DEVELOPMENT OF AN EMPIRICAL MODEL OF A COMPLEX, TIDALLY AFFECTED RIVER USING ARTIFICIAL NEURAL NETWORKS

Paul A. Conrads*, Edwin A. Roehl, William B. Martello

*Hydrologist, U.S. Geological Survey, 720 Gracern Road, Columbia, SC, 29210

ABSTRACT

The Beaufort River is a complex estuarine river system that supports a variety of uses including shellfish grounds, fisheries nursery habitats, shipping access to Port Royal, receiving waters for wastewater effluent, and an 18-mile reach of the Intracoastal Waterway. The river is on the Section 303(d) list of impaired waters of South Carolina for low dissolved-oxygen concentrations. The Clean Water Act stipulates that a Total Maximum Daily Load must be determined for impaired waters.

An empirical model was developed to simulate the impact of point-source discharges and rainfall on dissolved-oxygen concentrations in the Beaufort River. The model uses of water level, specific conductance, temperature, and dissolved-oxygen concentration data collected at 15-minute intervals from seven real-time gaging stations and effluent point-source data collected on a weekly basis for a 33-month period. The empirical model utilizes data mining techniques, including artificial neural network (ANN) models, to quantify the relations between the time series of three wastewater point-source discharges and the dissolved-oxygen concentrations recorded at seven real-time gages distributed throughout the system. Data mining is a new science that extracts knowledge from large volumes of data, and uses attributes from fields such as computer science, signal processing, advanced statistics, machine learning, and chaos theory. The data mining produced a high-fidelity water-quality model that can predict the impacts that point and non-point source loads have on the dissolved-oxygen concentration throughout the river system. The analysis included environmental factors such as tides, specific conductance, water temperature, and rainfall. The model is comprised of numerous sub-models that are based on ANN models.

The data analyses and model provided unique ways to evaluate complex tidal dissolved-oxygen effects from point-source discharges and rainfall. The model executes non-iteratively, making it amenable to very long-term simulation runs of 33 months. The model also included a non-linear, constraint-based numerical optimizer to determine the maximum allowable daily effluent loading without violating the State's water-quality standard. Insights were garnered from this technical approach that leveraged the full historical record in which assimilative capacity was found to be constantly changing. For example, critical conditions for effluent impacts on dissolved-oxygen concentrations occur during neap tides due to the streamflow characteristics and limited flushing of the system. The predictive model/optimizer allowed for a variety of wastewater treatment plant operating scenarios and regulatory options that can be quickly evaluated. Several 33-month time series of daily loadings were simulated utilizing an optimizer. Frequency distributions of the allowable loading were subsequently generated from the time series of optimal loading. Water- resource managers can use the frequency distribution to help predict the percentage of time water-quality standards may be violated. Model dissemination is facilitated by incorporating the ANN sub-models and point-source optimizers into an Excel spreadsheet application. This paper describes the data collection and analysis, model development and Excel application, point-source load optimization, and interpretation of model results from this unconventional approach to estuary water-quality modeling and regulatory control.

KEYWORDS

Total Maximum Daily Loads, data mining, artificial neural networks, water-quality model, Beaufort River, optimization

INTRODUCTION

The Beaufort River is on the South Carolina Section 303(d) list of impaired waters for low dissolved-oxygen (DO) concentrations due to natural conditions. Although monitoring by South Carolina Department of Health and Environmental Control (SCDHEC) has indicated a decreasing trend in total phosphorus and 5-day biochemical oxygen demand (BOD5) concentrations which suggests improving conditions, it also indicates a decreasing trend in DO (SCDHEC, 1997). The Clean Water Act stipulates that Total Maximum Daily Loads (TMDLs) must be determined for impaired waters. Prior to developing an effective TMDL for the Beaufort River, water resource regulators need to address the following critical questions concerning the hydrology and water quality of the estuary:

- What is the volume and direction of flow in the estuarine river system?
- What hydrologic and water-quality conditions contribute to the low dissolved- oxygen concentrations?
- What are the relative impacts of point-source and non-point source loads on dissolved-oxygen concentrations?

In order to establish a TMDL for DO in the river, Beaufort-Jasper Water and Sewer Authority (BJWSA) initiated and directed the development of a dynamic water-quality model for SCDHEC to use in determining the assimilative capacity of the Beaufort River system. In cooperation with BJWSA, a project team composed of scientists and engineers from Jordan, Jones & Goulding (JJG), Advanced Data Mining (ADM*i*) and the U.S. Geological Survey (USGS) developed the Beaufort River assimilative capacity model. Data computed with this model will enable water resources managers to estimate the effects that point- and non-point source loads have on the dissolved-oxygen concentration throughout the river system. The analysis included environmental factors such as water temperature, specific conductance, tides, and rainfall.

Figure 1. Study area.

This paper describes the development of the Beaufort River assimilative capacity model, including the results of reviewing real-time network data on the Beaufort River and applying data mining and artificial neural network models to the Beaufort River. The empirical model was built using over 1.5 million measurements from the continuous monitoring network and measurements from discrete sampling of the water reclamation facilities (WRF). The modeling scope of this effort consisted of two phases: (1) compiling, reviewing, and preparing hydrologic and water-quality data and subsequently predicting the impacts of tidal dynamics, rainfall, and point-source effluent (BOD and NH_3) on measured DO concentrations, and (2) constructing a predictive model of the river system by combining the predictions at each gaging station location.

DESCRIPTION OF STUDY AREA

The Beaufort River is a complex estuarine river system that connects Port Royal Sound to the south and St. Helena Sound to the north through Brickyard Creek and the Coosaw River (Figures 1 and 2). Crucial to the economic success of the region, the river and its tributaries support shellfish grounds, fisheries nursery habitats, shipping access to Port Royal, receiving waters for wastewater effluent, and an 18-mile section of the Intracoastal Waterway. The river experiences semi-diurnal tides of approximately 9 feet at its confluence with Port Royal Sound. The watershed consists primarily of sea islands and the tidally influenced creeks that separate them, with no significant drainage area providing fresh-water to the system. The Beaufort River assimilative capacity model study area essentially includes the entire basin and both SA (tidal saltwaters) and SFH (shellfish harvesting waters) water-use classifications.

Permitted Discharges

Four water reclamation facilities (WRFs) are permitted to discharge oxygen -consuming constituents into the Beaufort River and Albergottie Creek. BJWSA operates the Shell Point Plant and the Southside (SS) plants that are permitted at 0.8 and 4.0 million gallons a day (MGD) and wasteload allocations of 1,210 and 6,052 pounds per day (lbs/d) of ultimate oxygen demand (UOD), respectively. (UOD is the total, theoretical demand for oxygen from carbonaceous and nitrogenous sources.) The discharge location of the two BJWSA facilities is co-located just north of USGS station 2176611. In addition, there are permitted discharges from Parris Island (PI) and the Marine Corps Air Station (MCAS). The PI discharge is currently permitted at 3.0 MGD, with a waste load allocation of 4,539 lbs/d UOD. The MCAS discharge is currently permitted at 0.75 MGD, with a waste load allocation of 1,135 lbs/d UOD. These discharges represent a total permitted point-source loading of 13,843 lbs/d of UOD.

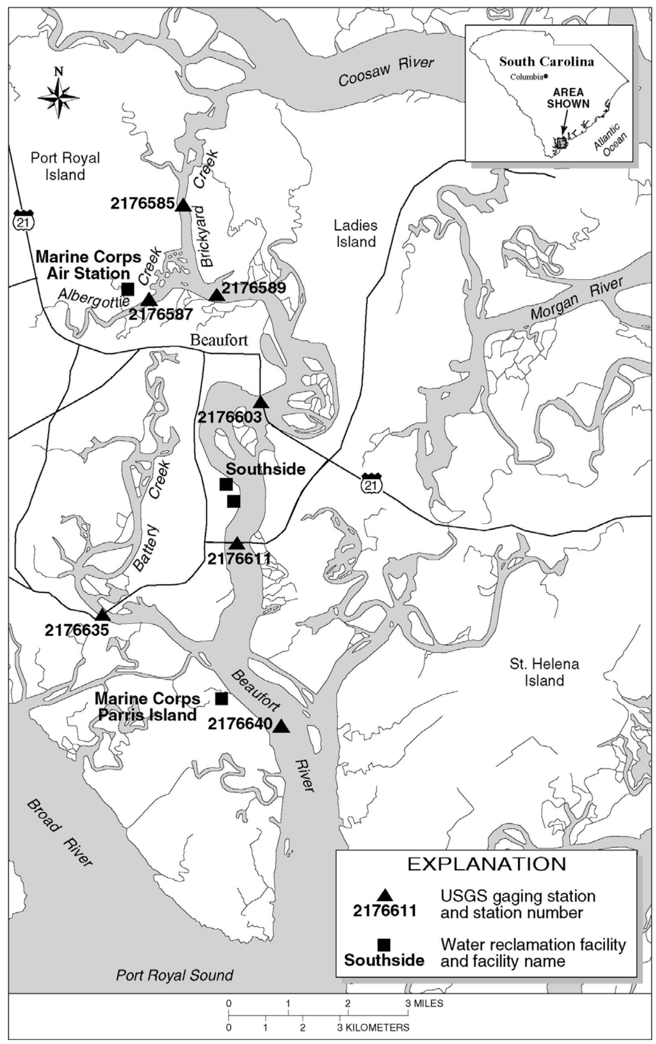

Figure 2. Beaufort River real-time gaging network and location of water reclamation facilities.

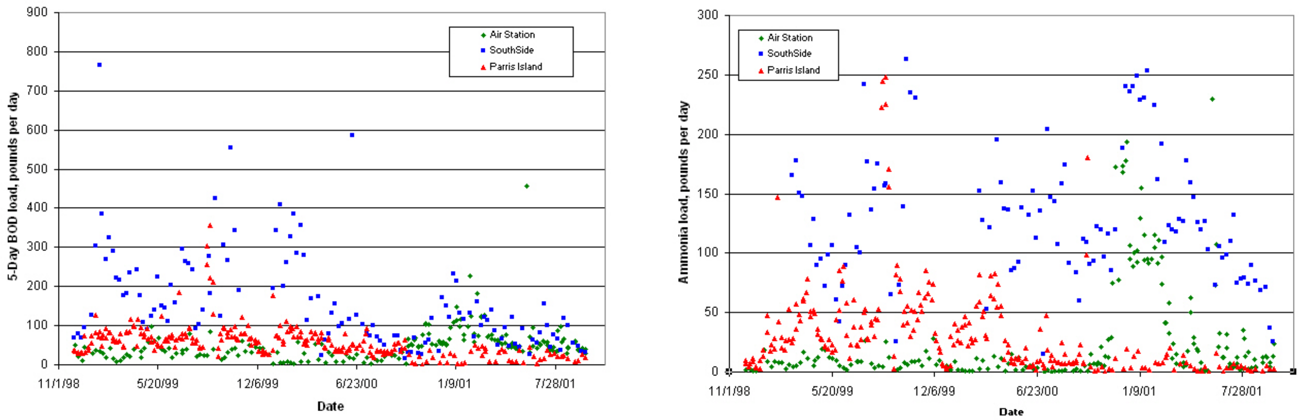

Figure 3. Biochemical oxygen demand loads (left) and ammonia loads (right) for the period January 1999 to September 2001.

Figure 3 shows the BOD5 and ammonia (NH_3) loads to the Beaufort River system from January 1999 to September 2001. BOD loads to the system were nearly 800 lbs/d in 1999, but generally have been below 200 lbs/d for the June 2000-July 2001 period. Ammonia loads to the system have been below 250 lbs/d. The WRFs collect weekly effluent samples, generally on different days of the week.

To estimate a time series of historical loading to the system in terms of UOD, the weekly loading values were interpolated. Due to the variability in effluent loads, data gaps greater than 7-days were not interpolated. Days with concurrent estimated UOD data were summed for an estimate of total UOD loading to the river. Figure 4 shows the estimated total UOD loading to the Beaufort River system. The greatest loading to the system for the period shown in the figure is approximately 2,900 lbs/d in early September 1999. The recent loading to the system has been less than 1,000 lbs/d.

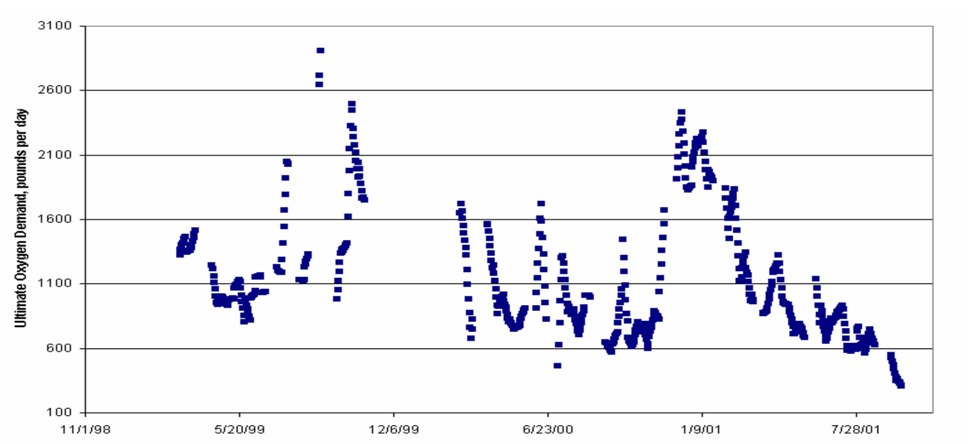

Figure 4. Ultimate oxygen demand load to Beaufort River January 1999 to September 2000.

