shown that no single and straight-forward relationship exists between bottom and surface salinity across multiple Bay stations (List, 1994), therefore a conversion from mid-depth and surface salinity (at a different location) was not performed for CAR. The data obtained at mid-depth for CAR were used directly in the training. The non-availability of surface salinity data at this station may contribute the uncertainty in X2 fits, especially during high flow periods when the X2 position is to the west, and when the salinity at CAR is used for interpolating X2 position.

The filling procedures applied here to the cleaned daily salinity data resulted in a continuous block of salinity data from 1964-2012 for the Western Delta stations, and from September 1990 to September 2008 for the Bay stations.

3.2.3.3 Bulletin 23 Data

Compilation of the 1921-1971 salinity data (referred to as Bulletin 23 data) is described in detail in Roy et al. (2014). The salinity data in these reports are grab samples collected at fixed locations typically every 4 days, one and one-half hours following higher high tide, which corresponds to the highest salinity for the day. There were exceptions in that on some dates data were not collected or not sampled at the higher high tide.

The development of the Bulletin 23 database included the following steps: converting observed data to a common salinity unit, accounting for tidal effects on grab samples, converting values to represent a daily average salinity, cleaning data, and filling data gaps. Once the Bulletin 23 data had been converted to daily average EC, a more sophisticated cleaning exercise was performed by comparing daily average EC values at pairs of stations. The data filling was conducted based on the salinity data of nearby stations. The same regression relationships between pairs of stations that were used for data cleaning in the previous section were repeated on the cleaned dataset. After the "neighbor station filling" was completed, any remaining short gaps (up to 8 days, inclusive) in each station's salinity record were linearly interpolated.

3.2.4 DSG Model Fitting

Fitting for the DSG model was conducted in Matlab using Equation 1 above. Three parameters were fitted on a daily basis based on the salinity data available for that day. These three parameters include: downstream salinity (S_o), upstream salinity (S_b), and a fitting parameter (Φ_2) for the shape of salinity as a function of distance. Numbers of salinity data stations available each day vary from 5-26 for the Sacramento River model and 5-30 for the San Joaquin River model. The salinity dataset used in the DSG model fitting include both the Bulletin 23 and the CDEC dataset. The three parameters were fitted using a range of 0-50,000 µS/cm for S_o, 0 – 2,640 µS/cm for S_b, and 0-20 for $-1/\Phi_2$ (for the training a new variable was defined that equated to $-1/\Phi_2$). The fitted daily parameters were used as model inputs in the ANN training for the DSG-based approach.

Table 3-1
Candidate ANN Model Structures Evaluated in Chen and Roy (2013)

Number	Flow	Tides	Antecedent salinity Data as Input	Time Delay
1	Net Delta Outflow	Astronomical tide, residuals with actual tide	No	30
2	Rio Vista Flow (Q_{Rio}),Qwest Flow	Astronomical tide, residuals with actual tide	No	30
3	Rio Vista Flow, residual of Qwest and Q_{Rio}	Astronomical tide, residuals with actual tide	No	30
4	Rio Vista Flow, residual of Qwest and Q_{Rio}	Three tidal terms (tidal range at Golden Gate and Martinez, and half tide at Mallard Island)	No	30
5	Rio Vista Flow, residual of Qwest and Q_{Rio}	Two tidal terms (tidal range at Golden Gate and Martinez)	No	30
6	Rio Vista Flow, residual of Qwest and Q_{Rio}	One tidal term (tidal range at Golden Gate)	No	30
7	Rio Vista Flow, residual of Qwest and Q_{Rio}	No tidal term	No	30
8	Rio Vista Flow, residual of Qwest and Q_{Rio}	Actual tide (MSL at Golden Gate)	No	30
9	Rio Vista Flow, residual of Qwest and Q_{Rio}	Astronomical tide, residuals with actual tide	Yes	30
10	Rio Vista Flow, residual of Qwest and Q_{Rio}	Astronomical tide, residuals with actual tide	NARX	30

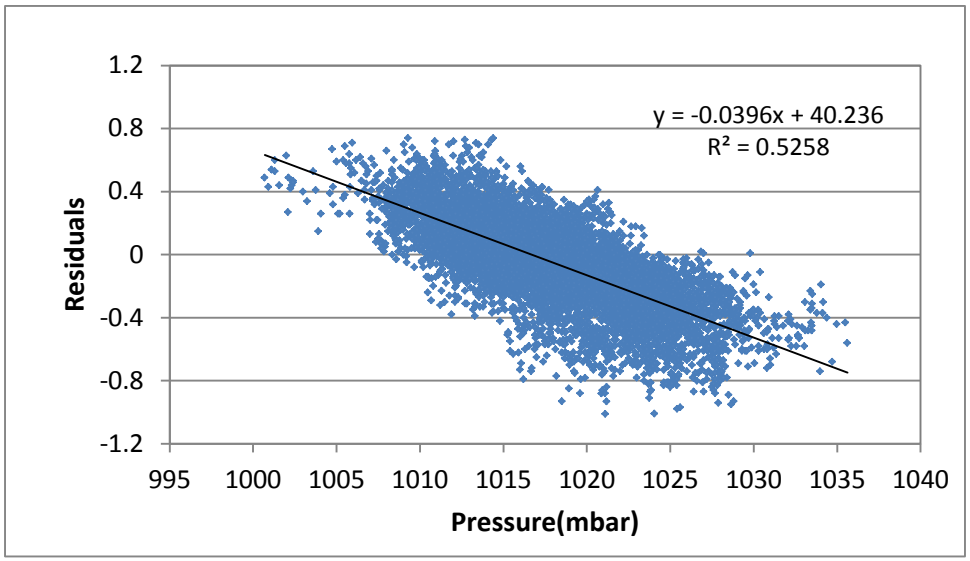

Figure 3-1 **Correlation between residuals (difference between actual tide and astronomical tide) and atmospheric pressure at Golden Gate (Data source: NOAA).**

4. RESULTS

This chapter presents the results of the ANN training for the two approaches presented in Chapter 3. In this chapter, the ANN models are compared to existing predictive models of salinity in the Delta (the DSG model with constant parameters and the K-M model). This chapter also explores ANN model sensitivity to changes in flow, sea level, and air pressure. Model diagnostics are presented in a framework similar to that used for evaluating models versus interpolated X2 data in Roy et al. (2014).

4.1 ANN NETWORK TRAINING RESULTS

4.1.1 Training Using Approach 1

ANN models were developed using the daily values for flow, astronomical tidal level and the residual between astronomical tide and actual tide, and with salinity data from a number of stations that have the most continuous data, for WY 1975-2012 (October 1974 to June 2012, all CDEC stations). The available data points by station are shown in Table 4-1. This approach resulted in a large number of daily salinity points for training (both observed and filled): 109,000 points for the Sacramento River and 151,000 points for the San Joaquin River. The cleaned and filled data for this analysis were obtained from Roy et al. (2014) as noted in Chapter 3. The stations identified are the primary sources of CDEC data. One station, Chipps Island, for which data were available for a limited time period (1976-1992) was not used directly, but was used to fill and provide a more complete record at a nearby station (Mallard Island).

The statistics associated with the training results are shown in Table 4-2. Results suggested a typical $R^2 > 0.95$ for the Sacramento River model with the exception of Emmaton (EMM) (0.85) and Rio Vista (RVB) (0.58). For the San Joaquin River model, the fit as R^2 was generally >0.92 with the exception of Blind Point (BLP) (0.87), Jersey Point (JER) (0.82), Three Mile Slough (TSL) (0.80) and San Andreas Landing (SAL)(0.3). Fits between the ANN and observed data are shown at the station level as scatterplots in Figure 4-1 and as time series plots in Appendix A. Fits were generally better for stations along the Sacramento River branch, for stations that exhibited a wide range of salinity. Thus, consistent with Table 4-2, the eastern stations with low salinity are not well fit by the overall model. This is addressed through the development of station-specific ANNs described in Chapter 4.6.

Results for the nudging are shown in Table 4-3 and Table 4-4. The nudging suggests a change in distance of -3 to +3 km will result slightly improved fits at some stations. However, the improved fits are marginally better than using the original distances, and stations with poor fits remained hard to fit. In some cases nudging by a larger amount (by more than 3 km) could have changed results, but there was no physical basis to modify the distance by such a large amount. Although the analysis provides some insights how much improvement can be achieved by modifying the distance, the changes were not

large enough to formally adopt these new "nudged" distances. These results are not used in the subsequent analysis.

4.1.2 Training using Approach 2 (DSG-based Approach)

The results of fitting a daily DSG model are daily values of three parameters (S_o, S_b, and $-1/\phi_2$) used in the DSG model. These values are used as outputs for the ANN model training. ANN models were developed using the daily flow, mean sea level and tidal range as inputs described above and the fitted parameters, for both the Bulletin 23 (WY 1922-1971) and the CDEC (WY 1965-2012) data. Available data for this approach are identified in Table 4-5.

Results for the training are a model that predicts these three parameters based on two flows (Qwest and Rio Vista flow), astronomical tide and tidal residual. Results of the Sacramento River ANN predictions for these values are shown in Figure 4-2 and Figure 4-3 as scatterplots and time series plots. There is no simple relationship between the three DSG model input variables, although for some, it appears that the 1922-1967 period ranges are distinct from the 1968-2012 values. The S_o and S_b values are related to the inflows, with lower values for both at higher inflows.

The ANN-derived parameters were then used to predict salinity at different locations based on the equation from the DSG model. The values of the results for the DSG fitted ANN model compared to the observed values are shown in Table 4-6.

Results suggested a general fit of R^2 near or above 0.9 for the Sacramento River model with the exception of a few stations (PTD, PTO, PSP, and RVB). The CDEC stations have better fits that the Bulletin 23 stations. For the San Joaquin River model, the fit is generally above 0.9 with the exception of a few stations (OPT, PTO, SAL). The comparison of the model-trained and observed salinity (for the Bulletin 23 and CDEC time periods) as scatter plots are shown Figure 4-4, and as time series plots in Appendix B.

4.2 Evaluation of ANN Model Performance With Respect to X2

Additional performance evaluation is presented for ANNs developed using Approach 1 and 2 using a sequence of plots that include scatterplots of observed/interpolated values and ANN-calculated values and residuals as a function of target variable value and over time. For context, the range of interpolated X2 values over the period of interest (WY 1922-2012) is shown as a function of month in Table 4-7. Typically, between 84% and 99% of the interpolated daily X2 values fall between 50 and 100 km. This is important to consider when any of the subsequent scatterplots are viewed; although the values that lie outside this range are of interest, the vast majority of points fall in the middle range and are less visually apparent because they overlap each other.

In Figure 4-5 through Figure 4-7 we show various comparisons of the ANN model and interpolated X2 (Approach 1). Figure 4-5 shows scatter plots of X2 conditioned on river and month on the scale of the X2 positions. The quality of these fits is assessed in Table 4-8. There is an overall R^2 of 0.92, ranging from 0.79 to 0.96 depending on the month and river. Second, plots of the residuals (interpolated X2 subtracted from model X2)

against interpolated X2 show the agreement between modeled and interpolated values for various salinity conditions (Figure 4-6). Typically, the values with high residuals fall in the <50 Km range of interpolated X2. The weakness of the ANN model in capturing these low values of X2 (during high flow months) may be related to the nature of the data in the bay (i.e., the availability only of mid-depth salinity at Carquinez Strait) and potentially changes in the stratification that occur during high flow periods and affect surface salinity values. Besides the excursions at low X2 values, there does not appear to be any pattern in the residuals over the remaining data range. The residuals are shown as a function of time in Figure 4-7. Although the presence of a large number of data points results in there being significant non-zero fits to the residual, for the most part, there does not appear to be a practically meaningful temporal trend in the residuals. In addition, we show the monthly averaged X2 (both ANN and observed) in Figure 4-8 and Table 4-9. Fits are considerably improved following monthly averaging, with R^2 values ranging from 0.91 to 0.99, and standard error of 2.2 km and 2.5 km for the Sacramento and San Joaquin River branches.

Using a similar set of plots, we show the performance of the ANNs developed using Approach 2 for WY 1922-2012 (Figure 4-9 through Figure 4-11). In these plots, the data for the 1922-1967 is shown with a different color; this was done to highlight some events with very high X2 (i.e., salt water intrusion into the Delta) that occurred prior to the completion of the State Water Project. Figure 4-9 shows scatter plots of X2 conditioned on river and month on the scale of the X2 positions. The quality of these fits is assessed in Table 4-10. There is an overall R^2 of 0.94, i.e., slightly better than Approach 1, ranging from 0.82 to 0.96 depending on the month and river. Second, plots of the residuals against interpolated X2 show the agreement between modeled and interpolated values for various salinity conditions (Figure 4-10). Typically, the values with high residuals fall in the <50 Km or >100 km range of interpolated X2. The inclusion of data prior to 1944 in the analysis also brings in periods with very high X2 values that are often not captured through the ANN. Because each point represents a single day, many of the outlier points on the east (high X2) actually correspond to a single period in the 1930s where the X2 position was much further east than has typically been the case in subsequent years. Importantly, this is not the case for the high X2 residuals when X2<50 km. In this case, there are points that belong to both the pre- and post-project periods. The weakness of the ANN model in representing low values of X2, as in Approach 1, also remains in this version of the ANN. Other characteristics of the residuals (over time) are similar to those noted for Approach 1 (Figure 4-11). Over this time there is no systematic pattern in the residuals although there are instances of high values in the dry months of the pre-project period. Of the 24 river-month combinations possible, only 3 show the presence of a slope that is different from zero, and even in these cases the magnitude of the slope is small. As with Approach 1, we show the monthly averaged X2 (both ANN and observed) in Figure 4-12 and Table 4-11. Fits are improved following monthly averaging, although not as good as for the monthly averaging for Approach 1.

Taken together, both ANN modeling approaches appear to represent the X2 isohaline reasonably well for all but the most extreme conditions over a large period of record. Their utility for future applications depends on the relative performance of other available models, which are discussed below.

4.3 EVALUATION OF DAILY DSG MODEL PERFORMANCE

The daily DSG model is evaluated using a set of plots and tables that are consistent with the approach used for the ANN model above. This approach assumes constant values of S_0, S_b, and Φ_2 [2]. Figure 4-13 shows scatter plots of X2 conditioned on river and month on the scale of the X2 positions. The quality of these fits is assessed in Table 4-12. There is an overall R^2 of 0.92, ranging from 0.83 to 0.95 depending on the month and river. Plots of residuals (interpolated X2 subtracted from model X2) against interpolated X2 show agreement between modeled and interpolated values for various salinity conditions (Figure 4-14). Similar to the ANN model, values with high residuals fall in the >100 km or the <50 Km range of interpolated X2. Visually, the performance of the two models is similar and the numeric values of the fit statistics in Table 4-12 are needed for more direct comparison (such as the fit for individual months).

4.4 EVALUATION OF DAILY K-M MODEL PERFORMANCE

The K-M equation has also been developed in daily form (Kimmerer and Monismith, 1992):

$$\text{Daily X2}(t) = 10.16 + 0.945*\text{X2}(t-1) - 1.487*\log_{10}(Q_{out}(t)) \quad \text{Eq. 4.1}$$

This equation was used for comparison with the other methods discussed above (ANN and daily DSG model). The X2 value was not calculated for days when the Net Delta Outflow (Q_{out}) was negative, which occurred at various times in the early part of the record. The scatterplots of the modeled and interpolated X2, by month and river, are shown in Figure 4-16, and the fit statistics are shown in Table 4-13. The R^2 of the overall fit is 0.89, somewhat lower than for the ANN and the daily DSG model, and ranges from 0.76 to 0.93. The scatterplots show considerably greater noise in the fit, consistent with these numbers.

Given the generally superior fits of the ANN model and the daily DSG model, additional statistics on the daily K-M model are not shown here.

4.5 DETAILED EXAMINATION OF PERIODS WITH HIGH RESIDUALS

To gain a better understanding of the differences between interpolated and ANN-calculated X2 (Approach 2), selected periods with large residuals were examined more closely. We noticed that many of the days with large disagreement between interpolated and modeled X2 occurred closely spaced in time. Therefore we placed all days where the magnitude of the residual was larger than fifteen kilometers into groups separated by less than thirty days. Plots of several variables during a time window surrounding these groups (starting one year before the beginning of the group and ending thirty days after the end of the group) were developed. All instances of large residuals occurred near either the western or eastern extreme of the range of X2 positions.

An example showing one of the most significant departures between the model and interpolated value is shown in Figure 4-17. The first plot (upper panel) shows daily

[2] ϕ_1 = 441 (Sacramento River) , ϕ_1 = 483 (San Joaquin River), ϕ_2 = -0.189 (Sacramento River), ϕ_2 = -0.198 (San Joaquin River); S_b = 200 uS/cm, S_o = 350 uS/cm/km * X2

outflow. The second plot (middle panel) shows the daily X2 series for the interpolated values and for the values modeled by the DSG model or the Approach 2 ANN. The last plot (lower panels), shows the same isohaline positions superimposed on the salinity field given by the filled and cleaned salinity data. The extent of the grouping of high residual days is indicated by the shaded region. All exceedances greater than 15 km are shown in Appendix C. The example here displays limitations of the approach and the overall challenge of describing the early 1930's period with relatively limited data.

4.6 ANNs FOR SPECIFIC STATIONS

Some stations of significance for compliance, notably Emmaton and Jersey Point, are not as well described by focusing on the entire salinity gradient rather than individual stations. Thus, fits to EC at these specific stations are poorer than at other stations, or the X2 isohaline.

To address these issues, we developed targeted ANNs using the same inputs (except for distance from Golden Gate) and focused on the CDEC period, similar to Approach 1. Additional evaluation supported the use of a longer lag time in these ANNs, and a 60-day time lag was used. The station-specific approach generally resulted in substantially improved fits: The R^2 values for Jersey Point and Emmaton were both 0.92 (Table 4-14 and Table 4-15). These single-station ANNs are therefore recommended for the eastern locations. Additional diagnostics on the fit, by data source are shown for the two stations in Figure 4-18 through Figure 4-24. In general, the data span 2 orders of magnitude (<100 to >1,000 µS/cm), and the relative fits are poorer at the very low end of the range (i.e, as a percent of the target value. When the residuals are compared as a function of target EC (Figure 4-19 and Figure 4-23), the higher EC values correspond to larger errors. Compared over time, the residuals indicate a periodicity in some months, likely corresponding to the type of water year, but do not show a continuous time trend (Figure 4-20 and Figure 4-24).

4.7 SENSITIVITY TO RIO VISTA FLOW AND SEA LEVEL RISE

The trained station-based ANN models (Sacramento River Model, Approach 1 ANN) were used to project sensitivity of EC over distance under different flow conditions from Rio Vista: 5,000, 10,000, and 25,000 cfs. For the sensitivity analysis, the inputs for the ANN models are: distance, Rio Vista flow, Qwest flow, astronomical tide, and the residuals between the actual and astronomical tide. The model was run in steady state where the flow inputs were held constant at different values. The Qwest flow was specified as a function of Rio Vista flow (Qwest = 0.2666*QRio - 834.64, based on data from WY 1975-2012), and allowing a variation of $\pm$2,000 cfs about this value. The values for distance used were specified at 10 km intervals. The astronomical tide and tide residuals used were averages of the most recent 10 years (2002-2012, calendar years). The results represent sensitivity of EC over distance to changes in flow conditions for a set condition of average tidal conditions. Simulated EC as a function of distance decreases from Golden Gate, with lower EC under higher flows (Figure 4-26). Following the presentation in Jassby et al. (1995), simulated EC values were plotted as a function of standardized distance (X/X2). The approach results in different curves of "self-similarity" as Rio Vista flow increases (e.g., 25,000 cfs; Figure 4-27). Although not shown in this

form, previous results have shown the changing of the horizontal salinity structure at higher Delta outflows (Monismith et al., 2002).

Sea level at Golden Gate Bridge has increased at a rate of 0.08 inches per year over the past century (Fleenor et al., 2008). In the coming decades, the rate of sea level rise at Golden Gate is projected to further increase. The CALFED independent science board (ISB) has recommended the Delta Vision effort use a mid-range of sea level rise of 8–16 inches by 2050 and 28–39 inches by 2100. The trained ANN models (using Approach 1) were used to test sensitivity of changes in EC due to sea level rise of 6 inches (0.5 foot) and 12 inches (1 foot; Figure 4-28). As noted in Chapter 1, data-driven tools such as ANNs are best applied within the envelope of the training data. In this regard, an increase of 1 foot is within the training range of sea level used, although higher values may not result in reliable extrapolations. The results show some mixed changes in high salinity zones near Golden Gate, and a complex relationship with distance, but for most distances showed increases in EC due to sea level rise.

4.8 SENSITIVITY TO AIR PRESSURE AND QWEST FLOW

The difference between actual tide and the astronomical tide to a large degree is explained by the air pressure. The air pressure can affect mean sea level. A sensitivity of predicted salinity to air pressure using the trained station based ANN models (Approach 1) was performed to evaluate effects of this variable for different values of Rio Vista flow. The sensitivity was performed at a range of air pressure at 1,000 mbar, 1,015 mbar, and 1,030 mbar, bounding the observed air pressure values. The results show a small sensitivity to air pressure, with a slightly greater effect at lower flows, with higher salinity occurring at the same distance for lower pressure (with increased tidal range; Figure 4-29). Changes, when computed from a baseline of 1,015 mbar, show an approximate mirror image relationship with distance. Salinities are higher for the higher pressure case over most of the gradient.

A sensitivity analysis was also conducted to evaluate effects of Qwest flow on salinity, again using the Approach 1 ANN. Qwest flow reflects diversions from Delta and is a function of Rio Vista flow. A sensitivity of salinity to Qwest flow was performed to evaluate the effects of this variable. The sensitivity was performed at a range of Qwest flow at mean levels predicted from the Qwest and Rio Vista flow relationship, and a range of ± 2,000 cfs. The results show a relatively large sensitivity to Qwest flow, particularly under low flow conditions (Figure 4-30). Similar to the relationship with pressure, changes from the mean Qwest (i.e., based on the correlation between Qwest and Rio Vista flow) show an approximate mirror image relationship with distance. Salinities are higher when Qwest is lower, indicative of lower overall freshwater flow into the bay.

