METHODOLOGY FOR FLOW AND SALINITY ESTIMATES IN THE SACRAMENTO-SAN JOAQUIN DELTA AND SUISUN MARSH

SEVENTEENTH ANNUAL PROGRESS REPORT TO THE STATE WATER RESOURCES CONTROL BOARD IN ACCORDANCE WITH WATER RIGHT DECISION 1485, ORDER 9

JUNE 1996

CHAPTER 2

Emulation of DWRDSM Using Artificial Neural Networks and Estimation of Sacramento River Flow

[Editor's Note: The following report is an electronic reproduction of the second chapter from the 17th annual progress report to the State Water Resources Control Board. The original text and structure of this chapter was left the same, however, the font styling and positioning of the figures within the report have been modified.]

The need for a fast, reasonably accurate method of modeling the relationship between flows and water quality in the Sacramento-San Joaquin Delta has been recognized for many years. The uses of such a method would be to estimate the flows required to achieve a certain salinity standard (the Minimum Delta Outflow routine, used in DWRSIM), to rapidly simulate many different flow and salinity scenarios for optimization, real-time control of operation, and so on.

Artificial Neural Networks are widely used and well-suited for multiple nonlinear regression. Department of Water Resources' Delta Simulation Model is a computer model that simulates hydrology and constituent transport in the Sacramento-San Joaquin Delta. Feed-forward ANNs are used to relate the flow conditions and gate positions in the Delta to the DWRDSM simulated salinity at interior and boundary locations in the Delta. The ANNs provide a fast and reasonably accurate method of modeling the relationship between flows and water quality. This relationship is then used to estimate the Sacramento River flow required to meet a salinity standard.

Previous Work

Most efforts at modeling the Delta and improving the MDO (Mininum Delta Outflow) routine involved using Net Delta Outflow, with perhaps some time-lagged NDO as memory, as the only input. However, a NDO only model cannot say anything about the effect of rim flow combinations and gate operations. Thus a model is needed that can handle multiple inputs and nonlinear behavior. DWRDSM has been used to investigate carriage water and related issues. Like most numerical models, it is extremely slow as compared to running time of DWRSIM and therefore not a serious candidate for a MDO replacement. Even with a reduced channel grid, DWRDSM run times would be far longer (on the order of 10,000) than the current MDO routine.

Statistically-based models have been tried and found lacking (Winkler 1985); traditional time-series analysis (Shumway 1993) is often linear or requires transformation to a stationary time-series, which renders the resulting model unsuitable as an MDO replacement. Recently, the California Department of Water Resources, Delta Modeling Section, has investigated a new mathematical technology called Artificial Neural Networks as an alternative to conventional techniques.

Why Use ANNs

Artificial Neural Networks are widely used and for this particular application they can be thought of as a multiple nonlinear regression technique. Neural networks are universal approximators, that is they are capable of modeling any function with a finite number of discontinuities to any desired degree of accuracy, given sufficient number of hidden neurons (Masters 1993). Minimal preprocessing is required and different types of inputs can be used in the same network. Once calibrated ANNs are fast and reasonably accurate (Sandhu and Finch 1995). The only two significant disadvantages are that they tend to require more data to calibrate than conventional methods and the only method to optimize the weights are gradient descent methods, which being iterative consum significant amounts of computing time. Once trained, however, simulation only involves matrix multiplication and is quite fast. Also in this case sufficient data can be generated to calibrate and validate ANNs.

Development of ANN Model

Software Used

SNNSv4.0 (Stuttgart Neural Network Simulator, version 4.0) was the package used in developing and training the ANNs. This is a public domain package which runs on X Windows on UNIX systems. It has a wide variety of architectures available and undergoes major improvements every six months.

Data Description

Salinity data, in the form of total dissolved solids, was available from a 20-year DWRDSM run. This run used historical flows and gate positions to simulate salinity at various locations in the Delta. Flow values were given in cubic feet per second and TDS in milligram per liter. Gate closed position was indicated as 0.1 and open position as 1.1. Years 1980-1990 were used to calibrate the ANNs and 1971-1980 was the period used to validate the ANNs.