Continuous Monitoring Network

To gain a better understanding of the Beaufort River and its tributaries, BJWSA, in cooperation with the USGS, established a network of seven real-time gaging stations in the Beaufort area (Figure 2). The gaging stations use satellite telemetry to transmit the data in "near" real-time (4-hour interval) to the USGS Office in Columbia. This network consists of four stations on the Beaufort River, and one station each on Brickyard, Albergottie, and Battery Creeks. Each station records water level (WL), water temperature (WT), specific conductance (SC), and DO concentration on a 15-minute interval. A precipitation gage is located at the Albergottie Creek gage. Three acoustic velocity meters were deployed in the spring of 2001 at the Brickyard Creek gage (station 2176585), the Beaufort River gage at Port Royal (station 2176611) and at the Battery Creek gage (station 2176635) to measure continuous (15-minute interval) tidal streamflow.

Tidal systems are highly dynamic and often daily data, rather than hourly data, are analyzed. The complex behaviors of the variables in a natural system result from interactions between multiple physical forces. The semi-diurnal tide is dominated by the lunar cycle which is greater 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. The filtering method of choice is frequency domain filtering. It is applied to a signal, or time series of data, after it has been converted into a frequency distribution by Fourier transform. This allows a signal component that lies within a window of frequencies (for example, the 12.4-hour tidal cycle lies between periods of 12.0 to 13.0 hours) to be excised, analyzed, and modeled independently of other components (Press and others, 1993). The filter for removing the high frequency tidal cycle is often referred as a “low-pass” filter. 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.

Two variables were computed from the field measurements of the physical parameters — tidal range (XWL) and dissolved oxygen deficit (DOD). Tidal dynamics are a dominant force for estuarine systems and the tidal range is a significant 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. Dissolved oxygen and WT are inversely related and highly correlated. Dissolved oxygen deficit is defined as the difference between the actual DO concentration and the saturated DO concentration. The computed variable, DOD, is derived using an algorithm that assumes a constant barometric pressure (USGS, 1981).

Water-Level and Streamflow Data

Tides enter the Beaufort River through the Coosaw River and Brickyard Creek in the north and the Broad River in the south. Generally, the physical properties measured at the gaging station (WL, WT, SC, and DO) fluctuate similarly. Figure 5 shows the gage heights at three stations on the Beaufort River and Brickyard Creek for August 2001. There is little change in the tidal amplitude and timing as the tidal wave propagates through the system. The maximum tidal range for the period shown in Figure 5 is 11.50 feet for station 2176640 and 11.22 feet for stations 2176603 and 2176585.

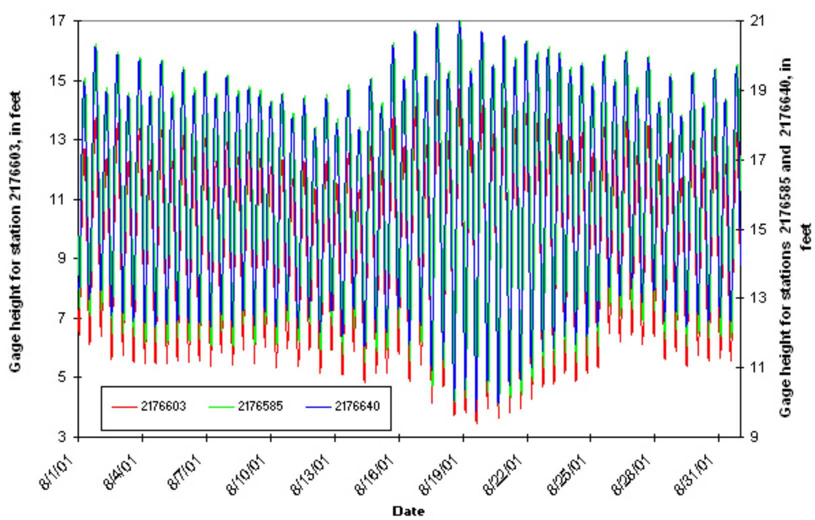

Figure 5. Beaufort River and Brickyard Creek gage heights for three stations for August 2001.

There is about an hour lag in the tide from the Beaufort River at Parris Island gage in the south (station 2176640) to the Brickyard Creek gage in the north (station 2176585). The 14-day semi-diurnal tidal cycle is also apparent in Figure 5. The neap tidal period, characterized by a relatively smaller amplitude in tidal range, occurs around August 12 and 26, and the spring tidal period, characterized by a larger amplitude in tidal range, occurs around August 6 and 20.

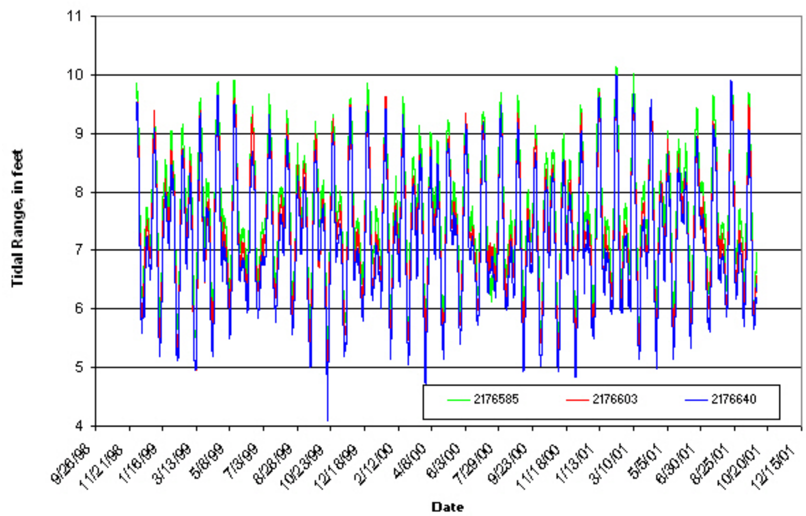

Figure 6. Beaufort River tidal ranges.

The tidal range for the three stations in Figure 6 are shown for the period January 1999 to October 2001, and clearly shows the longer term cyclic patterns in the tidal ranges. For example, a high spring tide range (greater than 9 feet) is followed by a low spring tide range (less than 9 feet). A similar pattern is apparent in the neap tides where a low neap tide range (less than 6 feet) is followed by a higher tidal range (greater than 6 feet). Also apparent are semi-annual cycles of minimum and maximum tidal ranges.

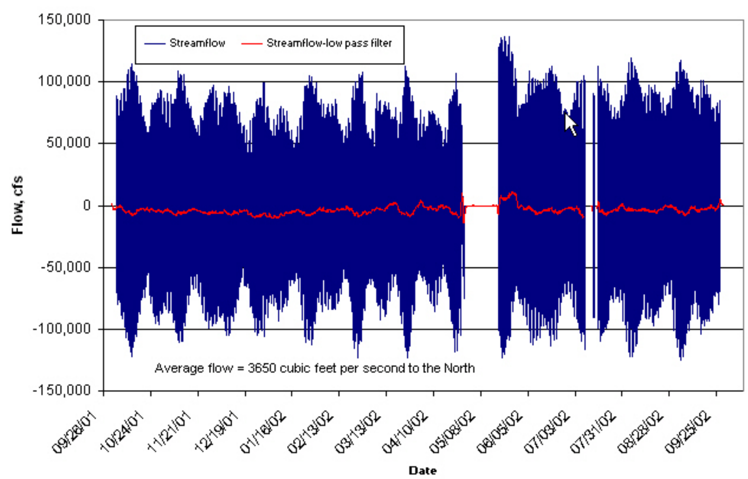

Figure 7. Beaufort River streamflows at station 2176611 for 2002 water year.

With a 9-foot tidal range and wide channel geometry, especially in the southern end of the system, the Beaufort River experiences large tidal streamflows of greater than 50,000 cubic feet per second (ft^3/s). Figure 7 shows hourly streamflows for the 2002 water year. The average positive streamflows (ebb flows or out-going tides) and negative streamflows (flood flows or in-coming tides) are 53,600 ft^3/s and 58,800 ft^3/s, respectively. The maximum flood and ebb flows are 125,000 ft^3/s and 136,000 ft^3/s, respectively. Filtering the streamflow data to remove the tidal variability, using methods described earlier, shows that the net streamflow is 3,650 ft^3/s to the north through Brickyard Creek (and other tidal creek connections) to the Coosaw River.Tidal-connections, such as the Beaufort River, usually experience a tidal node where the tidal waves entering the system from the two connections meet. For the Beaufort River, tides from St. Helena Sound travel up the Coosaw River and enter the northern end of the system through Brickyard Creek. From the south, tides from Port Royal Sound travel up the Broad River and into the Beaufort River. Figure 8 shows the tidal node in the velocity time series at the Brickyard Creek gage (station 2176585). Positive flow at the gage is to the north toward the Coosaw River (the sign convention is opposite of the station 2176611) and negative velocities are to the south towards Port Royal Sound. The figure shows that the flood tide (negative velocities) is significantly retarded as the tide from the south (Port Royal Sound) overpowers the tide from St. Helena Sound. The tidal node results in a distortion of the tidal velocities with ebb tides approximately double that of flood tides.

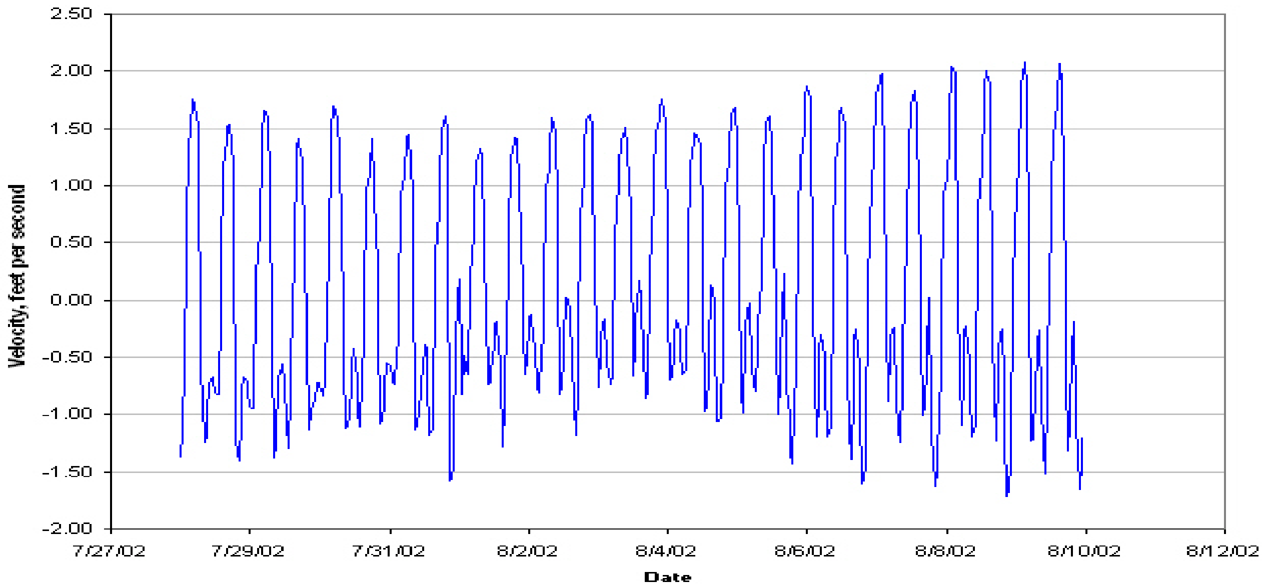

Figure 8. Stream velocities for Brickyard Creek (station 2176585) for July 15-22, 2001. Positive flow is to the Coosaw River to the north.

Precipitation and Water-Quality Data

The mean annual rainfall for the Beaufort area is approximately 50 inches per year for the period 1971 to 2000 (SCDNR, 2003) with the greatest monthly rainfall occurring in June, July, August and September. South Carolina experienced a severe drought from the period of the last El Niño event in the spring of 1998 until the increased rainfall during the late summer of 2002. During the continuous gaging period (1999-2001), three minimum and maximum monthly records were established. In June 1999, the Beaufort area experienced the wettest June on record (14.22 inches) despite the drought throughout the rest of the State. Monthly minimum precipitation records were set for August 1999 and October 2000 with monthly totals of 1.96 inches and 0.12 inches, respectively.

As a tidal connection between two sounds with little contributing drainage area to the system, the SC values of the system are similar to ocean values. Figure 9 shows the SC and 2-day average rainfall values for the period of December 1998 to September 2001, and the majority of values are greater than 45,000 microsiemens per centimeter (us/cm). Like the water-level data, the dynamic behavior of SC is similar at all the gages. The inland stations (Albergottie Creek, 2176587; Brickyard Creek, 2176585; and Beaufort River above Beaufort, 2176589) respond the most to input of freshwater from rainfall and show the greatest variability in SC.

The significant decrease in SC during July 1999 was due to the large rain event that the Beaufort area experienced on June 30, 1999. As noted above, June 1999 was the wettest June for the 30-year period of record and the majority of the rain for the month fell on June 30. Figure 9 shows the 2-day average rainfall and the SC response for the 7-month period from March to October 1999. It is interesting to note that the Beaufort River took approximately 75 days for SC values to return to levels prior to the rainfall event. This extended recovery indicates the long residence time and limited flushing of the system.

