Comparing Physics-Based and Neural Network Models for Simulating Salinity, Temperature, and Dissolved Oxygen in a Complex, Tidally Affected River Basin

Paul A. Conrads, Hydrologist, U.S. Geological Survey, Columbia, South Carolina
Edwin A. Roehl Jr., OptiQuest Technologies, LLC, Greenville, South Carolina

ABSTRACT

The U.S. Geological Survey participated in an evaluation of using neural networks for modeling estuarine systems, and selected physics-based, dynamic one-dimensional flow and water-quality models of the Cooper and Wando Rivers for comparison. The results showed that the neural network models were more accurate at simulating salinity, temperature, and dissolved oxygen of this very complex estuarine system than physics-based models. They were also far less costly to develop and can be easily integrated into commercial software, such as spreadsheets, and real-time data acquisition and control systems. However, the physics-based models may provide a better understanding of the system's behavior and provide a better methodology for extrapolating to environmental conditions that are not manifest in the historical data.

INTRODUCTION

Beginning in 1992, the U.S. Geological Survey (USGS), in cooperation with State agencies, initiated a study to apply the one-dimensional dynamic flow model BRANCH (Schaffranek and others, 1981) and the dynamic mass transport and water-quality model Branched Lagrangian Transport Model (BLTM) (Jobson and Schoelhamer, 1987) to the Cooper and Wando Rivers (Conrads and Smith, 1996, 1997). As part of the study (Conrads and others, 1997), the real-time gaging program for the Cooper River was expanded to include the Wando River (fig. 1). Additional gages were also added for the Cooper River. The water properties that were measured included water level, temperature, specific conductance (used to calculate salinity), and dissolved-oxygen concentration. Data from each station were collected at 15-minute intervals and transmitted via satellite communication to a database at the USGS District Office in Columbia, SC. The applied models were implementations of existing finite-difference computer codes that were calibrated and validated to match the flow, transport, and water-quality characteristics of the Cooper and Wando Rivers.

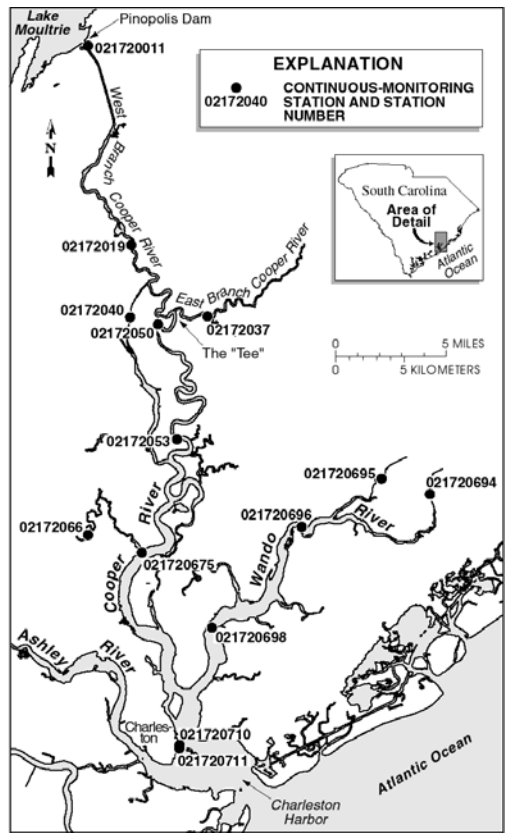

Figure 1. Study Area.

In early 1997, the USGS participated in an evaluation of using neural networks for modeling environmental systems. Unlike the finite-difference models, which are based on the principles of physics and chemistry (referred to as physics-based models), neural network models are inspired by the learning behavior of animals and humans (Jensen, 1994). The type of neural network model used in this study employs a set of mathematical relations with adjustable coefficients to map a set of input variables into a set of output variables. In a process called "training," neural networks implement a "learning algorithm" that, for a given data set, iteratively searches for coefficient values that provide a good generalized mapping of inputs to outputs.