Salinity predictions and analysis were made at two stations, Chipps Island (CHIPPS) and Contra Costa Canal Pumping Plant No. 1 (CCC). CHIPPS is representative of the salinities at the boundary of the Delta with the San Francisco Bay. CCC is representative of the salinities in the interior Delta. CCC is influenced by both land salts (due to leaching of fields by farmers) and ocean salts while CHIPPS is effected mainly by ocean salt.

Performance Indicators

The cost function used to train the neural networks was the sum-squared error. The mean sum squared error (MSE) criteria was used in evaluating the fit in calibration and validation data sets. Generalization was used as an indicator of network performance. This generalization was defined by the MSE of the ANNs on the validation set.

Preparation of Input and Some Criteria Used for ANN Architecture Selection

Net Delta Outflow is made up of many component flows, the major ones are Sacramento River flow, San Joaquin River flow, Eastside stream flows, Central Valley Project pumping, State Water Project pumping, and channel depletions. These component flows along with the Delta Cross-Channel Gate position were given as multiple inputs to ANNs in a time-lagged fashion.

Inputs and outputs were scaled to range (0.2, 0.8). A daily time step was used for the purpose of this analysis. All inputs were presented in a time-lagged manner to represent information of the memory of the phenomenon. For example, for Sacramento River flow (SAC) input, the present flow, previous seven days flow and previous to that seven days period, and 10 previous weekly flows were given as inputs to the ANN.

Feed-forward architecture has been used to due their simpler architecture and faster training times. Experimentation with the number of hidden neurons and number of hidden layers was done to determine a network with enough flexibility to do well on the calibration set and yet generalize well enough to do well on the validation set (Sandhu and Finch 1995). Two layers with four neurons in the first layer and two in the second layer were used for most locations.

SAC Flow Estimator

The goal was of predicting SAC, given other flow conditions and gate positions, such that a certain target salinity level is achieved. Historical hydrology from years 1970-1990 was used. In any month the target salinity was defined as the DSM simulated salinity for that month and the following month. ANNs serve as a function relating the input flows and gate positions to the salinity.

The above problem then became a root-finding problem with SAC as the variable which needed to be determined. The cost function was defined as the difference between ANN simulated salinity and DSM simulated salinity. A simple bisection search method was used to solve for SAC.

Discussion of Results

Validation plots for CHIPPS and CCC are shown in Figure 2-1 and Figure 2-2, respectively. The scatter for CCC is greater and this may be due to the influence of land salts (salts introduced into the channels due to leaching of fields by farmers).

Sacramento river flow (SAC) was solved for b using monthly DSM simulated salinity values (years 1970-1990) for CHIPPS and CCC along with other historical conditions (Figure 2-3 and Figure 2-4).

At CHIPPS <2000 TDS the relationship between flows and salinities gets blurred. Thus, a cutoff value of CHIPPS > 2000 TDS was chosen for CHIPPS. Also the regions of interest are the higher salinity periods. Percentage errors for predicted SAC were within +/-25 percent for 95 percent of data (CHIPPS > 2000 TDS).

The reason for choosing CHIPPS > 6000 TDS cutoff for CCC is that the influence of ocean salt on the interior Delta stations becomes dominant only at higher salinity levels at the mouth of the Delta. Percentage errors for predicted SAC from CCC were within +/-25 percent for 94 percent of data (CHIPPS > 6000 TDS).

On further investigation of sensitivity, it was found that DXC showed a unreasonable amount of influence on CHIPPS. It was reasoned that this influence is caused by the correlation of DXC with SAC. It was decided to exclude DXC from the set of inputs, until this behavior could be better understood.

EAST was seen to have minimal effect on salinity throughout the Delta. It was therefore decided to exclude EAST from the set of inputs. The resulting ANN models results are shown in Figures 2-5 to 2-8.

ANN with NDO as the only input has been compared before (Sandhu and Finch 1995) and the improvement of performance of the multiple input ANN over the NDO – only ANN is shown in the case of SAC predictions from CCC (Figure 2-9).