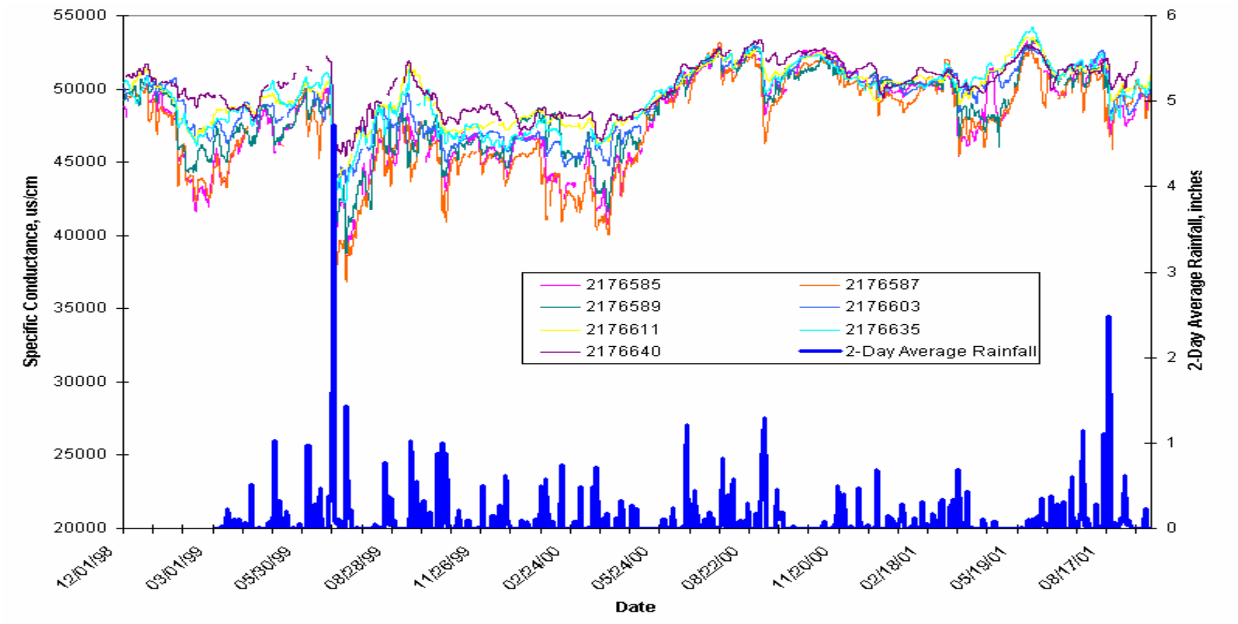

Figure 9. Specific conductance values and rainfall for the Beaufort River and two tributaries for December 1998 to September 2001.

The dynamic behavior of WT between all of the gages is similar. The inland gages experience the highest and lowest temperatures in the summer and winter, respectively. Inland gages generally are in smaller channels and not as buffered by the thermal mass as gages in reaches with larger channel geometry that are closer to Port Royal Sound and the ocean. The temperatures in the rivers reach 20 degrees Celsius (^{o}C) in early April and 30 ^{o}C by July, and do not fall below 20 ^{o}C until late October.

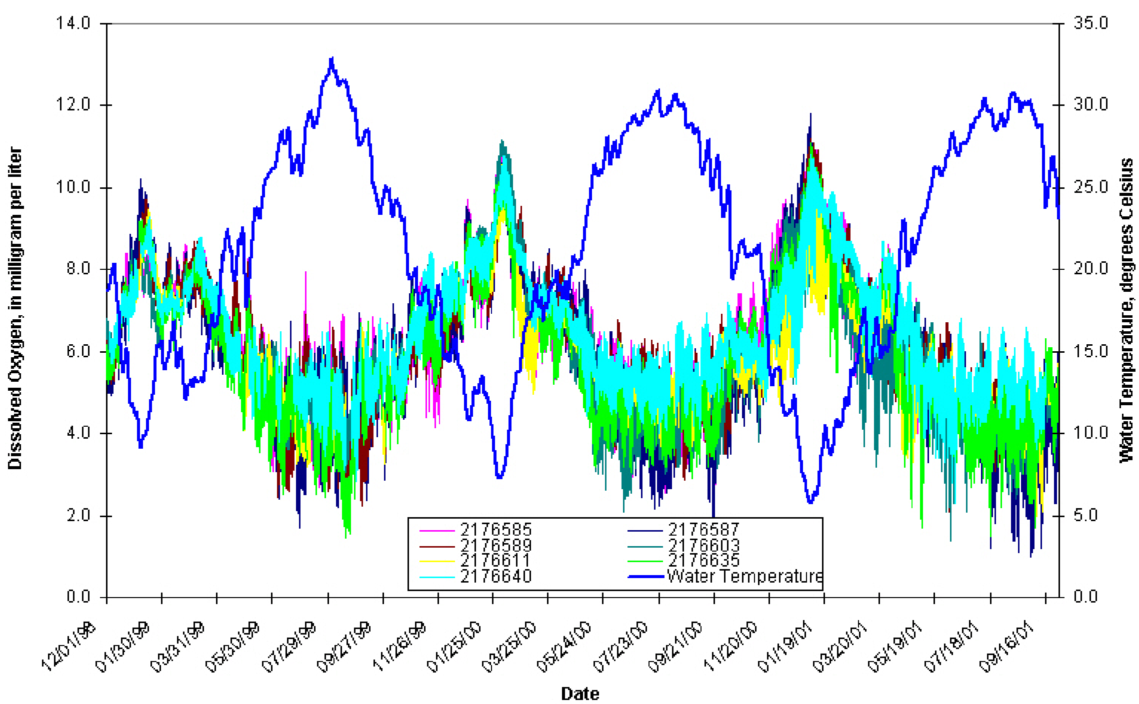

Figure 10. Time series of DO concentration for Beaufort River and two tributaries and water temperature at station 2176611 for December 1998 to September 2001.

The Beaufort River is on the SCDHEC 303(d) list of impaired waters for low DO concentrations because of natural conditions. The State water-quality standard is a daily mean of 5.0 milligrams per liter (mg/L) or a daily minimum of 4.0 mg/L. Figure10 shows the time series of hourly DO concentrations for the seven stations on the Beaufort River and two tributaries. During the summer months, the minimum DO is less than 4.0 mg/L for extended periods, and is generally higher in the southern segments of the river and lower for the upper reaches of the system. WT and DO are inversely related and highly correlated (Figure 10). As WT increases greater than 25 ^{o}C, DO concentrations decrease to the State water-quality standard of 5 mg/L or less.

Dissolved oxygen deficit (DOD) is a measure of the difference between the actual DO concentration and DO concentration for saturated conditions, and effectively "normalizes" DO to WT. Lower values of DOD indicate water of higher percent saturation, whereas higher DOD values indicate water of lower percent saturation or greater impairment.

For the seven gages on the Beaufort River and its tributaries, cumulative percentages of DOD were computed from the time series (Figure 11). The higher the DOD, the greater the increased impairment of DO from point or non-point sources. The figure shows that the Beaufort River at Parris Island gage (station 2176640) has substantially lower DOD than the other six stations. The net streamflow of the system is to the north, so the Parris Island gage is less affected by the point- and non-point source loading of oxygen-consuming constituents into the Beaufort River and more affected by the higher water quality of Port Royal Sound. The Brickyard Creek gage (station 2176585) has the next highest DOD. Although the net movement is to the north, the higher quality water from the Coosaw River and St. Helena Sound also affects the Brickyard Creek gage. The stations with the highest values of DOD are either in the tributary creeks (Battery Creek, station 2176635 or Albergottie Creek, station 2176587) or in the upper segment of the Beaufort River (Beaufort River above Beaufort, station 2176589).

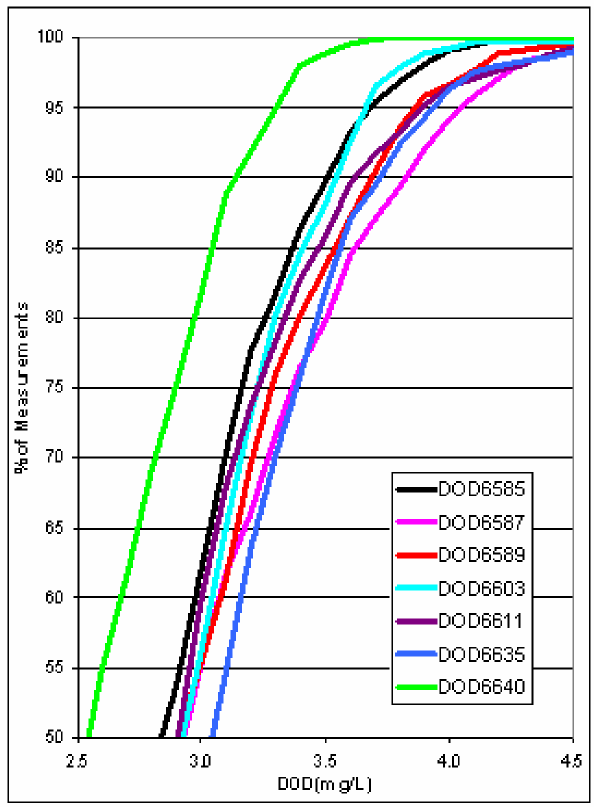

Figure 11. Cumulative percent of DOD for the gages on the Beaufort River and its tributaries for January 1999 to September 2001.

APPROACH

Simulating dissolved-oxygen concentration for estuarine systems is typically done using dynamic deterministic models that incorporate the physics of coastal hydrodynamics and the chemistry that describes the eutrophication process and its effects on dissolved oxygen. These one-, two-, or three-dimensional models often are expensive and time consuming to apply to complex coastal systems with satisfactory results. Although deterministic models have been the state of the practice for regulatory evaluations of point-source and non-point source impacts, developments in the field of advanced statistics, machine learning, and data mining offer opportunities to develop models that are more accurate. Conrads and Roehl (1999) compared the application of a deterministic model and an Artificial Neural Network model (ANN) to simulate DO on the tidally affected Cooper River in South Carolina. They found that the ANN models offer some significant advantages, including faster development time, utilization of larger amounts of data, incorporating optimization routines, and model dissemination in spreadsheet applications. With the real-time gaging network on the Beaufort River and the large database of hydrologic and water-quality data, BJWSA and SCDHEC realized an opportunity to develop an empirical model using data mining techniques, including ANNs.

The emerging field of data mining addresses the issue of extracting information from large databases. It is comprised of several technologies that include signal processing, advanced statistics, multi-dimensional visualization, chaos theory and machine learning. Machine learning is a field of Artificial Intelligence (AI) in which computer programs are developed that automatically learn cause-effect relationships from example cases and data. For numerical data, commonly used methods include ANN, genetic algorithms, multivariate adaptive regression splines, and partial and ordinary least squares.

Data mining can solve complex problems that are unsolvable by any other means. Weiss and Indurkhya (1998) define data mining as "…the search for valuable information in large volumes of data. It is a cooperative effort of humans and computers." A number of previous studies by the authors and others have used data mining to predict hydrodynamic and water-quality behaviors in the Beaufort, Cooper, and Savannah River estuaries (Roehl and Conrads, 1999; Conrads and Roehl, 1999; Roehl and others, 2000; Conrads and others, 2002a; 2002b) and stream temperatures in western Oregon (Risley and others, 2002). These studies have demonstrated the performance of data mining to predict WL, WT, DO, and SC, and for assessing the impacts of reservoir releases and point and non-point sources on receiving streams.

The ultimate goal of an effective model is to simulate the impact of the point- and non-point sources on dissolved oxygen. An effective water-quality model is able to link sources to impairments. If the goal was just to simulate DO, that can be done quite accurately with only temperature due to the strong inverse relationship between temperature and DO. The real necessity in a regulatory model is to be able to determine how much of the variability in DO is attributable to a point-source discharge. The variability of DO in the Beaufort River is a result of many factors including the quality of the water from Port Royal Sound and the Coosaw River, the loading of oxygen-consuming constituents from the tidal marshes and other non-point sources, effluent from four permitted point sources, and physical characteristics of streamflow, tidal range, salinity, and temperature.

The approach taken, which uses all available point-source and rainfall measurements in contexts of extraordinarily long time series of hydrodynamic and water-quality measurements at individual gages, provides an accounting of point-source and rainfall impacts. The modeling approach uses correlation functions that were synthesized directly from data to predict how the change in DOD at each gage location is affected by rainfall and each point-source discharge over time. The general idea is that BOD or NH_3 pulses at a discharge point will some time later modulate the change in DOD at some or all of the gaging locations. Prevailing conditions of hydrodynamic transport and de-oxygenation and nitrification reaction kinetics affect the timing and extent of the daily change in DOD.

Signal Decomposition and Correlation Analysis

The behaviors of the variables of a natural system result from interactions between multiple physical forces. For example, the WT at a fixed location is subject to annual and diurnal (24-hour) ambient temperature cycling, and also by tidally forced mixing of warmer and cooler waters. For the application of the ANN model to the Beaufort River, data mining methods are applied to maximize the information content in raw data while diminishing the influence of poor or missing measurements. Methods include digital filtering using fast Fourier Transforms, time derivatives, time delays, and running averages. Signals, or time series, manifest three types of behavior: periodic, chaotic, or noise. Examples of periodic behavior are the diurnal light and temperature patterns caused by the rising and setting sun or tidal water levels due to orbital mechanics. Noise refers to random components usually attributed to measurement error. Chaotic behavior is neither periodic nor noise, and always has a physical cause. Weather provides an example of chaotic behavior.

Signal decomposition involves splitting a signal into sub-signals, called “components,” that are independently attributable to different physical forces. Digital filtering can also 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. To analyze and model these time series, the periodic and chaotic components of the signals need to be separated. Filtered signals are comprised of the chaotic and noise components of the original time series.