This paper compares the results of the BRANCH/BLTM modeling effort, in which salinity transport, temperature, and dissolved-oxygen concentration were simulated, to a subsequent application of neural networks. Other aspects of the modeling approaches, such as time to apply the model, costs, and deployment options, are also discussed.

DESCRIPTION OF STUDY AREA

The Cooper River is formed by the confluence of the West and East Branches of the Cooper River at an area referred to as the “Tee” (fig. 1). The freshwater inflows to the West Branch are controlled by the releases from Pinopolis Dam on Lake Moultrie. The area above the Tee is characterized by meandering natural channels bordered by extensive tidal marshes and old rice fields in varying states of disrepair. Downstream of the Tee, industries are located along the west bank of the river. The east bank is dominated by extensive *Spartina alterniflora* salt marshes and contains numerous dredge-material disposal areas. Saltwater in the Cooper River extends from the Harbor upstream to several miles below the Tee. The Wando River is a tidal slough that tapers from a width of about ½ mile at its mouth to a narrow tidal creek 21 miles upstream from the confluence with the Cooper River. Saltwater extends throughout the Wando River. The banks of the river are dominated by extensive *Spartina alterniflora* salt marshes. The Cooper and Wando Rivers are tidally affected throughout their entire reach, and have a spring-tidal range of approximately 6 feet at their lower reaches.

MODELING APPROACHES

Similar procedures were used for both modeling techniques. Data from “boundary stations,” gaging stations located at the external boundaries of the study area (stations 021720011, 02172037, 02172040, 02172066, 021720694, 021720695, and 021720710), were inputs to the models. Data from “internal stations,” gaging stations that were represented as internal nodes on the finite-difference grid of the physics-based models (stations 02172019, 02172050, 02172053, 021720675, 021720696, and 021720698), were the model outputs. The models computed hydraulic properties and water-quality constituent values for internal stations that were then compared with measured values.

Physics-Based Models

Modeling was completed in two phases. The first phase calibrated and validated the unsteady-flow model (BRANCH) and the branched Lagrangian transport model (BLTM) to simulate the movement of a conservative constituent (salinity) in the system. The scope of the second phase was to calibrate and validate the water-quality model (BLTM) to simulate the fate and transport of non-conservative constituents, such as nutrients, biochemical oxygen demand (BOD), and dissolved oxygen.

Several types of data were required to apply, calibrate, and validate the transport and water-quality models. A large data-collection effort was completed during 1992-95 (Conrads and others, 1997). The data included: (1) continuous water level, specific conductance, temperature, and dissolved-oxygen concentration at the upstream and downstream boundaries and at an interior location; (2) tidal-cycle measurements of streamflow and nutrient concentrations at boundaries and selected interior locations; (3) channel geometry; (4) municipal and industrial discharge-flow rates and effluent concentrations; and (5) meteorological data including equilibrium temperature, solar radiation, and wind speed.

In the BRANCH and BLTM models, rivers are represented as a series of cross sections and channel lengths, which define segments, junctions, and branches. The BRANCH model computes hydraulic properties at each cross section. The BTLM model uses the water-quality reaction kinetics found in the QUAL2E model (Brown and Barnwell, 1987) to simulate the fate and transport of nutrients, BOD, and dissolved oxygen. It computes water temperature, nutrient concentration, BOD, algal biomas, and dissolved-oxygen concentration at every cross section. The schematization of the BLTM for the Cooper and Wando Rivers uses 30 branches, 7 external boundaries, 16 internal junctions, and 112 cross sections.

The BRANCH model was calibrated by adjusting flow-resistance coefficients, gage datums, cross-sectional areas, storage volumes, and dispersion-rate model parameters until simulated and measured (or calculated) values met a predetermined calibration goal.. The calibration and validation periods for the BRANCH model for water level, streamflow, and salinity were July 30-31 and September 24-25, 1993. Because the model was ultimately used to simulate the fate and transport of conservative and non-conservative constituents, emphasis was placed on the salinity-transport simulations during the calibration and validation. Months were spent getting satisfactory simulations of water levels, streamflow, and salinity transport to characterize the mass transport in the system.