Table 4-1
Available Data for Training (Approach 1)

Station Name	Code	Distance	Time Period	Data Count
Sacramento River				
Point San Pablo	PSP	22	1974-2012	13355
Carquinez	CAR	45.5	1974-2012	13490
Martinez (USBR)	MBR	55	1974-2012	7847
Martinez	MRZ	54	1974-2012	5926
Port Chicago	PCT	64	1974-2012	13773
Mallard Island	MAL	75	1974-2012	13773
Collinsville	CLL	81	1974-2012	13773
Emmaton	EMM	92	1974-2012	13762
Rio Vista Bridge	RVB	101	1974-2012	13762
San Joaquin River				
Point San Pablo	PSP	22	1974-2012	13355
Carquinez	CAR	45.5	1974-2012	13490
Martinez (USBR)	MBR	55	1974-2012	7847
Martinez	MRZ	54	1974-2012	5926
Port Chicago	PCT	64	1974-2012	13773
Mallard Island	MAL	75	1974-2012	13773
Pittsburg	PTS	77	1974-2012	13744
Antioch	ANH	85.75	1974-2012	13773
Blind Point	BLP	92.85	1974-2012	13752
Jersey Point	JER	95.75	1974-2012	13736
Three Mile Slough @ SJR	TSL	100.4	1974-2012	13689
San Andreas Landing	SAL	109.2	1974-2012	13740

Table 4-2
Performance of Trained Salinity ANN Model (Approach 1)
ANN Salinity (µS/cm) = C1 + C2*Observed Salinity (µS/cm)

Station	Daily				Monthly			
	C2	C1	R^2	SE(µS/cm)	C2	C1	R^2	SE(µS/cm)
Sacramento River								
PSP	0.9331	2453.8	0.9556	2232.6	0.9942	82.5	0.9741	1686.2
CAR	0.9526	1153.0	0.9565	2382.9	0.9902	129.0	0.9781	1789.4
MRZ	0.9079	1706.7	0.9786	1760.4	0.9342	1273.1	0.9927	911.1
MBr	0.8936	1714.7	0.9557	2707.0	0.9102	1427.1	0.9674	2101.3
PCT	0.9049	1044.6	0.9583	2099.3	0.9318	760.5	0.9749	1480.7
MAL	0.8877	427.5	0.9556	1427.9	0.9002	358.8	0.9744	960.5
CLL	0.9027	332.7	0.9353	1129.3	0.9242	268.1	0.9601	814.3
EMM	1.0341	-28.0	0.8528	573.7	1.0816	-68.8	0.9239	462.4
RVB	2.8000	-374.4	0.5755	134.7	2.9650	-414.7	0.7184	445.9
San Joaquin River								
PSP	0.894	4044	0.945	2481.1	0.963	1332	0.965	1905.5
CAR	1.007	-875	0.956	2584.8	1.061	-2414	0.974	2107.4
MRZ	0.920	2276	0.973	1960.1	0.952	1755	0.990	1072.9
MBR	0.907	2285	0.952	2821.2	0.925	1972	0.965	2217.7
PCT	0.825	1822	0.954	2197.2	0.852	1540	0.972	1432.5
MAL	0.794	794	0.955	1440.3	0.809	712	0.972	904.0
PTS	0.799	766	0.940	1464.4	0.822	657	0.960	973.8
ANH	1.019	186	0.929	752.8	1.040	146	0.953	645.1
BLP	1.182	-65	0.868	506.2	1.206	-91	0.912	527.4
JER	1.330	-145	0.820	401.2	1.360	-168	0.883	484.6
TSL	1.279	-215	0.795	295.7	1.302	-230	0.889	314.2
SAL	1.459	-234	0.274	98.3	1.580	-260	0.370	380.0

Table 4-3
Nudging results for distance at CDEC stations

Sacramento River			
Station	**Distance (km)**	**Distance change (km)**	**Best distance (km)**
PSP	22	0.3	22.3
CAR	45.5	0.8	46.3
MRZ	54	-0.7	53.3
MBR	55	-0.6	54.4
PCT	64	-0.4	63.6
MAL	75	0	75
CLL	81	0	81
EMM	92	0	92
RVB	101	-3	98
San Joaquin River			
Station	**Distance (km)**	**Distance change (km)**	**Best distance (km)**
PSP	22	3	25
CAR	45.5	2.4	47.9
MRZ	54	0.7	54.7
MBR	55	1	56
PCT	64	0.5	64.5
MAL	75	3	78
PTS	77	2.3	79.3
ANH	85.75	-0.4	85.35
BLP	92.85	-3	89.85
JER	95.75	-3	92.75
TSL	100.4	-3	97.4
SAL	109.2	-3	106.2

Table 4-4
Performance of Trained Salinity ANN Model (Approach 1) after Nudging of Distance
ANN Salinity (µS/cm) = C1 + C2*Observed Salinity (µS/cm)

Station	Daily				Monthly			
	C2	C1	R^2	SE	C2	C1	R^2	SE
Sacramento River								
PSP	0.9461	2222.5	0.953	2246.4	1.002	42.1	0.970	1799
CAR	0.9636	13.4	0.958	2502.1	1.004	-1107.2	0.973	1924
MRZ	0.8942	2786.7	0.976	1697.3	0.923	2306.7	0.991	993
MBR	0.8914	2335.6	0.956	2447.4	0.912	2000.1	0.968	2024
PCT	0.8727	1385.4	0.957	1896.0	0.900	1098.2	0.974	1438
MAL	0.858	593.4	0.959	1195.9	0.870	519.5	0.974	915.9
CLL	0.890	610.9	0.937	1036.3	0.917	498.1	0.959	801.2
EMM	0.969	353.9	0.877	603.4	1.049	239.4	0.915	454.5
RVB	2.077	204.4	0.722	443.1	2.252	120.6	0.737	332
San Joaquin River								
PSP	0.9688	677.3	0.941	2618.5	1.041	-2086.9	0.963	2106.8
CAR	0.9314	-834.2	0.957	2465.3	0.971	-1918.6	0.974	1902.2
MRZ	0.8771	1679.5	0.975	1715.9	0.905	1227.5	0.990	1011.6
MBR	0.8701	1193.8	0.953	2471.9	0.890	869.1	0.966	2047.2
PCT	0.8480	1353.3	0.950	1972.9	0.877	1056.5	0.970	1500.0
MAL	0.6632	602.7	0.951	1016.1	0.679	502.1	0.971	743.5
PTS	0.6765	702.0	0.932	1095.3	0.704	550.2	0.958	826.8
ANH	0.9206	456.8	0.915	812.9	0.966	324.1	0.946	615.4
BLP	1.2499	270.9	0.871	722.3	1.328	133.6	0.913	550.7
JER	1.4175	174.6	0.845	641.3	1.525	46.6	0.891	477.9
TSL	1.4698	106.0	0.831	490.5	1.569	2.7	0.889	357.6
SAL	3.0970	-149.4	0.637	401.9	3.500	-302.9	0.737	292.0

Table 4-5
Available Data for ANN Model (Approach 2)

Station Name	Code	Distance	Time Periods	Data Count
Sacramento River				
Point Orient	PTO	19.8	Bulletin 23	12637
Point Davis	PTD	40.6	Bulletin 23	13327
Crocket	CRK	44.6	Bulletin 23	12521
Benicia	BEN	52.3	Bulletin 23	12567
Martinez	MRZ	52.6	Bulletin 23	12621
Bulls Head Point	BHP	54.7	Bulletin 23	7680
West Suisun	WSN	59.5	Bulletin 23	12032
Bay Point	BPT	64.2	Bulletin 23	9100
Port Chicago	PCT	66	Bulletin 23	12429
O. and A. Ferry	OAF	74.8	Bulletin 23	13370
Collinsville	CLL	81.8	Bulletin 23	13609
Emmaton	EMM	92.9	Bulletin 23	13104
Threemile Slough Bridge	TSB	96.6	Bulletin 23	13097
Rio Vista Bridge	RVB	102.2	Bulletin 23	13229
Isleton Bridge	ITB	110.6	Bulletin 23	7491
Point San Pablo	PSP	22	CDEC	16839
Carquinez	CAR	45.5	CDEC	17010
Martinez	MRZ	54	CDEC	6033
Port Chicago	PCT	64	CDEC	17389
Mallard	MAL	75	CDEC	17505
Collinsville	CLL	81	CDEC	16985
Emmaton	EMM	92	CDEC	17420
Rio Vista Bridge	RVB	101	CDEC	17420

Table 4-5 (continued)
Available Data for ANN Model (Approach 2)

Station Name	Code	Distance	Time Periods	Data Count
San Joaquin River				
Point Orient	PTO	19.8	Bulletin 23	12637
Point Davis	PTD	40.6	Bulletin 23	13327
Crocket	CRK	44.6	Bulletin 23	12521
Benicia	BEN	52.3	Bulletin 23	12567
Martinez	MRZ	52.6	Bulletin 23	12621
Bulls Head Point	BHP	54.7	Bulletin 23	7681
West Suisun	WSN	59.5	Bulletin 23	12032
Bay Point	BPT	64.2	Bulletin 23	9100
Port Chicago	PCT	66	Bulletin 23	12429
O. and A. Ferry	OAF	74.8	Bulletin 23	13370
Antioch	ANH	88.4	Bulletin 23	13315
Antioch Bridge	ANB	93.7	Bulletin 23	12904
Jersey Point	JER	98.8	Bulletin 23	13272
False River	FRV	101.2	Bulletin 23	12171
Oulton Point	OPT	108.1	Bulletin 23	5395
San Andreas Landing	SAL	113.1	Bulletin 23	5395
Webb Pump	WBP	115.9	Bulletin 23	3659
Point San Pablo	PSP	22	CDEC	16839
Carquinez	CAR	45.5	CDEC	17010
Martinez	MRZ	54	CDEC	6033
Port Chicago	PCT	64	CDEC	17389
Mallard Island	MAL	75	CDEC	17505
Antioch	ANH	85.75	CDEC	17561
Blind Point	BLP	92.85	CDEC	17540
Jersey Point	JER	95.75	CDEC	17388
Threemile Slough Bridge	TSL	100.4	CDEC	17526
San Andreas Landing	SAL	109.2	CDEC	17405

Table 4-6
Performance of Trained Salinity ANN Model (Approach 2)
ANN Salinity (µS/cm) = C1 + C2*Observed Salinity (µS/cm)

Station	Source	Daily				Monthly			
		C2	C1	R^2	SE	C2	C1	R^2	SE
Sacramento River									
PTO	Bulletin 23	0.946	6604.1	0.5638	3636.6	0.998	5804.6	0.6582	2990.6
PTD	Bulletin 23	0.941	3210.6	0.8587	2835.3	0.947	3117	0.898	2293.8
CRK	Bulletin 23	1.04	860.9	0.8925	2689	1.031	1033.1	0.9198	2205.1
BEN	Bulletin 23	1.068	2165.9	0.915	2551.4	1.067	2172.8	0.9459	1960.8
MRZ	Bulletin 23	1.097	2786.6	0.9071	2670.8	1.104	2719.5	0.9419	2033
BHP	Bulletin 23	0.938	1099	0.9413	2149	0.939	1116.1	0.9664	1544.7
WSN	Bulletin 23	0.981	1634	0.9152	2282.1	1.004	995.7	0.9144	2282.6
BPT	Bulletin 23	0.923	787.9	0.9437	1811.1	0.949	217.6	0.9326	1935.3
PCT	Bulletin 23	0.905	-646.5	0.9505	1518.1	0.926	-1070.3	0.9524	1488.7
OAF	Bulletin 23	0.834	297.1	0.9689	914.5	0.832	275.6	0.9752	800.8
CLL	Bulletin 23	0.755	269.7	0.9711	607	0.756	256.2	0.981	492.4
EMM	Bulletin 23	0.645	160.6	0.9382	420.2	0.658	152.7	0.9661	323.6
TSB	Bulletin 23	0.684	196.8	0.9066	385.2	0.711	180.4	0.9505	298.4
RVB	Bulletin 23	0.723	201.9	0.8646	289.4	0.73	194.1	0.9126	260.7
ITB	Bulletin 23	0.789	158.5	0.8932	142	0.634	194.7	0.9162	157.5
PSP	CDEC	0.809	3631.2	0.6865	4086.3	0.833	2348.2	0.736	3655.7
CAR	CDEC	1.058	-2176	0.9001	3068.6	1.075	-2740.5	0.9221	2665.5
MRZ	CDEC	1.167	-175.2	0.9608	1974.3	1.21	-627.2	0.9894	1012.5
PCT	CDEC	1.106	143.9	0.9382	2021.5	1.128	-192.7	0.9655	1478.8
MAL	CDEC	0.98	9.7	0.9694	816.5	0.993	-15.9	0.9823	612.1
CLL	CDEC	0.932	188.5	0.9653	546.7	0.955	156.5	0.9742	466.2
EMM	CDEC	0.826	231.9	0.9817	116.4	0.833	212.2	0.9928	72.4
RVB	CDEC	1.155	198.2	0.653	127.3	1.198	179.1	0.7457	107

Table 4-6 (continued)
Performance of Trained Salinity ANN Model (Approach 2)
ANN Salinity (µS/cm) = C1 + C2*Observed Salinity (µS/cm)

Station	Source	Daily				Monthly			
		C2	C1	R^2	SE	C2	C1	R^2	SE
San Joaquin River									
PTO	Bulletin 23	1.045	4368.5	0.550	4030.7	1.097	3581.3	0.794	2463.5
PTD	Bulletin 23	0.917	3276.4	0.831	3009.6	0.924	3152.1	0.935	2503.7
CRK	Bulletin 23	1.002	1075.9	0.874	2787.6	0.993	1247.5	0.950	2219.2
BEN	Bulletin 23	1.032	2131.4	0.912	2448.6	1.028	2182.2	0.973	1741.5
MRZ	Bulletin 23	1.060	2738.8	0.903	2574.4	1.063	2713.9	0.971	1746.9
BHP	Bulletin 23	0.920	956.2	0.939	2040.8	0.914	1009.8	0.983	1559.0
WSN	Bulletin 23	0.955	1521.9	0.921	2096.8	0.972	966.6	0.958	2104.3
BPT	Bulletin 23	0.945	487.0	0.947	1636.8	0.965	-66.0	0.968	1790.1
PCT	Bulletin 23	0.900	-693.0	0.950	1460.2	0.916	-1067.2	0.975	1544.3
OAF	Bulletin 23	0.930	218.1	0.968	872.3	0.918	221.4	0.985	904.0
ANH	Bulletin 23	0.890	65.5	0.919	704.0	0.854	87.2	0.963	786.0
ANB	Bulletin 23	1.081	40.6	0.840	439.1	1.226	-27.2	0.896	445.5
JER	Bulletin 23	1.231	-16.5	0.867	516.1	1.120	31.4	0.922	485.0
FRV	Bulletin 23	0.536	216.1	0.753	93.1	0.867	133.5	0.577	280.8
OPT	Bulletin 23	0.528	210.5	0.636	82.4	0.501	220.5	0.823	121.4
SAL	Bulletin 23	0.691	196.4	0.352	72.5	0.706	196.2	0.666	57.7
WBP	Bulletin 23	0.976	-83.5	0.756	520.6	0.720	40.3	0.848	574.2
PSP	CDEC	0.921	542.9	0.717	4317.4	0.956	-815.0	0.875	3467.4
CAR	CDEC	1.023	-1684.4	0.909	2811.5	1.047	-2325.1	0.966	2158.2
MRZ	CDEC	1.104	-61.5	0.964	1795.7	1.119	-293.7	0.993	936.6
PCT	CDEC	1.033	250.0	0.938	1884.2	1.056	21.2	0.981	1309.6
MAL	CDEC	0.943	134.8	0.969	788.6	0.948	100.4	0.992	582.4
PTS	CDEC	0.931	210.7	0.947	906.3	0.944	149.3	0.984	702.9
ANH	CDEC	0.998	44.2	0.976	300.5	1.000	42.6	0.994	204.9
BLP	CDEC	0.953	35.5	0.962	183.2	0.946	42.9	0.987	146.6
JER	CDEC	0.971	49.5	0.944	156.6	0.961	56.2	0.981	122.5
TSL	CDEC	0.744	124.4	0.916	103.0	0.734	129.6	0.972	103.1
SAL	CDEC	0.716	169.0	0.430	79.8	0.755	161.3	0.738	61.1

Table 4-7
Percentage of Interpolated X2 Values that fall between different distance ranges

Branch	Month	Percent of X2 West of 50km	Percent of X2 between 50-100km	Percent of X2 East of 100km
SAC	Dec	3.8%	96.2%	0.0%
SAC	Jan	8.9%	91.1%	0.0%
SAC	Feb	15.4%	84.6%	0.0%
SAC	Mar	15.6%	84.4%	0.0%
SAC	Apr	9.2%	90.8%	0.0%
SAC	May	7.0%	93.0%	0.0%
SAC	Jun	3.1%	96.9%	0.0%
SAC	Jul	0.0%	97.8%	2.2%
SAC	Aug	0.0%	97.2%	2.8%
SAC	Sep	0.0%	95.6%	4.4%
SAC	Oct	0.0%	98.0%	2.0%
SAC	Nov	0.5%	99.2%	0.3%
SJR	Dec	3.9%	95.2%	0.9%
SJR	Jan	8.9%	91.1%	0.0%
SJR	Feb	15.4%	84.6%	0.0%
SJR	Mar	15.6%	84.4%	0.0%
SJR	Apr	9.2%	90.8%	0.0%
SJR	May	7.0%	93.1%	0.0%
SJR	Jun	3.1%	95.9%	0.9%
SJR	Jul	0.0%	97.7%	2.3%
SJR	Aug	0.0%	95.9%	4.1%
SJR	Sep	0.0%	93.7%	6.3%
SJR	Oct	0.0%	98.1%	1.8%
SJR	Nov	0.6%	98.6%	0.8%

Table 4-8
Scatterplot Statistics of ANN Model (Approach 1), Grouped by River and Month
ANN X2 (km) = C1 + C2*Observed X2 (km)

Branch	Month	C2	C1	R^2	Standard Error (km)
All	All	0.93	7.6	0.920	3.6
SAC	Jan	0.84	13	0.860	4.9
SAC	Feb	0.83	14	0.850	5.1
SAC	Mar	0.85	12	0.790	5.6
SAC	Apr	0.93	6.9	0.870	3.9
SAC	May	0.95	5.2	0.930	3.1
SAC	Jun	0.97	3.1	0.960	2.1
SAC	Jul	0.94	6.7	0.920	2.5
SAC	Aug	1	1.6	0.880	2.5
SAC	Sep	1	1.3	0.900	2.5
SAC	Oct	0.93	6.7	0.930	2.1
SAC	Nov	0.92	8.2	0.900	2.9
SAC	Dec	0.95	5.1	0.930	3.4
SAC	all	0.92	7.9	0.910	3.6
SJR	Jan	0.89	8.9	0.910	4.3
SJR	Feb	0.88	10	0.880	4.8
SJR	Mar	0.89	11	0.880	4.3
SJR	Apr	0.9	9.8	0.920	2.9
SJR	May	0.93	7.1	0.930	3.2
SJR	Jun	0.91	9.7	0.950	2.4
SJR	Jul	0.89	12	0.920	2.4
SJR	Aug	0.96	5.8	0.890	2.6
SJR	Sep	0.91	9.7	0.920	2.5
SJR	Oct	0.97	5.8	0.940	2.3
SJR	Nov	0.95	6.7	0.880	3.6
SJR	Dec	0.89	9.2	0.920	3.7
SJR	All	0.93	7.6	0.930	3.5

Table 4-9
Scatterplot Statistics of ANN Model (Approach 1) Averaged Monthly, Grouped by River and Month
ANN Monthly X2 (km) = C1 + C2*Observed Monthly X2 (km)

Branch	Month	C2	C1	R^2	Standard Error (km)
All	All	0.96	4.9	0.97	2.3
SAC	Jan	0.87	11	0.96	2.2
SAC	Feb	0.85	12	0.96	2.4
SAC	Mar	0.88	10	0.94	2.5
SAC	Apr	0.99	3.4	0.98	1.7
SAC	May	0.99	2.5	0.98	1.7
SAC	Jun	0.99	1.9	0.99	1.1
SAC	Jul	0.96	4.8	0.96	1.7
SAC	Aug	1	-1.7	0.93	1.9
SAC	Sep	1	-1.5	0.95	1.9
SAC	Oct	0.96	4.2	0.97	1.4
SAC	Nov	0.93	7.3	0.93	2.3
SAC	Dec	0.97	3.5	0.98	1.7
SAC	all	0.95	5.8	0.97	2
SJR	Jan	0.99	1	0.96	2.8
SJR	Feb	0.94	5.8	0.93	3.5
SJR	Mar	0.94	7.2	0.96	2.4
SJR	Apr	1	1.9	0.96	2.3
SJR	May	0.97	4.7	0.97	2.1
SJR	Jun	0.92	9.1	0.98	1.6
SJR	Jul	0.92	10	0.96	1.9
SJR	Aug	0.97	4.8	0.92	2.2
SJR	Sep	0.92	9.3	0.95	2
SJR	Oct	0.98	5.5	0.97	1.5
SJR	Nov	0.95	6.8	0.91	2.8
SJR	Dec	0.92	7.2	0.95	2.6
SJR	All	0.97	4.3	0.96	2.6

Table 4-10
Scatterplot Statistics of ANN Model (Approach 2), Grouped by River and Month
ANN X2 (km) = C1 + C2*Observed X2 (km)

Branch	Month	C2	C1	R^2	Standard Error (km)
All	All	0.94	4.4	0.94	3.5
SAC	Jan	0.89	7.1	0.92	3.7
SAC	Feb	0.9	6.1	0.9	3.8
SAC	Mar	0.86	8.6	0.88	3.7
SAC	Apr	0.89	6.6	0.9	3.5
SAC	May	0.94	4.2	0.92	3.3
SAC	Jun	0.95	3.5	0.96	2.6
SAC	Jul	0.94	4.8	0.94	2.4
SAC	Aug	0.94	5.1	0.91	2.6
SAC	Sep	0.95	4.2	0.93	2.5
SAC	Oct	0.89	9.1	0.91	2.5
SAC	Nov	0.93	5.8	0.92	2.7
SAC	Dec	0.93	5	0.93	3.1
SAC	All	0.95	3.7	0.95	3.1
SJR	Jan	0.9	6.1	0.9	4.2
SJR	Feb	0.92	5.1	0.9	4
SJR	Mar	0.88	7.6	0.88	3.9
SJR	Apr	0.9	6.1	0.89	3.8
SJR	May	0.93	4.5	0.92	3.4
SJR	Jun	0.95	3.3	0.95	2.8
SJR	Jul	0.95	4.2	0.94	2.7
SJR	Aug	0.93	5.7	0.91	3
SJR	Sep	0.9	8.7	0.89	3.9
SJR	Oct	0.78	18	0.81	4.1
SJR	Nov	0.84	13	0.85	3.9
SJR	Dec	0.85	11	0.88	4.4
SJR	All	0.93	5	0.93	3.8

Table 4-11
Scatterplot Statistics of ANN Model (Approach 2) Averaged Monthly, Grouped by River and Month
ANN Monthly X2 (km) = C1 + C2*Observed Monthly X2 (km)

Branch	Month	C2	C1	R^2	Standard Error (km)
All	All	0.96	2.9	0.95	3
SAC	Jan	0.94	3.7	0.94	2.8
SAC	Feb	0.96	1.6	0.94	2.9
SAC	Mar	0.94	3.6	0.95	2.1
SAC	Apr	0.95	2.9	0.94	2.6
SAC	May	0.97	1.7	0.95	2.6
SAC	Jun	0.96	2.9	0.96	2.4
SAC	Jul	0.95	4.1	0.97	1.9
SAC	Aug	0.91	7.7	0.92	2.5
SAC	Sep	0.96	3.8	0.94	2.3
SAC	Oct	0.89	9.3	0.93	2.1
SAC	Nov	0.93	5.9	0.94	2.1
SAC	Dec	0.97	1.8	0.95	2.5
SAC	All	0.98	1.5	0.97	2.5
SJR	Jan	0.96	1.7	0.94	3
SJR	Feb	0.99	0.31	0.94	2.9
SJR	Mar	0.97	2.2	0.95	2.3
SJR	Apr	0.98	0.95	0.94	2.8
SJR	May	0.97	2	0.95	2.6
SJR	Jun	0.93	5.3	0.91	3.8
SJR	Jul	0.92	6	0.96	2.2
SJR	Aug	0.91	8.2	0.92	3.3
SJR	Sep	0.88	10	0.88	4
SJR	Oct	0.67	27	0.79	4.3
SJR	Nov	0.81	15	0.86	3.5
SJR	Dec	0.88	9.1	0.89	3.8
SJR	All	0.94	4.1	0.94	3.5

Table 4-12
Scatterplot Statistics of Daily DSG Model, Grouped by River and Month.
DSG Daily X2 (km) = C1 + C2*Observed Daily X2 (km)

