The Role of Artificial Neural Network Models in Developing a Regional Wastewater Reclamation Facility for Beaufort, SC, USA

Paul A. Conrads[a], Edwin A. Roehl[b], William Martello[c], and Edward Saxon[d]

[a]*U.S. Geological Survey, South Carolina Water Science Center, Columbia, SC, USA*

[b]*Advanced Data Mining, Greenville, SC, USA*

[c]*formerly with Jordan, Jones, and Goulding, Athens, GA, USA*

[d]*Beaufort-Jasper Water and Sewer Authority, Beaufort, SC, USA*

Abstract: Beaufort and Jasper Counties are two rapidly growing coastal counties in South Carolina. According to the 2000 census, the region grew by 40 percent during the preceding 10 years. The population growth has increased the quantity of wastewater in the area. The Beaufort River is a complex tidal river system that is home to shellfish grounds, fisheries nursery habitats, and extensive recreation by the local community. The river also serves as receiving water for treated wastewater from military and civilian wastewater reclamation facilities (WRF). Although not uncommon for coastal areas, the river is on the South Carolina list of impaired waters for low dissolved oxygen. The WRFs discharging into the Beaufort River are old and operating close to their permitted capacity. A regional approach was necessary to address the complex river system issues and to protect the highly sensitive ecosystems of the Beaufort River.

A regional wastewater management approach was developed that called for consolidating the military and municipal wastewater discharges into a single, high-quality discharge. This plan, however, required extensive data collection and modeling to prove its validity. Toward that end, an artificial neural network (ANN) dissolved-oxygen model was developed for the river to predict the impacts from a regional WRF. The database, ANN models, model controls, streaming graphics, and simulation output were integrated into a spreadsheet-based Decision Support System (DSS) to facilitate the dissemination of the model. Regulatory agencies and stakeholders were able to use the application to analyze data and model simulations for the permitting process. Potential wastewater management scenarios were evaluated for developing the optimal WRF loading while protecting the water quality of the Beaufort River. The DSS confirmed that a regional approach proposed by the Beaufort-Jasper Water and Sewer Authority provided the most protective solution. An effort to consolidate the military wastewater to the new regional wastewater treatment plant is currently underway.

Keywords: Tidal River; Artificial Neural Network Modeling; Dissolved Oxygen; TMDL, Decision Support System

1. INTRODUCTION

Beaufort and Jasper Counties are two rapidly growing coastal counties in South Carolina. According to the 2000 census, the region grew by 40 percent during the preceding 10 years. The population growth has increased the quantity of wastewater in the area. The principle receiving stream for treated effluent is the Beaufort River. The river is a complex tidal river system that is home to shellfish grounds and fisheries nursery habitats in addition to receiving treated wastewater from four civilian and military water reclamation facilities (WRF, fig.1). Although not uncommon for coastal areas, the river is on the South Carolina 303(d) list of impaired waters for low dissolved oxygen (DO). The Clean Water Act stipulates that Total Maximum Daily Loads (TMDL) must be determined for all waters on the 303(d) list. Critical to the development of a defensible TMDL is the linkage between the impairment and the

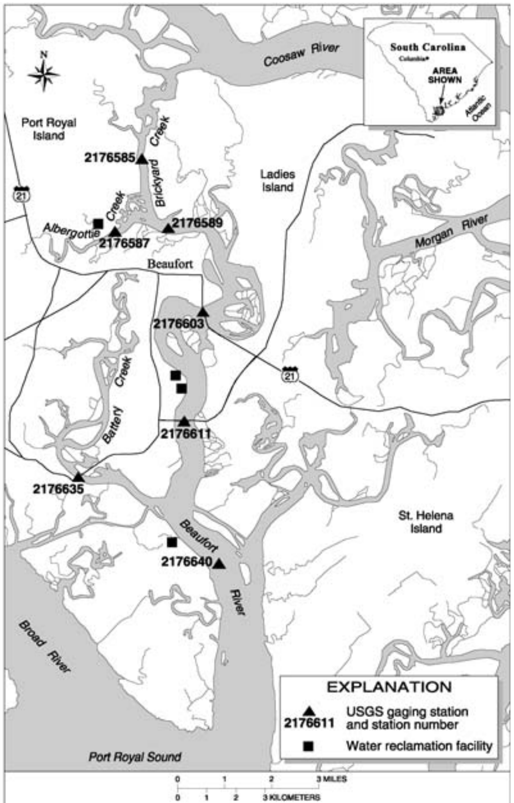

Figure 1: Beaufort River, S.C.

source of the impairment. The linkage is typically performed using a prediction model.

