Features of Advanced Decision Support Systems for Environmental Studies, Management, and Regulation

Edwin A. Roehl Jr., Advanced Data Mining, Greer, South Carolina, ed.roehl@advdatamining.com
Paul A. Conrads, U.S. Geological Survey, Columbia, South Carolina, pconrads@usgs.gov
Ruby C. Daamen, Advanced Data Mining, Greer, South Carolina, ruby@advdatamining.com

Abstract: Natural resource managers and users face difficult challenges when managing the interactions between natural and man-made systems. Even though the collective interests and computer skills of the community of managers, scientists, and other stakeholders are quite varied, there is an overarching need for equal access by all to the scientific knowledge needed to make the best possible decisions. A decision support system (DSS) can meet this need. DSS have been described as, *"computer-based systems (for) helping decision-makers to solve various semi structured and unstructured problems involving multiple attributes, objectives, and goals… Historically, the majority of DSSs have been either computer implementations of mathematical models or extensions of database systems and traditional management information systems."* This paper describes DSS developed for three different hydrologic systems in South Carolina. The goals of the three were – the regulatory permitting of wastewater plants on the Beaufort River, evaluating the environmental impact of a proposed deepening of Savannah Harbor, and regulating hydroelectric generation on the Pee Dee River to protect Myrtle Beach-area fresh water intakes from salinity intrusions. These DSS provide predictive models with real-time databases for simulation, graphical user interfaces, and streaming displays of results. Additional features include optimizers, integrations with other models and software tools; and color contouring of simulation output data.

Keywords: decision support, neural network, optimization

1. INTRODUCTION

Natural resource managers and users face difficult challenges when managing the interactions between natural and man-made systems. At considerable cost, complex mathematical (mechanistic) models based on first principles physical equations are often developed and operated by senior scientists to evaluate options for using a resource while minimizing harm. However, varying technical abilities and financial constraints among different stakeholders effectively restricts access to relevant scientific knowledge and tools. This can lead to distrust between *haves* and *have-nots* and stall important management processes for years. There is a need to provide equal access to the knowledge and tools required for informed decision-making. A decision support system (DSS) can help meet this need.

Dutta et al [1997] describe DSS as, *"computer-based systems helping decision-makers to solve various semistructured and unstructured problems involving multiple attributes, objectives, and goals…Historically, the majority of DSS have been either computer implementations of mathematical models or extensions of database systems and traditional management information systems…With the help of AI (Artificial Intelligence) techniques DSS have incorporated the heuristic models of decision-makers and provided increasingly richer support for decision-making."* This paper describes DSSs developed to address hydrologic issues in the three South Carolina estuaries shown in Figure 1. The issues are regulatory permitting of three wastewater treatment plants (WWTP) on the Beaufort River, evaluating the environmental impact of a proposed deepening of Savannah Harbor, and regulating hydroelectric generation on the Pee Dee River to protect coastal fresh water intakes from salinity intrusions.

2. MODELS

A DSS is often a software package built around a model, making the model the DSS' most important component because ostensibly it can correctly predict, *"What will happen if we do A instead of B?"* Models are often complicated and expensive to develop. While good packaging can broaden their usefulness, a model lacking scientific credibility can delay the resource management process indefinitely.

Calibrating a model is a process of fitting a line or surface (function) through data from two or more variables. This can be difficult when the data is noisy or incomplete, and the variables for which data is available may only be able to provide a partial explanation of the causes of variability. Functions are either prescribed or synthesized. The functions prescribed by mechanistic models are physical equations, which incorporate tunable coefficients that are adjusted by modelers to match calibration data. Linear regression is the most common empirical modeling technique. It prescribes straight lines, planes, or hyper-planes to fit calibration data. The insurmountable problem with prescriptive modeling techniques is that if their functions are inherently unable to fit the variable relationships that are manifested in the data, a representative model is unobtainable. In South Carolina, some mechanistic models that have cost millions of dollars and years of effort to develop were never accepted by the regulatory agencies and stakeholders.