Branch	Month	C2	C1	R^2	Standard Error (km)
All	All	0.94	5.5	0.92	4
SAC	Jan	0.8	12	0.87	4.2
SAC	Feb	0.79	13	0.86	4.1
SAC	Mar	0.78	15	0.83	4
SAC	Apr	0.83	12	0.87	3.6
SAC	May	0.86	10	0.91	3.2
SAC	Jun	0.86	11	0.95	2.7
SAC	Jul	0.92	7	0.95	2.2
SAC	Aug	1	0.53	0.94	2.2
SAC	Sep	1.1	-5.9	0.95	2.6
SAC	Oct	1.1	-6.8	0.91	3.1
SAC	Nov	1	-1.5	0.91	3.2
SAC	Dec	0.93	3.6	0.91	3.6
SAC	All	0.96	3.6	0.92	3.9
SJR	Jan	0.81	12	0.87	4.4
SJR	Feb	0.8	12	0.86	4.2
SJR	Mar	0.8	14	0.83	4.2
SJR	Apr	0.85	11	0.88	3.8
SJR	May	0.89	8.3	0.92	3.3
SJR	Jun	0.88	10	0.94	2.9
SJR	Jul	0.86	12	0.94	2.5
SJR	Aug	0.81	17	0.92	2.4
SJR	Sep	0.8	19	0.92	2.9
SJR	Oct	0.87	13	0.88	3.4
SJR	Nov	0.91	7	0.88	3.6
SJR	Dec	0.85	9.2	0.88	4.4
SJR	All	0.91	7.1	0.92	4

Table 4-13
Scatterplot Statistics of Daily K-M Model, Grouped by River and Month.
KM Daily X2 (km) = C1 + C2*Observed Daily X2 (km)

Branch	Month	C2	C1	R^2	Standard Error (km)
All	All	0.91	6.3	0.89	4.6
SAC	Jan	0.73	18	0.83	4.4
SAC	Feb	0.76	15	0.8	4.9
SAC	Mar	0.83	11	0.79	4.9
SAC	Apr	0.87	8.2	0.84	4.4
SAC	May	0.91	5.6	0.89	3.9
SAC	Jun	0.95	3.4	0.92	3.4
SAC	Jul	1.1	-4.2	0.89	3.6
SAC	Aug	1.1	-9.7	0.8	4.3
SAC	Sep	0.94	6.1	0.89	3.1
SAC	Oct	0.78	17	0.87	2.7
SAC	Nov	0.75	18	0.9	2.4
SAC	Dec	0.75	17	0.9	3.1
SAC	All	0.91	6.3	0.9	4.3
SJR	Jan	0.74	17	0.82	4.9
SJR	Feb	0.8	12	0.79	5.5
SJR	Mar	0.87	8.4	0.78	5.4
SJR	Apr	0.92	4.6	0.83	4.9
SJR	May	0.97	1.2	0.89	4.3
SJR	Jun	1	-1.1	0.93	3.6
SJR	Jul	1.1	-6.7	0.91	3.4
SJR	Aug	1	-2.1	0.84	4.1
SJR	Sep	0.79	18	0.86	3.8
SJR	Oct	0.72	22	0.81	3.8
SJR	Nov	0.7	23	0.84	3.3
SJR	Dec	0.69	21	0.84	4.2
SJR	All	0.91	6.3	0.89	4.9

Table 4-14
Scatterplot Statistics of Jersey Point ANN Model, Grouped by Month.
ANN Salinity (µS/cm) = C1 + C2*Observed Salinity (µS/cm)

Month	C2	C1	R^2	Standard Error (µS/cm)
Jan	0.93	58	0.9	190
Feb	0.9	27	0.94	140
Mar	0.89	29	0.94	110
Apr	1	1.9	0.95	100
May	1	15	0.93	110
Jun	0.94	33	0.95	120
Jul	0.96	40	0.94	150
Aug	0.94	77	0.91	190
Sep	0.89	150	0.9	220
Oct	0.88	130	0.9	210
Nov	0.91	130	0.88	270
Dec	0.9	140	0.86	300
all	0.94	55	0.92	190

Table 4-15
Scatterplot Statistics of Emmaton ANN Model, Grouped by Month.
ANN Salinity (µS/cm) = C1 + C2*Observed Salinity (µS/cm)

Month	C2	C1	R^2	Standard Error (µS/cm)
Jan	0.82	91	0.88	280
Feb	0.82	64	0.91	230
Mar	0.92	17	0.92	190
Apr	1	4.5	0.91	200
May	0.94	40	0.94	190
Jun	0.92	52	0.95	230
Jul	0.95	35	0.94	240
Aug	0.93	68	0.91	290
Sep	0.95	72	0.89	330
Oct	0.88	190	0.9	400
Nov	0.91	120	0.87	460
Dec	0.92	110	0.9	410
All months	0.92	63	0.92	310

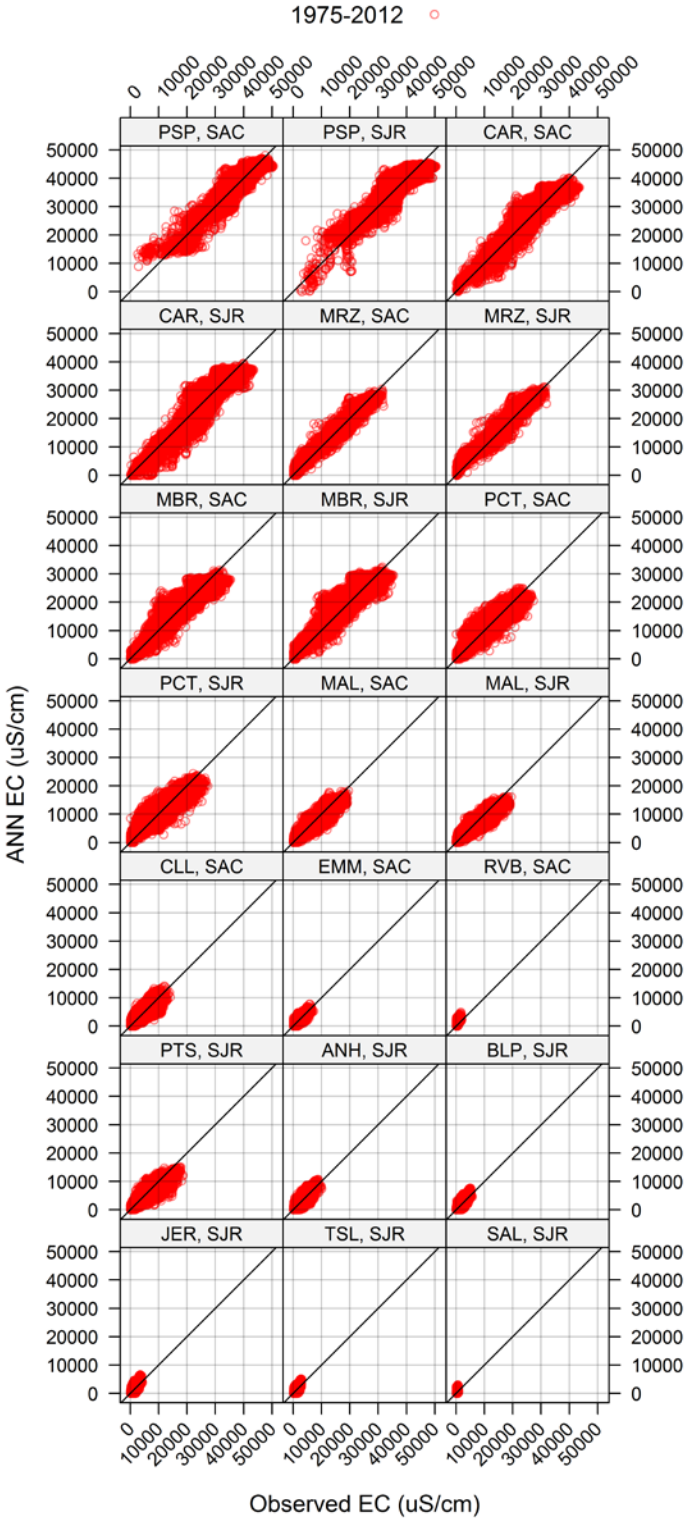

Figure 4-1 **Scatterplot of the stationwise EC calculated from the ANN model (Approach 1) and observed data for the Sacramento and San Joaquin River branches. Data from Oct 1974 to June 2012 (WY 1975-2012). The solid line is the 1:1 slope.**

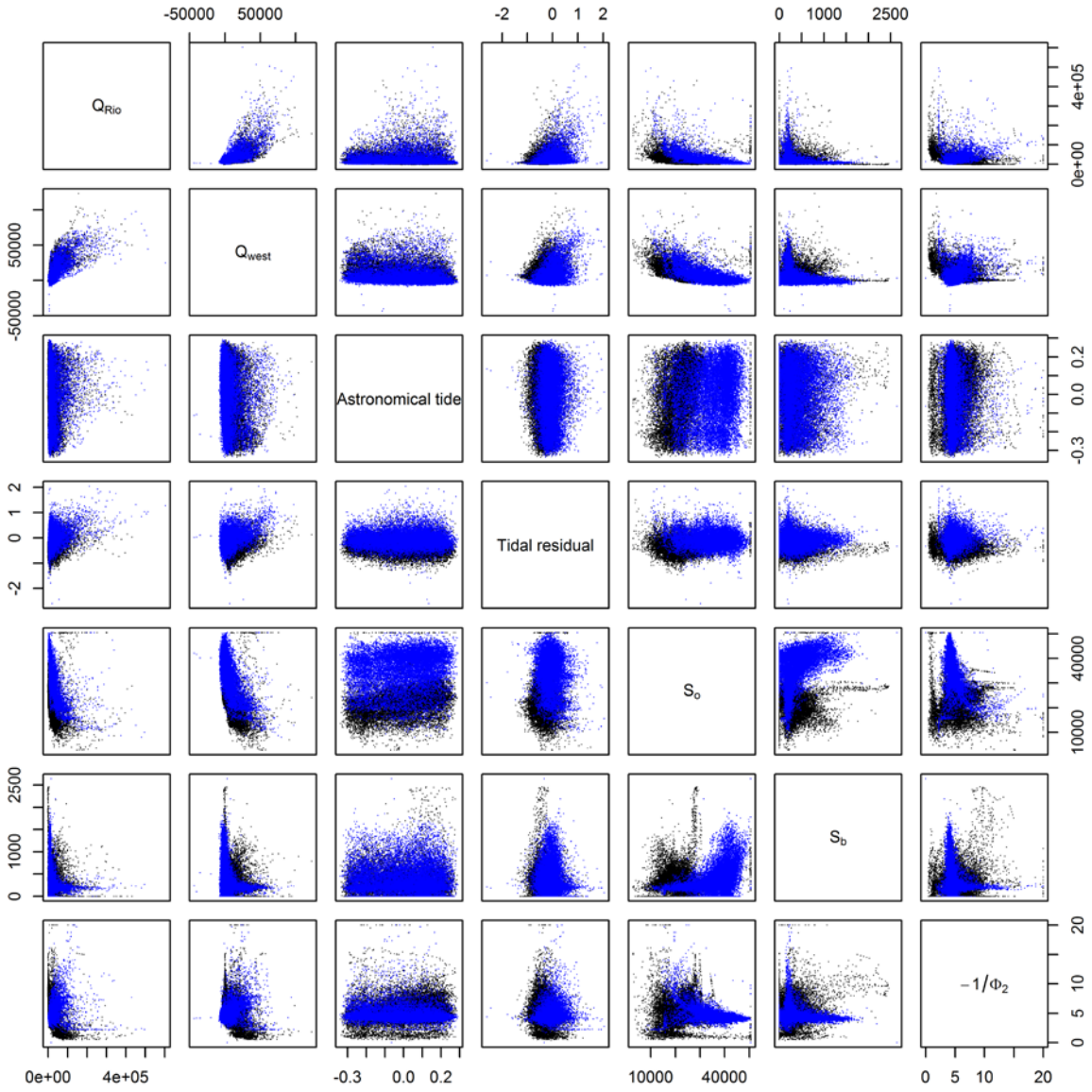

Figure 4-2 **Scatterplot of the DSG model parameters calculated from the ANN model for the Sacramento River stations. Blue points symbolize WY 1922-1967 data, and black points the WY 1968-2012 data.**

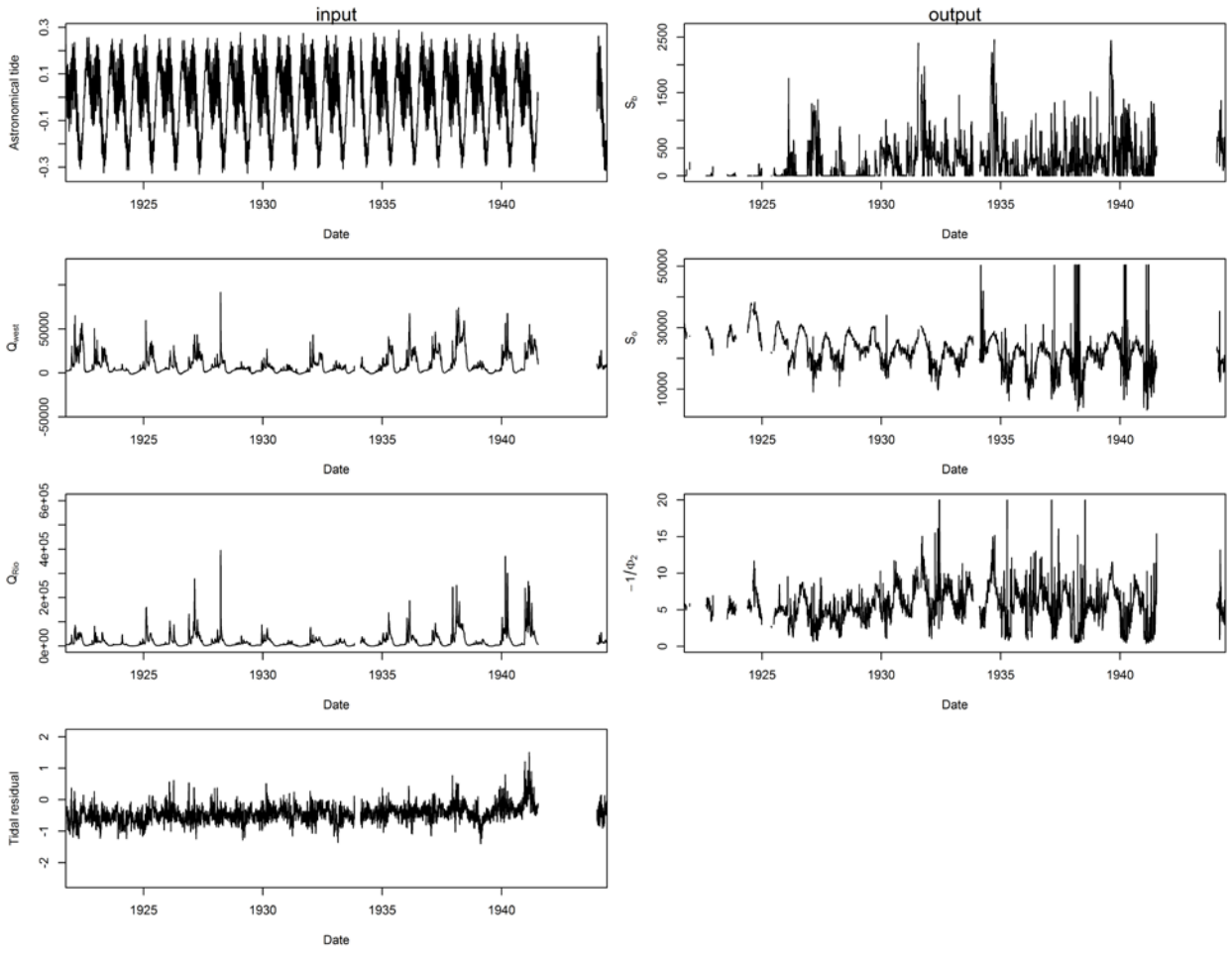

Figure 4-3 **Time series plot of the DSG model parameters calculated from the ANN model for the Sacramento River stations (each page shows a different time interval).**

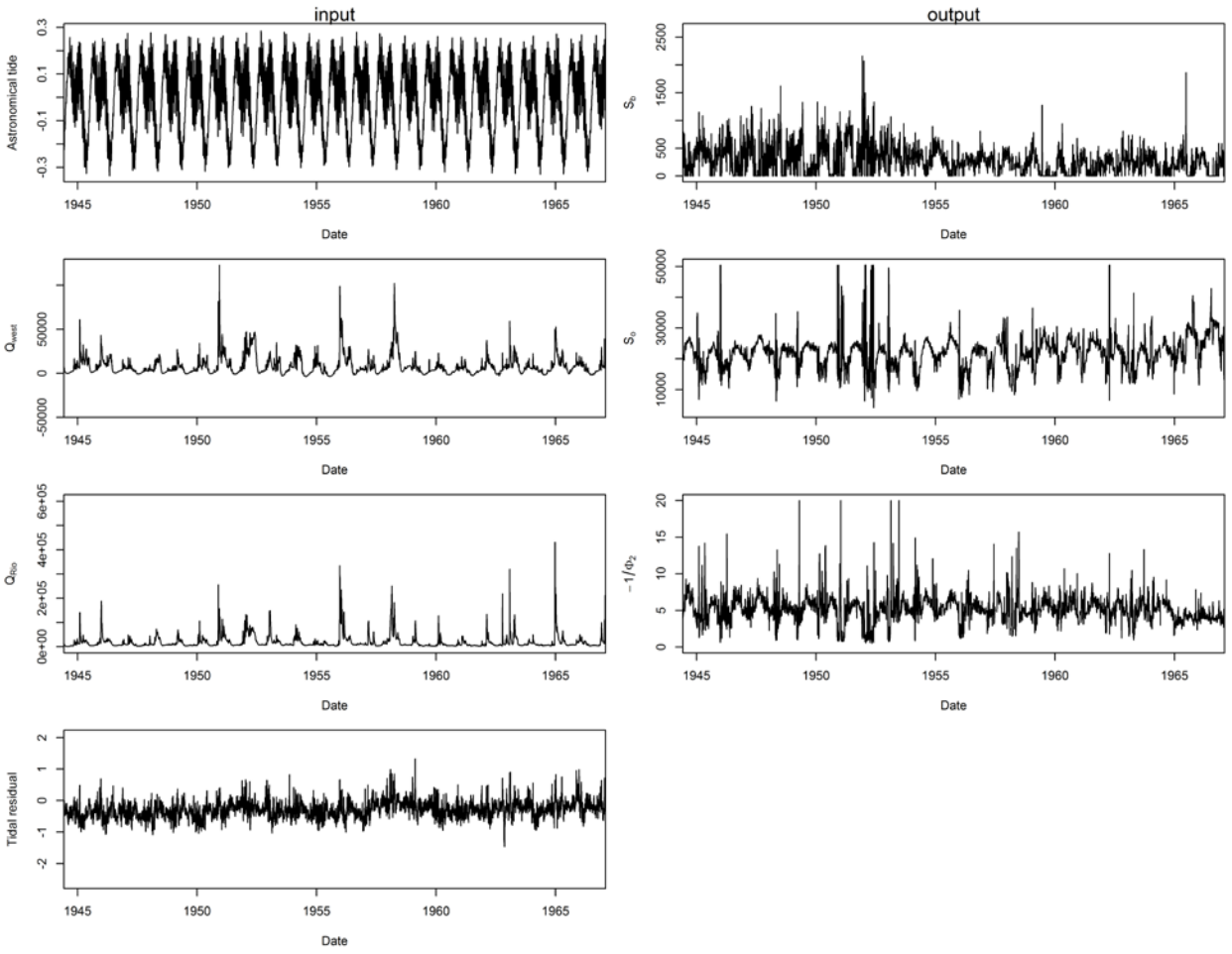

Figure 4-3 (cont'd). Time series plot of the DSG model parameters calculated from the ANN model for the Sacramento River stations (each page shows a different time interval).

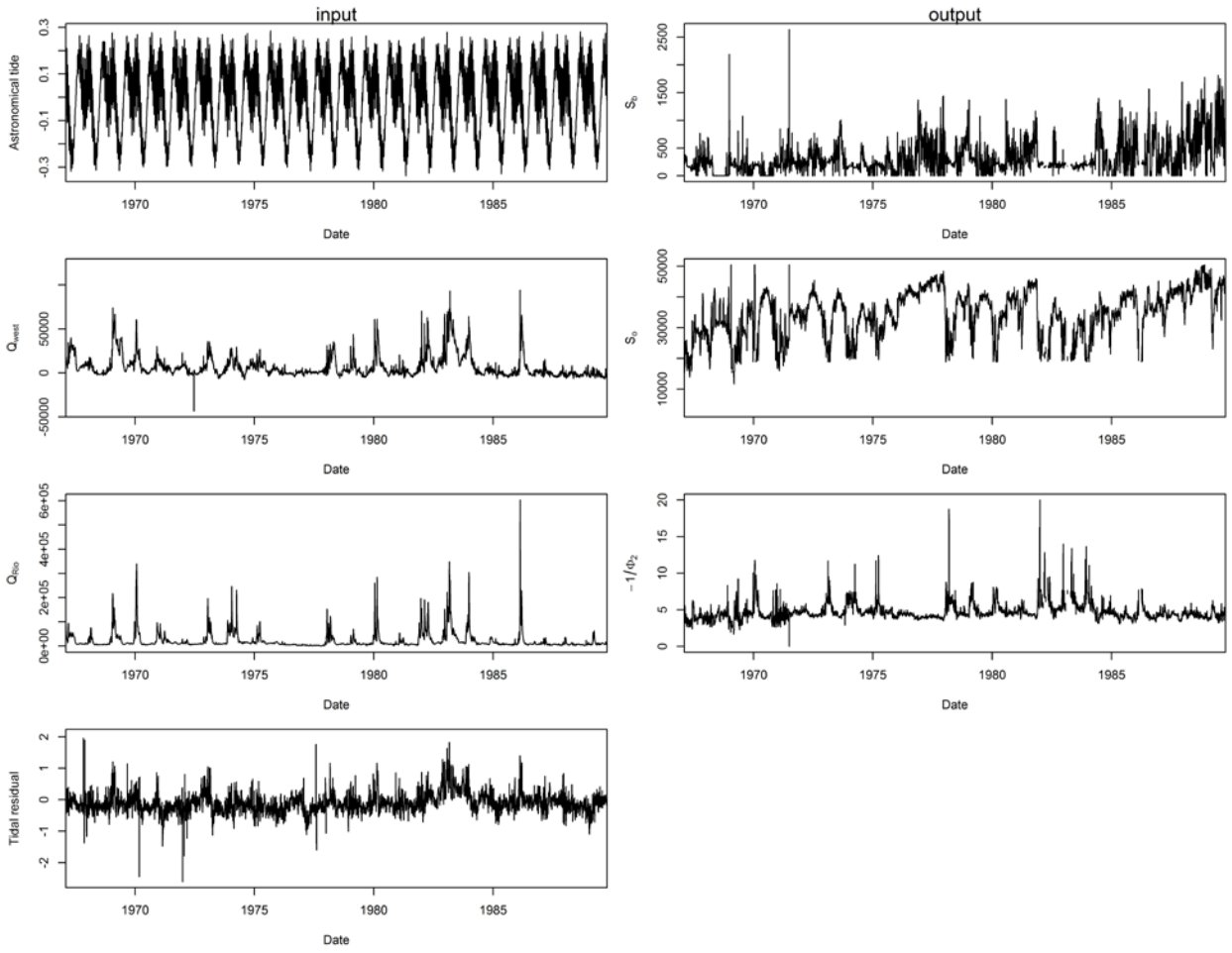

Figure 4-3 (cont'd). Time series plot of the DSG model parameters calculated from the ANN model for the Sacramento River stations (each page shows a different time interval).

