7th International Conference on Hydroinformatics
HIC 2006, Nice, FRANCE

An Artificial Neural Network-Based Decision Support System to Evaluate Hydropower Releases on Salinity Intrusion

PAUL A. CONRADS
U.S. Geological Survey, 720 Gracern Road, Suite 129
Columbia, SC 29210, United States

EDWIN A. ROEHL JR
Advanced Data Mining, LLC 3620 Pelham Road, PMB 351
Greenville, SC 29615, United States

Six reservoirs in North Carolina, USA, discharge into the Pee Dee River, which flows 260 kilometers through South Carolina to the coastal communities near Myrtle Beach. During the Southeast's record-breaking drought from 1998 to 2002, salinity intrusions inundated a coastal municipal freshwater intake, limiting water supplies. The North Carolina reservoirs are currently (2006) undergoing a re-licensing process by the Federal Energy Regulatory Commission for a 50-year operating permit. Stakeholders along the Pee Dee River formed a coalition to determine the necessary flows to protect the freshwater intakes in the future. Salinity intrusion results from the interaction of three principal forces—streamflow, mean tidal water levels, and tidal range. To analyze, model, and simulate hydrodynamic behaviors at critical coastal gages, data mining techniques were applied to more than 15 years of hourly streamflow, coastal water-quality, and water-level data. Artificial neural network (ANN) models were trained to learn the variable interactions that cause salinity intrusions. Streamflow data from the 47,900-square-kilometer basin are used as input to the model as time-delayed variables and accumulated tributary inflows. Tidal inputs to the models were obtained by decomposing tidal water-level data into a "multiply periodic" signal of tidal range and a "chaotic" signal of mean water levels. The ANN models were able to convincingly reproduce historical behaviors and generate alternative scenarios of interest. To make the models directly available to all stakeholders, a user-friendly decision support system was developed as a spreadsheet application that integrates the historical database, ANN models, user controls, streaming graphics, and simulation output.

INTRODUCTION

The Pee Dee Basin, with approximately 47,900 square kilometers of drainage area in eastern North and South Carolina, USA, supplies freshwater to the Grand Strand along the South Carolina coast from Little River Inlet to the north and Winyah Bay to the south [1] (fig. 1). Six reservoirs in North Carolina discharge into the Pee Dee River, which flows 260 kilometers through South Carolina to the coastal communities near Myrtle Beach. During the drought between 1998 and 2002, salinity intrusion forced a municipal intake to close until increased streamflow moved the freshwater/saltwater interface downstream from the intake.

The North Carolina reservoirs are currently being re-licensed by the Federal Energy Regulatory Commission (FERC) for a 50-year operating permit. The water has important commercial value for generation of electric power, waterfront property development, water supply, assimilative capacity, navigation, and recreation. A coalition of stakeholders including Alcoa Power, Progress Energy, the Pee Dee River Coalition, and the South Carolina Department of Natural Resources sought to model the system's hydrodynamics and determine the minimum flow needed to protect coastal intakes.

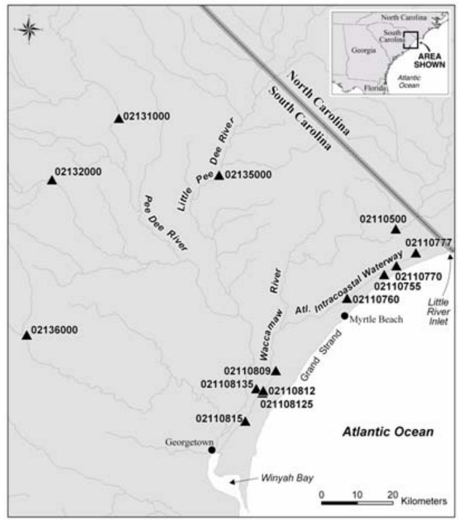

Figure 1. Location of study area and continuous gaging stations (triangles).

Water-resource managers and stakeholders face difficult challenges when managing the interactions between natural and man-made systems. Complex mathematical (mechanistic) models based on first principles physical equations are often developed and operated by scientists or engineers to evaluate options for using a resource while minimizing damage: however, the interests and computer skills of the actual decision makers and other stakeholders can be quite varied. To meet the needs of the technically diverse group of stakeholders, a decision support system (DSS) was developed in a spreadsheet application that integrates historical databases, simulation models, simulation controls, streaming graphics, and model outputs. The DSS met the need for equal access by all to the body of scientific knowledge needed to make the best possible decisions.