According to Conrads and Roehl [2005], calibrating mechanistic estuary models is *"particularly difficult due to low watershed gradients, poorly defined drainage areas, tidal complexities, and a lack of understanding of watershed and marsh processes."* Artificial neural networks (ANN) are a machine learning technique from AI. Rather than prescribe functions, ANNs synthesize non-linear functions to fit multivariate data. Conrads and Roehl [1999] found that ANN models had prediction errors that were significantly lower than those of a state-of-the-practice mechanistic model when predicting water temperature (WT), specific conductance (SC), and DO on Charleston's Cooper River estuary. Other benefits included shorter development time, fast execution that lets ANN to be coupled to numerical optimizers and embedded in spreadsheets, and integrating (ANN) with mechanistic models to improve predictions of how non-point source loading from rainfall and tidal marsh fluxing affect dissolved oxygen (DO) concentrations.

3. DSS FEATURES

All three of the DSS were developed as Microsoft Excel™/Visual Basic for Applications (VBA) programs. This allowed the DSS to be prototyped, easily modified, and distributed in a familiar form. The DSS are operated through a point-and-click graphical user interface (GUI) that requires no typing. This makes the DSS easy to use and eliminates the need to trap user errors. The GUI also provides graphical outputs that depict measured and predicted hydrologic behaviors. Other common elements of the DSS are described below.

Data: predictive models were developed to represent complex non-linear dynamic behaviors manifested in years of time series. Spectral filtering was applied to decompose the hydrodynamic, water quality, and meteorological signals into components that differentiate *periodic* and *chaotic* behaviors. Moving window averages (MWA) of varying window sizes are applied to augment these components with calculated variables that represent behaviors evolving on different time scales, for example, it takes months of data to represent a extended drought condition.

Modeling and Simulation: ANN *sub-models* are used to systematically decorrelate input variables and predict individual signal components of parameters of interest. The sub-models are then assembled into a *super-model* that predicts all of the parameters of interest throughout an entire system, customized to its unique circumstances and data. This "divide and conquer" approach allowed the statistical properties and behaviors of sub-models to be evaluated during model and DSS development by the various stakeholder's process, modeling, and regulatory specialists, making technology transfer and model validation an on-going, collective activity.

Each DSS has at least two instantiations of the super-model. One generates predictions using actual historical input conditions, which are used to compute prediction errors and graphically depict accuracy. The second instantiation generates "*What if?"* predictions using user-set *controllable* inputs. Two of the applications provide optimizers that modulate controllable inputs during simulations to obtain predictions that match user-set setpoints. A single simulation with optimization can replace numerous runs with fixed inputs. Each DSS incorporates a database of measured, filtered, 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 instantiations, post-processes and writes model output, and drives graphics.

4. BEAUFORT WATER QUALITY

The Beaufort River is a complex estuarine system that supports a variety of uses including fisheries, shipping, and the receiving of wastewater effluent. According to South Carolina Department of Health and Environmental Control [1998], the river was on the Section 303(d) list of impaired waters for low DO. The Clean Water Act stipulates that a Total Maximum Daily Load (TMDL) must be determined, so a data collection and modeling project was launched to support the permitting of two existing facilities and a new municipal WTTP to be constructed by the Beaufort-Jasper Water and Sewer Authority.

Data: a network of seven real-time gaging stations was operated by the U.S. Geological Survey (USGS) on the Beaufort River and its tributaries. Water level (WL), WT, SC, and DO were measured at 15-minute intervals for thirty-four months. The *DO-difference from saturation* (DOD) was calculated per U.S. Geological Survey [1981] to extract the component of DO variability that was unrelated to gas-in-liquid solubility. Three acoustic velocity meters were used to compute tidal streamflow. Daily rainfall was measured at two locations. Biochemical oxygen demand (BOD) and ammonia (NH_3) loads were generally measured only once per week at each WWTP, and not concurrently plant-to-plant. The low sample frequency of the BOD and NH_3 loads dictated a 1-day time step for the model.