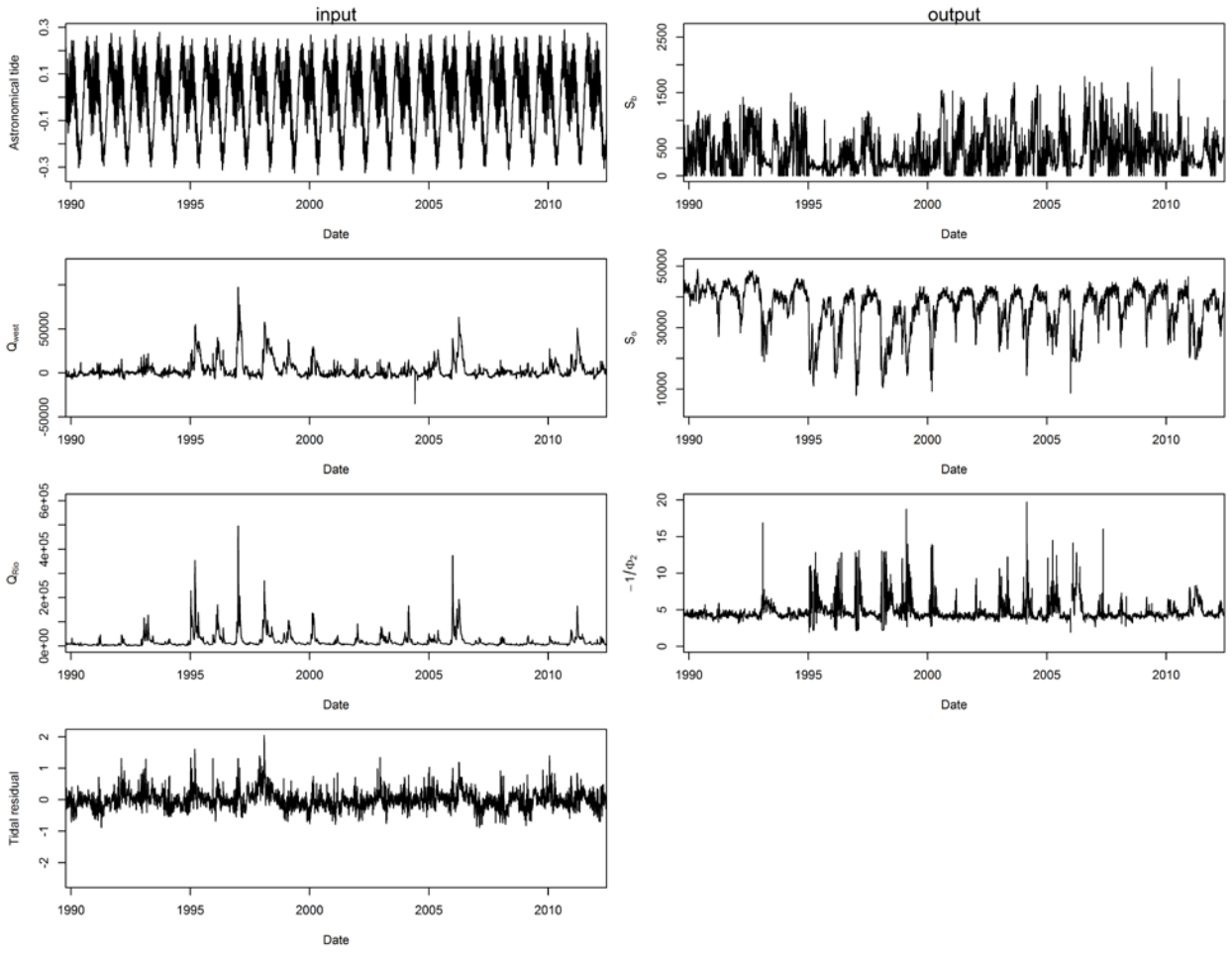

Figure 4-3 (cont'd). Time series plot of the DSG model parameters calculated from the ANN model for the Sacramento River stations (each page shows a different time interval).

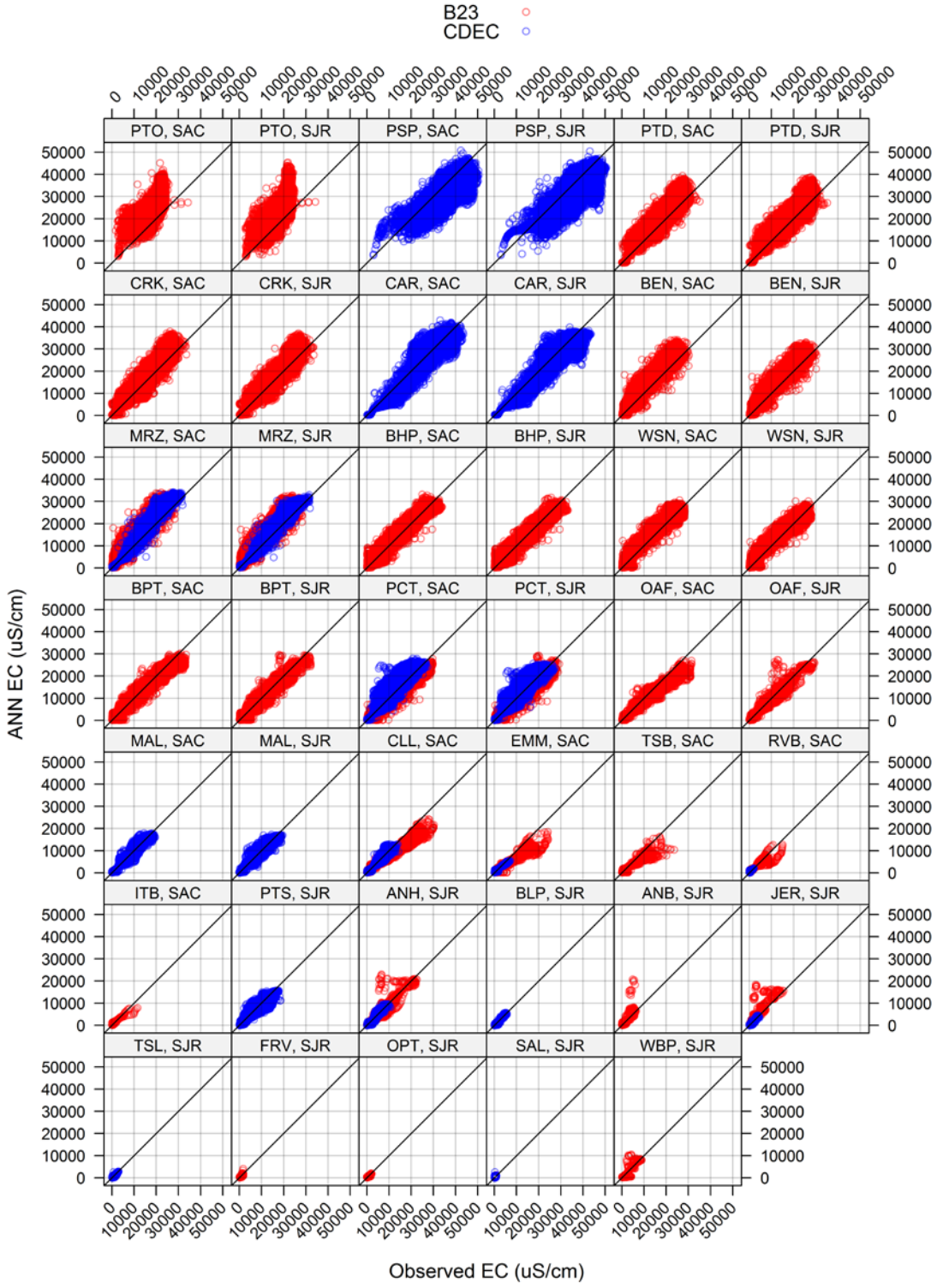

Figure 4-4 **Scatterplot of EC calculated from the ANN model (Approach 2) and observed data for the Sacramento and San Joaquin River branches. Data from Oct 1974 to June 2012 (WY 1975-2012). The solid line is the 1:1 slope. Red dots and blue dots represent Bulletin 23 and CDEC data, respectively.**

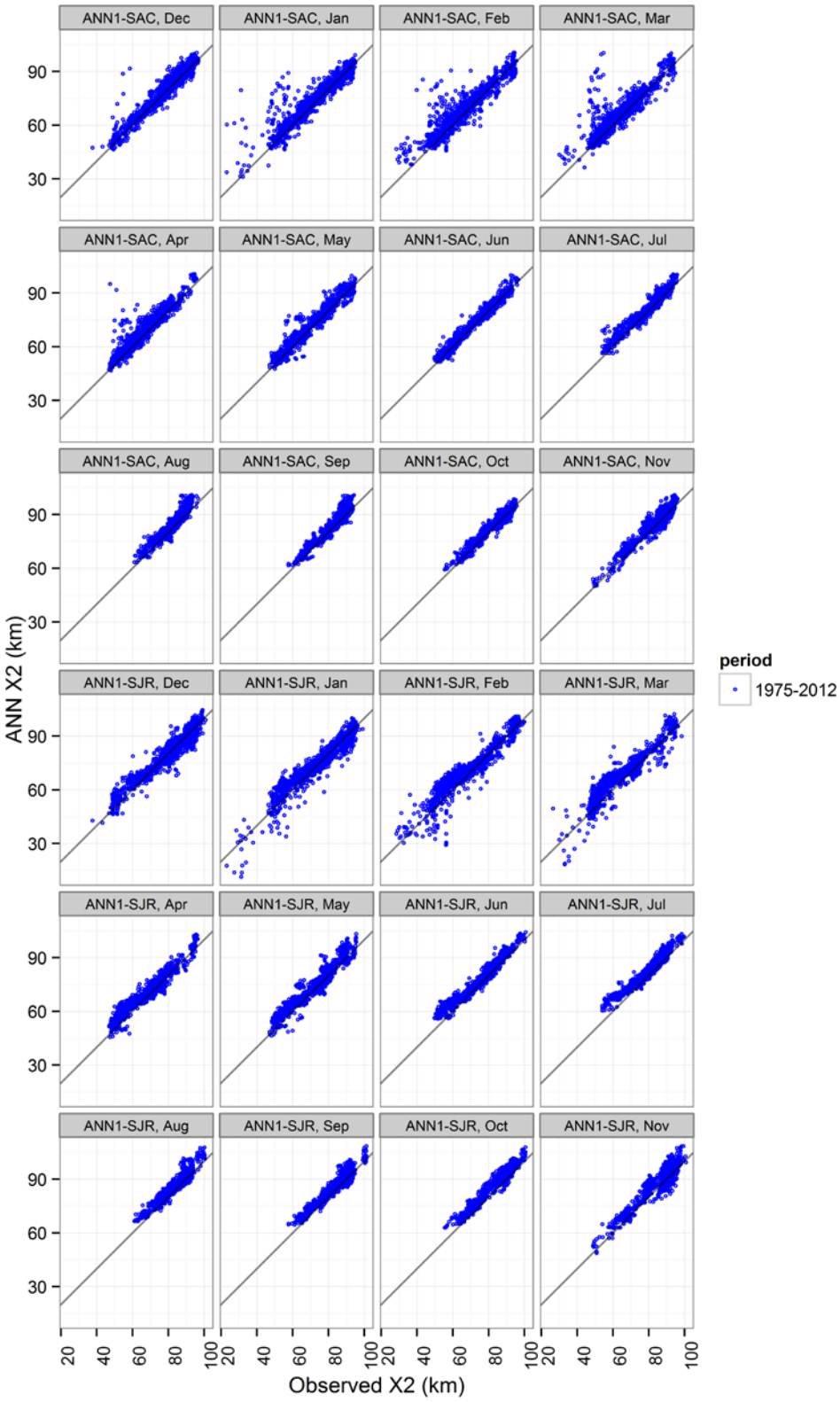

Figure 4-5 **Scatterplot of the interpolated daily X2 and X2 calculated from the ANN model (Approach 1) for the Sacramento and San Joaquin River branches, grouped by month.**

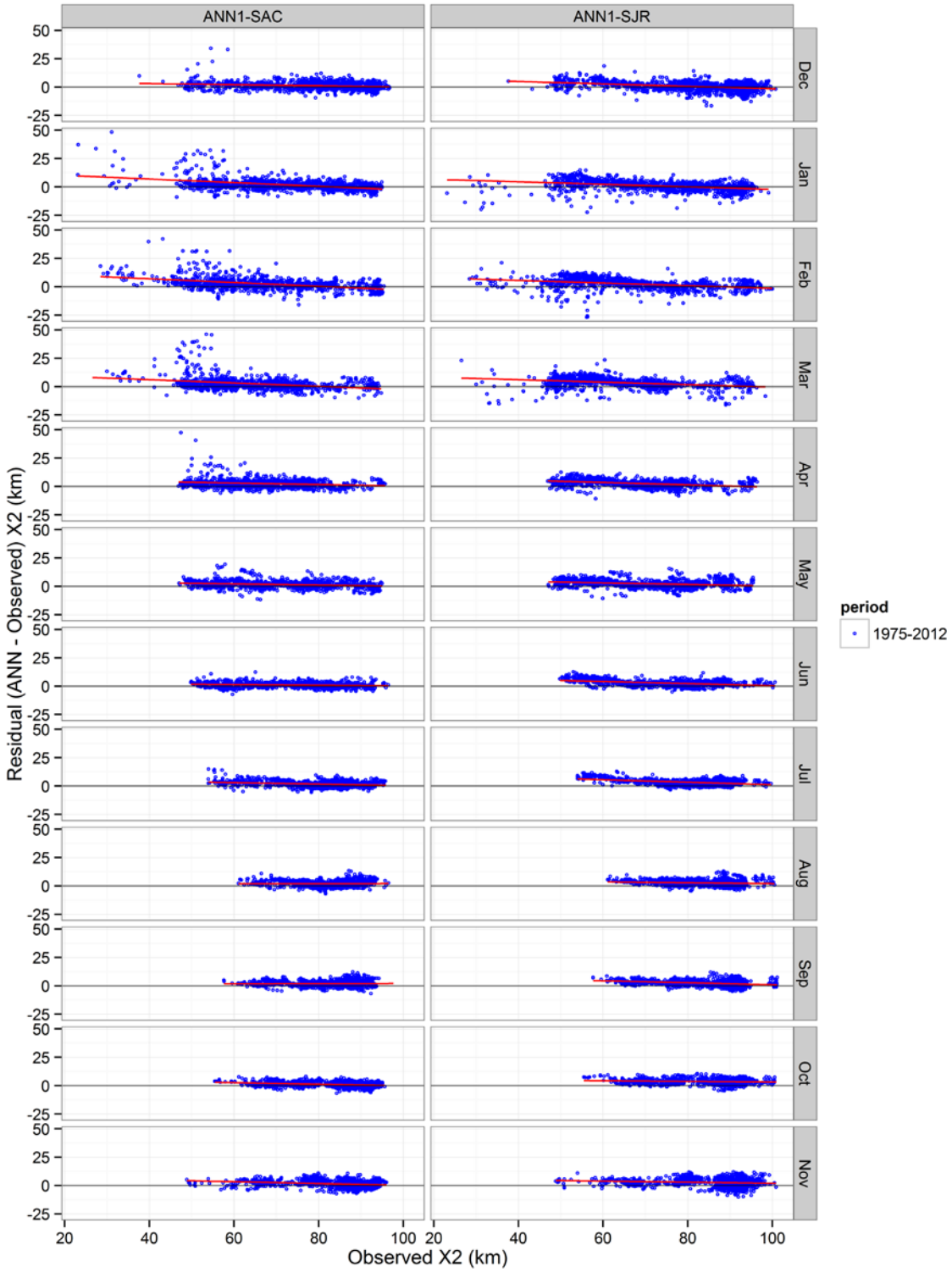

Figure 4-6 **Scatter plots of daily ANN model residuals (Approach 1), grouped by month and river.**

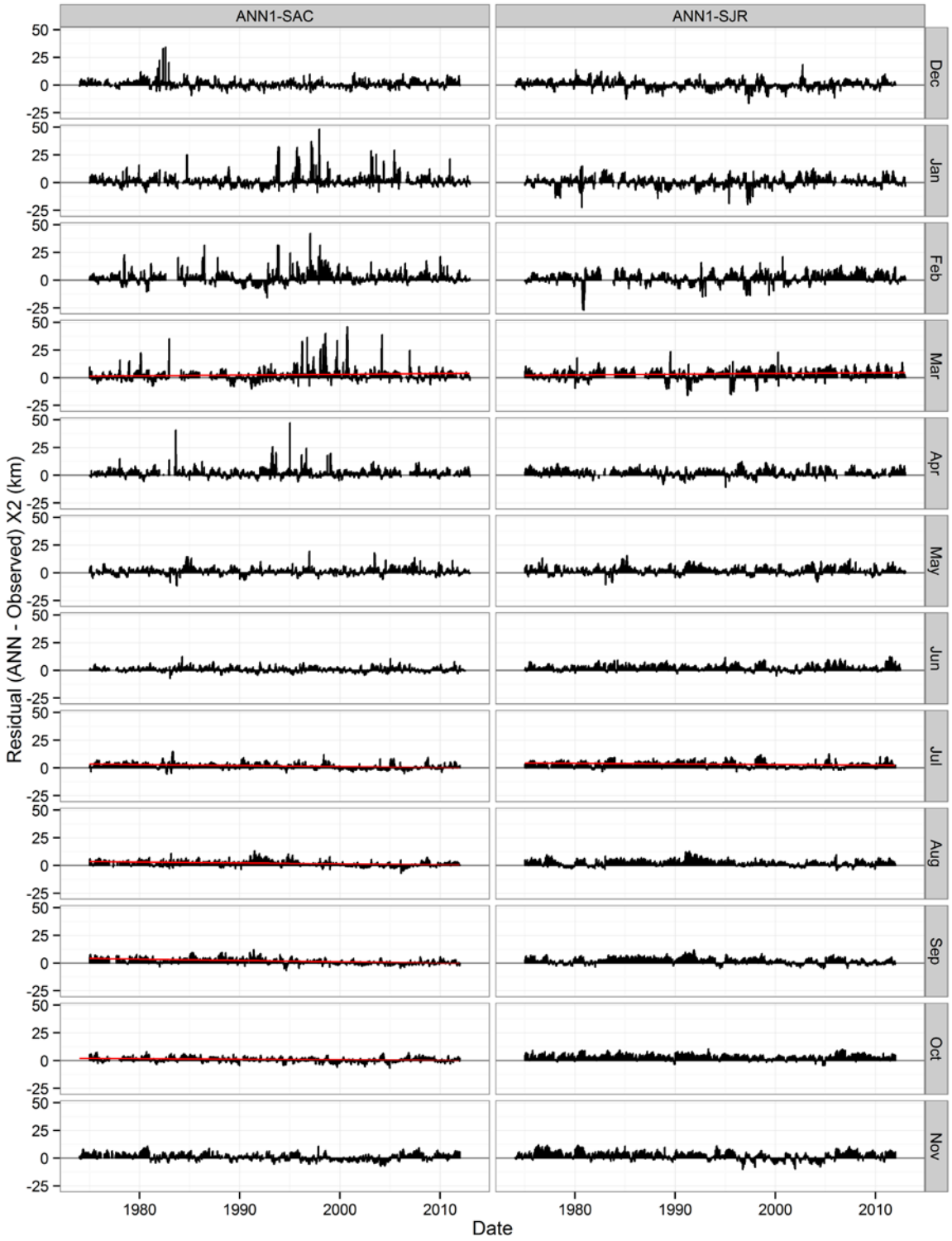

Figure 4-7 **Time series of daily ANN model X2 residuals grouped by month and river (Approach 1). Red lines, where present, indicate statistically different slopes from zero.**

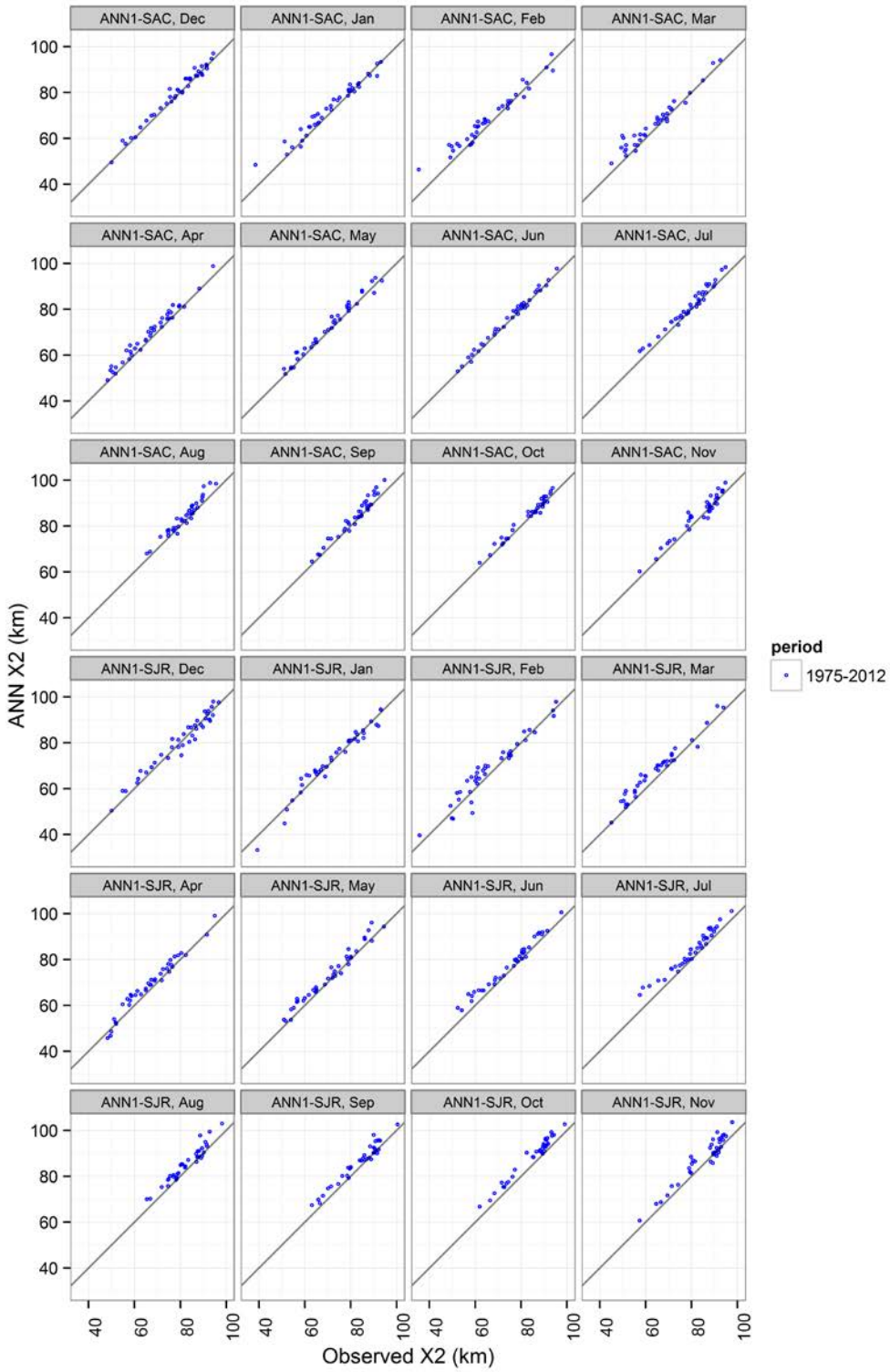

Figure 4-8 **Scatterplot of ANN model X2 (averaged to monthly level) grouped by month and river (Approach 1). Solid line represent 1:1 slope.**