Calibration of the BLTM model was accomplished by adjusting rate kinetics and model parameters that control the dynamics of photosynthesis, nitrogen and phosphorus cycling, BOD decay, sediment oxygen demand, and reaeration. The calibration and validation periods for the BLTM water-quality model were April 8 to May 8 and August 1-30, 1993.

The models were validated by using measured data different from those used for calibration. The parameters used to calibrate the hydraulic and mass-transport models were not changed in the validation process. The calibration and validation process took approximately two person-years to accomplish.

Neural Network Models

The approach of the neural network model evaluation followed the boundary and internal station input/output representation used for the BRANCH and BLTM models. However, unlike the BRANCH and BLTM models that were calibrated by systematically adjusting various model parameters, the neural network models were "trained" on the measured time-series data. The type of neural network employed is called a "back-propagation" network (Jensen, 1994) and is especially good at fitting high dimension, continuously valued functional approximations to data.

A requirement for training and evaluating neural networks is that model data be arranged into input/output vector pairs representing the variables of interest. A vector pair having one or more missing measurements cannot be used for training a model and an input vector with a missing measurement cannot be evaluated by a developed model. Therefore, the efficacy of implementing a neural network model is highly dependent on the quantity, historical range, and quality of the data used for their development.

Bifurcating the modeled data set into "training" and "test" sets provides verification of a model's accuracy. A model trained on the training set can be evaluated by comparing its predictions to the measured values of the test set. The resulting empirical model is subject to the same statistical characterizations of accuracy and applicability as physics-based models. For this study, the data were bifurcated into 80% training data and 20% test data by random number selection.

The first neural network implementation focused on developing a model that would apply all of the 3½ years of available time-series data to evaluate the feasibility of performing long-term simulations (Roehl and Conrads, 1999). This departed from the more temporally restricted calibration procedure used for BRANCH and BLTM models, which had been focused on modeling "critical conditions" that cause dissolved-oxygen concentrations to be low. The 1992-95 data were converted from 15-minute to 30-minute intervals by simply not using the extra values. The time-series for all but a few variables contained numerous large gaps where data was not taken or sensors had failed, resulting in a need to fill in the gaps to obtain a satisfactory number of complete input/output vector pairs spanning the 3½ years. This was accomplished by developing more than 90 neural network models to generate synthetic data, which were appended to the actual data. Variables that had relatively large numbers of measurements were used as inputs to models whose outputs were variables with fewer measurements. A model was developed using the "appended data set" which was then compared to the physics-based models. This work required approximately two person-months with most of the time spent on formatting the raw data received from the USGS and building the appended data set.

A second implementation was performed to directly compare the prediction accuracy of the two approaches. This involved developing a series of models that focused on the same critical condition periods that were used to evaluate the BRANCH and BLTM models. The training and test sets omitted the use of synthetic data, and covered periods as follows: water level and salinity - July 1 to October 21, 1992, water temperature - April 1 to October 15, 1993, and dissolved-oxygen concentration - August 1 to October 1, 1993. These periods were much longer than those used to calibrate the physics-based models to insure that the neural network models would learn the range of hydrodynamic and biochemical behaviors exhibited by the estuarine system. This work required approximately one person-week.

RESULTS

The neural network model (second implementation using the datasets that include the calibration and validation periods of the BRANCH and BLTM models) more accurately simulated the salinity, water temperature, and dissolved-oxygen concentrations of the Cooper and Wando Rivers than the BRANCH/BLTM models. Typical results are shown in figures 2 and 3. The reach on the Cooper River between station 021720675 and 02172053 is of particular water-quality concern. It is the location of the freshwater and saltwater interface and dissolved-oxygen "sag" in the system. Computed and measured salinity concentrations for two stations on the Cooper River are shown in figure 2. At the downstream station (021720675), both models satisfactorily simulate the salinity concentrations; however, the BRANCH/BLTM simulation has a timing error of 60 minutes. The root mean square error (RMSE) of the neural network simulation was 1.25 parts per thousand (ppt) as compared to 3.14 ppt for the BRANCH/BLTM simulations. Farther upstream at station 02172053, the neural network model more accurately simulates the range of salinity values (RMSE 0.36 ppt) than the BRANCH/BLTM, which is only able to simulate the lower range of salinity (RMSE 1.97 ppt).