METHODS

The authors had previously developed ANN-based models of estuaries in Georgia and South Carolina. The type of ANN used was the multi-layered perceptron (MLP) described by Jensen [2], which is a multivariate, non-linear regression method based on machine learning. In a side-by-side comparison, Conrads and Roehl [3] found that ANN models had prediction errors 60-82% lower than those of a state-of-the-practice mechanistic model when predicting water temperature, specific conductance (SC), and dissolved oxygen on the Cooper River in South Carolina. In a regulatory application, Conrads and others [4] describe an ANN-based model for the permitting of three wastewater treatment plants that discharge into the Beaufort River estuary. In general, high-quality predictive ANN models can be obtained when:

- The data are well distributed throughout the state space (historical range of conditions) of interest,
- The input variables selected by the modeler share a lot of “mutual information” about the output variables, and
- The form “prescribed” or “synthesized” for the model used to “map” (correlate) input variables to output variables is a good one. Machine learning techniques like ANNs synthesize a best fit to the data.

Data Sets

The U.S. Geological Survey (USGS) maintains a real-time stream-gaging network of water level (WL) and SC recorders on the Pee Dee and Waccamaw River Basins. For the streamflow stations, there is greater than 50 years of record at the majority of the stations. For the coastal water-quality stations, there is greater than 15 years of WL and SC data. (Specific conductance is a field measurement that is commonly used to compute salinity concentration.) Data from the Grand Strand network are a valuable resource for addressing the critical conditions for salinity encroachment on the Pee Dee and Waccamaw Rivers. During the past 15 years of data collection, the estuarine system has experienced various extreme conditions including large rainfalls in a 24-hour period, the passing of major hurricanes offshore, and drought conditions.

For this study, a subset of the USGS data was used including nine coastal gaging stations that provided WL and SC data and five upland gaging stations that provided streamflow data sets. The data spanned 17½ years, but not all of the gaging stations were operating concurrently. The database for the study was augmented with rainfall data from six regional meteorological stations, and coastal wind speed and direction data from one additional meteorological station. The resulting database comprises 17½ years of hourly data (150,000+ time stamps) for 27 measured variables.

Data Preparation

Tidal systems are dynamic and exhibit complex behaviors that evolve over multiple time scales. The hydrodynamic and water-quality behaviors observed in estuaries are superpositions of behaviors forced by periodic planetary motions and chaotic meteorological disturbances. Theoretically, periodic behaviors are perfectly predictable, and chaotic behaviors are only somewhat so; therefore, the real problem with modeling estuaries is empirically synthesizing chaotic output signals from multiple chaotic input signals. Signals are easily decomposed into periodic and chaotic components using filtering techniques. The primary chaotic inputs to this system are the flows and the chaotic oceanic disturbances represented in the chaotic component of water level in Little River Inlet and Winyah Bay. The primary periodic input to the system is the tide.

The semi-diurnal tide is dominated by the lunar cycle, which is more influential than the 24-hour solar cycle; thus, a 24-hour average is inappropriate to use to reduce tidal data to daily values. For analysis and model development, the USGS data were digitally filtered to remove semi-diurnal and diurnal variability. To filter the semi-diurnal tidal signal, nested 24- and 25-hour moving window averages (MWA) were applied to the WL and SC time series. The resulting time series represents the daily change in the tidal signal for SC and WL on a 60-minute time increment. Digital filtering also can diminish the effect of noise in a signal to improve the amount of useful information that it contains. Working from filtered signals makes the modeling process more efficient, precise, and accurate.

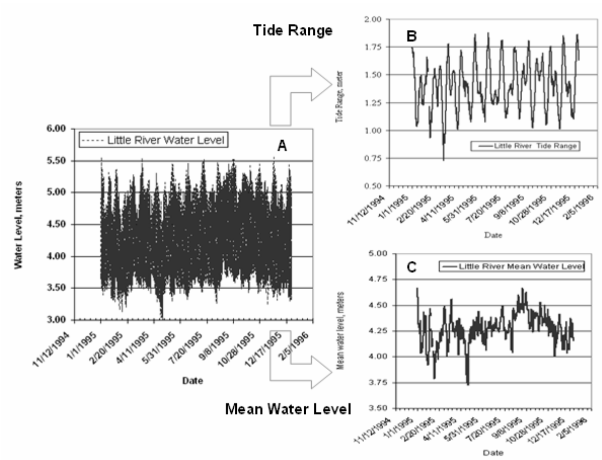

Figure 2. Tidal water-level signal (A), decomposition into a periodic signal of tide range (B), and a chaotic signal of mean water level (C).