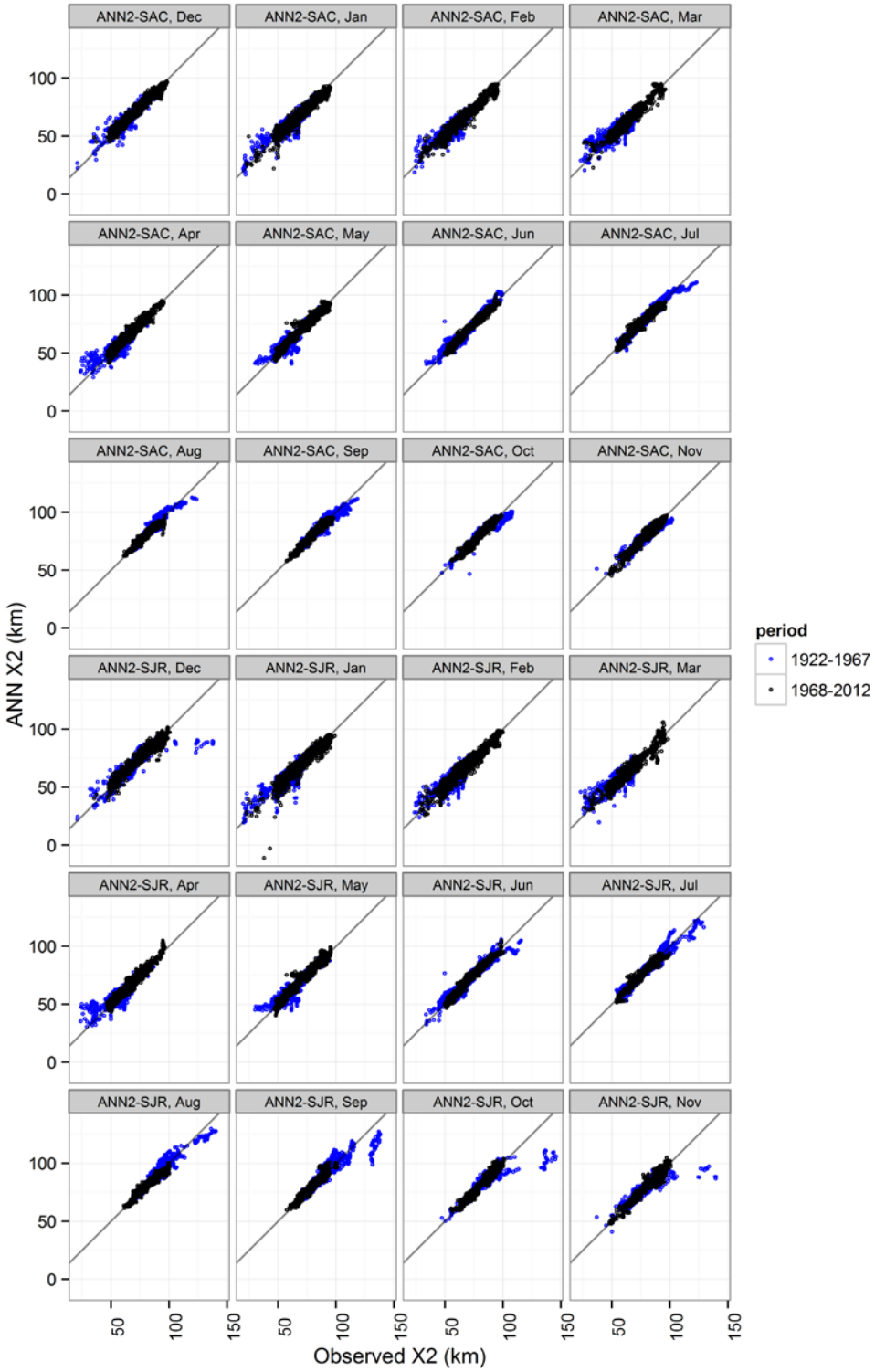

Figure 4-9 **Scatterplot of the interpolated daily X2 and X2 calculated from the ANN model (Approach 2) for the Sacramento and San Joaquin River branches, grouped by month. Data for WY 1922-1967 are shown in blue symbols, to identify the period prior to the completion of the State Water Project.**

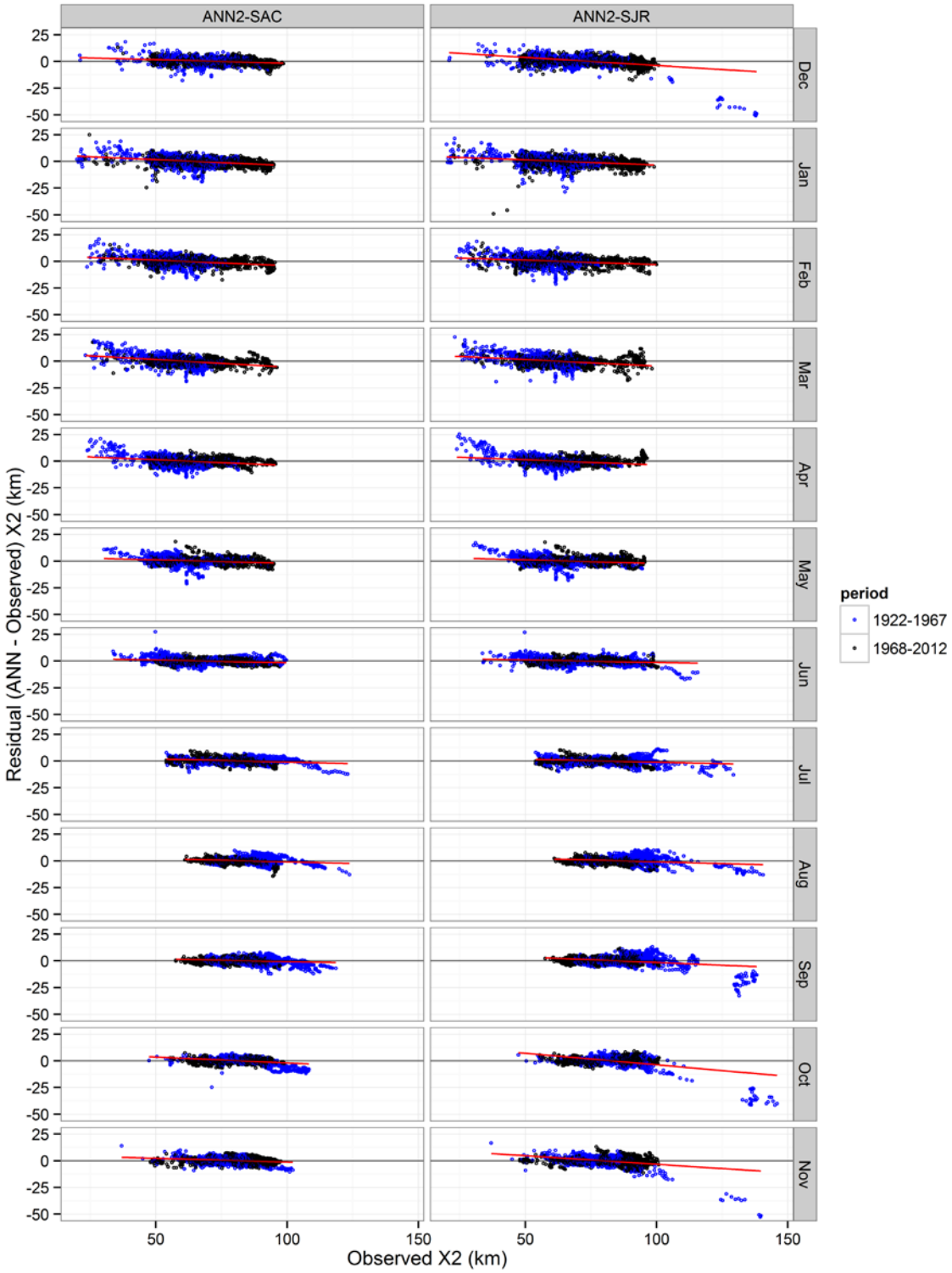

Figure 4-10 **Scatter plots of daily ANN model residuals (Approach 2), grouped by month and river. Data for WY 1922-1967 are shown in blue symbols, to identify the period prior to the completion of the State Water Project. Red lines, where present, indicate statistically different slopes from zero, at the 5% significance level.**

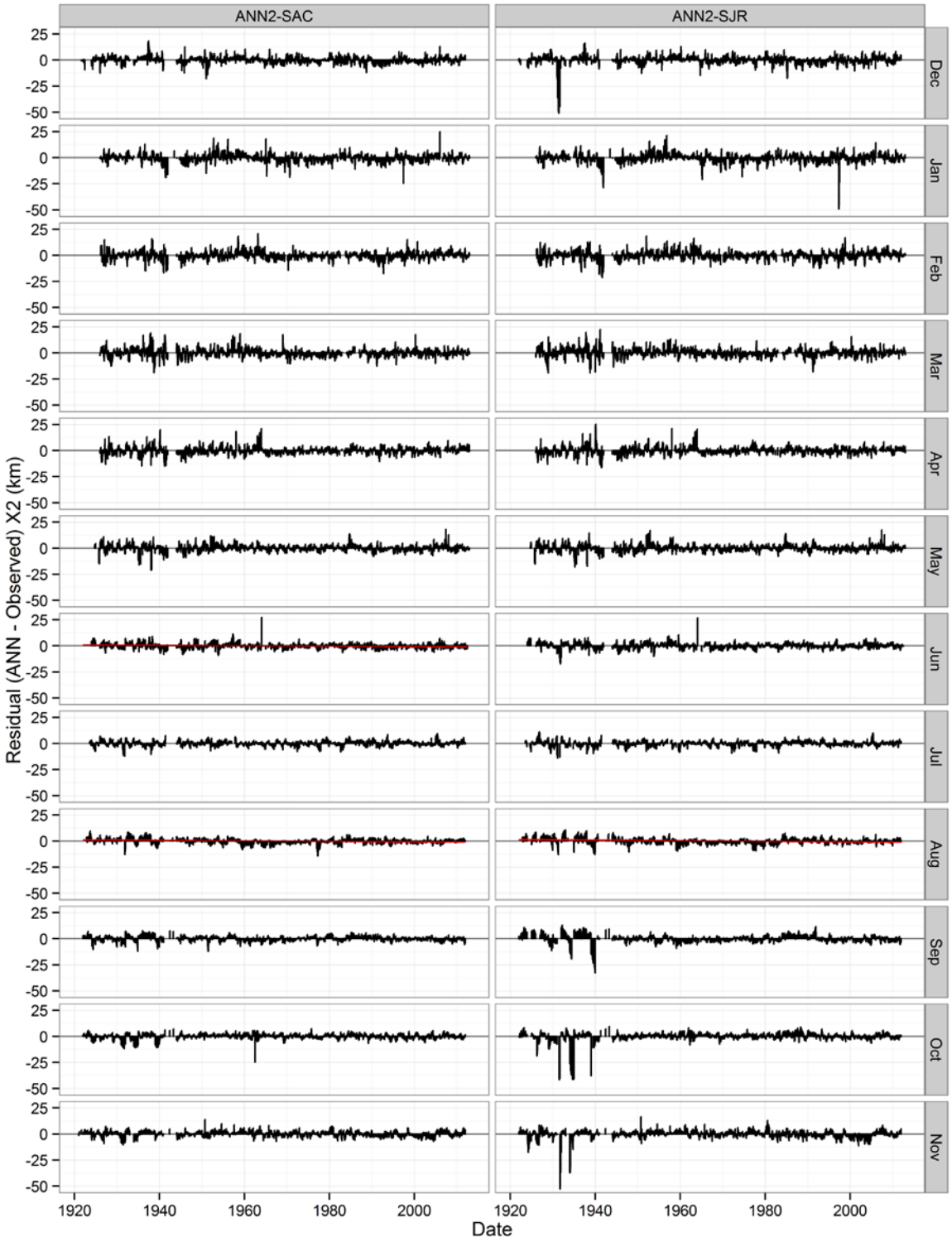

Figure 4-11 Time series of daily ANN model X2 residuals grouped by month and river (Approach 2). The red line, when present, indicates a linear time trend with a slope significant at the 5% level.

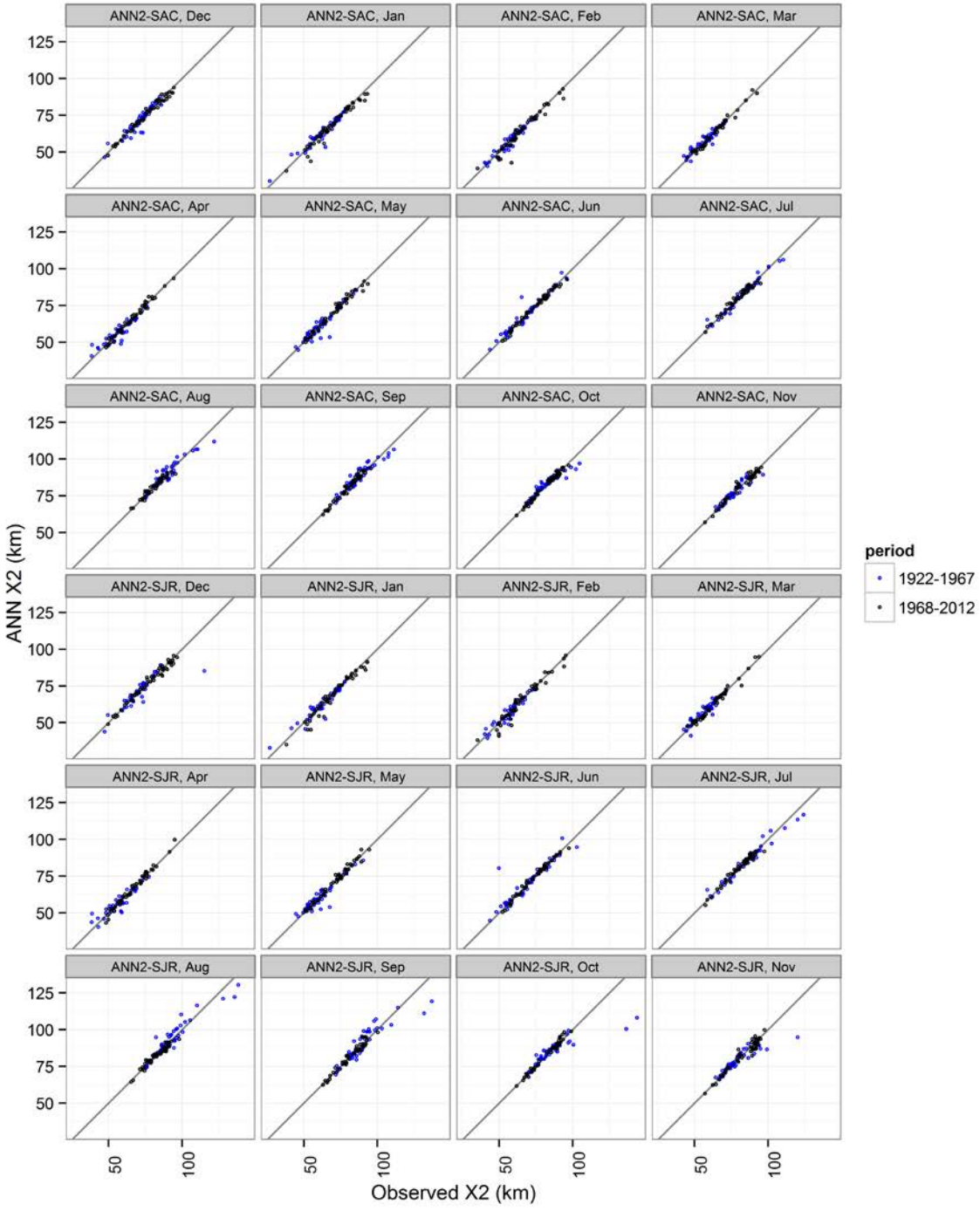

Figure 4-12 **Scatterplot of ANN model X2 residuals (averaged to monthly level) grouped by month and river (Approach 2). Solid lines represent 1:1 slope. Data for WY 1922-1967 are shown in blue symbols, to identify the period prior to the completion of the State Water Project.**

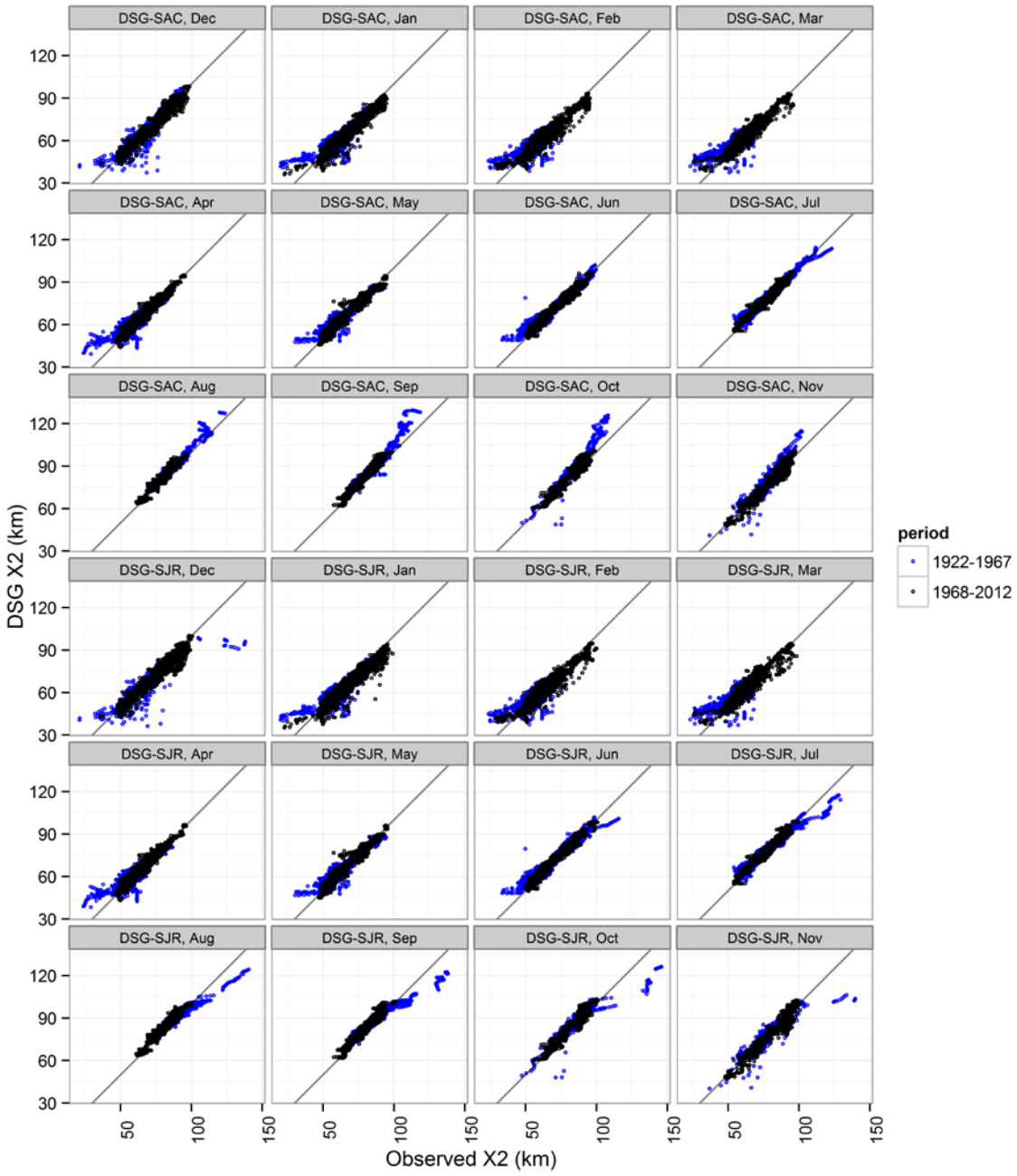

Figure 4-13 **Scatterplot of the interpolated daily X2 and X2 calculated from the daily DSG model for the Sacramento and San Joaquin River stations, grouped by month. Data for WY 1922-1967 are shown in blue symbols, to identify the period prior to the completion of the State Water Project.**

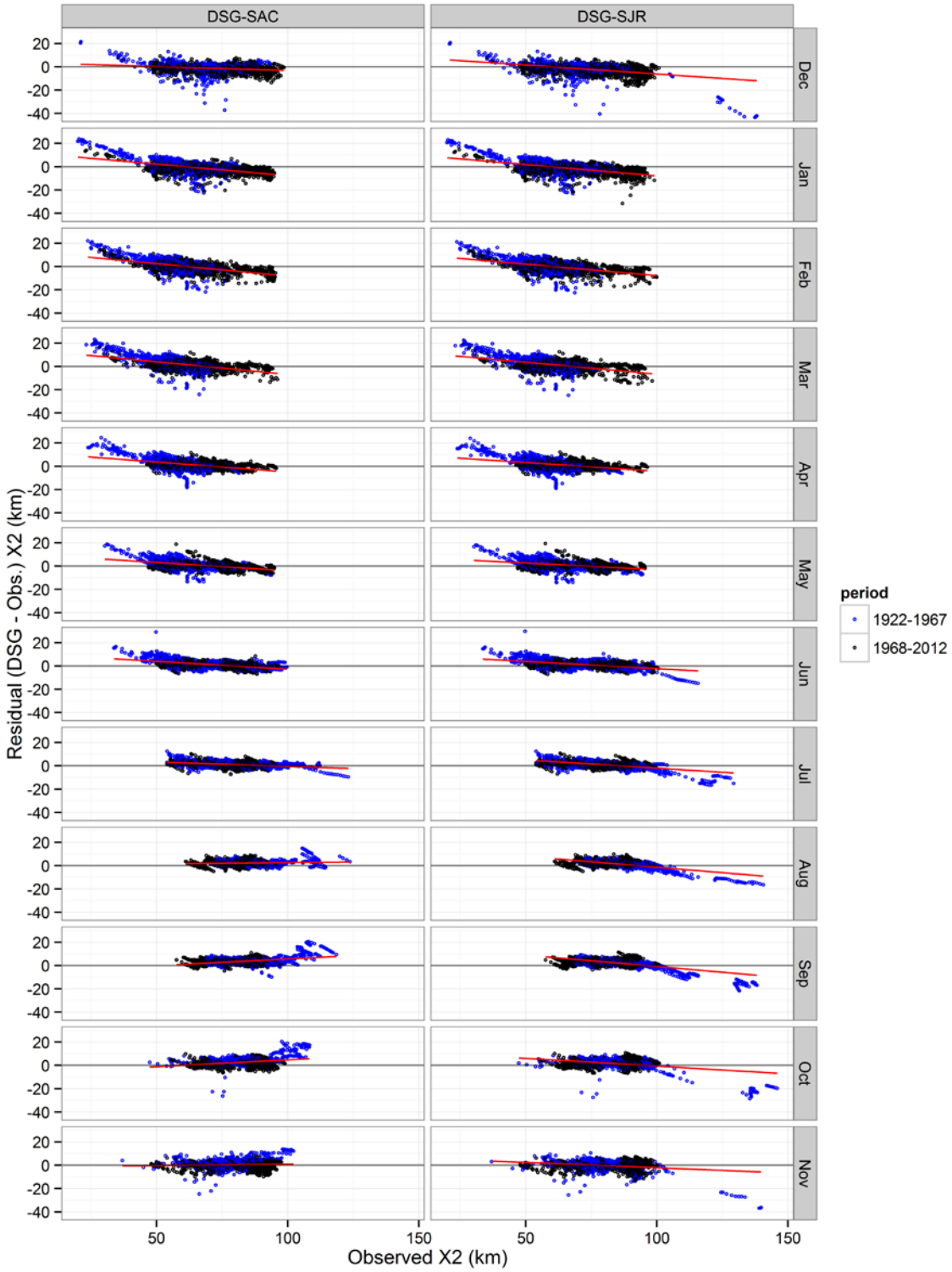

Figure 4-14 **Scatter plots of DSG model residuals. Data for WY 1922-1967 are shown in blue symbols, to identify the period prior to the completion of the State Water Project. Red lines, where present, indicate statistically different slopes from zero, at the 5% significance level.**