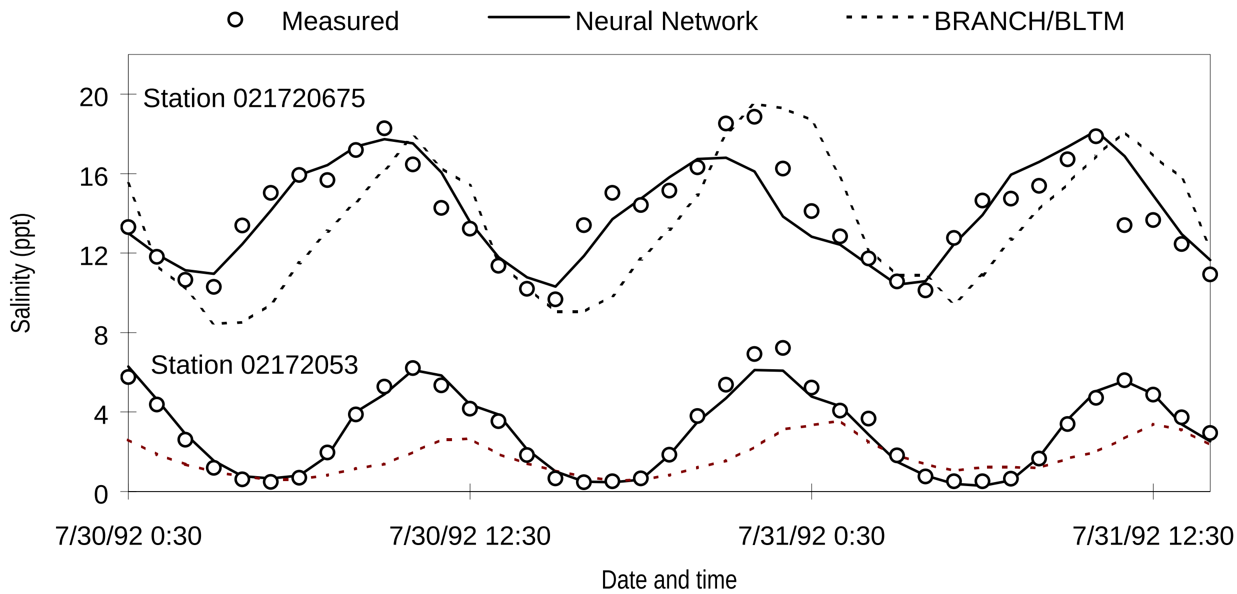

Figure 2. Computed and measured salinity (in parts per thousand) for two stations on the Cooper River.