The Beaufort-Jasper Water and Sewer Authority (BJWSA) operates two WRFs on the Beaufort River. The facilities are operating at 70 percent of capacity and must be replaced in 2006 to handle increased wastewater flows for the growing coastal community. The Director of BJWSA envisioned a plan to build a regional WRF and to consolidate the civilian and military wastewater discharges into a single, high-quality effluent at the location of one of the current outfalls. The ambitious plan required an expedited permitting effort that included developing a predictive DO model of the Beaufort River to evaluate the effect of existing and future WRFs.

Two previous modeling and permitting efforts along the SC coast (Myrtle Beach and Charleston) were lengthy processes taking more than 10 years from the initiation of data collection to issuance of permits. To meet the schedule for a new permit and the construction of a new WRF, a new approach to developing a predictive DO model was required. BJWSA assembled a team of the U.S. Geological Survey (USGS), Advanced Data Mining, LLC (ADM), and Jordan Jones and Goulding (JJ&G) to analyze existing data, build an empirical DO model, and coordinate the permitting process with the South Carolina Department of Environmental Control (SCDHEC).

The team was successful in developing an accurate predictive DO model of the Beaufort River, disseminating the study results in a user-friendly Decision Support System (DSS), and obtaining the required permits to initiate construction of a new regional WRF. This paper summarizes the technical aspects of the study, and describes how proactive project management was able to use the new modeling approach to reduce the time from the initiation of data collection to the issuance of permits by 50 percent.

2.0 Modeling Approach

The authors had previously developed artificial neural network (ANN) models of a tidal river in the vicinity of Charleston, South Carolina. The type of ANN used was the multi-layered perceptron (MLP) described by Hinton (1992) and Jensen (1994), which is a multivariate, non-linear regression method based on machine learning. In a side-by-side comparison, Conrads and Roehl (1999) found that ANN models had prediction errors 60-82% lower than those of a state-of-the-practice mechanistic model when predicting water temperature (WT), specific conductance (SC), and DO in the Cooper River, South Carolina. Conrads and others (2002) went on to use ANNs to estimate the effects of nutrient loading from rainfall runoff and tidal marsh inundation on DO in the same waterway.

The variability of DO in the Beaufort River is a result of many factors including the quality of the water from Port Royal Sound to the south and the Coosaw River to the north, the loading of oxygen-consuming constituents from tidal marshes and other non-point sources, effluent from four permitted point sources, and the physical characteristics of streamflow, tidal range, salinity, and temperature.

The following discussion is a brief summary of the data sets, data preparation, and ANN modeling. More detail descriptions of these technical aspects of the study can be found in Conrads and others (2002, 2003).

2.1 Data

The data used for analysis and modeling consisted of continuous (1-hour interval) tidal and water-quality data, daily total precipitation data, and weekly effluent data. In 1999, BJWSA, in cooperation with the USGS, established a network of seven gaging stations (fig. 1) on the Beaufort River that monitor water level (WL), WT, SC, and

DO. Three of the stations also record tidal streamflow. Precipitation data were obtained from the National Weather Service and two of the WRFs. Effluent data (sampled once a week) of 5-day biochemical oxygen demand (BOD_5) and ammonia (NH_3) were obtained from the WRFs.

Two calculated variables were derived — tidal range (XWL) and DO deficit (DOD). Tidal range is an important variable for determining the flushing dynamics of the tidal rivers. Tidal range is calculated from water level and is defined as the water level at high tide minus the water level at low tide for each semi-diurnal tidal cycle. The DOD is the measure of the difference between actual DO measurement and DO for fully saturated conditions. The DOD was computed using an algorithm that assumes a constant barometric pressure over the data collection period (USGS, 1981). The DOD was adjusted for salinity.

2.2 Data Preparation

Tidal systems are highly dynamic and exhibit complex behaviors that occur over a range of time scales. The semi-diurnal tide is dominated by the lunar cycle, which is more influential than the 24-hour solar cycle; thus, a 24-hour average is inappropriate to use to reduce tidal data to daily mean values. For analysis and model development, the USGS hourly data were digitally filtered using a low-pass filter (Press and others, 1993) to remove semi-diurnal and diurnal variability. (Filtered variables are denoted by an "F" prefix, for example, FDO). The resulting filtered time series were then averaged over a 24-hour period to represent the daily mean for each parameter.

Explanatory variables for a particular response variable are often themselves correlated. It is difficult, if not impossible, to identify the individual effects of these variables (sometimes known as confounded or correlated variables), on a response variable. Empirical models have no notion of process physics, nor the nature of interrelations among 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. The precipitation, XWL, WL, and SC were systematically, non-linearly decorrelated using ANN models.