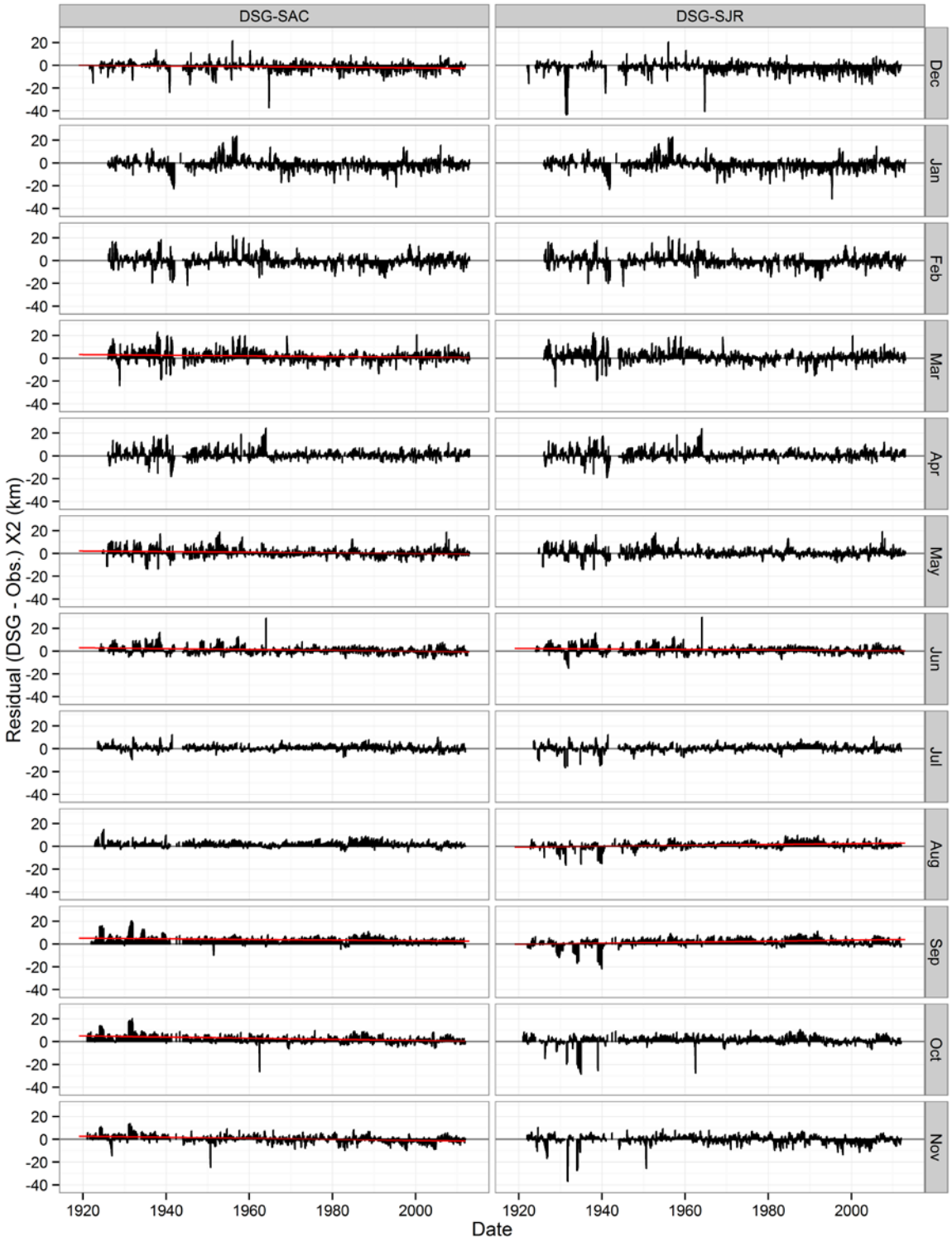

Figure 4-15 **Time series of daily DSG model residuals. The red line, when present, indicates a linear time trend with a slope different from zero significant at the 5% level.**

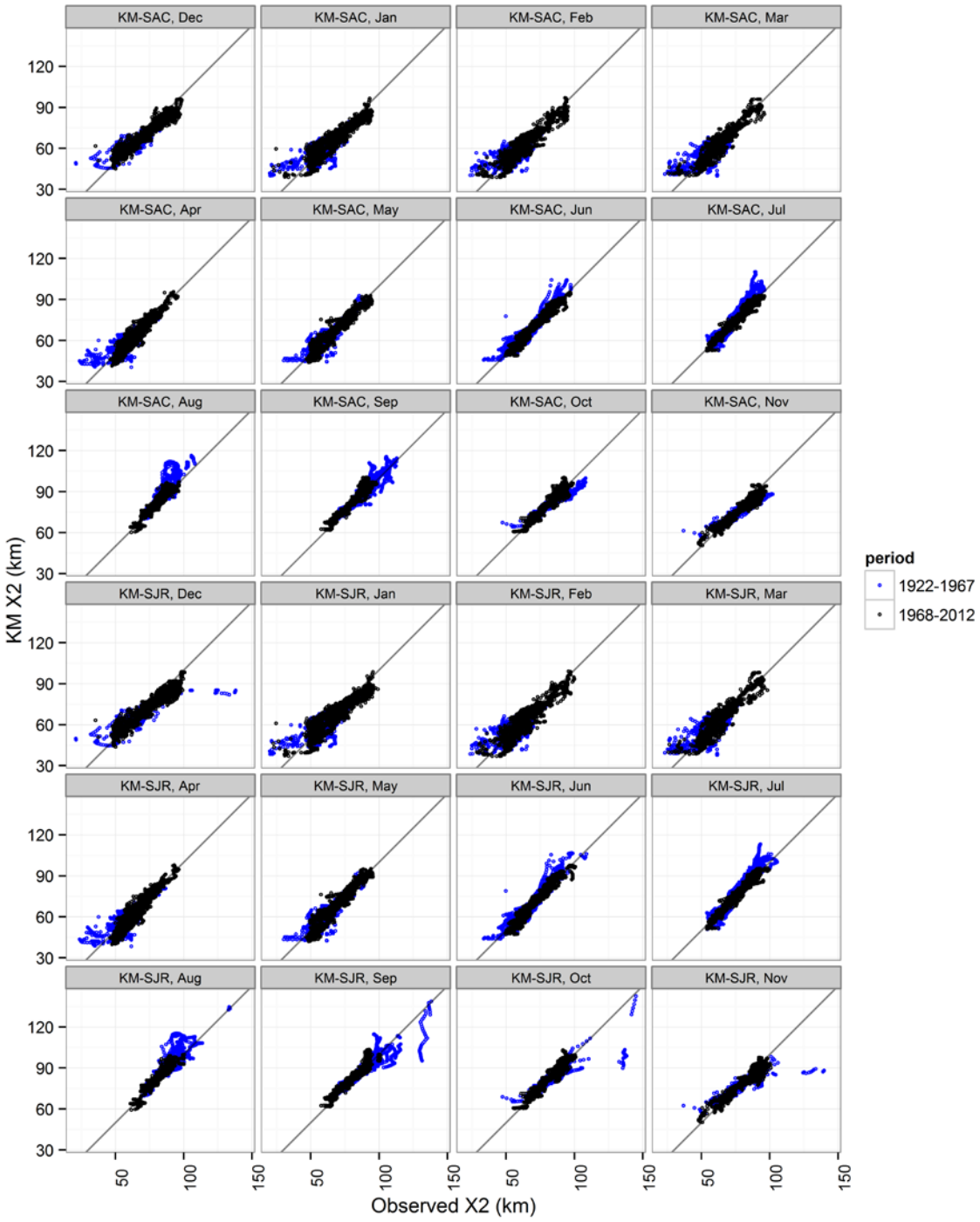

Figure 4-16 **Scatterplot of the interpolated daily X2 and X2 calculated from the daily K-M model for the Sacramento and San Joaquin River stations, grouped by month. Data for WY 1922-1967 are shown in blue symbols, to identify the period prior to the completion of the State Water Project.**

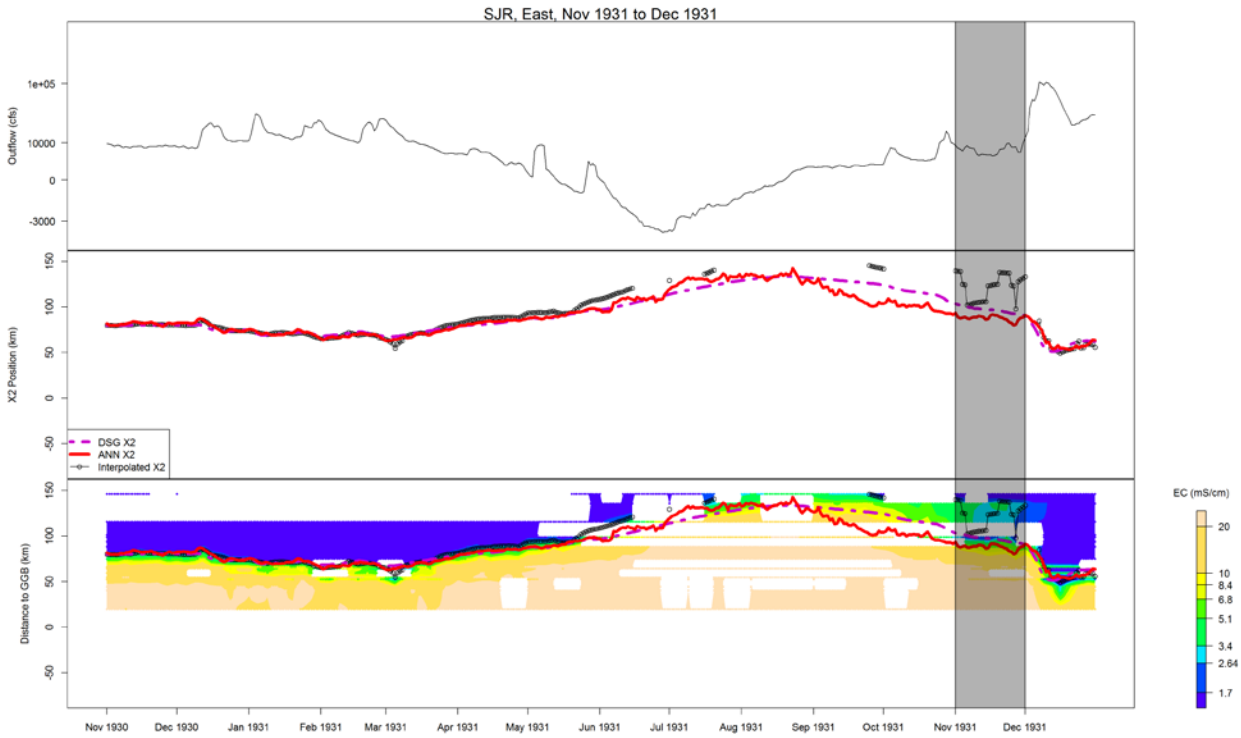

Figure 4-17 **Closer examination of variation of flow, X2, and salinity data for the Sacramento River, 1931 and preceding twelve months. The gray bar represents the area with high residuals between the interpolated X2 and the Approach 2 ANN and daily DSG models. Plots of similar exceedances are shown in Appendix C.**

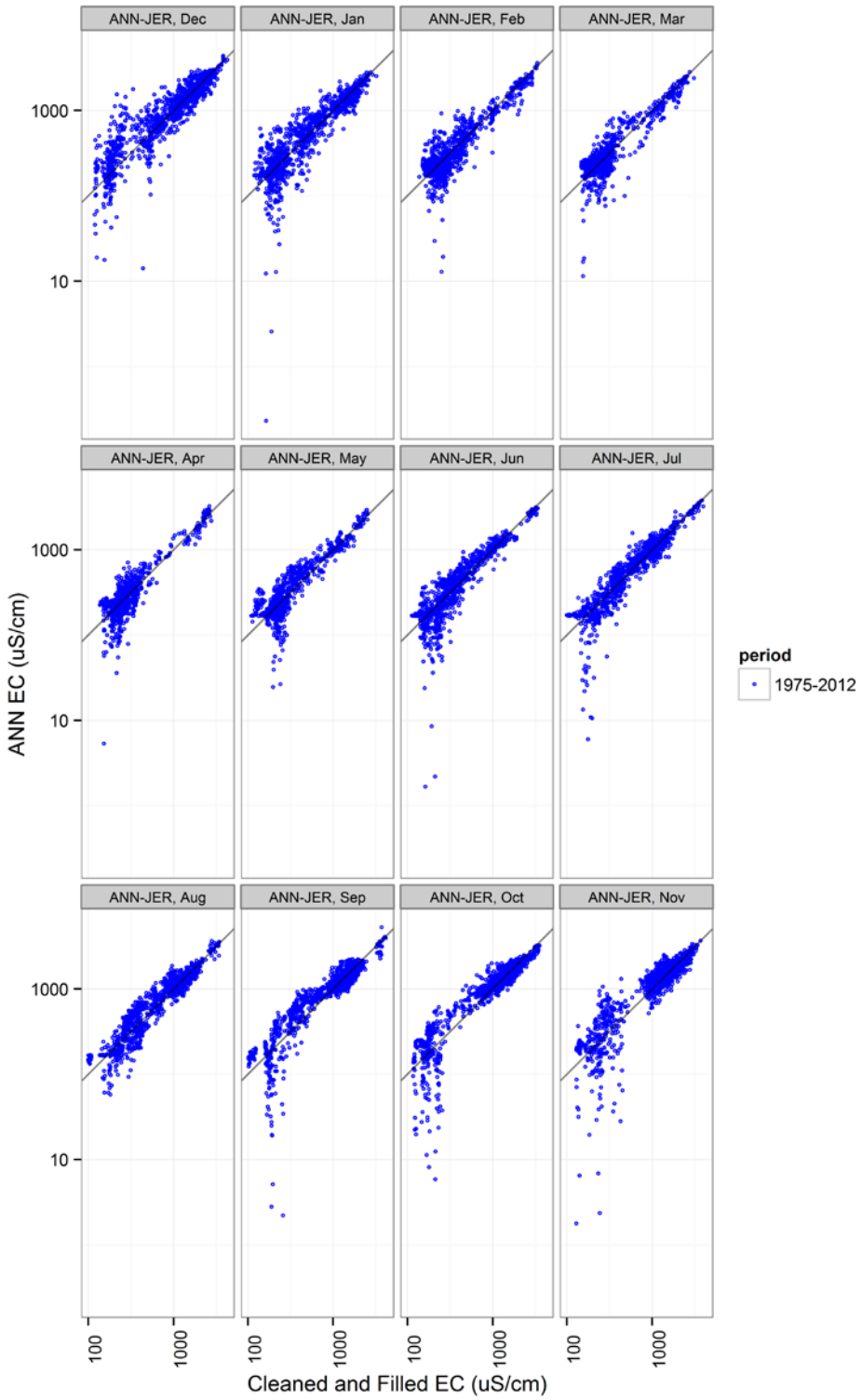

Figure 4-18 Scatterplot of Jersey Point EC, observed and ANN calculated daily values.

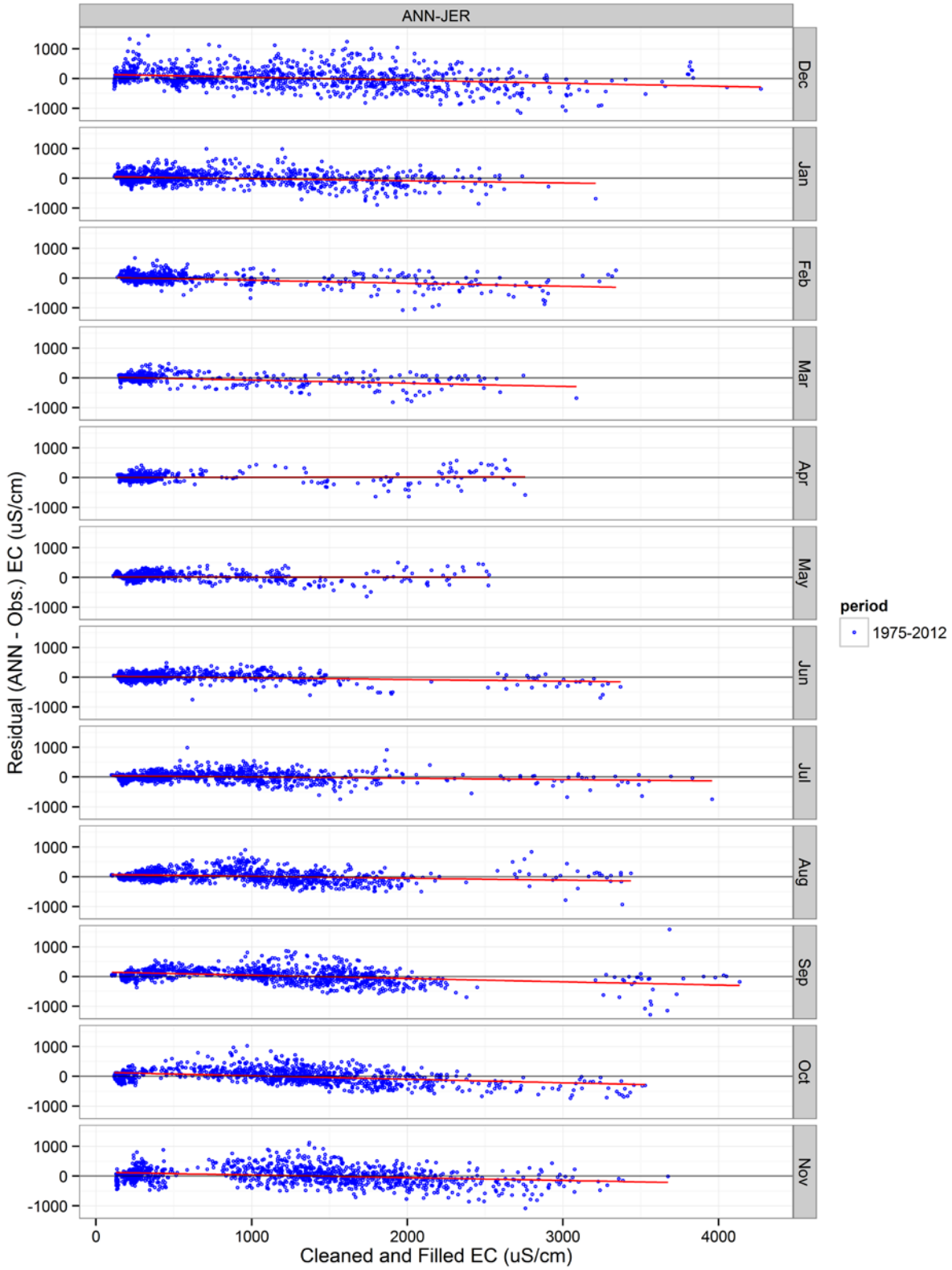

Figure 4-19 **Scatter plots of Jersey Point ANN model residuals. The red line, when present, indicates a slope different from zero significant at the 5% level.**

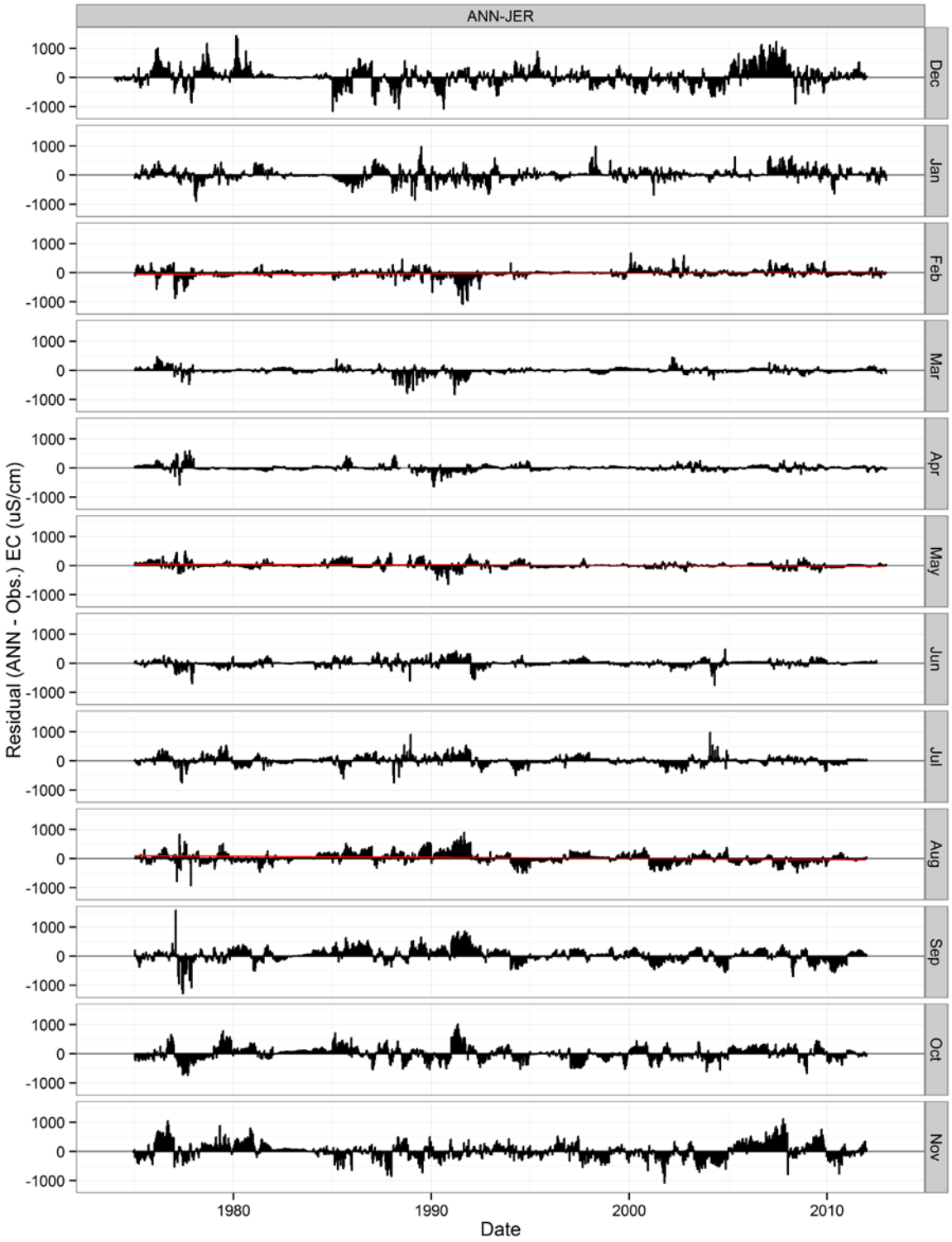

Figure 4-20 **Time series of Jersey Point ANN model residuals. Red lines, when present, indicate slopes that are statistically different from zero.**

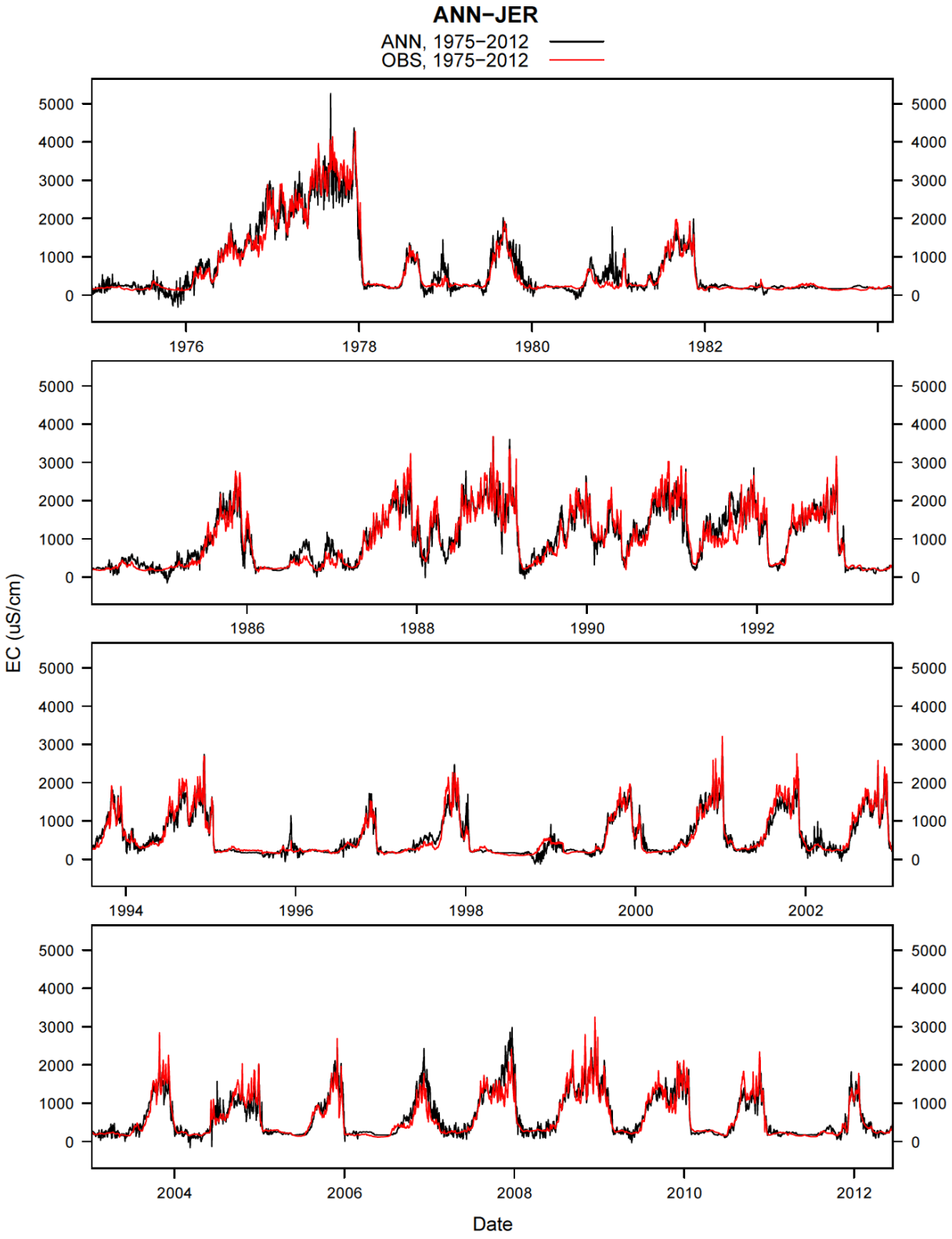

Figure 4-21 Time series of Jersey Point EC, observed daily values and ANN calculated values.

