Integrating 3D Hydrodynamic Transport and Ecological Plant Models of the Savannah River Estuary Using Artificial Neural Network Models

Edwin Roehl, Advanced Data Mining, Greer, SC, ed.roehl@advdatamining.com
Paul A. Conrads, U.S. Geological Survey, Columbia, SC, pconrads@usgs.gov
Ruby C. Daamen, Advanced Data Mining, Greer, SC, ruby@advdatamining.com
Wiley M. Kitchens, Florida Coop Unit, Gainesville, FL, kitchensw@wec.ufl.edu

Abstract: The Savannah Harbor is one of the busiest ports on the East Coast of the USA and is located just downstream of the Savannah National Wildlife Refuge (SNWR), which is one of the nation's largest freshwater tidal marshes. The Lower Savannah River estuary has been studied for years by governmental agencies, water users, universities, and consultants having an interest in controlling water quality and predicting the potential impacts of a proposed harbor deepening. Consequently, many different databases have been created that describe the natural system's complexity and behaviors. Variables having particular relevance include those describing bathymetry, meteorology, water level, and specific conductance. To evaluate the environmental impacts of the deepening, a three-dimensional hydrodynamic model (3DM) and a "marsh succession model" (MSM) were developed by different scientific teams. The 3DM predicts changes in riverine water levels and salinity in the system in response to potential harbor geometry changes. The MSM predicts plant distribution in the tidal marshes in response to changes in the water-level and salinity conditions in the marsh. To link the riverine predictions of the 3DM to the MSM, a "model to marsh" (M2M) model was developed using data mining techniques that included artificial neural networks (ANN). The ANNs simulated riverine and marsh water levels and salinity in the vicinity of the SNWR for the full range of 11½ years of data from riverine and marsh gaging networks. The 3DM, MSM, and M2M were integrated in a decision support system to make it easy for various regulatory and scientific stakeholders to use.

Keywords: hydrodynamic, plant, ecological, model, estuary, marsh, neural network, decision support

1. INTRODUCTION

Under sponsorship from the U.S. Army Corps of Engineers (USCOE) and the Georgia Ports Authority (GPA), the Lower Savannah River estuary and the surrounding freshwater tidal marshes of the Savannah National Wildlife Refuge (SNWR) have been studied for years by a variety of governmental agencies, water users, universities, and consultants. Their interests are in controlling water quality and predicting the potential impacts on the estuary and tidal wetlands of a proposed harbor deepening. Two major initiatives were the development of a three-dimensional hydrodynamic model (3DM) by a team of hydrologists, and the development of a marsh succession model (MSM) by a team of plant ecologists. The 3DM predicted changes in riverine water levels and salinity in the system in response to potential harbor geometry changes. The MSM predicts plant distribution in the tidal marshes in response to changes in the water-level and salinity conditions in the marsh. A mechanism for linking riverine and marsh behaviors was needed.

To support 3DM and MSM development, many disparate databases had been created that described the natural system's complexity and behaviors, but they had not been compiled into a usable form. Variables having particular relevance include those describing bathymetry, meteorology, water level (WL), specific conductance (SC), water temperature (WT), and dissolved oxygen concentration (DO). Most of the databases were comprised of time series that varied by variable type, periods of record, measurement frequency, location, and reliability. It was recognized that data mining techniques, which include artificial neural networks (ANN), could be used to link riverine and marsh behaviors.

The authors had previously developed ANN-based models of estuaries in Charleston and Beaufort South Carolina. The type of ANN used was the multi-layered perceptron (MLP) described by 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 WT, SC, and DO on Charleston's Cooper River. Conrads et al [2002] went on to use ANNs to estimate the impacts of nutrient loading from rainfall runoff and tidal marsh inundation on DO in the same waterway. In a regulatory application, Conrads et al [2003] describe an ANN-based model for the permitting of three wastewater treatment plants that discharge into the Beaufort River estuary. In terms of acceptance by stakeholders, the Beaufort model was particularly successful when compared to other similar initiatives in South Carolina that used state-of-the-practice mechanistic models. Permits were issued only 26 months after the program's development began, as compared to 10 or more years for similar modeling projects in Myrtle Beach and Charleston. This was due, in part, to demonstrably better prediction accuracy, and packaging of the model and databases as a decision support system (DSS) that made it easy for decision makers to use directly.