Time derivatives are a common analytical method used in the sciences to analyze the dynamics of a system. Time derivatives can also be computed for the measured (and computed) data on the Beaufort River to further understand the dynamics of the system. In Figure 12, the 1-day derivative of the low-pass filtered DO time series for a 90-day period is plotted with the original time series and the low-pass filtered data. The 1-day derivatives show the rate of change of the chaotic component of the DO time series. For the 90-day period, the daily change is as high as 0.6 mg/L.

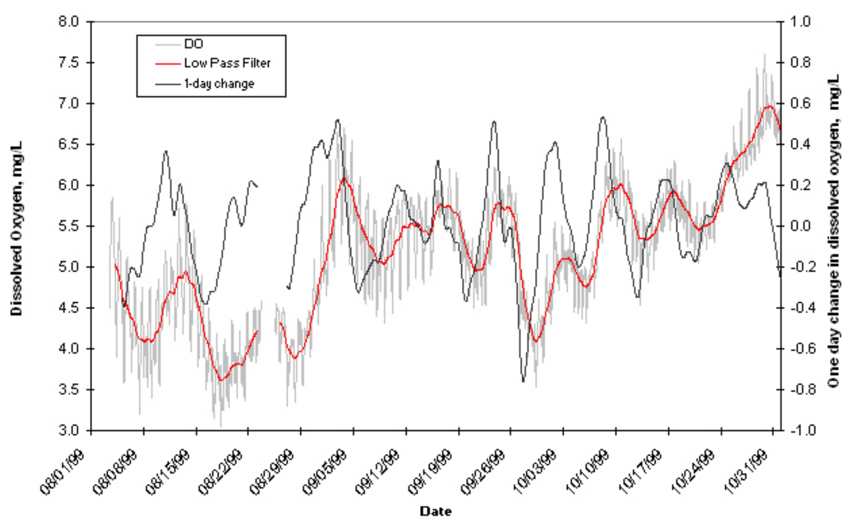

Figure 12. Plots showing time series of hourly and measured values, filtered values, and 1-day time derivatives of the low-pass filtered values for station 2176603. Note y-axis for 1-day derivative time series on the right side of the plot.

Often there are time delays between when an event is measured and the time that the response is observed in a system. Modeling a system is more complicated when two events of interest, a cause and an effect, do not occur simultaneously. The time between cause and effect is called the “time delay” or “delay.” Each input variable of a model has its own delay. Determining the correct time delays for pulses and system response is critical to accurately simulating a dynamic system.

For the Beaufort River, pulses of effluent from the three WRFs are entering the system at different loading rates. Time delays between when the load enters the system and the river response of the DO deficit were determined for each WRF and each gage.

Averages and running averages are commonly used to remove the variability of measurements. Rather than taking one measurement, multiple measurements are made and an average value is used. Precipitation data often exhibit large temporal and spatial variability. For the development of the Beaufort River ANN model, the average rainfall for two gages and running averages for 2, 4, 6, 8, and 10-days were determined.

The relations between the many variables and their various components are ascertained through correlation analyses to provide deeper understanding of system dynamics. The computer systematically correlates factors that most influence parameters of interest (e.g., water quality) to candidate combinations of controlled and uncontrolled variables (e.g., discharges and ambient temperatures). Correlation methods based on statistics and ANN are applied in combination. Promising results found by the computer are validated by comparing them to known patterns of behavior.

Artificial Neural Networks

An artificial neural network model (ANN) is a flexible mathematical structure capable of describing complex nonlinear relations between input and output data sets. The architecture of ANN models is loosely based on the biological nervous system (Hinton, 1992). Although there are numerous types of ANNs, the most commonly used type of ANN is the multi-layer perceptron (MLP) (Rosenblatt, 1958). As shown in Figure 13, MLP ANNs are constructed from layers of interconnected processing elements called neurons, each executing a simple "transfer function." All input layer neurons are connected to every hidden layer neuron and every hidden layer neuron is connected to every output neuron.
There can be multiple hidden layers, but a single layer is sufficient for most problems.

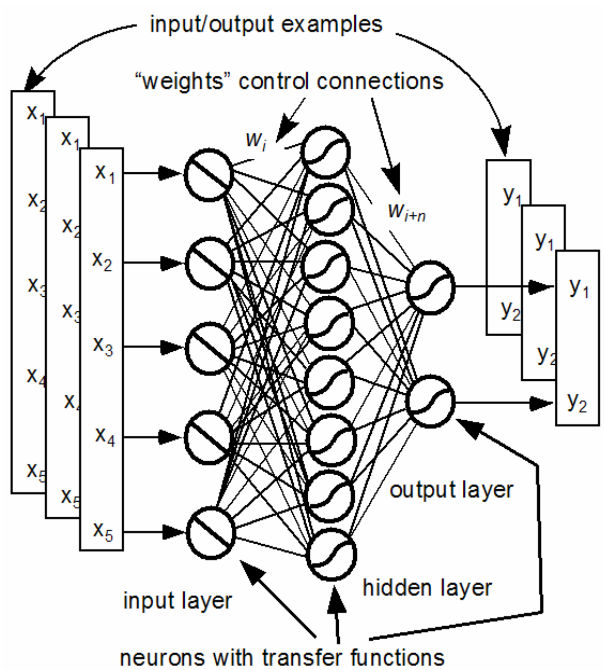

Figure 13: Multi-layer perceptron artificial neural network architecture.

Typically, linear transfer functions are used to simply scale input values to fall within the range that corresponds to the most linear part of the s-shaped sigmoid transfer functions used in the hidden and output layers. Each connection has a "weight" w_i associated with it, which scales the output received by a neuron from a neuron in an antecedent layer. The output of a neuron is a simple combination of the values it receives through its input connections and their weights, and the neuron's transfer function.

An ANN is "trained" by iteratively adjusting its weights to minimize the error by which it maps inputs to outputs for a data set comprised of "input/output vector pairs". Prediction accuracy during and after training can be measured by a number of metrics, including coefficient of determination (R^2) and root mean square error (RMSE). An algorithm that is commonly used to train MLP ANNs is the back error propagation (BEP) training algorithm (Rumelhart and others, 1986).

Experimentation with a number of ANN architectural and training parameters is a normal part of the modeling process. For correlation analysis or predictive modeling applications, a number of candidate ANNs are trained and evaluated for both their statistical accuracy and their representation of process physics. Interactions between combinations of variables also are considered. Finally, a satisfactory model can be exported for end-user deployment. In general, a high-quality predictive model can be obtained when:

- The data are well distributed throughout the space of interest.
- The input variables selected by the modeler share a lot of “mutual information” about the output variables.
- The form “prescribed” or “synthesized” for the model used to “map” (correlate) input variables to output variables is a good one. Techniques such as ordinary least-squares (OLS) and physics-based finite-difference models prescribe the functional form of the model’s fit of the calibration data. Machine learning techniques like ANNs synthesize a best fit to the data.

Input/Output Mapping and Problem Representation

Water-quality models are often used to simulate streams to assess the amount of point-source and non-point source loading of oxygen-consuming constituents that a receiving stream can assimilate. The domain of the model is defined to include the river segments of impaired waters and segments upstream and downstream from the impairment to establish boundary conditions that clearly define inputs to the model domain that will not bias the simulations for the area of concern. Estuarine systems, with tidally affected water levels, reversing streamflows, and large tidal excursions, present a unique set of challenges for establishing good boundary conditions.

Defining boundary conditions for the Beaufort River is particularly challenging. The entire length of the river is affected by large tidal exchanges, effluent discharges of the WRFs, semi-diurnal exchanges with the extensive tidal marshes, and loading of oxygen-consuming constituents during rainfall events. The Beaufort River is not a closed system between the Broad and Coosaw Rivers, but has substantial exchanges with many tidal creeks. Using a traditional deterministic water-quality model, input boundary conditions for the Beaufort River would be gaging stations in Port Royal and St. Helena Sounds with a model domain that included the Broad and Coosaw Rivers, and the many tidal creeks in between. Obviously, for addressing the impaired waters of the Beaufort River and linking the affect of effluent discharges on dissolved oxygen, gaging stations in the two Sounds and a model of such a large domain would increase the scope of the study to make it economically and logistically impractical.

To address the problem of boundary conditions, a different approach was taken to simulate the effect of point-source and non-point source loading to the system. Two problems or concerns had to be addressed. The first concern is that all the gages in the Beaufort River network are affected by point- and non-point source loadings, including the Parris Island gage (station 2176640), which is closest to open water. The second concern is the various sampling frequency of the input data (15-minute data for the continuous monitoring data and weekly sampling data for point-source flow and concentration data).

To further challenge the problem representation, the point-source data are collected weekly and the three dischargers collect their data on different days of the week. Empirical models are built on the measured data, and it is critical that the data used in the correlation analysis and construction of the model is built on measured, not interpolated data. The consequence of the weekly point-source data can be seen in the reduction in the size of the available data set of measured input conditions from the facilities and instream physical conditions. The instream gaging network is recording data every 15 minutes. From that data, high fidelity daily "averages" can be computed using Fast Fourier Analysis, as described previously. Of the daily time series from the gaging network of 365 data points a year, the inputs from the point sources are known only for 52 days.

There are many factors affecting the dynamics of the DO including tidal exchange, diurnal cycling, point-source loading, tidal marsh exchange, rainfall impacts, and benthic demands. For the factors where there are measured data (WL, SC, WT, point-source BOD and NH_3 loads, and precipitation), an accounting of the contribution to the variability of DO can be made at the location of each gage. Knowing the point-source load from a facility for a particular day, the pulse of oxygen-consuming constituents can be correlated to the change in DO at a gage. A similar analysis is also done for analyzing the effect of rainfall on DO concentration. For example, at any gage in the network, the variability of the DO due to the instream physical properties, point-source loads, and rainfall effects can be determined independent of any boundary conditions describing input data to the system. By modeling the DO variability at each gage, the problem representation can address the issues of model boundary and varying sampling frequency of the data.

Decorrelation of Variables

Often, explanatory variables share information about the behavior of a response variable. It is difficult, if not impossible, to understand the individual effects of these variables (sometime known as confounded or correlated variables), on a response variable. Empirical models have no notion of process physics, nor the nature of interrelations between input variables. To be able to clearly analyze the effects of confounded variables, the unique informational content of each variable must be determined by "de-correlating" the confounded variables.

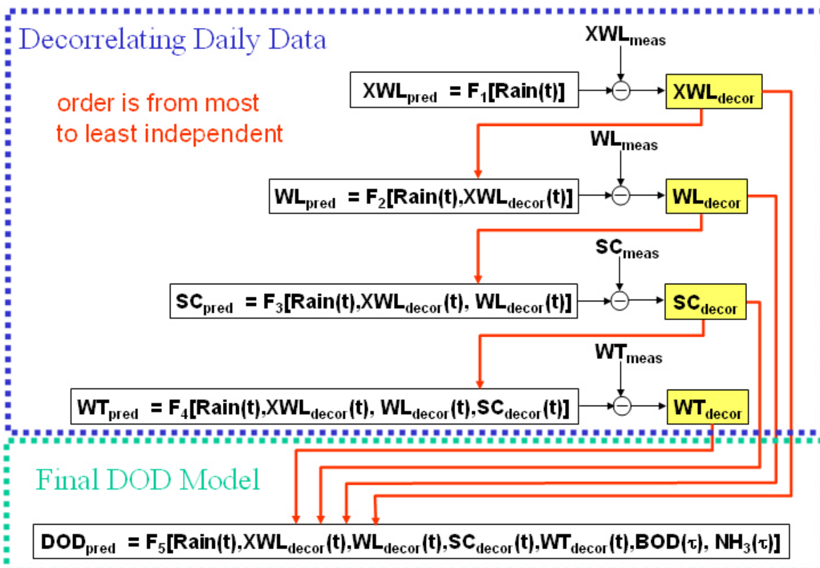

Figure 14. ANN sub-model execution sequence for decorrelating variables. The model has seven instances of the decorrelation sub-model sequence shown in the blue dotted box at top, one for each gage location. There are seven DOD sub-models for computing Rainfall impact on DOD, also one per gage. There are 83 DOD sub-models for computing the impacts of BOD and NH_3 from the Water Reclamation Facilities at different time delays.

The physical properties of WL, SC, WT, and DO measured at the seven gaging stations in the Beaufort River all exhibit the 14-day and 28-day lunar and annual solar periodicity, and therefore, are significantly cross-correlated. Their use in the construction of empirical correlation functions, such as the ANN models used here, requires that the variables are systematically decorrelated in order of their relative independence from each other. Decorrelation is accomplished by generating an empirical correlation function and computing its residual error by subtracting the function's predicted values from actual measurements (Figure 14). The blue dotted box shows that tidal range (XWL_{decor}) is computed from correlation function F_1 having Rain inputs. In turn, WL (WL_{decor}) is computed from correlation function F_2 having Rain and XWL_{decor} inputs, and so on for

SC and WT. For a given gage location, DOD then can be modeled by a function F_5 using Rain and decorrelated XWL, WL, SC, WT, BOD, and NH_3.

Estimating Point-source Discharge Impacts on a Single Time Series

The following discussion (from Conrads and others, 2002b) explains how the point-source discharge influences on the DO time series at a gage are determined. The example uses the gage at Beaufort River at Port Royal (station 2176611), which is 500 feet south of the Southside WRF (see Figure 2). For this example, the data were comprised of 15-minute measurements for WL, SC, WT, and DO at station 2176611 and computed variables of DOD and XWL. Rainfall data were collected from the Albergottie Creek station (station 2176587) and one of the WRFs near the Beaufort River at the Port Royal gage. Two years of weekly data of flow rates, BOD, and NH_3 were obtained from the two WRFs (Southside and Shell Point) that discharge to the Beaufort River to the north of station 2176611. The outfalls for the two facilities are located beside one another. The effluent data for these two facilities were combined and treated as one point source in the analysis. The effect of the oxygen-consuming constituents in the WRF effluent on DO transpires on a time scale of several days. This effect can be difficult to discern when coupled to high frequency forces such as diurnal and semi-diurnal tidal water level, tidal flow, and ambient temperature. Therefore, the time series were filtered and systematically decorrelated.