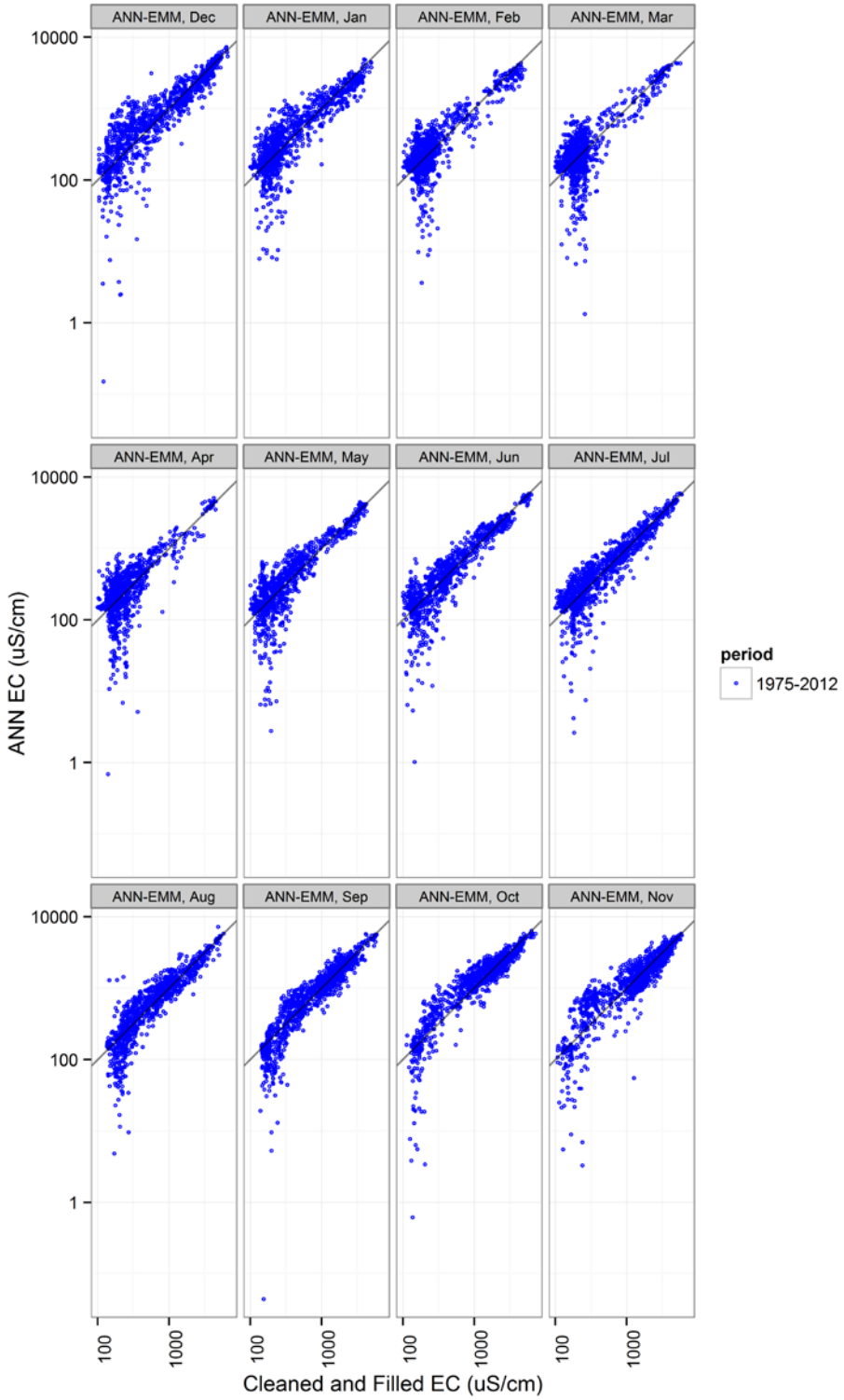

Figure 4-22 Scatterplot of Emmaton EC, observed and ANN calculated daily values.

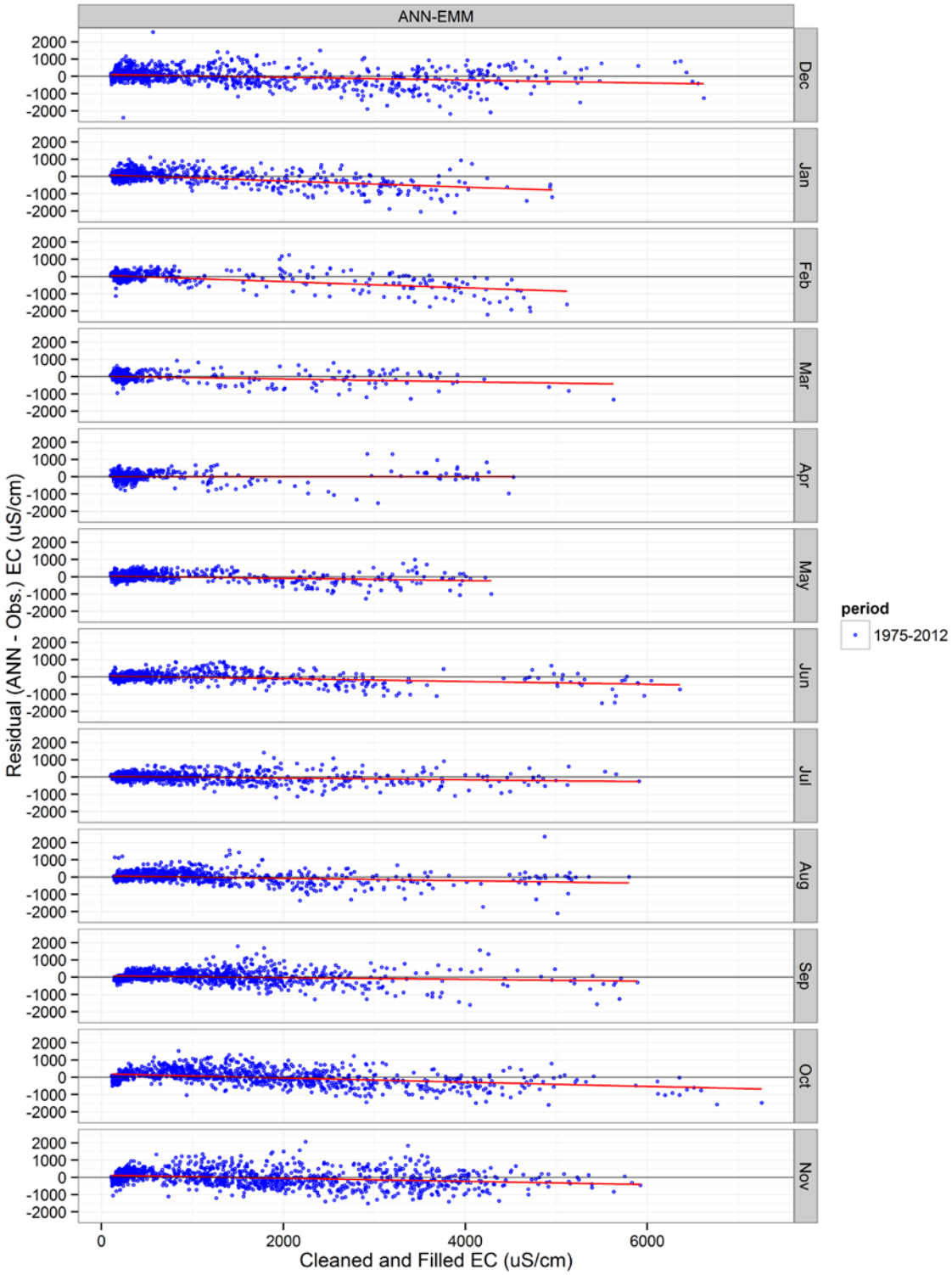

Figure 4-23 **Scatter plot of Emmaton ANN model residuals. Red lines, when present, indicate slopes that are statistically different from zero.**

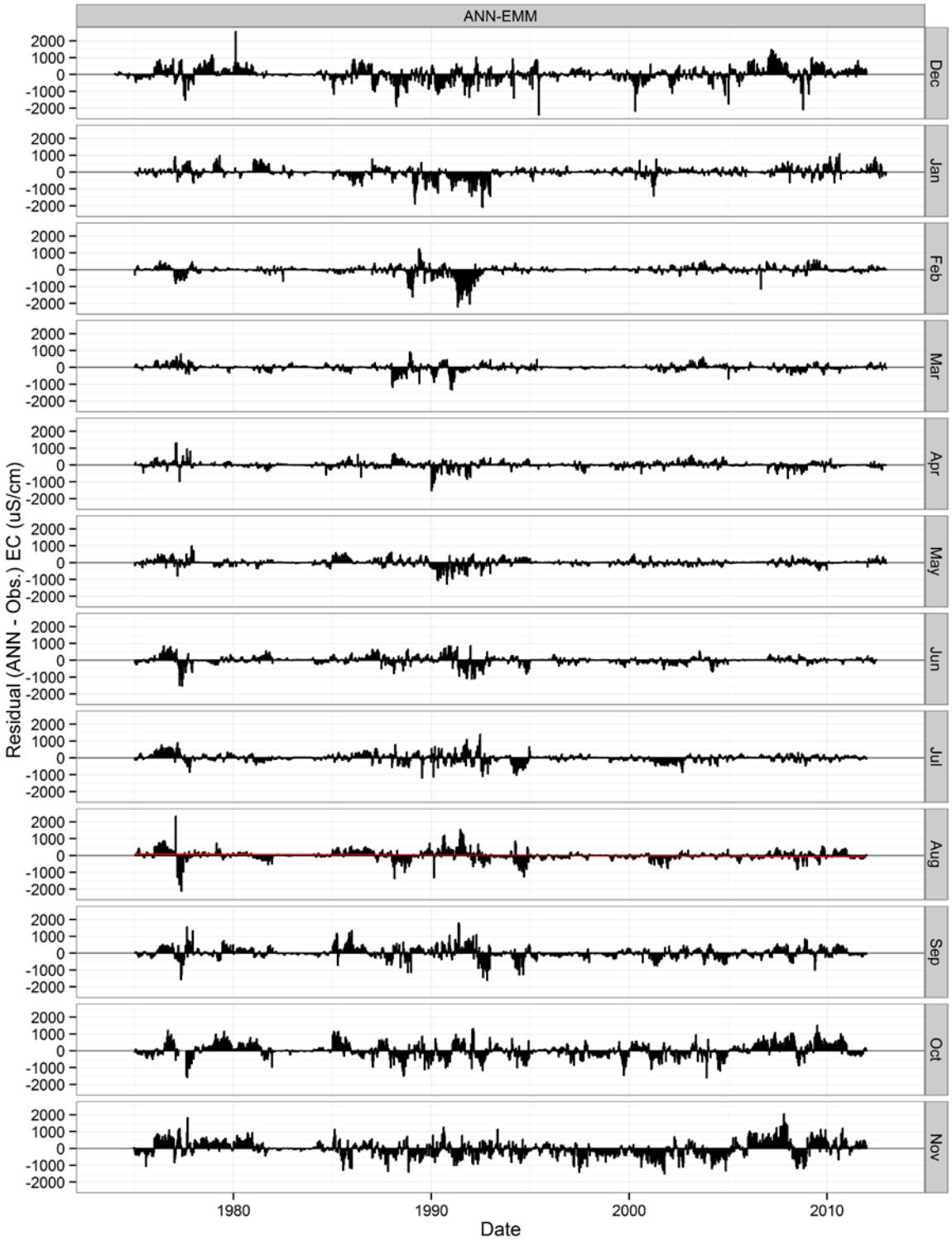

Figure 4-24 **Time series of Emmaton ANN model residuals. Red lines, when present, indicate slopes that are statistically different from zero.**

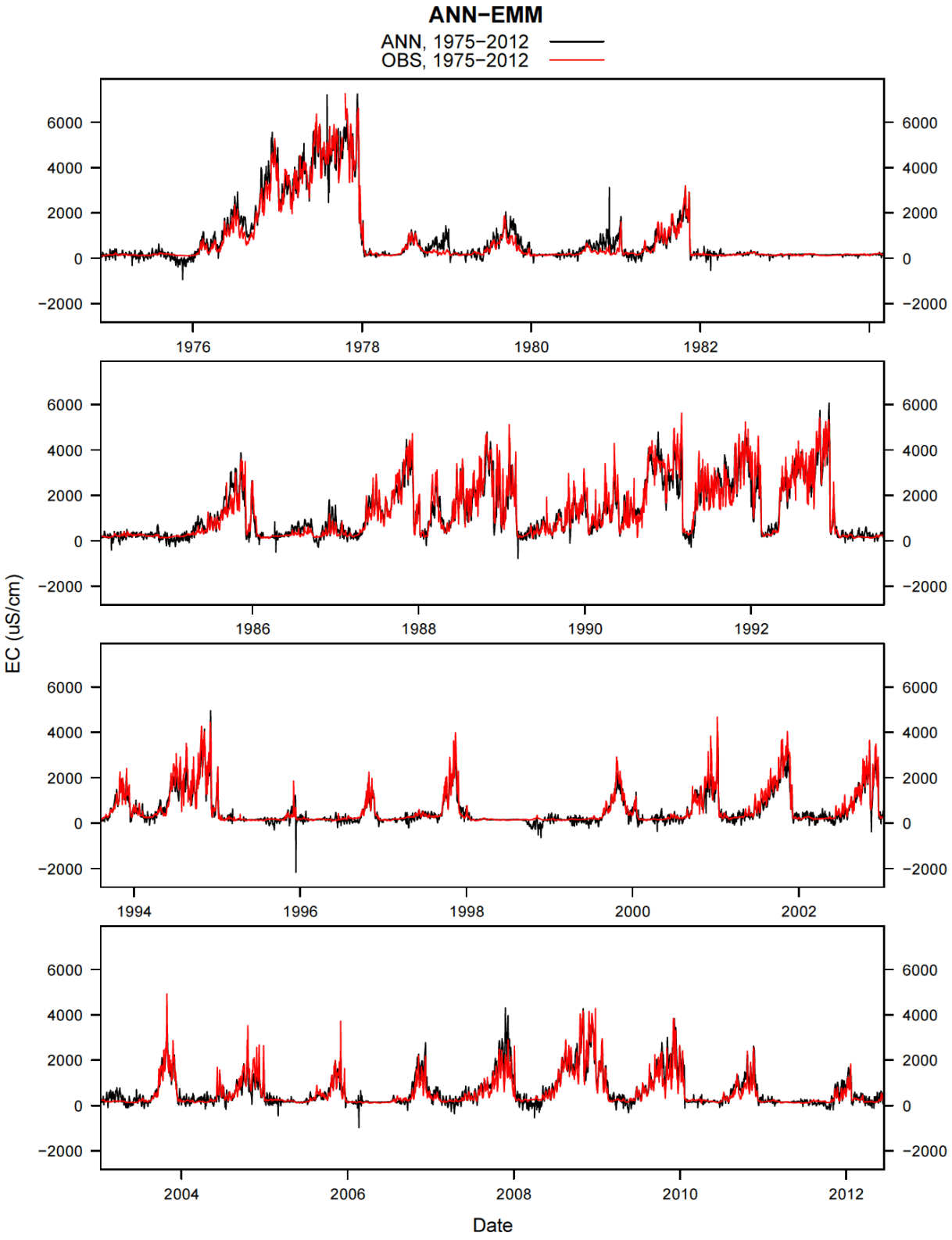

Figure 4-25 Time series of Emmaton EC, observed and ANN calculated daily values.

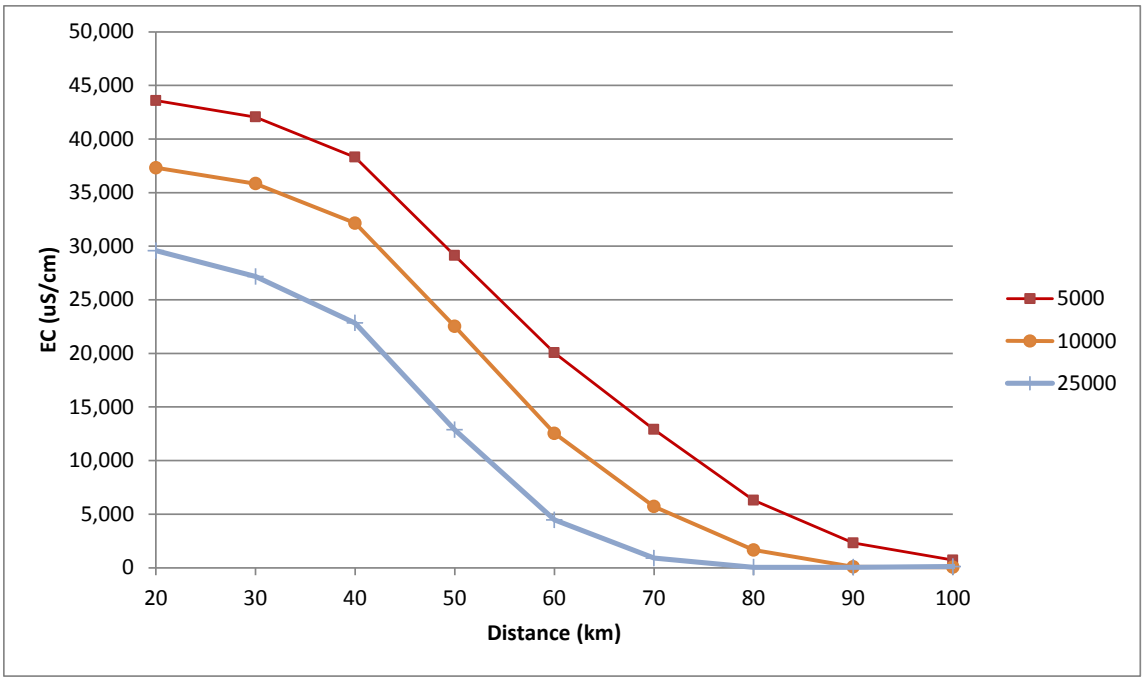

Figure 4-26 **Projected EC over distance under different Rio Vista flow conditions (5,000, 10,000, and 25,000 cfs). Sensitivity computed using the Approach 1 ANN for the Sacramento River branch.**

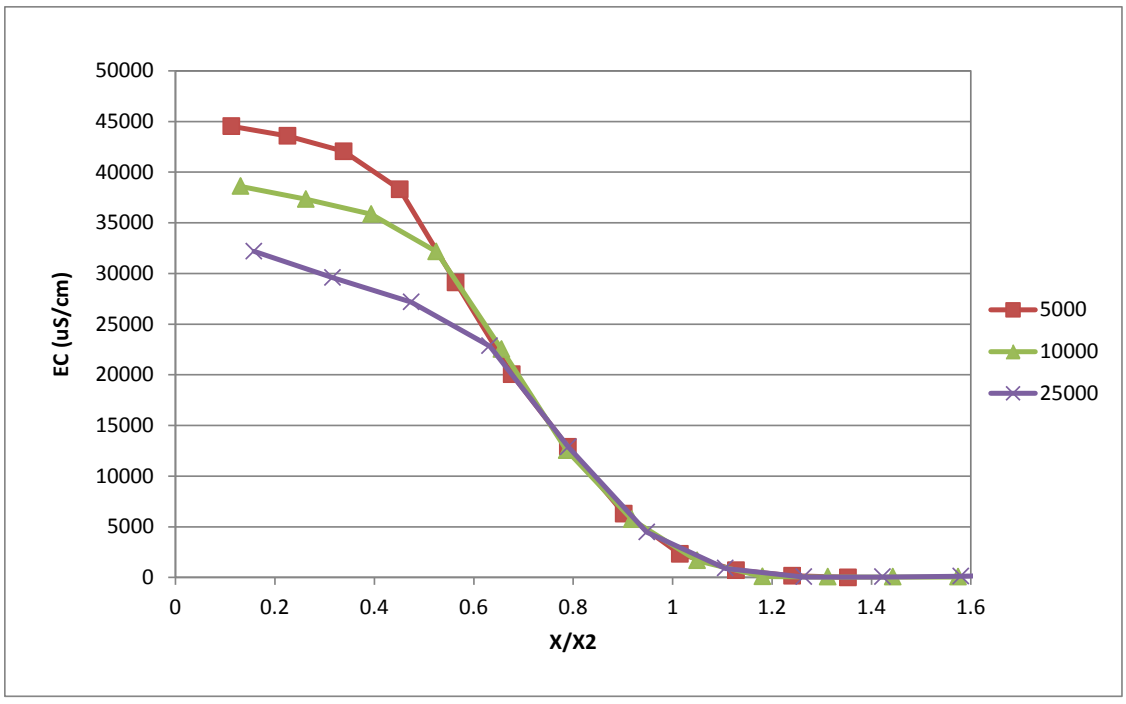

Figure 4-27 **Projected EC as a function of standardized distance (X/X2) under different Rio Vista flow conditions (5,000, 10,000, and 25,000 cfs). Sensitivity computed using the Approach 1 ANN for the Sacramento River branch.**