Also shown (Figure 2-10) are the SAC predictions from G-model predictions for CCC. The G-model is calibrated for Jersey Point salinity and then transformed to predict CCC. The G-model was calibrated for Jersey Point salinity using the sum squared criteria.

Future Directions

Historical flows and gate operations are correlated. Thus individual effects of inputs may not be properly ascertained. It is proposed to run DWRDSM varying one input at a time and then training ANNs on the data generated.

Further sensitivity analysis of ANNs has to be done to ascertain if responses to various inputs are reasonable. This sensitivity analysis can be done using the graphical simulator built for this purpose.

One of the important issues in modeling is the confidence in the output of the model. To get an estimate of the errors involved with the ANNs and an indication of the confidence interval, bootstrapping or some similar technique would be used.

References

Masters, Timothy. 1993. Practical Neural Network Recipes in C++, San Diego: Academic Press, Inc.

Sandhu, Nicky, and Ralph Finch. 1995. Artificial Neural Networks with Applications to the Sacramento-San Joaquin Delta. Proceedings of the 4th ASCE International Conference on Estuaine and Coastal Modeling. San Diego, California. 490-504.

Shumway, Robert H., and Rahman S. Azari. 1993. Statistical Analysis of Historical Delta Water Quality Data. Report prepared for Delta Modeling Section, Division of Planning, California Department of Water Resources under Contract No. B-59259.

Weigant, Andreas S., and Neil A. Gershenfeld. 1992. Time Series Prediction: Forecasting the Future and Understanding the Past. Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis. Santa Fe, New Mexico. May 14-17, 1992. pp. 25-29.

Winkler, Edward D. 1985. A Statistical Approach to Salinity Modeling in the Western Sacramento-San Joaquin Delta and Suisun Bay. California Department of Water Resources, Technical Information Record, Study Code No. 1463-CD-04.