Because of the limited number of data points of the effluent sampling concentrations data as compared to the river gaging data, a subset of the dataset was excised that included only the daily digitally filtered data. In addition, time derivatives of the hydrodynamic (WL and XWL) and water-quality variables were computed and added to the dataset. The derivatives of filtered variables are denoted by an E prefix, for example, ESC.

2.3 Simulating Dissolved-Oxygen Deficit

The goal of the model was to predict the impact of the point- and non-point sources on DO. Had the goal only been to predict DO and not the effects of the WRFs, this could have been done easily and accurately using only WT owing to their strong inverse relation. Linear regression produces a coefficient of determination (R^2) of 0.88, indicating that approximately 88 percent of the variability of FDO is explained by FWT alone (fig. 2), and that only approximately 12 percent of the variability is caused by other factors.

The real goal of a regulatory model is to predict how much of the variability in DO is attributable to point-source discharges. The use of DOD rather than DO as the response variable normalizes the DO signal with respect to WT to emphasize the effects of external loadings on DO. The response of DOD to BOD_5, NH_3, rainfall, and the other explanatory variables was predicted using ANN models that were trained using the back-propagation and conjugate-gradient algorithms.

Visual inspection of the BOD_5 loading data from the WRF and the 1-day change in DOD (EDOD) at station 02176611 (fig. 3) suggests a potential relation (note that the EDOD scale has been inverted so decreases in daily DO rise on the scale). The number of coincident peaks, for example observations 6, 31, 35, 39, and 58, indicate that BOD_5 loading may account for a significant part of the remaining 12 percent of the variability in DO.

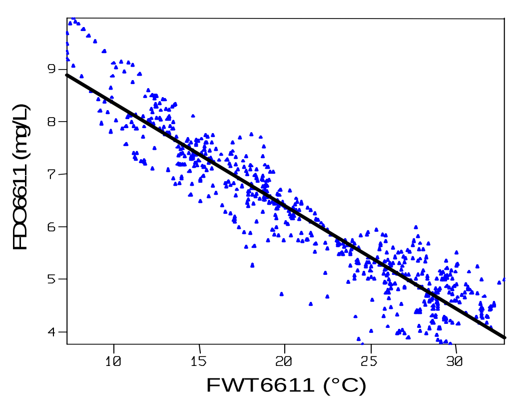

Figure 2: Scatter plot of filtered DO (FDO) and filtered water temperature (FWT) and least-squares regression line (R^2=0.88) for station 02176611.

For each station, 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 relation between the BOD_5 and the DO. Figure 4 shows that the model fits most of the higher peaks in the EDOD. The $R^2_{ANN} = 0.57$, indicating that approximately 57 percent of the variability in the EDOD is accounted for by variability in the input variables. A similar approach was used for modeling the impact of NH_3 on DOD.

The Beaufort River DO Model is a super-model composed of 118 cascaded sub-models. Separate sub-models were constructed for each combination of river gage location, discharge type (BOD_5 or NH_3), and relative time delay. The impacts of BOD_5 and NH_3 were computed by sub-models for each river gage that used decorrelated XWL, WL, SC, and WT and their 2-day time derivatives as inputs. Also included are inputs for each WRFs BOD_5 and NH_3 at appropriate time delays. The output of each sub-model is a prediction of the 1-day EDOD

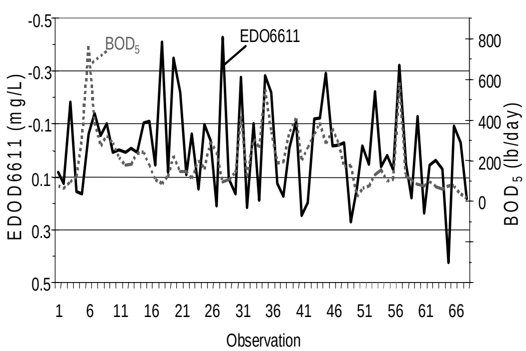

Figure 3: One-day change in DO (EDO) and BOD_5 (at a 1 day time delay).