Modeling and Simulation: the Beaufort super-model was composed of 118 separate ANN sub-models. DOD at each gage was modeled using inputs representing WL, SC, WT, rainfall, BOD and NH_3. Conrads et al [2003] detail how cascaded sub-models were used to decorrelate input variables and predict dynamic point and non-point source load responses. A cubic-spline was used to predict DOD at river locations between the gages. Bathymetric data was used to construct a geometric model having 90x90m cells. A medial axis transform was fitted to a 2D plan view of the waterways to provide the lengthwise spatial coordinate. The DSS' simulation database contained 1,035 daily records (34 months).

Special Features: rather than heuristically guided decision-making, a constrained optimizer was configured to represent South Carolina state law that governed the maximum allowable impact that nutrient loads from the three WWTPs could have on riverine DO. Water-resource regulators evaluate receiving waters for seasonally different impact limits, and segment rivers for volume-averaging impacts. Users can allocate the TMDL load among the six BOD and NH_3 discharges. At each time step the optimizer iterates load inputs as assimilative capacity changes. The GUI provides controls for exploring different load and segmentation scenarios. It was found that the overall TMDL was very sensitive to these parameters. The DSS also allows the impact of rainfall as a percentage of actual to be evaluated. It was found that historically the impacts of rainfall and the WWTP loads have been similar.

Status: In terms of acceptance by stakeholders, the Beaufort DSS was particularly successful when compared to similar coastal initiatives in South Carolina that used state-of-the-practice mechanistic models. Permits were issued only 26 months after model development began, as compared to 10 or more years for similar modeling projects in Myrtle Beach and Charleston. This was partly due to demonstrably better prediction accuracy, a modeling process that continuously engaged stakeholders, and DSS packaging that directly addressed the permitting problem.

4. SAVANNAH HARBOR DEEPENING

The Savannah Harbor is one of the busiest ports on the U.S. East Coast. It is located just downstream of the Savannah National Wildlife Refuge (SNWR), an important freshwater marsh. Under sponsorship from the U.S. Army Corps of Engineers and the Georgia Ports Authority (GPA), the Lower Savannah River estuary has been studied for years to evaluate the potential impacts of a proposed harbor deepening. Many databases have been created that describe the natural system's complexity and behaviors. A *three-dimensional finite-element hydrodynamic model* (3DM) is being developed to predict changes in riverine WL and salinity (S) in response to harbor geometry changes. A *marsh succession model* (MSM) is also being developed to predict how plant distributions in the marshes would respond to WL and S changes. This created a need for a third model, the *model to marsh* (M2M), to link river and marsh WL and S behaviors. There was an additional need for a DSS to integrate all of the models and data for stakeholders.

Data: Figure 1 shows the extensive network of real-time gages operated for the Savannah study. The

WL and SC data included: 11½ years from five USGS gages in the harbor and river; 4½ years from seven USGS marsh gages; three months from 14 riverine backwater gages operated on behalf of the Georgia Ports Authority (GPA) in 1997 and in 1999; and 19 months from 10 GPA marsh gages. 11½ years of flow (Q) from an upstream river gaging station was also obtained. The resulting database was composed of 11½ years of half-hourly data (200,000+ time stamps) for 110 measured variables. Further processing extracted chaotic signal components and calculated the tidal range and various MWA.

Modeling and Simulation: The M2M super-model comprises 127 sub-models. Chaotic sub-models predicted chaotic WL and SC at four USGS gages in the main channels using inputs for Q and harbor WL. These outputs were input to *high frequency* (HF) sub-models that also used HF harbor WL inputs to obtain HF WL and SC predictions at the four gages. The chaotic predictions in the main channel were input to sub-models for the remaining riverine and marsh gages. This provided one set of ANNs that linked the river's main channel behaviors to tidal forcing and fresh water flows, and a second set that linked main channel behaviors to those in backwaters and the marsh.