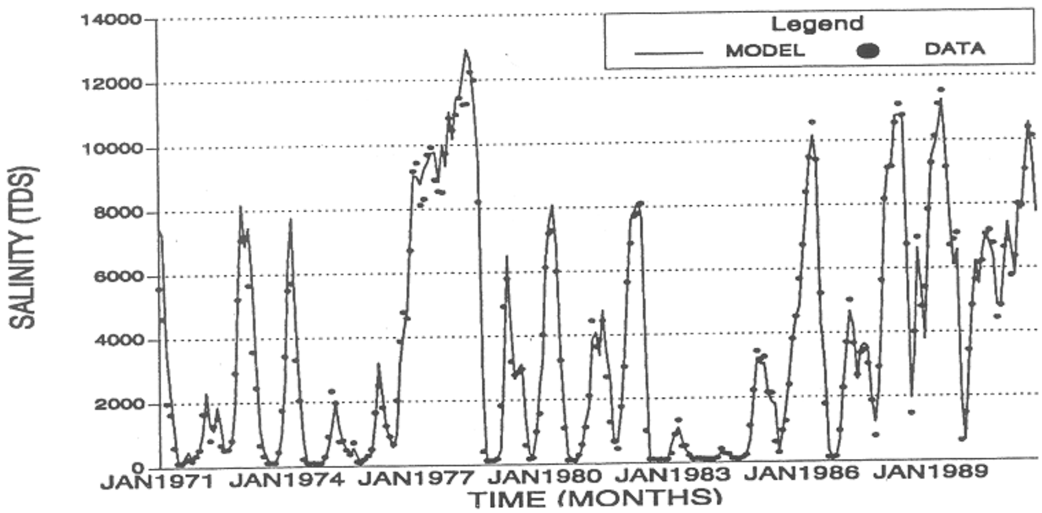

Figure 2-1. CHIPPS Time Series Plot

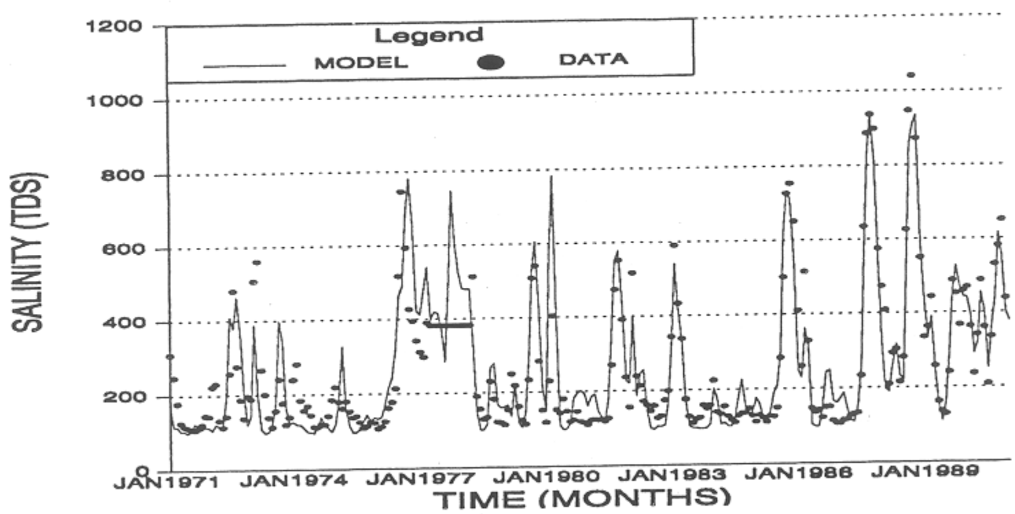

Figure 2-2. CCC Time Series Plot