2. MODELING

The modeling approach used to link Savannah riverine and marsh behaviors, is similar to one used previously for the Beaufort River DSS. The Beaufort model incorporated 118 separate ANNs *sub-models* that predicted both point and non-point source impacts on water quality throughout the natural system. Sub-models were used for different purposes: decorrelating input variables, which is an endemic problem in tidally forced systems where all hydrodynamic and water quality variable tend to move together; predicting point source impacts at each of seven real-time gages over several time delays; and predicting non-point source impacts at each gaging site. Sub-models were cascaded together to assemble a complete prediction for each gaging site. The completed application constituted a *super-model* comprised of sub-models. When combined with the multivariate, non-linear regression capability of ANNs, this divide-and-conquer problem solving approach produces models that optimize the use of all of the available data.

Linking the riverine predictions of the 3DM to the MSM required that another model be developed, called the M2M for "model-to-marsh". The M2M needed to simulate riverine and marsh water levels and salinity in the vicinity of the SNWR for the full range of historical conditions using data from the riverine and marsh gaging networks.

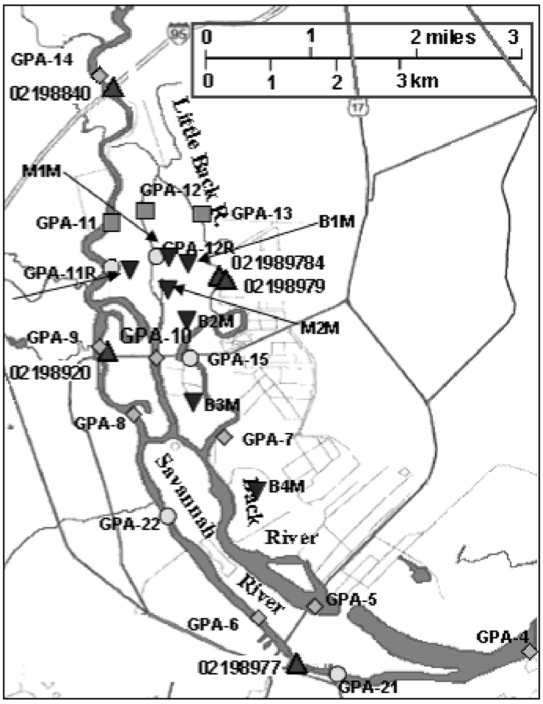

Figure 1. Study area showing some of the 56 gaging sites. Symbols designate gage operators and period of data collection.

2.1. Historical Databases

The study area is shown in Figure 1. The available data required extensive clean up for problems such as erroneous and missing values and phase shifts. The resulting database was comprised of 11½ years of half-hourly data (200,000+ time stamps) for 110 variables. The original sources of data were:

- Q_{Clyo} and WL_{Harbor} – 11½ years of half-hourly WL signals in Savannah Harbor and river flows measured 50 miles inland at Clyo by the U.S. Geological Survey (USGS).

- USGS riverine WL and SC – 11½ years of half-hourly signals collected from four stations in the Lower Savannah River by the USGS.
- GPA riverine WL and SC - half-hourly signals collected on behalf of the GPA from 14 stations over three months each in 1997 and 1999. Some stations recorded both surface and bottom SC measurements (SC_{top}, SC_{bottom}).
- USGS marsh WL and SC – 4½ years of half-hourly signals from 7 stations.
- GPA marsh WL and SC – 19 months of half hourly SC and WL from 10 stations.

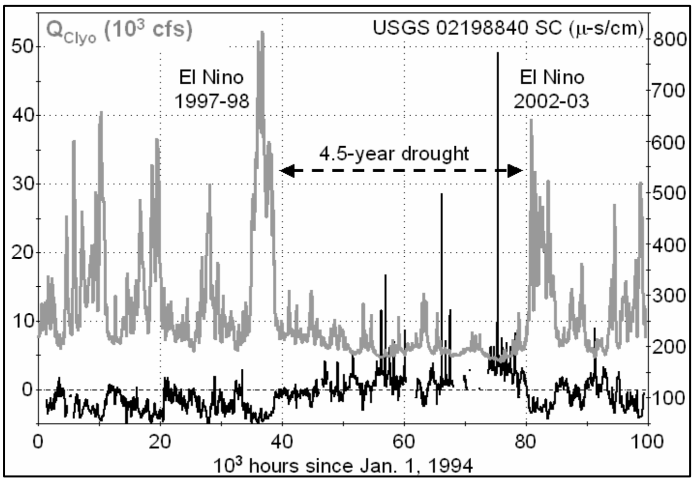

Figure 2. 11½ years of hourly Q_{Clyo} and SC at USGS 02198840, which was the farthest inland riverine gage. The SC spikes at center right occur at 28-day intervals, and are coincident peaking of the tidal range during the lowest flows of the drought.