The Savannah DSS provides for simulations of up to 11½ years at daily, hourly, or half-hourly time steps. Q can be set by the user to be a constant or a percent of the historical flow. User-defined hydrographs can also be run.

Special Features: the 3DM is a complicated program, limiting its accessibility. However, the impacts of different harbor change scenarios can be evaluated using a file generated by the 3DM and imported into the DSS. The file contains WL and SC biases that are calculated by subtracting 3DM predictions representing proposed channel geometries from predictions generated using historical conditions.

A custom post-processor imports simulation output and interpolates predictions at gaged sites to generate a 2D contour map of S on a grid of the study area. The interpolation is performed using heuristic hydrology rules written for each grid cell that accommodate the area's topological features and the different transport mechanisms of channels and marshes. The post-processor provides options for time-averaging the predictions, and writes interpolated values to an output file that can be imported into the MSM.

Status: the Savannah DSS was first prototyped in 2002 and a production version was delivered in 2004. Delays in the completion of the 3DM and MSM have postponed its widespread deployment. Most of the original marsh data was collected during a record setting 4½-year drought between 1998 and 2002; therefore, the M2M's ANN were recently retrained with an additional 2½ years of non-drought data.

5. RELICENSING PEE DEE DAMS

Six reservoirs in North Carolina (NC) discharge into the Pee Dee River, which flows 160 miles through South Carolina to the coastal communities near Myrtle Beach. During the drought between 1998 and 2002, salinity intrusions inundated a coastal municipal freshwater intake near Myrtle Beach, South Carolina. The NC reservoirs are currently being re-licensed by the Federal Energy Regulatory Commission (FERC) for a 50-year operating permit. The water has significant commercial value for generating electric power and for waterfront property development. A coalition composed of Alcoa Power, Progress Energy, the Pee Dee River Coalition, and the South Carolina Department of Natural Resources sought to model the system's hydrodynamics and determine the minimum flows needed to protect coastal intakes.

Data: nine USGS gaging sites provided the WL and SC data used in the study. The data spanned 17½ years, but not all of the gages were in operation at the same time, and data quality improved with time. Inflows were obtained from an additional seven USGS gages, and rainfall was obtained from six regional meteorological stations. Coastal wind speed and direction were obtained from another meteorological station. The resulting database comprises 17½ years of hourly data (150,000+ time stamps) for 27 measured variables. Further processing extracted chaotic signal components and calculated the tidal range and various MWA.

Modeling and Simulation: The Pee Dee super-model is similar to that of the Savannah DSS; however, only SC is predicted. It employs 18 sub-models, a chaotic and HF sub-model pair for each gage. Tidal forcing was input from the easternmost gage along the Atlantic Intracoastal Waterway, which was found to be largely unaffected by river flows. The controllable input to the model is the flow from the most downstream dam (Q_d) on the Pee Dee River, which is summed with the other measured flows with adjustments made for transport delays. Q_d is generally much larger than the other combined flows in the Pee Dee basin. It was found that rainfall is well accounted for in the inflows, and that wind speed and direction are influential at the southernmost gages.

The Pee Dee DSS provides for simulations corresponding to the most recent and higher quality 6½ years of data, at daily or hourly time steps. Q_d can be set by the user to be a constant or a percent of the historical measurements. User-defined hydrographs can also be run.

Special Features: the Pee Dee DSS also provides a constrained optimizer that automatically modulates Q_d to match user-set maximum-SC setpoints. The setpoints can be applied on a daily or hourly basis. Higher Q_d is required to suppress hourly SC intrusions. The Pee Dee DSS also provides built-in documentation that describes the variables and user controls. It appears in pop-ups as the mouse is moved in the GUI.

Status: a number of technical review sessions were held where data and model issues were detailed, and successive prototypes were distributed to stakeholders. Feedback from the sessions dictated the DSS's final form, which was completed in 2005.