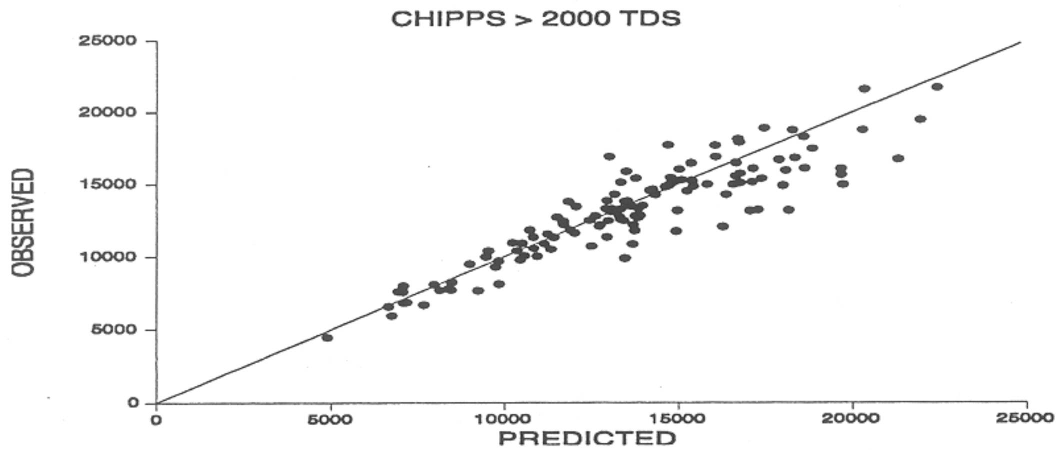

Figure 2–3. SAC Prediction from CHIPPS

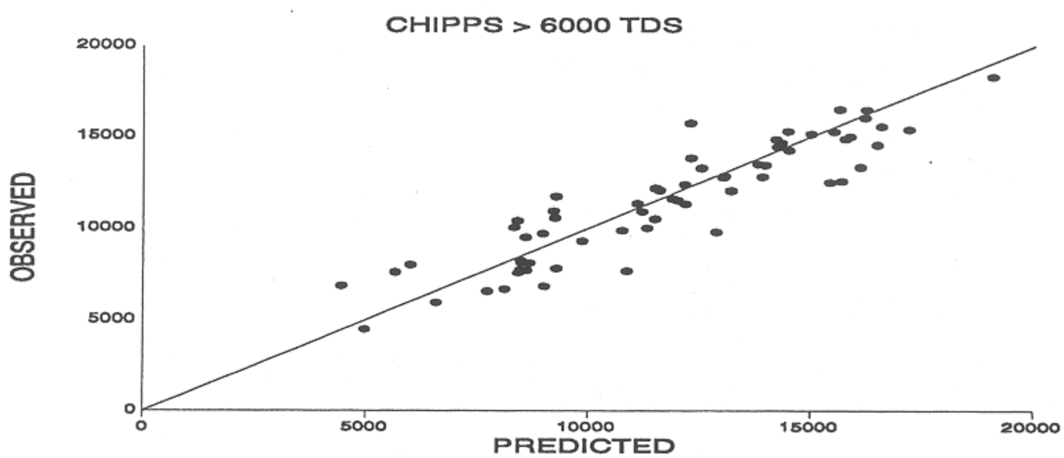

Figure 2–4. SAC Prediction from CCC

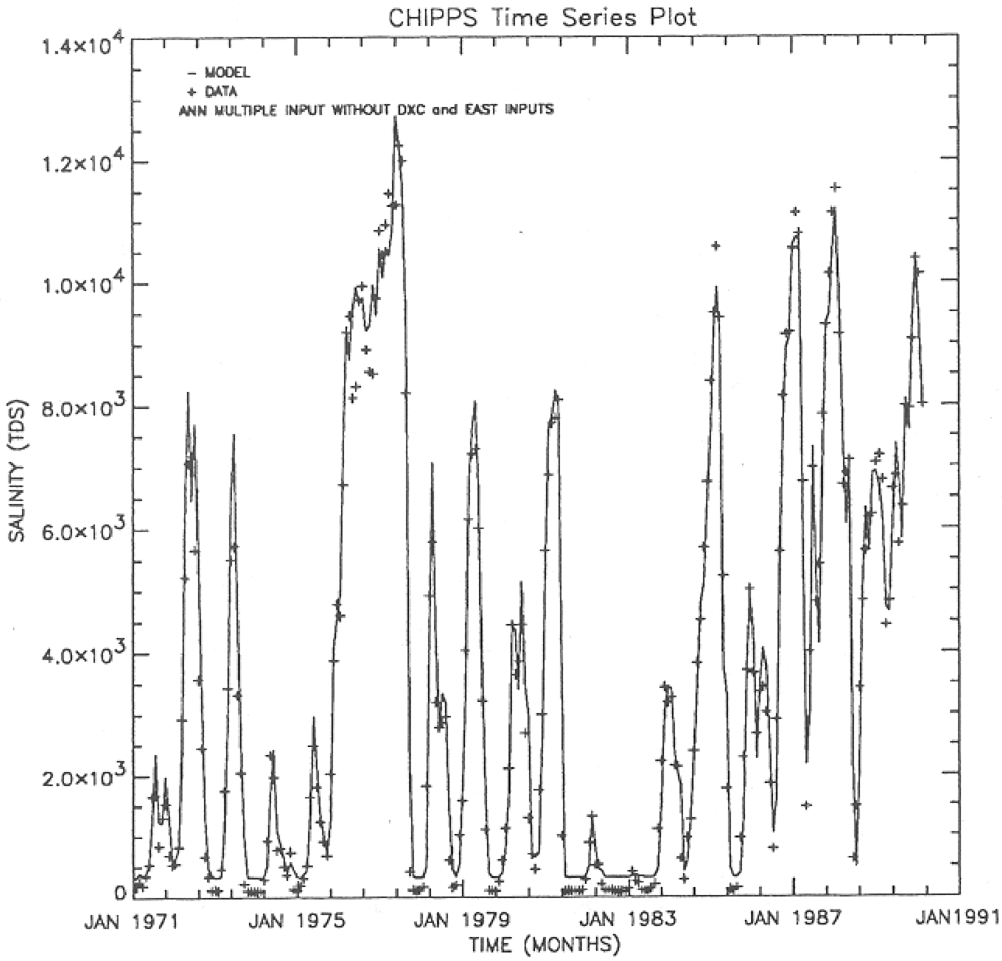