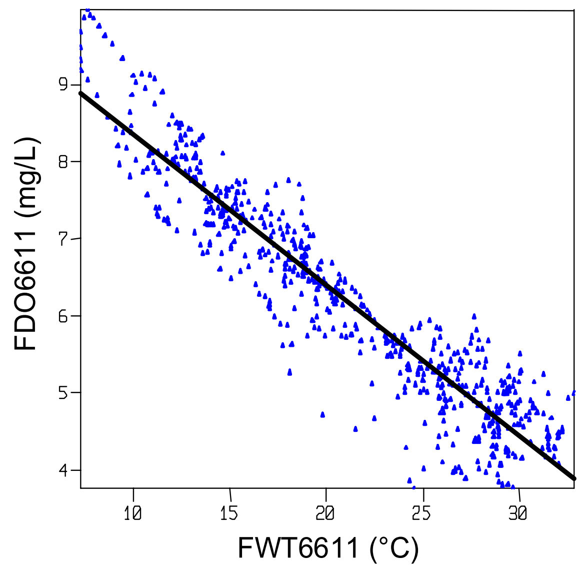

Figure 15. Scatter Plot of filtered dissolved oxygen (FDO) and filtered water temperature (FWT) and least-squares regression line (R^2=0.88).

Typically, DO and WT are inversely related, and the majority of the variability in DO is due to WT. Linear regression produces a coefficient of determination (R^2) of 0.88, indicating that approximately 88 percent of the variability of DO is explained by WT alone (Figure 15), and the remaining 12 percent is caused by other factors. WT has two effects. The first is that dissolved-oxygen saturation decreases with WT, and the second is that microbial activity that consumes DO also increases with WT (given sufficient DO and nutrients). The use of DOD rather than DO as the response variable of interest, effectively normalizes the DO signal with respect to temperature and emphasizes the microbial activity effect in the DO signal.

The goal of the Beaufort River assimilative capacity analysis model is to quantify the effect that point-source discharges of oxygen-consuming constituents have on instream DO. Due to the limited number of the effluent concentrations data points as compared to the gaging data (weekly values as compared to 15-minute data), a subset of the dataset was excised and included only the digitally filtered data of DO, WT, WL, and XWL for the day of the effluent sampling. In addition, the 1-day derivatives of the DO and WT were computed and included in the dataset (1-day derivative of the filtered variables are denoted by an E prefix, for example, EDO or ESC). The sensitivity of the response variables, DO and DOD, to the explanatory variables of BOD_5, NH_3, rainfall, and tidal range were determined using ANN models. The type of ANNs used were the multi-layer perceptrons described by Hinton (1992) that were trained using the back propagation and conjugate gradient algorithms.

Visual inspection of the BOD_5 loading from the WRFs and the daily change in DO concentration at station 2176611 (Figure 16) shows a relation between the two variables (note that the EDO scale has been inverted so decreases in daily DO rise on the scale). The number of coincident peaks in the daily change in DO and BOD5 loading (for example observations 6, 31, 35, 39, and 58) indicate that the BOD_5 loading may account for a significant part of the remaining 12 percent of the variability in DO. An ANN model of the EDOD, having BOD_5, rainfall, and decorrelated filtered WL, XWL, SC and WT as inputs, was generated to provide a more comprehensive assessment of the relations between the BOD_5 and the DO. Figure 17 shows that the ANN fits most of the higher peaks in the EDOD. The $R^2_{ANN} = 0.57$, indicating that approximately 57% of the variability in the EDOD is accounted for by variability in the input variables. The impact of the NH_3 discharge can be similarly evaluated. Figure 18 shows that predictions made by an ANN model of the EDOD, having NH_3, rainfall, and decorrelated filtered WL, XWL, SC and WT as inputs, generally runs through the middle of the actual data. The $R^2_{ANN} = 0.31$, indicating that approximately 31% of the variability in the EDOD is accounted for by variability in the input variables. It should be noted that the NH_3 input was delayed relative to the EDOD by 3 days, versus 1 day for the BOD_5, in the first model described above. The delays were chosen by testing different delay configurations and selecting those that produced the highest R^2 ANNs.

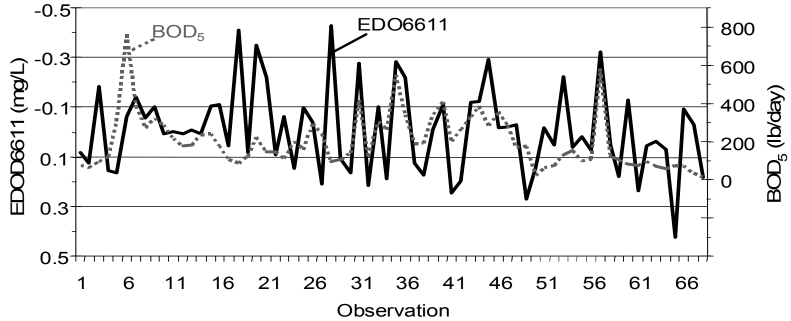

Figure 16. One-day change in DO deficit (EDOD6611) and BOD_5 (at a 1 day time delay) at station 2172211. Linear R^2 = 0.13.

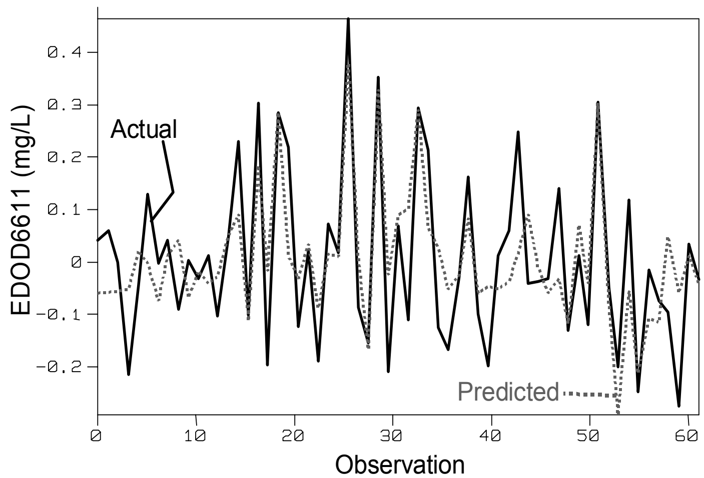

Figure 17. Measured and predicted EDOD ANN used BOD_5 as an input at a time delay of 1 day. R^2_{ANN} = 0.57.

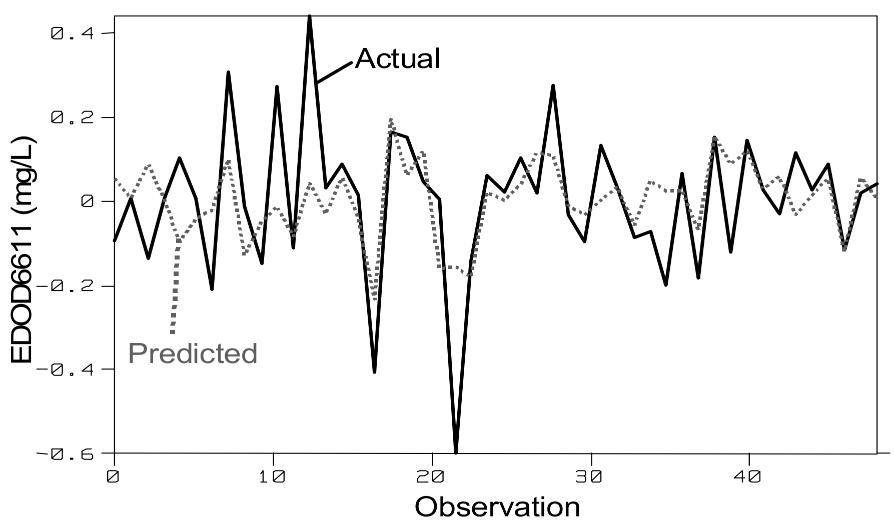

Figure 18. Measured and predicted EDOD ANN used NH_3 as an input at a time delay of 3 days. R^2_{ANN} = 0.31.

Construction of the Beaufort River Model

The Beaufort River assimilative capacity model is composed of many sub-models. The impacts of BOD and NH_3 are computed by sub-models that use decorrelated XWL, WL, SC, and WT and their 2-day time derivatives as inputs. Also included are BOD and NH_3 inputs at appropriate time delays, τ, for each WRF. The output of each sub-model is a prediction of the 1-day time derivative of DOD due to point-source discharges and rainfall. Each sub-model is a separate ANN file. The naming convention used for sub-models helps explain what they do and how their predictions are combined. The convention is:

WRF+ τ + load + gage; WRF = MCAS, PI, SS
τ = time delay of 0 (time of input),1,2,3,4,5,6,7,or 8 days
load = bod, nh3
gage = last 2 digits of station numbers

For example, the sub-models used to predict on-day change of DOD at station 2176585 are:

1. as0bod85, as1bod85 - predict impact of BOD from MCAS at gage 6585 1, 2 days after discharge.
2. as3nh385, as4nh385, as5nh385, as6nh385, as7nh385 as1bod85 - predicts impact of NH_3 from MCAS at gage 6585 4,5,6,7,8 days after discharge.
3. pi0bod85 as1bod85 - predict impact of BOD from PI at gage 6585 1, 2 days after discharge.
4. pi6nh385, pi7nh385, pi8nh385 - predict impact of NH_3 from PI at gage 6585 7,8,9 days after discharge.
5. ss0bod85, ss1bod85 - predict impact of BOD from SS at gage 6585 1, 2 days after discharge.
6. ss6nh385, ss7nh385 - predict impact of BOD from MCAS at gage 6585 7,8 days after discharge.
7. pfdoa6585 - prediction of rainfall impact based on 2-, 4-, 6-, 8-day rainfall averages.

The total predicted impact from all point and non-point sources is the sum of the averages of sub-models 1-6 above, plus prediction 7. The averaging is a convention established to make an accommodation for uncertainty in the development of a very complex model of a very complex system from relatively small numbers of point-source measurements.

Sub-models were constructed by a trial and error procedure. A single input for a BOD or NH_3 load from a WRF at one time delay, τ was added to a fixed input combination of decorrelated XWL, WL, SC, and WT and their 2-day time derivatives for the gage being modeled. The sub-model then was carefully trained such that the sensitivity of the output d/dt DOD (the time derivative of the DO deficit) was largely linear and positive with respect to the load variable. Sub-models having input configurations that met the sensitivity requirement were then subjected to additional criteria that evaluated the τ's at which the positive sensitivity was observed. Sub-models meeting all criteria were then included in the final model. Some observations:

- BOD was observed to have an impact only 1-2 days after discharge, in most cases.
- NH_3 impacts appeared 3-9 days after discharge, depending on travel distances from discharge points to gages.
- For some WRF gage combinations, positive NH_3 sensitivities were not all at consecutive τ's. This was especially true for SS. Because there were 2 ½ times as many data points for PI as for SS, a convention was established for the model to use τ's as close as possible to those determined for the PI sub-models of NH_3.

- Noting the small amount of discharge data, results indicate that each BOD and each NH_3 load is unique in the way the natural system responds to its discharge.

In the Beaufort River assimilative capacity model, separate sub-models are constructed for each combination of gage location, discharge type (BOD or NH_3), and relative time delay. For example, the observed onset, peak, and ebbing of the impact that the NH_3 from SS has on the DOD at station 2176635 on Battery Creek is most pronounced 4, 5, 6, and 7 days after a major discharge. Therefore, four separate sub-models were constructed, one for each time delay of NH_3 load. The need for so many sub-models arises from the very spotty and discontinuous nature of the WRF discharge measurements. BOD and NH_3 values have been measured at most once per week at MCAS and SS, and at most twice per week at PI.

Training of Artificial Neural Network Models

For a behaviorally complex system, with only 100-200 non-concurrent and non-consecutive data points available for each point source, it was deemed too risky to set aside data for independent testing of ANN performance. To do so would prevent the ANNs from learning from data representing unique and possibly important behavioral states. In applications where there are sufficient data, it is customary to set aside "test" data to provide an independent evaluation of model performance. There are many strategies for partitioning data into training and tests sets, but the most common is by random selection of a specified percentage of the total population of measurements. Randomly selected test data, usually 20-30% of the total, were used for all of the decorrelation and rainfall impact prediction sub-models, but not for the BOD and NH_3 impact sub-models because of the sparseness of point-load data.

To mitigate the extrapolation and sparseness issues, the sub-models were conservatively trained using a method called "Stop Training" to both fit the data and extrapolate in a minimally non-linear, and therefore predictable, fashion. Stop Training simply means stopping the training process before the ANN has fit the data to the maximum extent possible. Architectural and training parameters allow the modeler to control the geometric complexity of the surface that the ANN fits to the data. Sparse or noisy data are prone to over-fitting if surface fits are made overly complex. The data mining software (now iQuest™) used for this application writes R^2 and RMSE to the graphical user interface (GUI) during training, and an inflection in the rate of change in these parameters indicates a transition from a generally linear, multivariate surface fit to an progressively non-linear fit. This inflection point was used to trigger Stop Training.