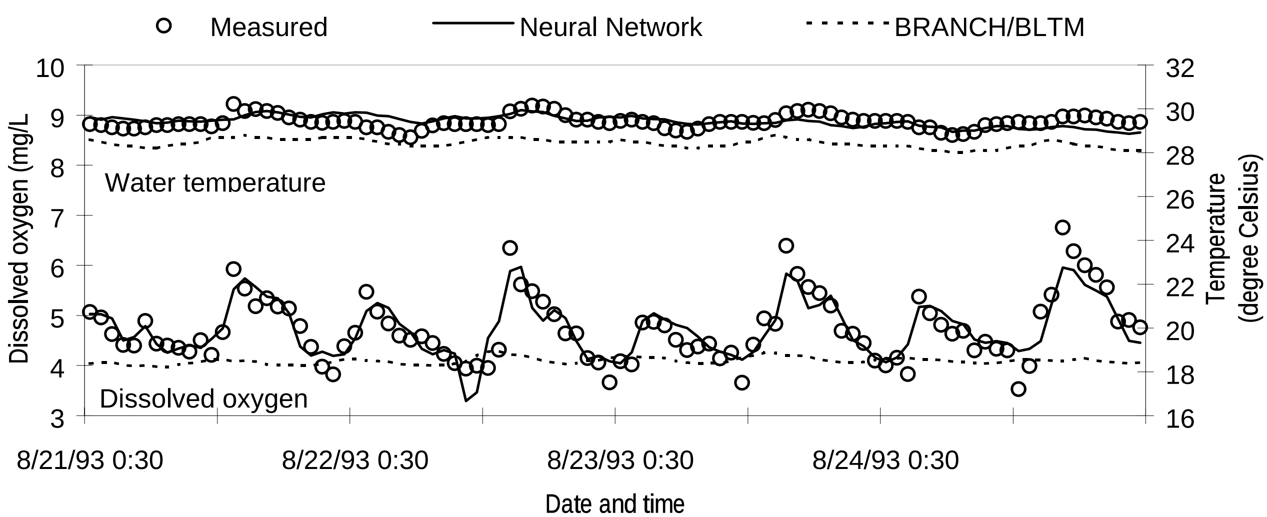

Figure 3. Computed and measured water temperature and dissolved-oxygen concentration at station 021720675 on the Cooper River.

Similar results are seen in the water temperature and dissolved-oxygen concentration simulations at station 021720675 (fig. 3). The BRANCH/BTLM water-temperature simulations are within two degrees Celsius of the measured temperatures (RMSE 1.48 degree Celsius). The neural network is able to simulate the water temperature to within a degree or less (RMSE 0.31 degree Celsius). The BRANCH/BLTM dissolved-oxygen simulations are within the range of the measured values but do not simulate the dynamics of the measured concentration (RMSE 0.81 milligram per liter (mg/L)). The neural network model is able to quite accurately simulate the tidal dynamics of the measured dissolved-oxygen concentration (RMSE 0.27 mg/L).

DISCUSSION

The evaluation of the BRANCH and BLTM models and neural network models highlighted the strengths and weaknesses of the two modeling approaches in a number of categories:

- *Data Requirements* - neural network models are developed directly from data; therefore, for a system to be adequately modeled there must be sufficient data available to fully describe the behaviors of interest. In the Cooper and Wando River systems, a neural network trained only for "critical conditions" will have little ability to perform for other conditions. In a physics-based model, knowledge and information contained in the modeling program may in some circumstances provide for a lower data requirement; however, the data must still be adequate for calibration and testing. The physics-based models used 30-day datasets for calibration and validation; whereas, the neural network model used datasets of four to seven months for training and testing. Alternatively, the neural network models required only 12 to 15 variables to achieve the very accurate predictions shown above. This was far fewer than the number used for the physics-based models and underscores how highly correlated the boundary station measurements were with the internal station measurements, and the trade off between numbers of variables and time-series length.

- *Temporal and Spatial Interpolation* - the BRANCH and BLTM models are able to generate time series of constituents of concern at any point along their finite difference grid. This connotes both temporal (time) and spatial (location) interpolative capability. The neural network models as configured for this study were unable to interpolate spatially and could only make predictions at the internal stations. However, other work by the authors in modeling 3D ground-water flows indicates that neural network models can perform combined temporal and spatial interpolation as well.

- *Extrapolating Beyond Range of Data* – extrapolation is using a model to make a prediction for a "new" process state that is appreciably different from those manifest in the data used to calibrate and test the model. Neural network models can be designed to extrapolate to some degree; however, collecting data that characterizes a new state of interest and retraining the model is the preferred course. Physics-based models may be better at extrapolating if the physics of the new state is similar enough to the physics of the model.