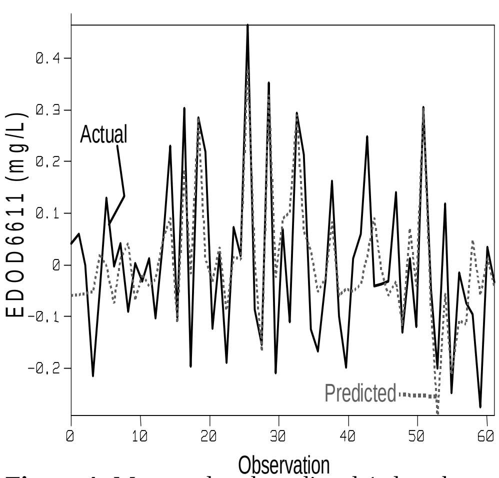

Figure 4: Measured and predicted 1-day change in DO deficit (EDOD). ANN used BOD_5 as an input at a time delay of 1 day. $R^2_{ANN} = 0.57$.

3.0 Development of the Decision Support System

Commonly, a DSS is often a software package built around a model, making the model the DSS's most important component because ostensibly it can correctly predict, *"What will happen if we do A instead of B?"* Models are often developed at considerable expense; therefore, the packaging is done only to maximize the usefulness of the model to the broadest possible community of users.

Complex mathematical (mechanistic) models based on first principles physical equations are often developed and operated by senior scientists or engineers: however, the interests and computer skills of the actual decision makers and other stakeholders are quite varied. A DSS was developed to meet the needs of the technically diverse group of stakeholders for equal access by all to the body of scientific knowledge needed to make the best possible decisions.

The DSS for the Beaufort River was developed as Microsoft®Excel/Visual Basic for Applications[1] (VBA) programs. This allowed the DSS to be prototyped, easily modified and also allowed the DSS to be distributed in a familiar form. The DSS is operated through a graphical user interface (GUI) composed of menus and controls that require no typing. This makes the DSS easy to use and eliminates the need to trap user input errors.

The DSS incorporates a database of measured and calculated time-series variables for running long-term simulations. Under user control, a VBA program loops through database records, assembles input vectors, executes super-model instructions, outputs model predictions, and drives graphics. The DSS incorporated the following in addition to the database and ANN sub-models:

- *Simulation Controls:* Model controls, including start/stop dates, user-defined setting, and optimizations run (discussed below), are set with the point and click GUI. The DSS executes multiple model simulations simultaneously such as the no-discharge load, actual (historical) loading, user-defined, and optimized conditions.
- *Spatial Interpolation:* The DO predictions are spatially interpolated by the modeling application using a "natural cubic

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

spline" algorithm (Burden and others, 1981). The longitudinal DO response in the system for the no-load, actual, optimized load is dynamically displayed during the model simulation.

- *Volumetric Averaging:* Regulatory analysis of impacts to a system typically is done on river segments over a specified period rather than on an "any time and any place" basis. The DSS allows the user to subdivide the estuary into a maximum offour segments for volumetric averaging of historical and computed hydrodynamic and water-quality variables.
- *Constrained Optimization:* a constrained optimizer was configured to represent South Carolina state law that governed the maximum allowable impact that nutrient loads from the WRFs could have on riverine DO. The optimizer modulates controllable inputs during simulations to obtain predictions that match user-set set points. Users can allocate the TMDL load among the BOD and NH_3 discharges from each WRF. At each time step the optimizer iterates load inputs as assimilative capacity changes. Output from the optimizer is time-series of allowable loadings for WRFs that meet the water-quality standard.

4.0 Model Results

The Beaufort River DSS enabled stakeholders and regulators to use new approaches to analyzing critical conditions and allowable loading to coastal systems (Conrads and others, 2003). The assimilative capacity of a system (the amount of effluent that can be discharged without violating the State water-quality standard) 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 regulator, the question becomes one of selecting the steady-state load that the WRFs will be permitted.

For the Beaufort River the critical condition occurs during neap tides and has a recurrence interval of 14 days (Conrads and others, 2003). Rather than select one neap tide to use as a critical condition, the allowable loading was computed for the full 3-year period of record. A frequency distribution of the allowable loading (ultimate-oxygen demand, in pounds per day) was generated to better understand the range and occurrences of the predicted loads. Figure 5 shows the load frequency distribution and the cumulative percentile plot.