Tidal dynamics are a dominant force for estuarine systems, and tidal range (XWL) is an important variable for determining the lunar phase of the tide. Tidal range is calculated from WL and is defined as the WL at high tide minus the WL at low tide for each semi-diurnal tidal cycle. As shown in Figure 2, the measured water level at Little River Inlet (Station 02110777) (fig. 2a) was decomposed into its periodic signal of XWL time series (fig. 2b) and its chaotic signal of mean WL time series (fig. 2c).

Conceptual Model and Data Analysis

The estuarine portions of the Pee Dee and Waccamaw Rivers are constantly integrating the changing streamflow, changing tidal conditions of the Atlantic Ocean, and changing meteorological conditions including wind direction and speed, rainfall, low and high pressure systems, and hurricanes. The location of the saltwater/freshwater interface is a balance between upstream river flows and downstream tidal forcing (fig. 3). During periods of high streamflow, it is difficult for salinity to intrude upstream and the saltwater/freshwater interface moves downstream towards the ocean. During periods of low streamflow, salinity is able to intrude upstream and the saltwater/freshwater interface

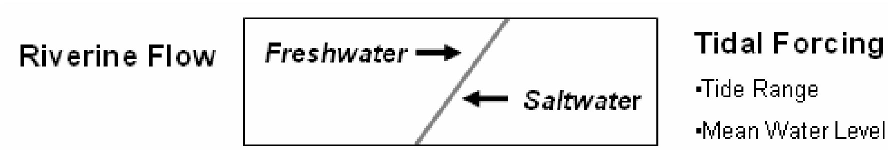

Figure 3. Diagram showing simplified conceptual model of the location of the freshwater/saltwater interface in estuarine rivers.

moves upstream.

Historically, streamflow in the Pee Dee River varies between 20 and 6,000 cubic meters per second (m^3/s) [5]. Salinity in the lower Pee Dee River is constantly responding to changing streamflow and tidal conditions. Figure 4 shows the daily mean SC values for the Hagley Landing gaging station (Station 02110815, fig. 1) and daily mean streamflow for Pee Dee River at Pee Dee (Station 02131000, fig. 1) for the 1995 to 2003 water years[1]. The period includes a full range of flows for the system from high flows of the El Niño in 1998 and 2003 (approximately, 1,500 and 3,500 m^3/s, respectively) to the low flows of the extended drought in the Southeast from 1998 to 2002. During periods of medium and high flows (streamflow greater than 200 m^3/s), the SC values are low. During periods of low flow (streamflow less than 80 m^3/s), values of SC values increase with increased salinity intrusion. During the low-flow periods prior to the high-flow El Niño of 1998, salinity intrusion with SC values ranging from 10,000 to 15,000 microsiemens per centimeter (µS/cm) were not uncommon. After the high flow of 1998 and during the extended drought, flows were even lower and remained lower for extended periods, which resulted in greater salinity at Hagley Landing with daily mean SC values greater than 15,000 µS/cm.

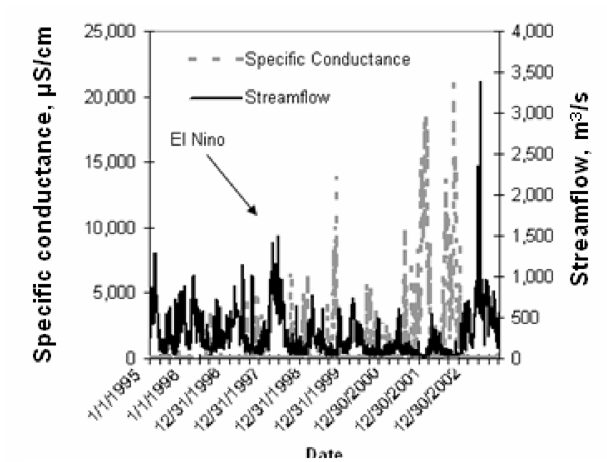

Figure 4. Graph showing Pee Dee River flow at Station 02131000 and specific conductance response at Hagley Landing (Station 02110815) for the period 1995 to 2003.