Spreadsheet Application

The 118 ANN sub-models that comprise the Beaufort River assimilative capacity model are incorporated in an Excel/VBA (Visual Basic for Applications) program that integrates a large historical database, streaming graphics, and a graphical user-interface (GUI). This approach provided a number of features and benefits that are new to estuary modeling applications, including:

- Excel has a large number of built-in functions, controls, graphics, systems integration, and programming features. Their use shortened development time and costs.
- The application's sophisticated capabilities are easy to use through a point and click GUI and supporting graphics (Figure 19). No typing is required to input data or control the model's operation.
- Incorporation of the historical database allows the user to run long-term simulations that are permutations of the historical record. The database is comprised of the USGS gaging data, BOD and

NH_3 data from the MCAS, PI, and SS WRFs, and rainfall data, which are an average of measurements taken at MCAS and SS.

- The model has an integrated optimizer that automatically computes the maximum allowed discharges for each simulation time step. This feature reduces the number of simulations needed to evaluate discharge scenarios by over 90%.

Perhaps the greatest benefit of the application development approach is that it produced a program that can be readily distributed and understood by a wide range of end-users who already know Excel. A complete explanation of the program's features and operation are provided in a companion Beaufort Model User Guide (Jordan, Jones & Goulding, 2003a).

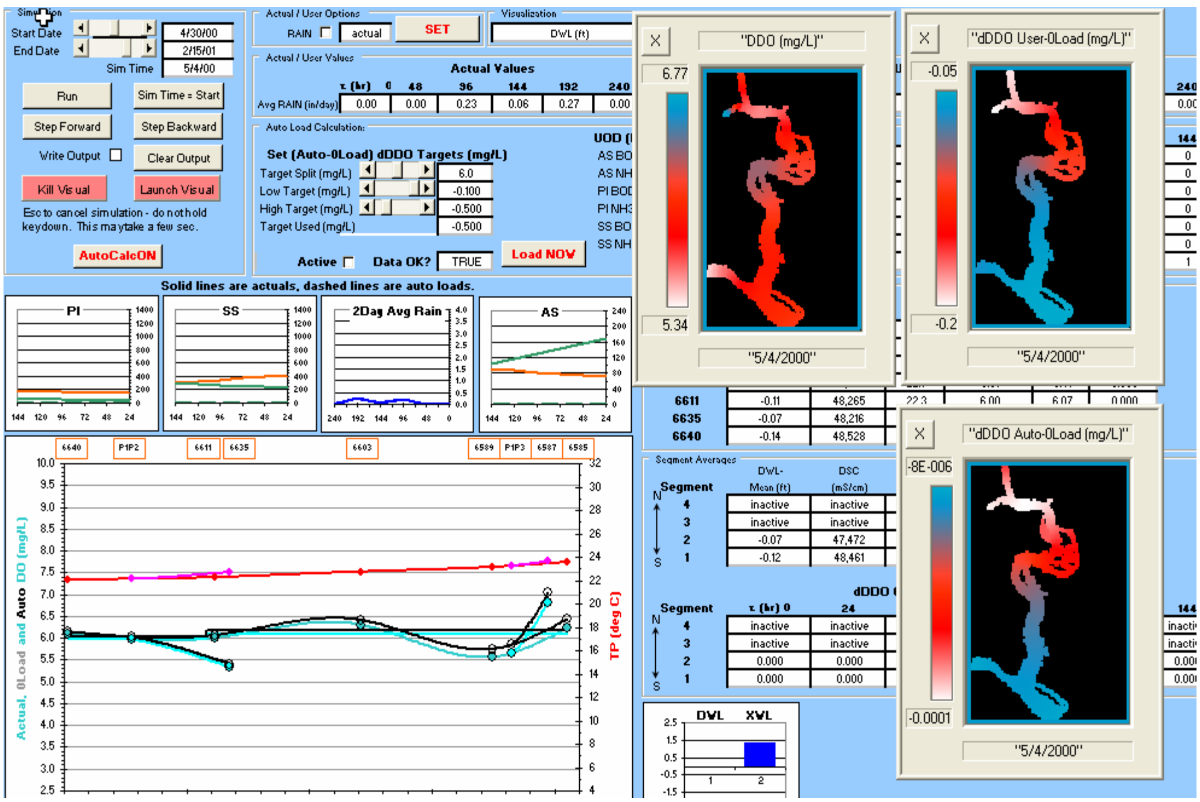

Figure 19. Elements of a GUI worksheet for the Beaufort River assimilative capacity model. Note the extensive decision support graphics and user controls in the form of check boxes, buttons, and scroll bars. The three panels at right show plan- view color-gradient renderings of DDO- and DO-related calculated variables.

Statistical Measures of Prediction Accuracy

Model accuracy is often reported in terms of R^2 and is commonly interpreted as the "goodness of the fit" of a model. A second interpretation is one of answering the question, "How much information does one variable or a group of variables have about the behavior of another variable?" In the first context, an $R^2 = 0.3$ might be disappointing, while in the latter it is merely an accounting of how much information is shared by the variables being used. While the developers believe that the Beaufort River assimilative capacity model is unusually accurate relative to 1D, 2D, and 3D finite-difference models developed for comparably complex estuaries, its predictions are knowingly made with missing variables, most notably those that would provide "pristine" boundary condition information.

Despite the non-continuous, non-concurrent BOD and NH_3 sampling, the model's special architecture made it possible to employ all of the available data without introducing error through the use of interpolated values. The character of the load data makes it impossible to use actual measurements to assess model accuracy. There is not a single time stamp in the historical record that contains all the load data needed to make a model prediction at any gage. For example, computing a prediction for station 2176640, which has only six NH_3 sub-models, requires four consecutive NH_3 measurements from SS and two from PI, and requires that the SS and PI measurements at $\tau = 3$ and 4 days be concurrent. But SS is sampled only once per week, with no coordination with PI.

A "rough estimate" of model accuracy was made by evaluating two data sets. Case 1 included interpolated load data (for up to 7 days), even though it was known that interpolation would heavily mask the high variability in the "spikey" WRF discharges. Case 2 used the 10 days in the historical record having concurrent BOD and NH_3 from all three WRFs, however, load values for preceding τs were still mostly interpolated. Table 1 shows results for the two cases that together give insight about the model's accuracy. In Case 1, the R^2 and RMSE range from 0.05 to 0.09 and 0.44 to 0.16 mg/L respectively, and increase to 0.23 to 0.75 and 0.42 to 0.13 mg/L, respectively for Case 2. The number of records N used to compute the statistics is roughly 1/3 of the 1035 total for Case 1, and 5 to 8 of the available 10 Case 2 records. It is likely that the model is more accurate than the Case 1 results, with $R^2 \approx$ 0.3 or better, and RMSE $\approx$ 0.3 mg/L or better relative to a gage average one-day change in dissolved-oxygen deficit range = 1.5 mg/L.

Table 1. Estimated model accuracy statistics for cases 1 (left) and 2 (right). Arrows match like statistical measures for the two cases. RMSE in units of mg/L.

[DOD', 1-day change in dissolved-oxygen deficit; N, number of records in data set; R, correlation coefficient; R^2, coefficient of determination; RMSE, root mean square error]

	Predict DOD' w Interpolated Discharges				Predict DOD' w Concurrent Discharges			
GAGE	N	R	R^2	RMSE	N	R	R^2	RMSE
6585	269	0.30	0.09	0.35	5	0.63	0.40	0.25
6587	357	0.27	0.07	0.44	5	0.61	0.38	0.42
6589	335	0.23	0.05	0.38	8	0.87	0.75	0.27
6603	344	0.24	0.06	0.27	7	0.62	0.39	0.24
6611	452	0.26	0.07	0.19	7	0.48	0.23	0.13
6635	413	0.25	0.06	0.21	6	0.80	0.64	0.16
6640	428	0.30	0.09	0.16	6	0.48	0.23	0.14

A more detailed discussion of the estimation of the model accuracy, including statistics and plots of sub-model performance, can be found in the report "Assimilative Capacity Analysis for the Beaufort River Water Quality Report" (Jordan Jones & Goulding, 2003b)

MODEL APPLICATIONS

The Beaufort River assimilative capacity model was used to analyze the tidal river for various issues pertaining to responsible water-resource management of the system. The Beaufort River model offers different opportunities for addressing coastal regulatory issues. The model has the ability to simulate 33 months of data, and utilizes a constrained optimization routine to determine allowable point-source loading to maintain a water-quality standard. Some of the coastal regulatory issues addressed in this report are critical conditions for DO concentration, impacts of precipitation on DO, determination and analysis of allowable point-source loading using an optimization routine, and alternative point-source loading scenarios. The results from these scenarios are intended to demonstrate the utility of the model in making water-resource management decisions and the intended use as a TMDL model with regulatory applications.

Critical Conditions for Dissolved Oxygen

The procedure for determining the assimilative capacity of an upland, unregulated stream is well established. The procedure involves a statistically computed steady-state low-flow value, such as a 7Q10 flow, often referred to as the critical flow, which is used in conjunction with a critical water temperature in a simulation model. The results are interpreted in accordance to the State water-quality standards. Applying a similar approach to coastal system is often difficult due to the complexity of DO due to tidal dynamics. For example, are low flows critical for DO along the coast? During low flow, ocean water with higher DO concentration would propagate farther upstream which may not be a critical condition. The Beaufort River offers some particular challenges for determining the assimilative capacity:

- What are the critical conditions to use for DO concentration for evaluating the affect of point-source loading?
- How do you determine a critical flow for a coastal river that is just a tidal connection with no substantial inflow?
- For a system where the water temperatures are greater than 25 ^{o}C for 4 or 5 months, what is the critical temperature?
- Is there a different approach that needs to be taken for coastal waters that are naturally low in DO?

Most of the coastal waters of South Carolina are considered naturally low in DO. SCDHEC regulations allow a maximum deficit of 0.1 mg/L where waters do not meet the numeric standard for dissolved oxygen because of natural conditions. This is known as the "0.1 mg/L rule." To evaluate the effect of point-source loading, conditions for point-source loading are compared with a condition where there are no point-source discharges into the system (a no-load condition). The effects of the point-source loading can be evaluated by computing the differences in the DO concentrations, the delta DO, between the load and no-load condition to determine whether the impact exceeds the maximum deficit of 0.1 mg/L.

It is difficult to analyze the DO from the Beaufort River gaging network to determine the critical conditions. The DO is below a daily mean of 5 mg/L and a minimum of 4 mg/L every year for extended periods. During these periods, the DO is constantly stressed by variable point and non-point sources; the WRF loads ranged between 400 and 2,900 lbs/d of UOD and precipitation varied from the wettest to driest months on record. The critical condition is not necessarily during the period when DO concentrations are low or below the standard; more accurately, the critical condition represents a regulatory period when the assimilative capacity is limited.

To determine the assimilative capacity of the Beaufort River, the approach was taken to determine the hydrologic and water-quality conditions that were most sensitive to point-source loading. Under these conditions, the delta DO would be the greatest. It was difficult to determine this critical period, due to the variability of tides, temperatures, point-source discharges, and rainfall. The solution was to hold certain variable inputs constant, principally point-source loads. By holding the effluent loading constant, the changing delta DO would be due to the other changing conditions in the system. An arbitrary constant load was put into the system that was high enough to have a response in the system.

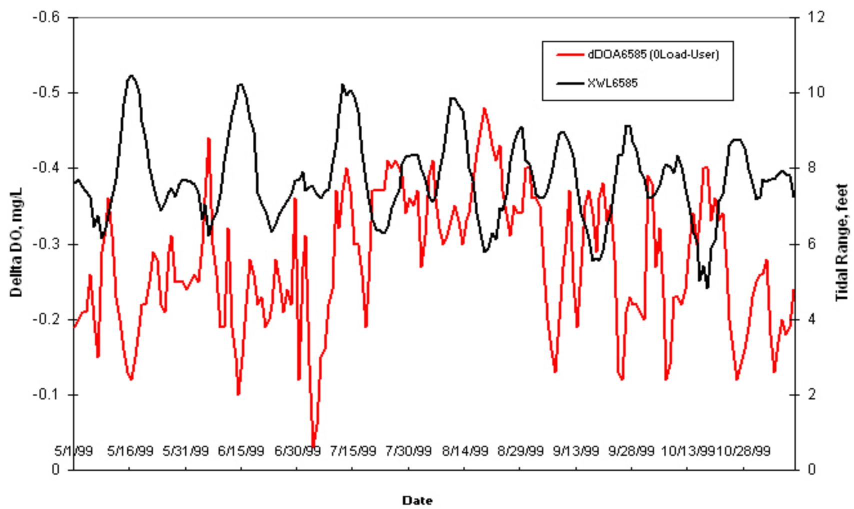

Figure 20. Delta DO concentration and tidal range for the period May to October 1999. Note the inverted delta DO y-axis scale.

The delta DO was plotted with other measured variables to understand the conditions controlling effect of the point-source load. Figure 20 shows the delta DO (dDOA6585) and the tidal range (XWL6585) for station 2176585 (Brickyard Creek) for the 6-month period from May to October 1999. The magnitude of the delta DO is not significant because an arbitrary constant load was used as an input to the model. What is important is the relative change in delta DO over the 6-month period. During the spring tides (when the tidal amplitude is higher) is when the point-source impact is the lowest (note the inverted scale). During neap tides (lower tidal ranges), the point-source impacts can increase significantly. There is a similar periodicity of the delta DO as with the semi-diurnal tide range. The large effects appear to be occurring on a 14-day cycle with the greatest effects occurring on a 28-day cycle.