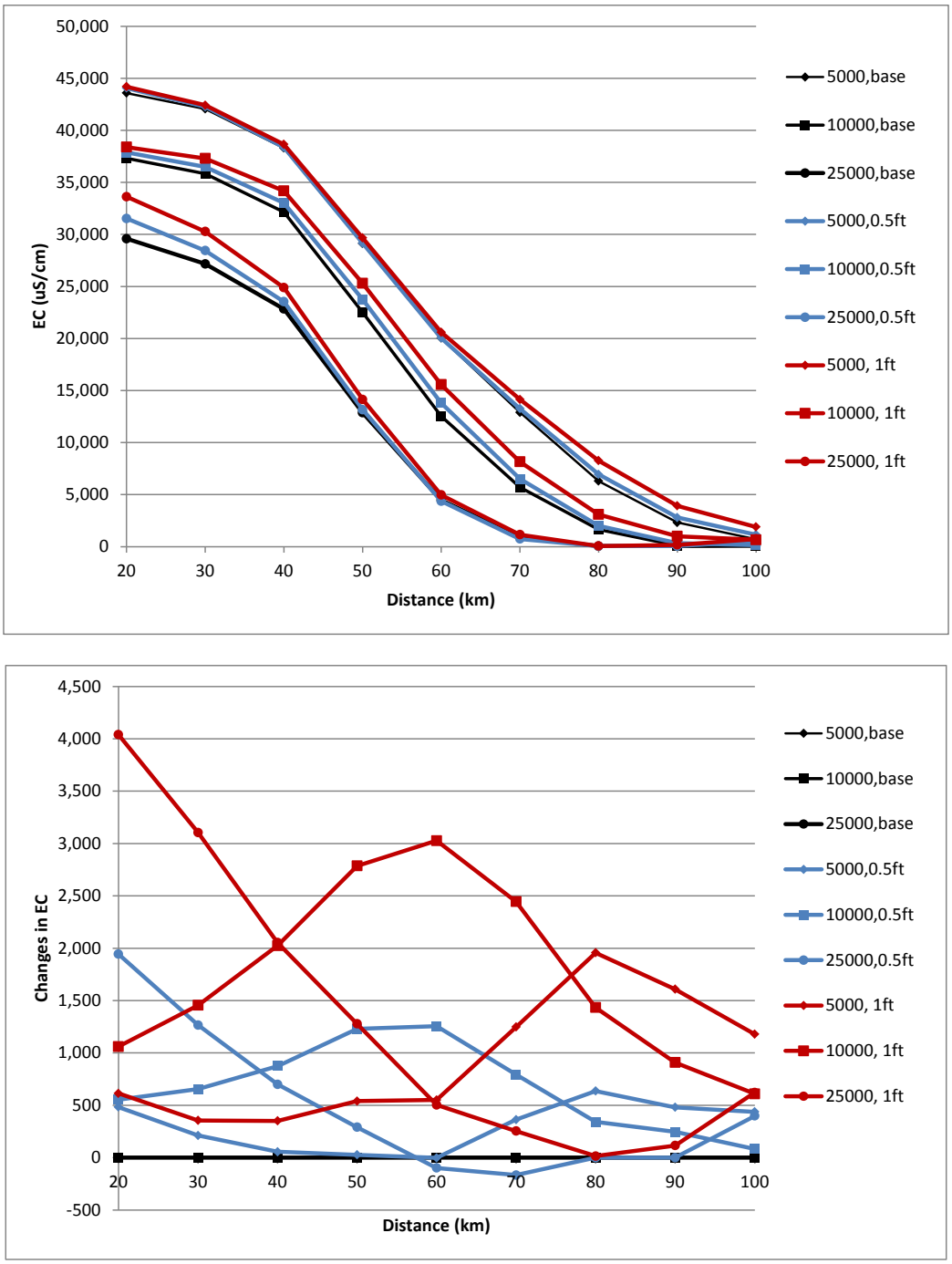

Figure 4-28 **Projected EC over distance due to sea level rise of 0.5 feet and 1 foot under different Rio Vista flow conditions (5,000, 10,000 and 25,000 cfs). Black: base, blue: 0.5 ft rise, red: 1 ft rise. Sensitivity computed using the Approach 1 ANN for the Sacramento River branch. Upper panel shows EC values, lower panel shows change from the 0 ft case.**

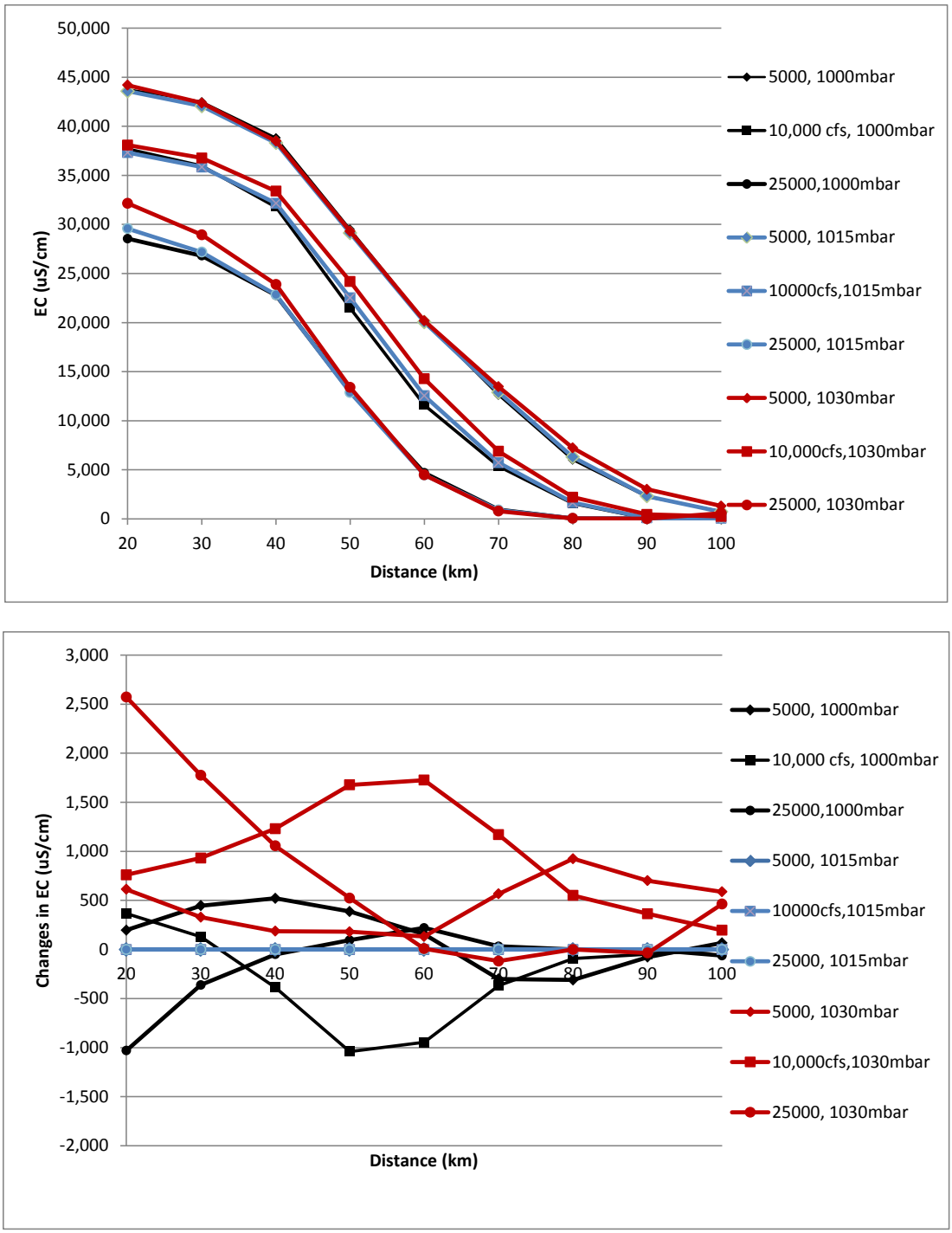

Figure 4-29 **Sensitivity of salinity to air pressure (1000, 1015, 1030 mbar), for three different values of Rio Vista flow (5,000, 10,000, and 25,000 cfs). Sensitivity computed using the Approach 1 ANN for the Sacramento River branch. Upper panel shows EC values, lower panel shows change from the 1,015 mbar case.**

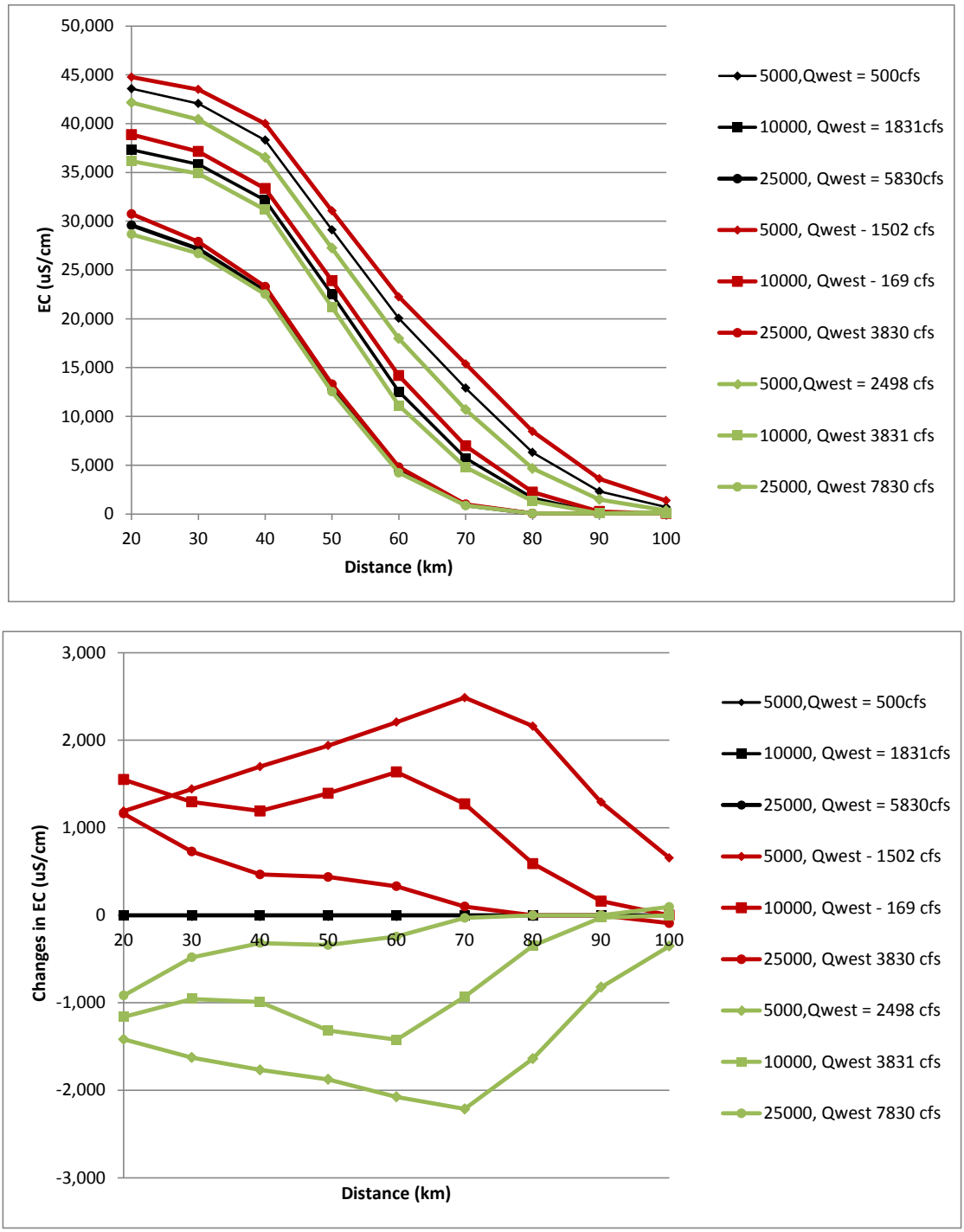

Figure 4-30 **Sensitivity of salinity to Qwest flow (mean using relationship between Qwest and Rio Vista flow, mean Qwest = QRio * 0.2666 - 834.64; mean - 2,000 cfs, mean + 2000 cfs), for three values of Rio Vista flow (5,000, 10,000, and 25,000 cfs). Sensitivity computed using the Approach 1 ANN for the Sacramento River branch. Upper panel shows EC values, lower panel shows change from the mean Qwest case.**

5. SUMMARY AND RECOMMENDATIONS

This work used the nine-decade-long record of observed daily salinity in Suisun Bay and the western Delta to develop ANNs using various boundary inputs. Over this period, the system has experienced a wide range of hydrologic, development, and regulatory conditions, which are embodied in the trained ANNs. There is confidence that the trained ANNs should represent salinity behavior over a similar range of conditions, although the behavior under conditions that are well outside the training envelope is not well-defined. This limitation applies to all data-driven tools, and the availability of the extensive data set in this region is central to the future utility of the ANN approach. Importantly, this work differs from prior salinity ANN development in the Delta region, where the training has been performed on synthetic data generated from the DSM2 model (Wilbur and Munevar, 2001; Mierzwa, 2002; Seneviratne et al., 2008).

This work built on an initial effort with training salinity ANNs for a more limited time period, October 1974 to June 2012,that was reported in Chen and Roy (2013). In particular the ANN structure and inputs were based on the previous work, where a wide variety of potential inputs was explored, and ANNs that were parsimonious and provided good fits to the observed data were identified. Based on that assessment, the specific ANNs developed here used the following inputs: flows past Rio Vista on the Sacramento River and past Jersey Point on the San Joaquin River (identified as Qwest in the DAYFLOW model); astronomical tide at Golden Gate; and the residual between astronomical and actual tide at Golden Gate. Two model output forms were considered: one where the data were used directly and the second where the data on a given day were first fitted with a model (the Delta Salinity Gradient, or DSG model), and then the parameters of that model were estimated using an ANN. In each case, two models were developed: one for the Sacramento River and one for the San Joaquin River. The comparison at individual stations as time series and scatter plots was used to evaluate the performance.

The outcomes of the training may be summarized as follows:

- Initial training using the station-based approach for data from 1930-2012 showed relatively poor results, and this approach was modified to use the data from 1974-2012, which corresponds to the availability of continuous EC sensor data, and which was shown to have a high quality ANN fit in previous work.
- For the longer-term data (1922-2012), the DSG-based ANN provided better fits than obtained at the station level over 1930-2012 and was considered valid because it incorporated prior knowledge of the salinity gradient in the estuary. Also, the use of the DSG model allowed interpolation in space such that a continuous record was not needed at all stations of interest. This was of special benefit for the Bulletin 23 data where there were many gaps in the data record.

- For comparisons of observed and ANN-predicted EC by Approach 1, the results suggested a typical $R^2 > 0.95$ for the Sacramento River model, and $R^2 > 0.92$ for the San Joaquin River model. In both cases, some eastern stations were not fit as well (such as Blind Point (BLP) (0.87), Jersey Point (JER) (0.82), Three Mile Slough (TSL) (0.80) and San Andreas Landing (SAL)(0.3)).
- For comparisons of observed and ANN-predicted EC by Approach 2, results were not as good as for Approach 1. Results suggested a general fit of R^2 near or above 0.9 for the Sacramento River model with the exception of a few stations (PTD, PTO, PSP, and RVB). The CDEC stations have better fits that the Bulletin 23 stations. For the San Joaquin River model, the fit is generally above 0.9 with the exception of a few stations (OPT, PTO, SAL).
- The overall fit of the station-based ANN (Approach 1) to interpolated X2, represented as R^2 was 0.94 and 0.91 for the Sacramento and San Joaquin River models. The overall fits of the DSG-ANN model (Approach 2) to interpolated X2 values, were slightly better ($R^2 = 0.95$ and 0.93 for the Sacramento and San Joaquin River models).
- The DSG-ANN approach (Approach 2) is a compact way to represent the entire salinity gradient, but some eastern locations, often with lower salinity, were poorly fit. Because of the importance of some of these stations for compliance with existing regulation, targeted station-specific ANNs were developed for two stations (Emmaton and Jersey Point). For future application, some combination of distance-salinity and station-specific ANNs may be suitable to best represent the observed values across a wide range of distances from Golden Gate.
- While station-specific ANNs could potentially fit data better at all stations (even western stations), they are not suggested as a replacement for the distance-salinity ANNs. This is because, (a) the distance-salinity ANNs are a more efficient representation of the data and encapsulate the behavior across multiple stations, and (b) they can be used to provide insight into the response of the horizontal salinity structure as a function of flow and tidal inputs in a more flexible manner than possible for individual ANNs for each station.
- The DSG-based ANN approach fit the X2 values over the 50-100 km range well, and this represented the vast majority of the data. There were some outliers at high and low X2 that were challenging to fit. At extreme conditions, either very high flow or low and negative flows into the Delta, the underlying hydrodynamic and mixing processes may be different. Also, the ANN may have difficulty in representing the most extreme conditions due to the limited number of observations. Depending on the need for fully capturing this behavior, future work may need to address the small subset of days that are poorly described by the inputs we have used. The ANN model outputs for these conditions may be further evaluated against surface salinity results produced from other mechanistic models of the Bay-Delta, although these evaluations were beyond the scope of the present work.
- An evaluation was performed of the sensitivity of individual station fits by nudging the distance from Golden Gate. This is potentially useful if it could be shown that

some stations are better represented by a different distance (an "effective" distance) from Golden Gate than currently applied. However, the improvements in fits, where obtained, were small, and the overall process is affected by the noise in the observed salinity. Although the concept of an effective distance is useful, other approaches, especially mechanistic hydrodynamic models may be more useful tools to explore the effects of local-scale distance on salinity.

- The ANNs developed in this work may also be compared to existing models, particularly the daily DSG model operated with constant parameters, and the daily K-M model. For the daily DSG model, applied to fit X2, the results were as follows: $R^2 = 0.92$ for both the Sacramento and San Joaquin River models). For the daily K-M model, the R^2 was 0.90 and 0.89 for the Sacramento and San Joaquin River branches. Thus, in aggregate, the DSG-ANN model is slightly better than the DSG model with constant parameters, and the station-based ANN is marginally better. Additional testing may need to be performed to examine whether the ANN models are better over specific ranges of interest for future application.
- The trained ANNs, when operated in steady-state mode, provide some insight into the sensitivity of the model to specific inputs, such as sea level rise and various inflows. The ANNs provided results that were generally intuitive, in terms of the effects of salinity at specific locations and/or the location of the X2 position. The sensitivity analysis must be cognizant of the limited ability of ANNs to extrapolate beyond the training data set, and for truly significant departures from training conditions, such as sea level changes of several feet, other modeling tools are more appropriate.

The future use of ANNs for predicting salinity must be based on the relative performance of this tool with other available tools, and its relative complexity. In this work we looked at the daily DSG model and the daily K-M model. The interpolated data were fit quite well with the daily DSG model, although the R^2 values were slightly lower than the ANN model. The K-M fit was not as good, and was also limited by the ability of this model to represent negative Delta flows. The daily DSG model fits can be improved with bias correction as an alternative to future applications. For station specific application, the ANN approach was found to be a useful alternative to methods that are focused on the entire gradient. Future application may also consider more than one modeling approach to calculate X2 for specified conditions to obtain more robust estimates of X2 and salinity at specific stations.

6. REFERENCES

American Society of Civil Engineers (ASCE). (2000). Artificial Neural Networks in Hydrology, 1. Preliminary Concepts, Vol. 5, No. 2, pp. 115–123.

BDCP EIR/EIS (2013). Appendix 29A Effects of Sea Level Rise on Delta Tidal Flows and Salinity, on the Internet at: http://baydeltaconservationplan.com/Libraries/Dynamic_Document_Library/EIR-EIS_Appendix_29A_%E2%80%93_Effects_of_Sea_Level_Rise_on_Delta_Tidal_Flows_and_Salinity_5-10-13.sflb.ashx.

Chen, L. and S. B. Roy (2013). Modeling salinity in Suisun Bay and the western Delta using artificial neural networks, report prepared for Metropolitan Water District of Southern California, April.

Denton, R.A. (1993). Accounting for antecedent conditions in seawater intrusion modeling – applications for the San Francisco Bay-Delta. Hydraulic Engineering 93 (1): 448–453. Proceedings of ASCE National Conference on Hydraulic Engineering, San Francisco.

Denton, R.A. (1994) Predicting Surface 14-day EC from NDO, memo to Paul Hutton, April 12.

Finch, R. and N. Sandhu. (1995). Artificial neural networks with application to the Sacramento – San Joaquin Delta. California Department of Water Resources Delta Modeling Section, Division of Planning.

Fleenor, W., E. Hanak, J.R. Lund, J.R. Mount. (2008). Delta hydrodynamics and water salinity with future conditions. Technical Appendix C. in Comparing Futures for the Sacramento–San Joaquin Delta, Public Policy Institute of California.

Gross, E. S., Nidzieko, N. J., MacWilliams, M. L., & Stacey, M. T. (2007). Parameterization of Estuarine Mixing Processes in the San Francisco Estuary based on Analysis of Three-Dimensional Hydrodynamic Simulations. Journal of Estuarine and Coastal Modeling, ASCE, pp. 322–338.

Gross, E. S., MacWilliams, M. L., & Kimmerer, W. J. (2010). Three-dimensional modeling of tidal hydrodynamics in the San Francisco Estuary. San Francisco Estuary and Watershed Science, 7(2).

Harder, J.A. (1977). Predicting estuarine salinity from river inflows, Journal of the Hydraulics Division, ASCE, Vol. 103, pp. 877–888.

Hutton, P.H. (2013) A New Empirical Bay-Delta Salinity Model, presentation at California Water and Environmental Modeling Forum Annual Meeting, Folsom, CA, April 24.

Jassby, A.D., W.J. Kimmerer, S.G. Monismith, C.Armor, J.E. Cloern, T.M. Powell, J.R. Schubel and T.J. Vendlinski. (1995). Isohaline position as a habitat indicator for estuarine populations. Ecological Applications 5: 272–289.

Kimmerer, W. and S. Monismith. (1992). An estimate of the historical position of 2ppt salinity in the San Francisco Bay Estuary. Issue Paper prepared for the fourth technical workshop on salinity, flows and living resources of the San Francisco Bay/Delta estuary. August.

List, J. (1994). X2 and the X2/Net Delta Outflow Relationship. Prepared for California Urban Water Agencies, Sacramento, California.

Maier, H.R., A. Jain, G.C. Dandy, K.P. Sudheer. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. Environmental Modeling and Software 25: 891–909.

Mierzwa, M. (2002). Chapter 4: CALSIM versus DSM2 ANN and G-model Comparisons. Methodology for flow and salinity estimates in the Sacramento–San Joaquin Delta and Suisun Marsh. 23rd Annual Progress Report. June.

Monismith, S. G., Kimmerer, W., Burau, J. R., & Stacey, M. T. (2002). Structure and flow-induced variability of the subtidal salinity field in northern San Francisco Bay. Journal of Physical Oceanography, 32(11), 3003-3019.

Roy, S.B., J. Rath, L. Chen, M.J. Ungs, and M. Guerrero. (2014). Salinity trends in the Suisun Bay and the western Delta (October 1921– September 2012). Final report to San Luis and Delta Mendota Water Authority and Metropolitan Water District of Southern California, January.

Sandhu, N., D. Wilson, and R. Finch. 1999. Modeling flow-salinity relationships in the Sacramento- San Joaquin Delta using artificial neural networks. Technical Information Record OSP-99-1. California Department of Water Resources. Sacramento, CA.

Schemel, L. (2001). Simplified conversion between specific conductance and salinity units for use with data from monitoring stations. IEP Newsletter Volume 14, Number 1, Winter.

Senevirante, S., S.Wu, and Y. Liang. (2008). Chapter 3: Impacts of Sea Level Rise and Amplitude Change on Delta Operations. Methodology for flow and salinity estimates in the Sacramento – San Joaquin Delta and Suisun Marsh. 29th Annual Progress Report. June.

State Water Resources Control Board (2006). Water Quality Control Plan for the San Francisco Bay/Sacramento-San Joaquin Delta Estuary. Division of Water Rights.

Wilbur, R. and A. Munevar. (2001). Chapter 7: Integration of CALSIM and Artificial Neural Networks Models for Sacramento–San Joaquin Delta Flow-Salinity Relationships. Methodology for flow and salinity estimates in the Sacramento–San Joaquin Delta and Suisun Marsh. 22nd Annual Progress Report.