Figure 2–5: CHIPPS Prediction Without DXC and EAST

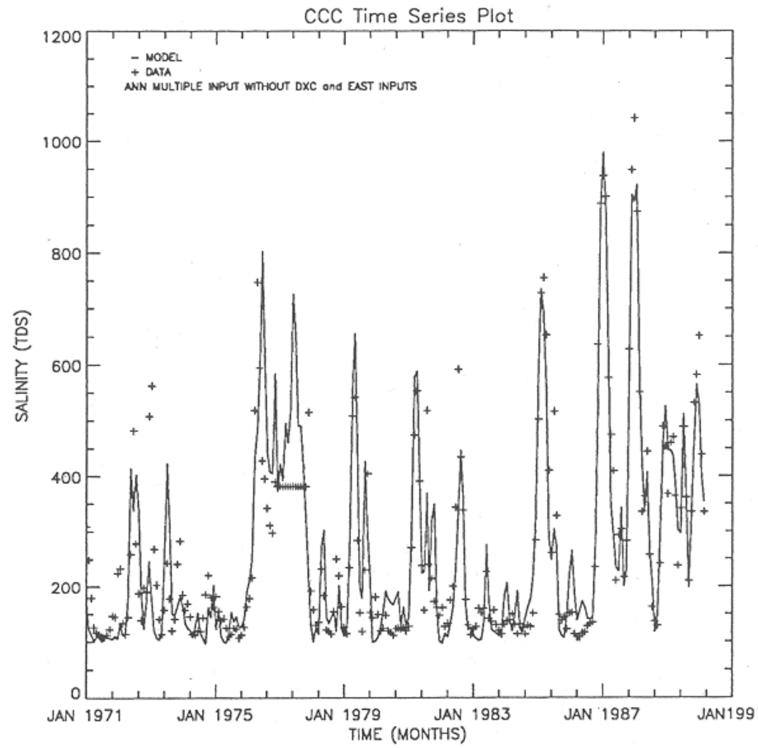

Figure 2–6. CCC Predictions Without DXC and EAST