The tidal range is the controlling factor for the assimilative capacity of the system. The 14-day and, especially, the 28-day spring tides, transport and mix water with higher DO from Port Royal Sound, while transporting lower DO water from the upper segment of the system to the Coosaw River. During the 14-day and 28-day neap tide cycles, there is not as much mixing and transport of higher DO water into the system or lower DO water out of the system, so there is an increase in the impact of the point-source loads. This phenomenon also was seen in the relatively long time (75 days) that it took the SC to recover after a large rainfall (see Figure 9). During that event, the system was loaded with a slug of freshwater and it took approximately three 28-day tidal cycles to move the large amount of freshwater out of the system.

Impact of Precipitation on Dissolved Oxygen

Non-point source loading during rainfall events may be a large source of oxygen-consuming constituents to receiving streams. Often, the ultimate oxygen demand of a load pulse during a rain event can be greater than the fully permitted point-source load. A critical element of TMDLs is a determination of the non-point-source impacts on an impaired stream. A lot of research and development has taken place over the last 20 years to improve watershed models and the coupling of watershed models with riverine models. Applications of these models to coastal areas are particularly difficult due to the low gradient watersheds, poorly defined drainage areas, tidal complexities, and lack of understanding of watershed and marsh processes. Despite these challenges, good water-resource management requires that there is an understanding of the contribution that non-point sources are having on impaired waters.

In the Beaufort River DO model, the effect of rainfall on DOD is estimated at each gage. Rainfall inputs to the model can be modified as a percent (0 to 150%) of the historical rainfall in the model database. To evaluate the impact of precipitation on dissolved oxygen, the model was run setting the rainfall inputs to zero (and point-source loads to the actual condition) and comparing the results to the simulations with the actual rainfall condition. The results for the two tributaries to the river, Battery Creek (station 2176635) and Albergottie Creek (station 2176587) and the upper gages on the Beaufort River (station 2176603 and station 2176589) are shown in Figure 21. (Note: the large rainfall of June 30, 1999, was not used in the training of the ANN models for the Beaufort River DO model so that more accurate models could be trained for normal rainfall events. The results for this period have been deleted from the plot.) The largest impact on DO in the system is in Albergottie Creek where rainfall increases DOD concentration by as much as 1.5 mg/L. The smallest impact is on Battery Creek where the DOD concentration increases by less than 0.5 mg/L. Although the riparian tidal marsh of Battery Creek drains the western side of the City of Beaufort, the impact may be low because the gage is located in the lower reaches of the creek where the channel geometry is large and there is good exchange with the lower reaches of the Beaufort River. Of the two river stations, Beaufort River at Beaufort (station 2176603) had the greater increases in DOD from rainfall.

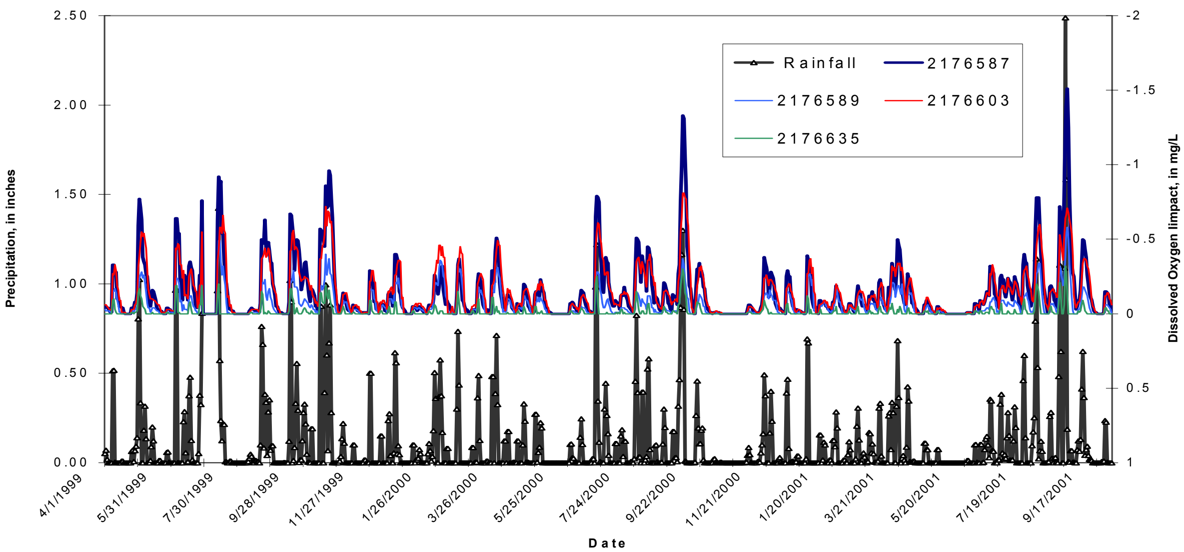

Figure 21. Two-day average rainfall and the dissolved oxygen impact due to precipitation for four stations on the Beaufort River and two tributaries for January 1999 to September 2001. Note the range of the second y-axis has been set to 1 to offset the dissolved oxygen impact for clarity.

To put the results from the evaluation of precipitation on DO in perspective, a similar simulation was run to evaluate the historical impact of point-source loads on DO. The model was run to compare the actual dynamic point-source loads to a no-load condition. The results of the precipitation and point-source loads on DO station 2176603 are shown in Figure 22. Point-source loads decreased the DO concentration by as much as 0.4 mg/L at the site. Rainfall decreased the DO concentrations by as much as 0.8 mg/L. The behavior of the two types of loading is quite different. The point-source loading is sustained throughout changing hydrologic conditions, and can be quite variable, but point-source loading is rarely discontinued for any length of time. Rainfall loads are pulse loads to the system that are not sustained for long periods. Although non-point-source loads due to rainfall can be higher than the point-source loads, they are transient and have no impact during periods of no rainfall. Although the maximum impact of precipitation is twice as great as the point-source loads for the simulation shown in Figure 22, the average impact of the point source and precipitation were coincidentally both 0.14 mg/L.

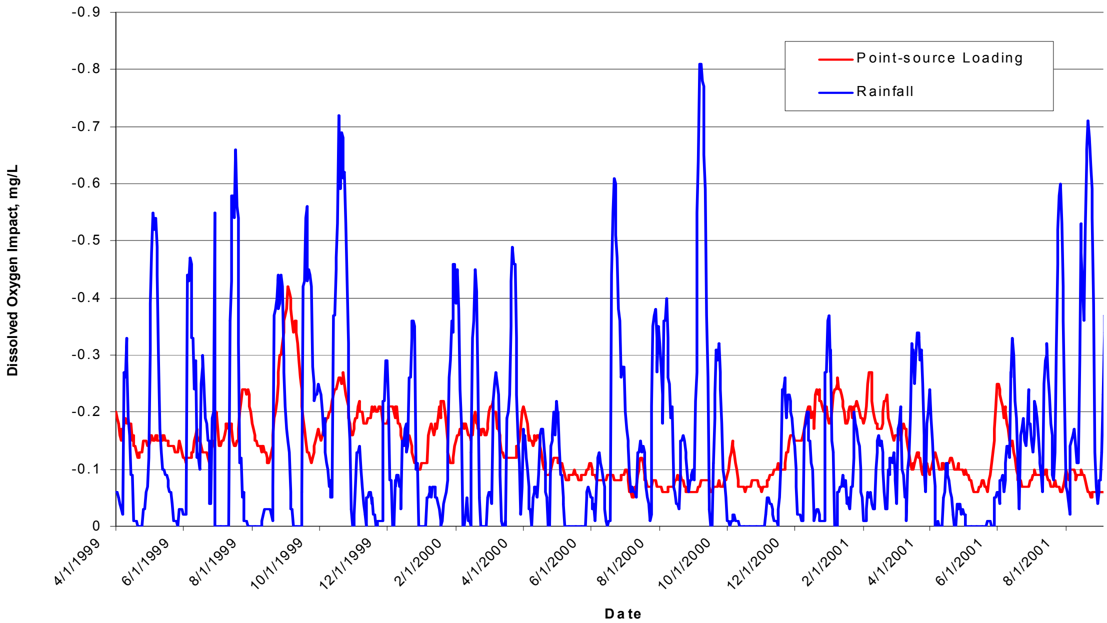

Figure 22. Graph showing the impact of precipitation and point-source loads on dissolved oxygen at station 2176603 for January 1999 to September 2001.

The results from the evaluation are not surprising. There is great concern by regulatory agencies to put in controls to minimize the non-point source impacts. Coastal waters are naturally low in DO due in part to the non-point source loading from rainfall and tidal exchanges with the marshes. The impact to DO predicted by the Beaufort River DO model does not differentiate between natural loading (tidal marshes, mudflats, etc.) and anthropogenic loading (impervious surfaces or altered landscapes, such as golf courses). For waters that are naturally low in DO, the impact of rainfall needs to be partitioned into the natural portion and the anthropogenic portion, which is controllable.

Allowable Point-source Loading Using Constraint Optimization

The assimilative capacity for oxygen-consuming constituents typically is accomplished as an iterative process. A proposed loading for a WRF is input into a model and the impact is evaluated with respect to the 0.1 mg/L rule. If the 0.1 mg/L target is exceeded, the load is reduced and the new load is evaluated. If the 0.1 mg/L target is not exceeded, the load is increased until the target is met. For rivers with a clearly understood critical condition and a limited number of WRFs, the process is manageable. For more complex systems, like an estuarine-receiving stream with multiple WRFs, the process can become very time consuming.

An alternative to the iterative approach to determine assimilative capacity is to utilize an optimization routine that allows the computer to determine the loading amount to meet a prescribed target. ANNs lend themselves to the use of optimization routines. Roehl and Conrads (1999) describe how optimization routines and ANNs can be integrated for real-time determination of assimilative capacity. Unlike deterministic models that must iterate for a solution for every time step and result in long run times, trained ANNs execute without iteration and execute very quickly. To utilize an optimization routine, the model is "inverted" where the output is prescribed or known. For the Beaufort River model, the output of concern is the delta DO for evaluating the 0.1 mg/L rule. With the output of the model set, or "constrained," one or more of the input variables must be allowed to vary in order to meet the desired output. Of the inputs to the Beaufort River model (WL, XWL, WT, SC, rainfall, effluent load), the only realistic variables for modulating are the effluent loads from the WRFs.

To utilize the optimization routine in the Beaufort River model, the user specifies the high and low target delta DO and the DO concentration for differentiating the two targets. For example, the user can specify that for DO concentrations less than 6.0 mg/L, the low target is a delta DO of 0.1 mg/L. For DO concentrations greater than 6.0 mg/L, the high target is set at a delta DO of 0.5 mg/L. For effluent loads to the system, the user specifies the relative ratios between the BOD and NH_3 loads. For example, if one WRF is simulated and the relative ratio for BOD is set at 1 and NH_3 is set at 0.5, then for every pound of BOD (in units of UOD) that is discharged to the system, a half of pound of NH_3 (in units of UOD) is discharged. For every time step during an optimization simulation, the model determines the specified delta DO target (high or low target depending on instream DO concentration) and increases loads to the system, while maintaining the specified relative ratios between BOD and NH_3, until the target is met. Output from the simulation is a time series of allowable loading for the simulation period. More information on the optimization routine in the model can be found in the User's Manual (Jordan Jones & Goulding, 2003a).

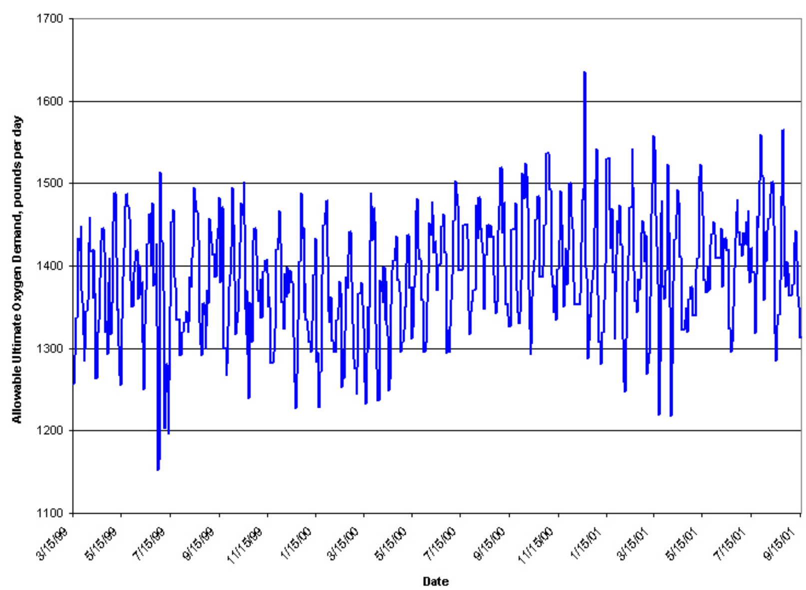

Figure 23. Time series of allowable loading: Southside WRF for March 1999 to September 2001.

The assimilative capacity of a system is a dynamic phenomenon that is changing with the changing hydrologic and water-quality conditions. For regulatory purposes, the assimilative capacity is a fixed quantity representing the allowable loading as determined by the critical conditions for the system. For the Beaufort River, the critical condition is occurring with the frequency of the neap-tide cycle of the semi-diurnal tide. A time series of allowable loading is shown in Figure 23. For this simulation, only the Southside WRF is discharging to the system and there was no difference between the high and low delta-DO target. The dynamic nature of the assimilative capacity, or allowable loading, is clearly seen. A seasonal or annual cycle is apparent in the time series. The variability of allowable loading on smaller time scales is due to the inter-tidal variability between spring and neap tides.