- *Providing Process Understanding* – preparing data, and configuring, calibrating, and testing computer models is a highly analytical endeavor from which fundamental understanding is derived. This is true for both types of models because their performance cannot be evaluated without a qualitative understanding of the physics of the modeled system. The model with the best accuracy over the broadest range of process conditions will best augment qualitative understanding with quantitative details that are necessary for regulating, controlling, and optimizing a process.

- *Development Time and Costs* - the time involved to apply the neural network model to the Cooper and Wando Rivers was approximately a tenth of the time to apply the BRANCH and BLTM models. The time difference is significant in terms of cost and timeliness of obtaining results. There are two large costs in applying models—(1) collecting, processing, and archiving data and (2) paying personnel to apply the models. Neural network models offer a significant saving to the personnel costs. The implementation of the neural network to simulate 3½ years of data demonstrated a cost-effective approach to performing long-term simulations that maximized all the data, not just the calibration and validation periods of the physics-based models. Often water-resource managers need answers to their water-quality questions in months, not years. Neural networks can often provide answers more quickly than physics-based models.
- *Deployment Options* - the neural network models execute several orders of magnitude faster than the BRANCH and BLTM models. Unlike physics-based models that perform multiple iterations every time step to compute the output values, trained neural networks execute without iteration. Therefore, the neural network models can be deployed as compact programs that are suitable for integrating with optimization routines and real-time information and control systems. The neural network models can be readily disseminated to a large number of water-resource managers by integrating into commonly used spreadsheet software.

CONCLUSIONS

Applying neural network models to highly complex hydrologic systems offers great potential for analyzing large amounts of time-series data and for assessing the impact of various hydrologic conditions. The results of this study indicate that neural network modeling can be an effective modeling tool under the right circumstances. This technique was well matched to modeling the Cooper and Wando Rivers because of the large amount of high-quality data provided by the USGS real-time monitoring program. The 3½ years of data used for this effort increased the likelihood that the range of behaviors modeled would be representative of the environmental system currently and for the foreseeable future. In addition, the execution time of the neural network models is several orders of magnitude faster than the BRANCH and BLTM models, and the neural network models can be readily integrated into spreadsheet programs to provide an extremely cost effective and user-friendly deployment vehicle.

REFERENCES

Brown, L.C., and Barnwell, T.O., Jr., 1987, The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS--Documentation and user manual: Athens, Georgia, U.S. Environmental Protection Agency, Environmental Research Laboratory, EPA/600/3-87/007, 189 p.

Conrads, P.A., Cooney, T.W., and Long, K.B., 1997, Hydrologic and water-quality data from selected sites in the Charleston Harbor Estuary and tributary rivers, South Carolina, water years 1992-95: U.S. Geological Survey Open- File Report 96-418, 987 p.

Conrads, P.A., and Smith, P.A., 1996, Simulation of water level, streamflow, and mass transport for the Cooper and Wando Rivers near Charleston, South Carolina, 1992-95: U.S. Geological Survey Water-Resources Investigations Report 96-4237, 51 p.

---- 1997, Simulation of temperature, nutrients, biochemical oxygen demand, and dissolved oxygen in the Cooper and Wando Rivers near Charleston, South Carolina, 1992-95: U.S. Geological Survey Water-Resources Investigations Report 97-4151, 58 p.

Jensen, B.A., 1994, Expert systems - neural networks, instrument engineers' handbook third edition: Chilton, Radnor PA, p. 48-54.

Jobson, H.E., Schoelhamer, D. H., 1987, Users manual for a Branched Lagrangian Transport Model: U.S. Geological Survey Water-Resources Investigations Report 87-4163, 73 p.

Roehl, E.A., Conrads, P.A., 1999, Real-time control for matching wastewater discharges to the assimilative capacity of a complex, tidally affected river basin: 1999 South Carolina Environmental Conference, Myrtle Beach, March 15-16, 1999.

Schaffranek, R.W., Baltzer, R.A., Goldberg, D.E., 1981, A model for simulation of flow in singular and interconnected channels: U.S. Geological Survey Techniques of Water Resources Investigations, Book 7, Chap. C3, 100 p.