Much of the field data was collected during a record setting 4½-year drought, raising concerns that it was not representative of "normal" hydrodynamic conditions. Figure 2 shows that the record low river flows during the drought led to unprecedented salinity intrusions far inland, even without a deepened harbor. It was expected that the ANNs could reasonably extrapolate from the field data by "learning" the full range of behaviors exhibited over 11½ years, which also included two El Nino events when flows were significantly above average, and presumably periods of normal conditions.

2.2. Signal Decomposition

The hydrodynamic and water quality behaviors observed in estuaries are superpositions of behaviors forced by periodic planetary motions and chaotic meteorological disturbances. Theoretically, periodic behaviors are perfectly predictable, and chaotic behaviors are only somewhat so; therefore, the real problem with modeling estuaries is to empirically synthesize chaotic output signals from multiple chaotic input signals. Signals are easily decomposed into periodic and chaotic components using spectral filtering. The primary chaotic inputs to the Lower Savannah River are the flows released from the dam at Clyo and the chaotic oceanic disturbances represented in the chaotic component of WL_{Harbor}.

The empirical representations of the dynamical behaviors that underlie periodic and chaotic signals are different. Multiply periodic signals are superpositions of individual periodic signals that are represented by three constants - phase, amplitude, and frequency. Abarbanel [1996] describes how chaotic univariate systems can be optimally represented by *dynamical invariants* - characteristic *time delays* and *dimensions*. Roehl and Murray [2005] describe an ANN model that forecasted a utility's water user demand, which incorporated signal decomposition and extended the univariate representation of chaotic behaviors to a multivariate system.

As shown in Figure 3, chaotic components were extracted from raw signals by applying a low pass spectral filter to remove high frequency (HF) diurnal and semi-diurnal variability. The important, multiply periodic tidal range XWL was computed from WL_{Harbor}. The chaotic component of Q_{Clyo} was further processed with moving window averages (MWA) of up to two weeks, so that when input to an ANN with multiple time delays, flow histories of up to 44 days were represented.

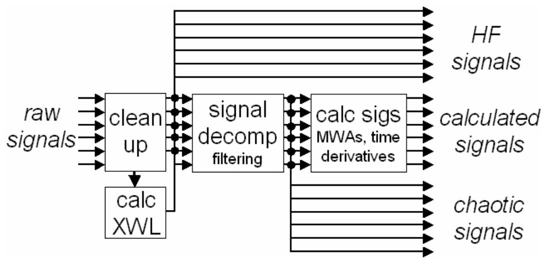

Figure 3. Signal processing and decomposition.

3. M2M

The M2M super-model was comprised of 127 sub-models. Figure 4 shows that cascading sub-models predicted chaotic WL and SC signal components at riverine and marsh gaging sites. Using low pass filtered Q_{Clyo}, WL_{Harbor}, and XWL signal components for inputs, "chaotic sub-models" predicted chaotic WL and SC behaviors at four USGS gages in the main channel. These outputs were input to "HF sub-models" that also used HF WL_{Harbor} and XWL component inputs to obtain HF WL and SC predictions at the four gages.

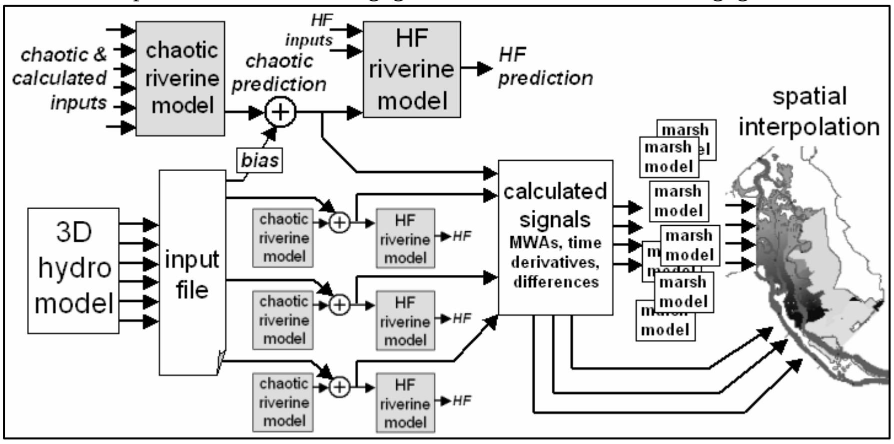

Figure 4. Data flow through the super model. Separate sub-models were used for each WL and SC prediction.