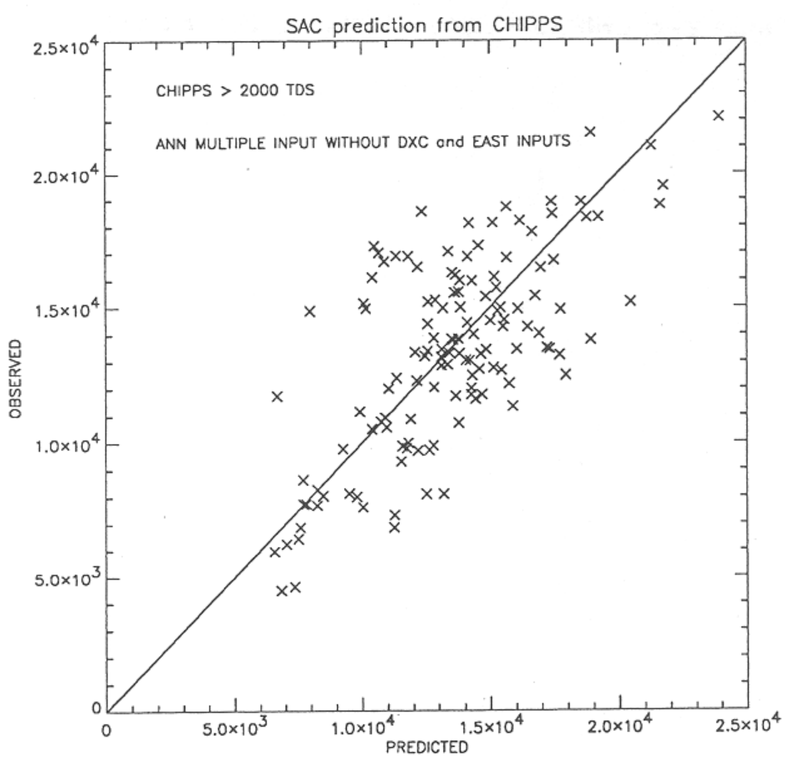

Figure 2–7. SAC Predictions from CHIPPS

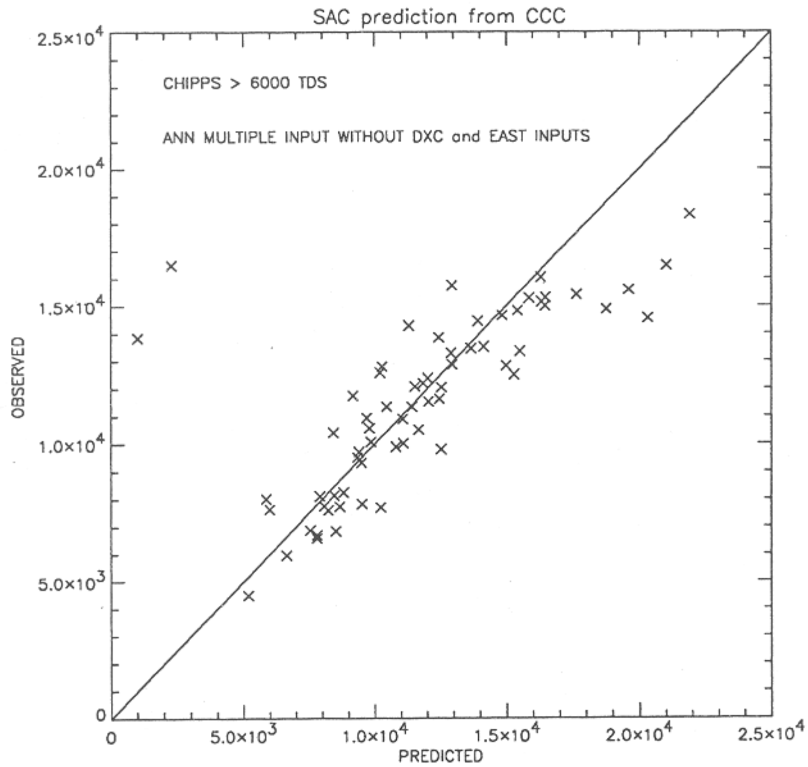

Figure 2−8. SAC Prediction from CCC