Model Approach

Subdividing a complex modeling problem into sub-problems and then addressing each is a means to achieving the best possible result. For the Pee Dee study, individual ANN models for SC were developed for nine continuous coast stream gages. The models were developed in two stages. The first stage modeled the chaotic, lower-frequency, portion of the signal, as represented by the filtered SC signals. The second stage modeled the periodic, higher frequency, hourly SC, using the predicted SC as a carrier signal. Each model uses three general types of signals, or time series: streamflow signal(s), WL

[1] Water year is the 12-month period October 1 through September 30. The water year is designated by the calendar year in which it ends.

signal(s), and XWL signal(s). The signals may be of the measured series values, filtered values, and/or a time derivative of the signals. The available dataset for developing the models were randomly bifurcated into training and testing datasets. Some small datasets were not bifurcated to maximize the information content in a signal. All ANN models were carefully evaluated to ensure the models did not "overfit" the data.

Eighteen models were developed: nine daily models and nine hourly models. Generally, the daily models used from one to four hidden-layer neurons and had coefficients of determination (R^2) ranging from 0.62 to 0.96. The hourly model used from two to six hidden-layer neurons and had R^2 values ranging from 0.69 to 0.92. An example of the measured and predicted daily and hourly SC response models are shown in Figure 5. The daily model is able to simulate the sharp SC spikes (fig. 5a) and the hourly model is able to simulate the high-frequency SC response (fig 5b).

Development of Decision Support System

Commonly, a DSS is a software package built around a model, making the model the DSS's most important component because ostensibly, it can correctly predict, *"What will happen if we do A instead of B?"* Models often are developed at considerable expense; therefore, the packaging is done only to maximize its usefulness to the broadest possible community of users. The DSS for the Pee Dee River was developed as Microsoft® Excel/Visual Basic for Applications[2] (VBA) programs. This allowed the DSS to be prototyped and easily modified, and also allowed the DSS to be distributed in a familiar form. The DSS operates through a graphical user interface (GUI) composed of point-and-

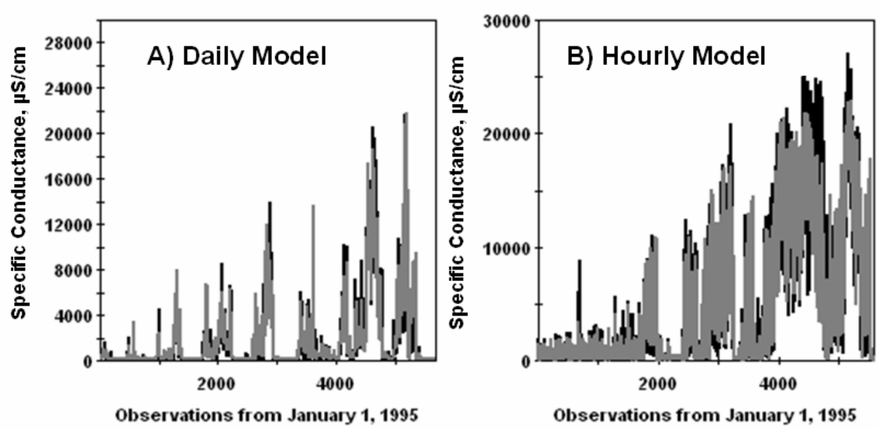

Figure 5. Graphs showing measured (black trace) and predicted (gray trace) specific conductance for Hagley Landing (Station 02110815). Results for the daily model are shown on the left (A) and the hourly model on the right (B).

---

[2] Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

click menus and controls (fig. 6). This makes the DSS easy to use and eliminates the need to trap user errors. The GUI also provides graphical outputs that depict measured and predicted hydrologic behaviors.

The model in the DSS is a “super model” that represents the whole system. The super model is composed of the 18 “sub-models” of the daily and hourly models for each gaging station. These sub-models are then incorporated into a “super-model” application that integrates the model controls, model database, and model outputs. This produces predictive models that are customized to the unique circumstances and data of a particular system. The DSS has at least two executions of the super-model. One generates predictions using actual historical input conditions, which are used to compute prediction errors and graphically depict accuracy. The second execution generates *“What if?”* predictions using user-established controllable inputs.