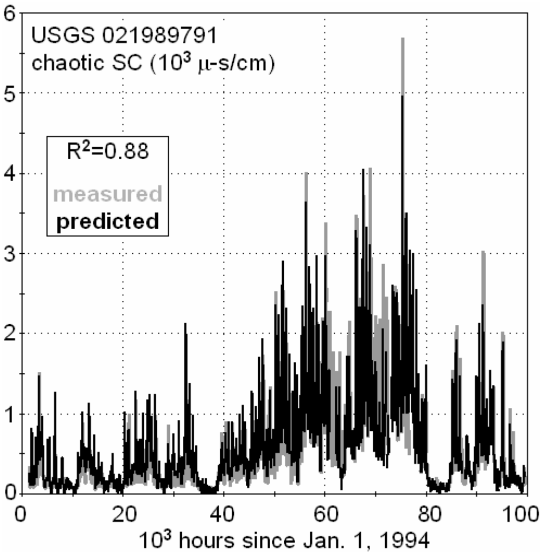

Figure 5. Measured and predicted chaotic riverine SC. Increased SC at center right occurred during the drought.

The chaotic predictions at the main channel gages were then transformed into calculated signals to decorrelate them and to represent dynamical behaviors that evolve over weeks. The calculated signals were used as inputs to model the historically shorter signals at the many remaining riverine and marsh stations. 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. Figures 5, 6, and 7 show SC predictions at a riverine gage and a nearby marsh gage. R^2 of the SC predictions at most of the gages were between 0.8 and 0.9. The R^2 of the WL predictions were generally above 0.9.

3D response surfaces can be used to visualize the multivariate interactions as learned by ANNs. A surface is generated by selecting and stepping two inputs across their historical ranges, while "unshown" inputs are set to values of interest, e.g., minimums, maximums, or means. Figures 8 and 9 show surfaces representing the behaviors at a riverine gage and a nearby marsh gage. While the behavior at the riverine gage is highly non-linear with respect to freshwater flows and tides, the marsh response to the riverine SC is relatively linear. This suggests the reasonableness of using ANNs trained with river and marsh data collected only during the drought, but driven by riverine predictions from ANNs trained over widely ranging conditions, to extrapolate to non-drought conditions.

3.1. Simulation and Decision Support

The execution of the large number of Savannah area sub-models was orchestrated by a custom decision support system (DSS). The DSS integrates the super-model with an 11½-year database, comprising more that 200k records of half-hourly measurements, for running long-term simulations. It also provides a graphical user interface, streaming graphics, several freshwater flow input options, and output file generation to allow stakeholders of varying technical backgrounds to evaluate alternative scenarios under the widely ranging conditions manifest in such a long historical record.

3.2. 3DM Integration

Figure 3 shows that the 3DM is linked to the M2M super-model through an output file. The file contains WL and SC biases for the main gaging sites. The biases are calculated by subtracting 3DM predictions representing proposed channel geometries from predictions generated using the actual historical conditions.

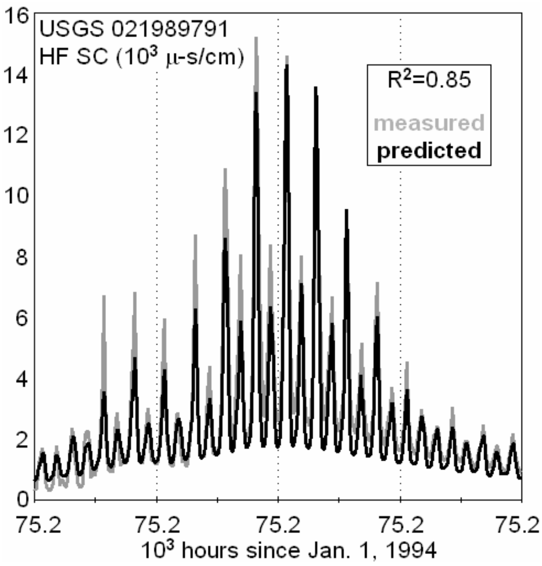

Figure 6. Measured and predicted HF riverine SC. 16.6 days are shown during the drought.