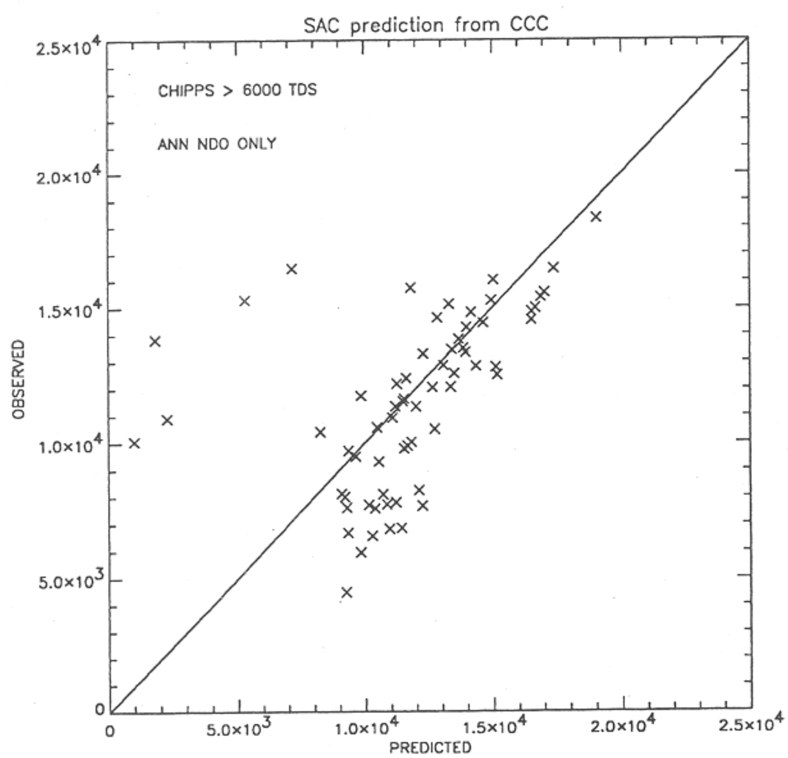

Figure 2−9. SAC Prediction from CCC Using ANN NDO Only

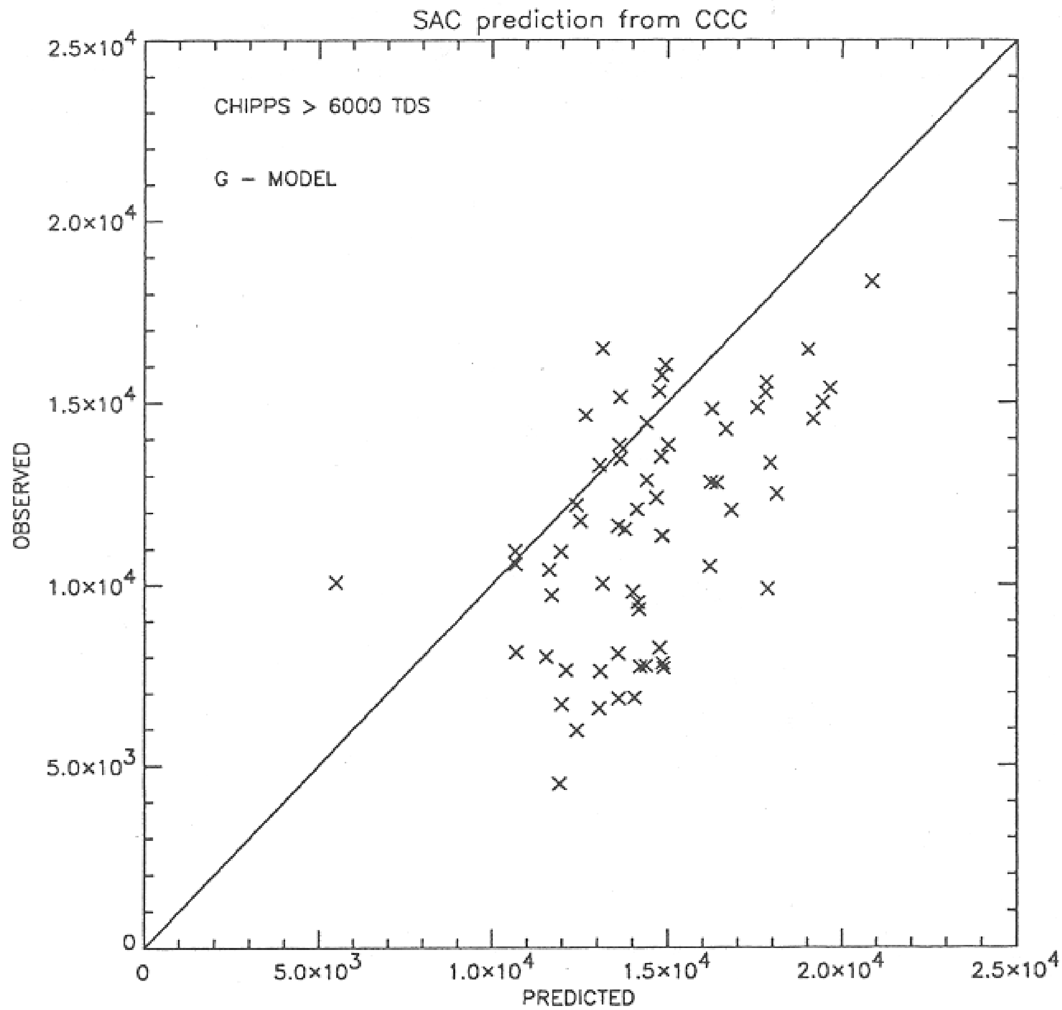

Figure 2−10. SAC Prediction from CCC Using G−Model