The Pee Dee DSS provides for simulations corresponding to the most recent and 6½ years of higher-quality data, at daily or hourly time steps. Streamflow inputs can be set by the user to be a constant or a percentage of the historical measurements. User-defined hydrographs can also be run. The Pee Dee DSS also provides a constrained optimizer that automatically modulates streamflow to match user-established maximum-SC setpoints. The setpoints can be applied on a daily or hourly basis. The Pee Dee DSS also provides built-in documentation that describes the variables and user controls. Documentation appears in pop-ups as the mouse is moved in the GUI.

Ultimately, stakeholders wanted to understand what causes the large salinity intrusions and what are the minimum flows that should be require in the FERC license to protect coastal intakes. Using the DSS, stakeholders were able to analyze the historical data and run model scenarios to determine that extreme salinity intrusions occur when there is a convergence of high mean WLs, usually the result of tropical storms, and low flow in the river. During these periods, large unrealistic flows from the reservoirs would be required to protect the intakes. The 14-day tidal cycle had insignificant impact during periods of extreme salinity intrusion.

SUMMARY

Effective environmental management of water resources relies on the information available from various sources including monitoring data, data analysis, and predictive models. Stakeholders of various technical backgrounds and ability need to be able to transform data into usable information to enhance understanding and decision making. To facilitate the technical transfer of historical data and predictive models for the Pee Dee Basin, a DSS was developed that would allow stakeholders to have equal access to the analytical tools. Stakeholders were able to determine a minimum flow to protect the intakes for a large range of hydrologic conditions. Stakeholders also realized that during extreme hydrologic conditions, the municipalities should have contingency plans rather than required excessive flows from the reservoirs to protect their intakes.

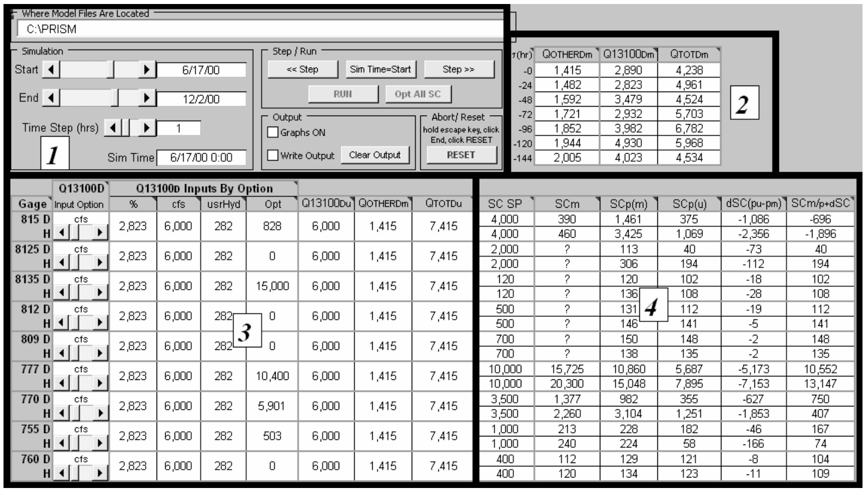

Figure 6. Graphical User Control Panel from Pee Dee River DSS - (1) simulation start/end/step; (2) input streamflow values and 1- to 7-day time delays; (3) user option for modified flows and their inputs; and (4) specific conductance values – measured, predicted, user defined, and impact.

REFERENCES

[1] U.S. Geological Survey, "*National Water Summary 1985- Hydrologic Events and Surface-Water Resources*", U.S. Geological Survey Water-Supply Paper 2300.

[2] Jensen, B.A., "*Expert systems - neural networks, Instrument Engineers' Handbook*" Third Edition, Chilton, Radnor PA, 1994.

[3] Conrads, P.A., and Roehl, E.A., "Comparing physics-based and neural network models for predicting salinity, water temperature, and dissolved-oxygen concentration in a complex tidally affected river basin", paper presented at the South Carolina Environmental Conference, Myrtle Beach, March 15-16, 1999.

[4] Conrads, P.A., E.A. Roehl, and Martello, W.P., "*Development of an empirical model of a complex, tidally affected river using artificial neural networks,*" Water Environment Federation TMDL Specialty Conference, Chicago, Illinois, November 2003.

[5] Cooney, T.W., Drewes, P.A., Ellisor, S.W., Lanier, T.H., and Melendez, F., "*Water Resources Data South Carolina Water Year 2002,*" U.S. Geological Survey Water-Data Report SC-02-1