Using the percentile plots, regulators could select a constant allowable loading based on a frequency of occurrence. Once selected, the allowable load

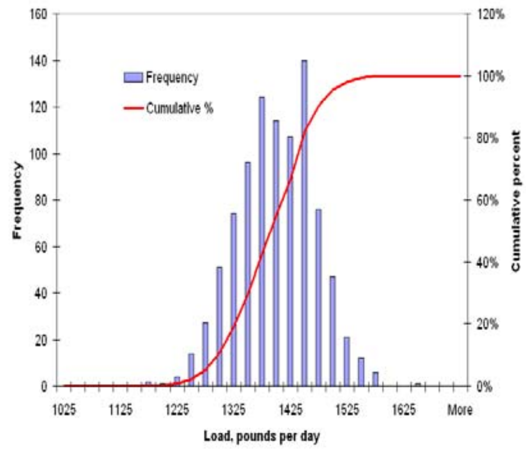

Figure 5. Frequency distribution for allowable loading (pounds per day) for WRF near station 02176611 for March 1999 to September 2001.

was simulated in the model as a constant load and the frequency of meeting the water-quality standard was evaluated. For the Beaufort River, a 90-percent reduction in loading was required to obtain the new permit for the regional WRF plant for the Beaufort River.

5.0 Project Management

Typically, water-resource managers and regulators are more comfortable using state-of-the-practice analytical tools than innovative state-of-the-art tools that have not been previously applied for determining a TMDL. To undertake a study using an innovative approach to meet the accelerated permitting schedule required proactive project management to ensure that all the stakeholders in the process clearly understood and supported the study approach. Important elements of the project's implementation and management are discussed below.

- *Study plan clearly understood by all parties:* Prior to initiation of the project, the Director of the BJWSA met with Bureau Chiefs at the SCDHEC to clearly articulate their need for a new WRF, their concern with the history of determining coastal TMDLs, and the typically long times needed to issue permits. Given the circumstances, SCDHEC and BJWSA saw an opportunity to try a new approach for developing a predictive DO model that was mutually beneficial for both agencies.
- *Technical team with well-defined roles:* It is critical that individuals collaborating on a study have clearly defined roles and a mutual respect for the integrity of the team. The technical team assembled by BJWSA was composed of: the USGS, which provided expertise in water-quality

and hydrodynamic data, riverine and estuarine dynamics, and coastal TMDL modeling; ADM, which provided expertise in data mining, ANN modeling, and software development; and, JJ&G, which provided expertise in wastewater regulatory issues and project management.

- *Open process with State and Federal regulators:* To implement a study using new technology, it is essential that the State and Federal regulators understand the technical aspects of the new technology. The technical dissemination of the project began at the inception of the study. Technical team members held a series of technical exchange meetings with the State and Federal regulators to discuss all aspects of data mining, ANN modeling, and the development of the DSS application.
- *Readily disseminated study results with the DSS:* The development of the DSS facilitated the dissemination of study results and made them accessible by a broad range of end-users. The DSS allowed stakeholders to easily simulate various loading scenarios and compare these results with historical conditions. Rather than spending a large amount of time on getting users past a steep learning curve, as is typical with resource-intensive mechanistic models, the DSS allowed users to focus on simulation results rather than struggle with model operation.
- *In-depth review process:* Success in any modeling study requires a thorough technical review to assure stakeholders that the model is technically sound. Building consensus on the model's technical viability gives stakeholders the confidence they need to address environmental and regulatory issues rather than debate the model. The technical team met with representatives of the affected military WRFs for an exhaustive review of every aspect of study.

6.0 Current Status of Regional WRF

BJWSA received their permit in May 2004 (35 months after the start of the modeling study) for the regional WRF. The new facility is currently under construction and should be online by June 2006. BJWSA has submitted a proposal to take over the operation of the two military WRFs and has been selected as the best-value offerer. A final decision on the offer will be made as early as October 2006. BJWSA has completed the preliminary engineering and planning process for installing new force mains to convey wastewater from the military installations to the new regional plant. It is anticipated that the consolidation of the separate discharges to the river to one high quality discharge will occur by November 2009.

7.0 Discussion

The successful role played by the Beaufort River ANN model in developing a regional WRF demonstrates that an innovative modeling approach can be undertaken if the river system is well characterized by continuous data and there is a cooperative relationship between the permitted dischargers and the regulatory agencies involved. Advances in environmental monitoring over the past 20 years have made it cost effective to acquire tremendous amounts of hydrologic and water-quality data and large databases exist for many riverine and estuarine systems. Empirical models of complex river systems can be developed directly from the data. These models often can be developed faster than traditional modeling methods and easily disseminated to meet the needs of a broad range of stakeholders.

8.0 References

Burden, R.L., Faires, J.D. and Reynolds, A.C., Numerical Analysis 2nd Edition, Prindle, Weber & Schmidt, Boston, pp. 112-113, 1981.

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

Conrads, P.A., E.A. Roehl, and J. Cook, Estimation of tidal marsh loading effects in a complex estuary, American Water Resources Association Annual Conference, New Orleans, May 2002.

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

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

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

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.

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