The optimization routine can determine the allowable loading to meet a specified water-quality target for changing hydrologic and water-quality conditions. The time series of allowable loading shows the dynamic behavior of the amount of oxygen-consuming constituents that the system can assimilate. For the regulator, the question becomes one of selecting the steady-state load that the WRFs will be permitted. If the minimum from Figure 23 is selected, regulators may be perceived as being overly protective and restrictive regarding a community's demographic and economic needs. If the maximum is selected, the regulators may be perceived as not being sufficiently protective of the State's water resources. The solution is somewhere between these two extremes.

Time-series Frequency Distribution of Allowable Point-Source Loading

For the typical upland river when a statistical flow such as a 7Q10 is used, the determined assimilative capacity would be considered protective of a low-flow critical condition that has a recurrence interval of 10 years. For the Beaufort River, the critical condition has a recurrence interval of every 14 days. Rather than select one neap tide to use as a critical condition, the allowable loading can be computed for the period of record of the model database (33 months). A histogram of frequency distribution of the allowable loading can be generated to better understand the range and occurrences of the predicted loading levels. Figure 24 shows the frequency distribution of the time series in Figure 23. The range of the allowable loading is between 1,100 lbs and 1,700 lbs with the highest occurrences between 1,350 and 1,450 lbs. The cumulative percentile plot also is shown in Figure 24. Using the percentile plots, regulators can select the constant allowable loading, based on the frequency and the percentage of time occurrence of a loading amount. Once selected, the chosen allowable load can be simulated in the model as a constant load and the frequency of meeting the 0.1 mg/L rule can be evaluated.

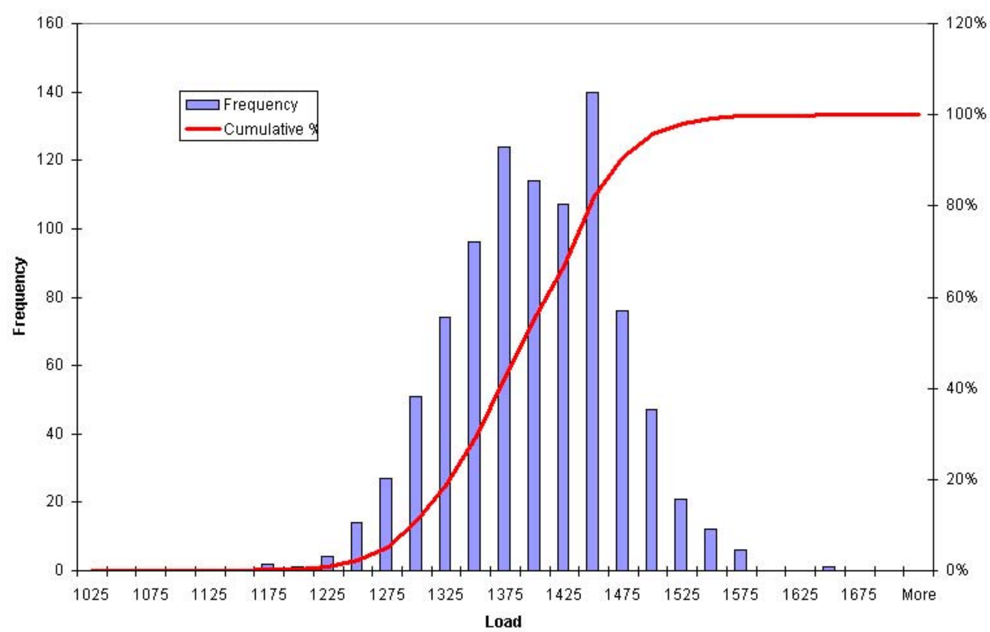

Figure 24. Frequency distribution for allowable loading: Southside WRF for March 1999 to September 2001. Frequency distribution is based on time series of predicted allowable loading shown in Figure 23.

Alternative Point-source Loading Scenarios

Given these current permits, the model simulations were developed assuming that the total waste load to the Beaufort River would be maintained at the current ratios between the WRFs. Currently (2003), BJWSA is permitted at 59% of the total load to the river (Table 2). Parris Island and MCAS are permitted at 33% and 8% of the total load, respectively. Inputs to the Beaufort River model were developed to analyze three discharge scenarios. With the imbedded optimizer function, the model seeks an optimal solution for the specified discharges within the modeling constraints. This allows the modeler to select the pass/fail criteria (0.1 mg/L DO deficit) along with specific BOD and NH_3 ratios between each specific treatment facility, and allow the model to determine the maximum allowable load.

Models were developed for the following scenarios:

1. Maintain all three existing discharges.
2. Eliminate the MCAS discharge and relocate those flows to the BJWSA discharge location.
3. Eliminate the MCAS and Parris Island discharges, and relocate those flows to the BJWSA location.

Constraints for the optimization routine were set with a low delta DO target of 0.1 mg/L for instream DO concentrations less than 6.0 mg/L. The high delta DO target was set at 0.5 mg/L. The relative ratios were set at current permits. The model simulations assumed that the Beaufort River was divided into two segments with the segment boundary at a point south of station 2176611. This model segmentation provides for two equal volume river segments. Model results for DO were compiled using volumetric averaging within these two segments.

Model pass/fail criteria were based on the SCDHEC standards for waters naturally low in DO. This standard allows for DO concentrations to be lowered by 0.1 mg/L below natural conditions. For this scenario, it is assumed that SCDHEC will apply this standard during the critical months when WT are high and DO is low. Based on the observed data, these critical months are from May through October. During the non-critical months of November through April, WT are low and DO concentrations are consistently above standard of a daily mean concentration of 5.0 mg/L. It is assumed that the 0.1-mg/L standard would not apply during the winter months and allow for a minimum DO concentration of 5.0 mg/L. The model scenarios conservatively assumed a 0.5-mg/L impact (delta DO) from the WRFs when DO concentrations are greater than 6.0 mg/L.

Given the model constraints described above, the following results were determined for each of the three discharge scenarios, as shown in Tables 2-4. These loadings are approximate and are considered preliminary, pending final review and SCDHEC concurrence. Based on these loads, the following critical period (summer) permit limits may be anticipated and are summarized in Table 5.

As noted above, the modeling analysis indicates that allowable loadings are *decreased* when the current discharge at MCAS remains in service. The MCAS discharge to Albergottie Creek near the confluence with the Beaufort River results in a substantial effect on DO concentrations in this vicinity. The optimal discharge scenario would be to combine all three discharges at a single point. Table 4 shows that Scenario 3, with a single discharge at the current BJWSA discharge location, allows for 24% more loading than Scenario 2 and 40% more loading than Scenario 1. For TMDL purposes, it appears that the three existing dischargers could consider a single discharge location near the current BJWSA discharge.

Table 2. Scenario 1 – Three discharges proposed UOD allocations to the Beaufort River summer limits (critical period)

Dischargers	**Total UOD Allocation (pounds per day)**				
	% of Allocation	**UOD**	**MGD**	**CBOD5 (mg/L)**	**NH_3 (mg/L)**
MCAS WRF	8	122	0.75	5.0	1.0
Parris Island WRF	33	490	3.0	5.0	1.0
Port Royal WRF	59	941	10.0	3.0	0.5
Totals	100	1,553	13.75		

Table 3. Scenario 2 – two discharges proposed UOD allocations to the Beaufort River

Dischargers	**Total UOD Allocation (pounds per day)**				
	% of Allocation	**UOD**	**MGD**	**CBOD5 (mg/L)**	**NH_3 (mg/L)**
MCAS WRF	0				
Parris Island WRF	33	527	3.0	5.5	1.0
Port Royal WRF	67	1,146	10.75	3.5	0.5
Totals	100	1,673	13.75		

Table 4. Scenario 3 – one discharge proposed UOD allocations to the Beaufort River

Dischargers	**Total UOD Allocation (pounds per day)**				
	% of Allocation	**UOD**	**MGD**	**CBOD5 (mg/L)**	**NH_3 (mg/L)**
MCAS WRF	0				
Parris Island WRF	0				
Port Royal WRF	100	2,154	13.75	5.5	0.5
Totals	100	2,154	13.75		

Table 5. Preliminary model results total UOD loading

Model Scenario	**Summer UOD (lbs/day)**	**Winter UOD (lbs/day)**
Scenario 1 – three discharges	1,500	4,800
Scenario 2 – two discharges	1,700	6,000
Scenario 3 – one discharge	2,100	6,500

SUMMARY

The Beaufort River assimilative capacity model was built using water-quality data collected during the period December 1998 through September 2001. This substantial database (over 1.5 million data points) includes a full array of time-series data for precipitation, dissolved oxygen (DO), water level, tidal streamflow, tidal stream velocity, water temperature, specific conductance, and wastewater discharge.

Seven data collection platforms installed by the USGS have provided 33 months of continuous record data for the ambient water-quality parameters. This extensive water-quality database, coupled with precipitation data and discharge data from the permitted wastewater facilities (BOD5, NH_3, and flow), provided the foundation for the modeling effort.

The data mining techniques successfully calculated the sensitivity of the point-source loads on instream DO concentrations. Sensitivities were performed for BOD and NH_3 loads from each treatment plant and the impact on the DO concentrations at each gaging station location. From this information, an empirical model of the system was created that makes accurate predictions of the response of instream DO concentration due to changing hydrologic, meteorological, and point-source loading conditions. Model findings indicate that regulators may consider reducing wastewater loadings to the Beaufort River below current permit levels. The model indicates that the DO response is sensitive to the location of the loads to the Beaufort River, particularly loads near the Albergotti Creek and Brickyard Creek confluence. Reductions in loading capacity could be minimized with the MCAS discharge relocated at or near the existing Parris Island and BJWSA discharge locations. It also appears that the reductions in loading to the Beaufort River could be further minimized with a single discharge combining all three of the current discharges.

The Beaufort River assimilative capacity model will be used by SCDHEC to develop waste-load allocations consistent with the TMDL for DO. Wastewater management scenarios have been evaluated by BJWSA for developing the optimal loading from the various dischargers while protecting the integrity of the Beaufort River. Further analysis and discussion with SCDHEC is warranted to confirm the critical period loads and to develop seasonal limits that would apply during periods of low temperatures and high instream DO concentrations. The TMDL allocation process will ultimately involve the WRFs operated by BJWSA, MCAS, and Parris Island and will be reviewed and approved through the designated 208 Planning Agency, the Lowcountry Council of Governments.

The result of this comprehensive data mining approach is an empirical model that is well trained and capable of predicting DO values with a high level of accuracy. The Beaufort River assimilative capacity model shows how the application of data mining techniques of large environmental data sets can be used to develop accurate empirical models that are an alternative to the development of mechanistic models to complex systems.

ACKNOWLEDGEMENTS

Beaufort-Jasper Water and Sewer Authority is responsible for providing water and wastewater services to the majority of Beaufort County and Jasper County in South Carolina and has been intimately involved in water-resource planning issues along coastal South Carolina. In conjunction with these responsibilities, BJWSA initiated and provided significant funding for the development of a dynamic water-quality model for SCDHEC to use in determining the assimilative capacity of the Beaufort River system. The authors thank Dean Moss, Director of BJWSA, for his leadership during this project and Ed Saxon and Chris Petry of the Engineering Section of BJWSA, for their support and assistance. In addition, the authors thank the Marine Corp Commands for their assistance in obtaining data on their facilities.

REFERENCES

Conrads, P.A., Roehl, E.A, and Cook, J.B., 2002a, "Estimation of Tidal Marsh Loading Effects in a Complex Estuary," American Water Resources Association Annual Conference, New Orleans, May 2002.

Conrads, P.A.; Roehl, E.A., and Martello, W. P., 2002b, "Estimating Point-Source Impacts on the Beaufort River Using Neural Network Models," American Water Resources Association Annual Conference, New Orleans, May 2002.

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

Hinton, G.E., 1992, "How Neural Networks Learn from Experience," *Scientific American*, September 1992, p.145-151.

Jordan, Jones & Goulding, Inc., 2003a, "Assimilative Capacity Analysis for the Beaufort River." Prepared for Beaufort-Jasper Water and Sewer Authority, February 2003.

Jordan, Jones & Goulding, Inc., 2003b, "Assimilative Capacity Analysis for the Beaufort River: Water Quality Model Report." Prepared for Beaufort-Jasper Water and Sewer Authority, July 2003.

Press, William H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P., 1993, Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, 1993.

Risley, J.C., Roehl, E.A, and Conrads, P.A., 2002, "Estimating Water Temperatures in Small Streams in Western Oregon Using Neural Network Models,' U.S. Geological Survey Water Resources Investigations Report 02-4218, 2002.

Roehl, E.A., Conrads, P.A., Roehl, T.A.S., 2000, "Near Real-Time Control of the Salt Front in a Complex Tidally Affected River Basin," Smart Engineering System Design: Volume 10, Proceedings of the Artificial Neural Networks In Engineering Conference, ASME Press, New York, pp. 947-954.

Roehl, E.A. and Conrads, P.A., 1999, "Near 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.

Rosenblatt, F., 1958, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review*, 65, 1958, p. 386-408.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J., 1986, "Learning Internal Representations by Error Propagation," Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1, 318-362, Cambridge, MA: The MIT Press.

South Carolina Department of Health and Environmental Control, 1997, "Watershed Water Quality Assessment - Savannah and Salkehatchie River Basins," Technical Report No. 003-97, December 1997.

South Carolina Department of Natural Resources, 2003, Southeast Region Climate Center web site: www.dnr.state.sc.us/climate/sercc/

U.S. Geological Survey, 1981, Technical Memorandum 81.11, Reston, Va.

Weiss, S.M. and Indurkhya, N., 1998, Predictive Data Mining – A Practical Guide, Morgan Kaufmann Publishers, Inc., San Francisco, p. 1.