6. CONCLUSIONS

DSS provide a means to make arcane databases and models more accessible to all stakeholders for informed decision-making. Important features that the DSS have in common include:

- *Predictive Models* - that reliably predict relevant behaviors.
- *Databases* – that contain data describing important historical behaviors and provide a baseline for evaluating proposed changes.
- *Simulation* – programmatically time-step models to generate output representing input scenarios.
- *GUIs* – that conceptually unite the DSS components with intuitive user controls and graphical output.

Features that are more specialized include:

- *Constrained Optimization* - greatly reduces the number of simulations needed to answer a question. The Beaufort DSS optimizer computes the TMDL as assimilative capacity changes. The Pee Dee DSS optimizer computes the minimum dam flows needed to prevent salinity intrusions as rainfall and tidal conditions vary.
- *Tool Integration* – the Savannah DSS integrates the 3DM and MSM, allowing alternative harbor deepening scenarios to be evaluated.
- *Expert Knowledge* – the Beaufort DSS implements South Carolina environmental law in the form of optimization constraints. The Savannah DSS spatially interpolates a limited number of S predictions using heuristic hydrology rules.

7. REFERENCES

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 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.

Conrads, P.A., and E.A. Roehl, Integration of data mining techniques with mechanistic models to determine the impacts of non-point source loading on dissolved oxygen in tidal waters, South Carolina Environmental Conference, Myrtle Beach, March 2005.

Dutta, S., Wierenga, B., Dalebout, A., Case-based reasoning systems: from automation to decision-aiding and stimulation," *IEEE Transactions on Knowledge and Data Engineering*, 9(6), 911-922, 1997.

South Carolina Department of Health and Environmental Control, State of South Carolina 303d List for 1998 – Priority Ranked Waterbodies Targeted for Water Quality Management Action, Bureau of Water, EPA Approved June 1998

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

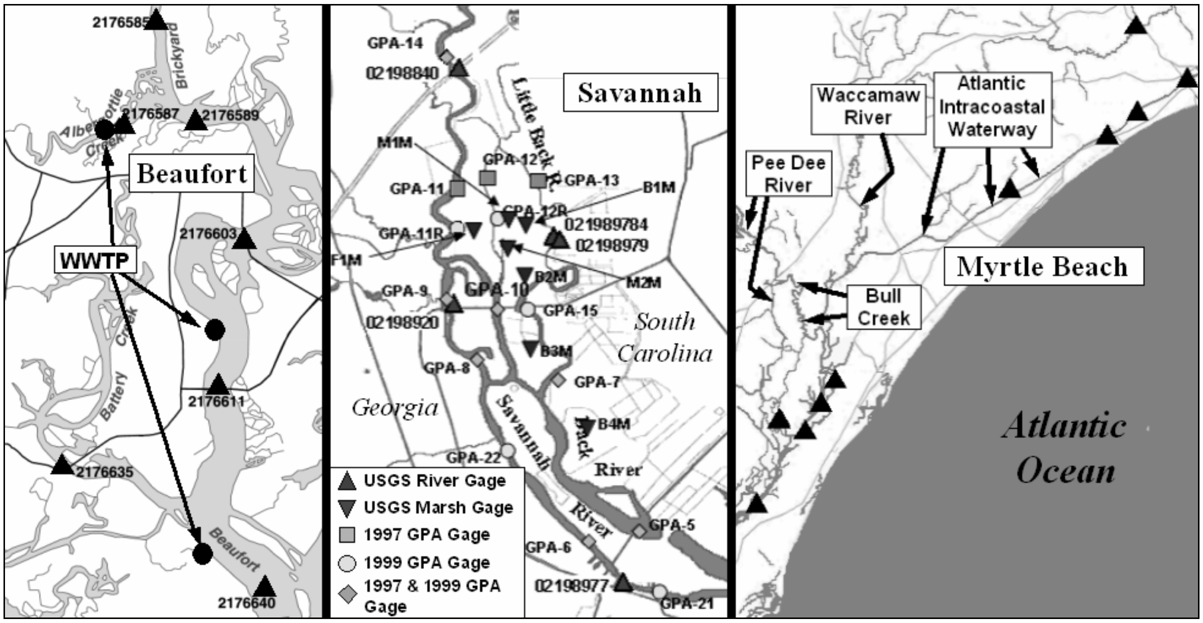

Figure 1. Beaufort, Savannah, and Myrtle Beach study areas. Markers denote gaging sites or wastewater treatment plant (WTTP) outfalls.

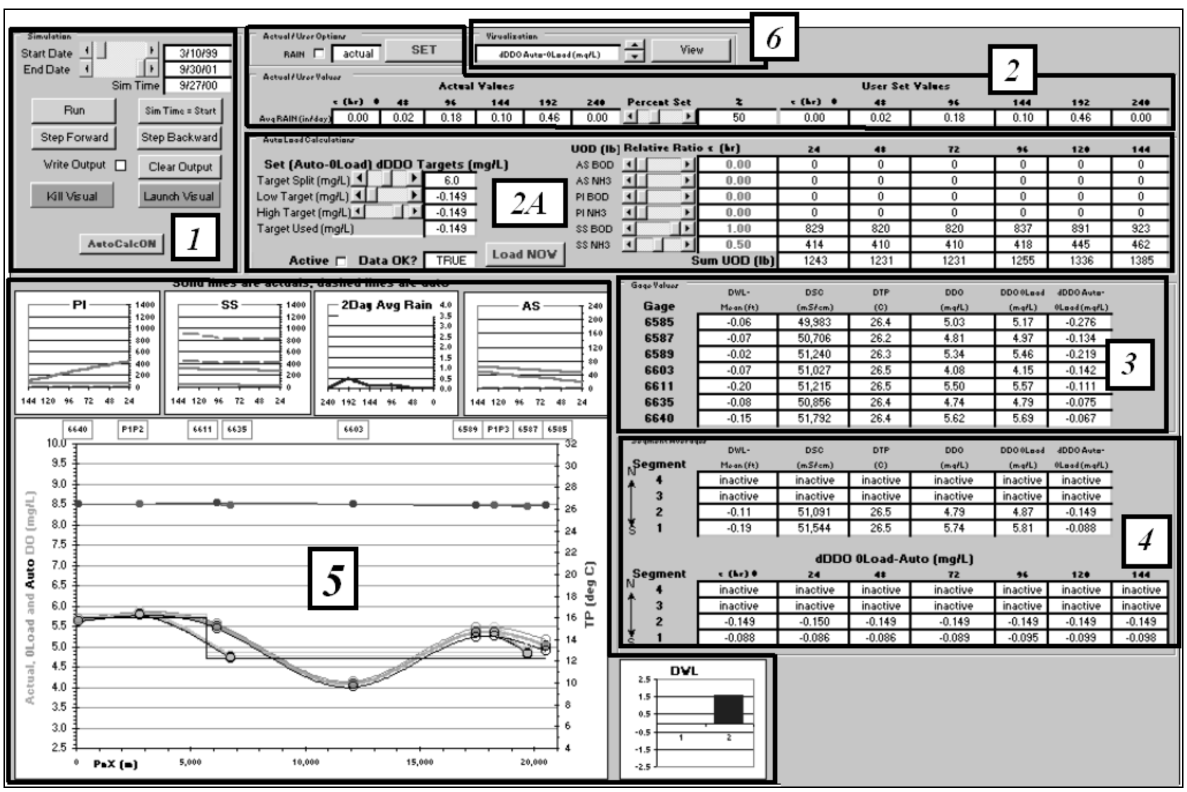

Figure 2. GUI Control Panel from Beaufort DSS - (1) simulation start/end/step; (2) manual loading - constant or %actual, streams left to right; (2A) "auto-loading" (optimizer) settings with streaming; (3) measured WL, WT, SC, DO, and DO predictions; (4) volume-averaged WL, WT, SC, DO, and DO predictions; (5) spatially interpolated measured and predicted WT and DO; and (6) color gradient visualization options (not shown).