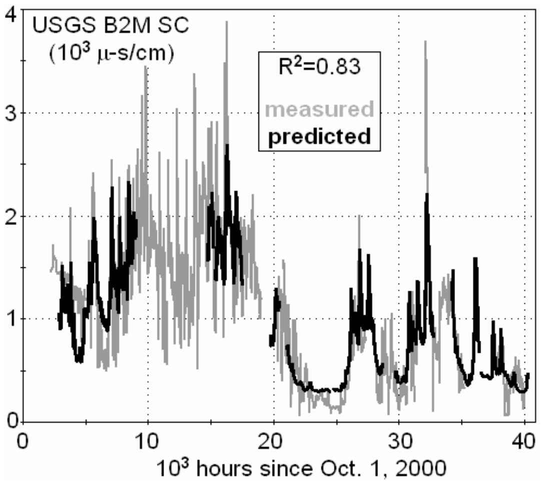

Figure 7. Measured and predicted marsh SC. Gaps mark missing input data. Marsh parameters are very difficult to monitor for extended periods because of the physical instability of gaging sites.

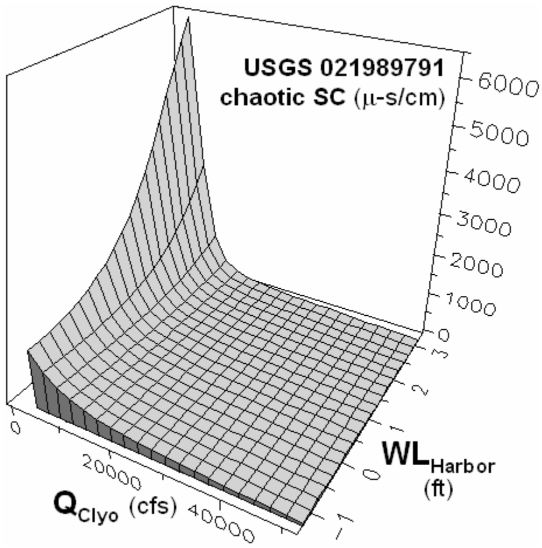

Figure 8. 3D response surface generated with a chaotic model of SC. The spikes in Figure 5 occur at low Q_{Clyo} and high WL_{Harbor}.

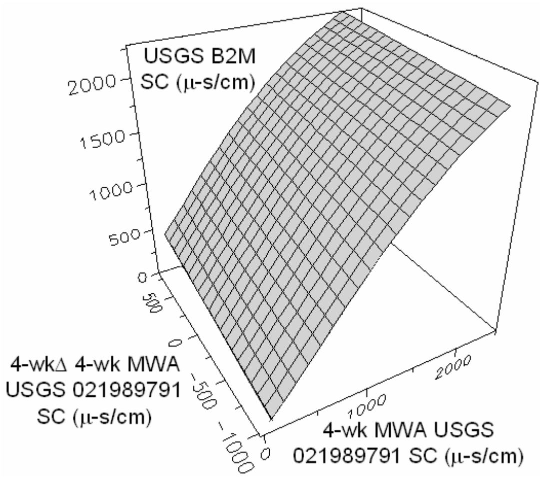

Figure 9. 3D response surface generated with a model of marsh SC at USGS B2M. The response at B2M to long-term (4-week MWA) SC at nearby riverine 021989791 is nearly linear. Not surprisingly, marsh SC increases if riverine SC has been high for some time, as indicated by the 4-week change (Δ) in the 4-week MWA of the riverine SC.

3.4. MSM Integration

Figure 10 shows that riverine and marsh predictions at the gaged sites shown in Figure 1 are interpolated to generate a 2D contour map of SC on a grid of the study area. The interpolation is performed using rules written for each grid cell. The rules accommodate the area's topological features and the different transport mechanisms of channels and marshes. The interpolation and visualization are performed in a custom post-processor that imports output from the DSS and writes interpolated values to an output file. The post-processor converts SC's to salinities, and provides different options for time-averaging the predictions. Output from the post-processor can be imported into the MSM so that plant ecologists can evaluate the impacts of predicted salinity changes.

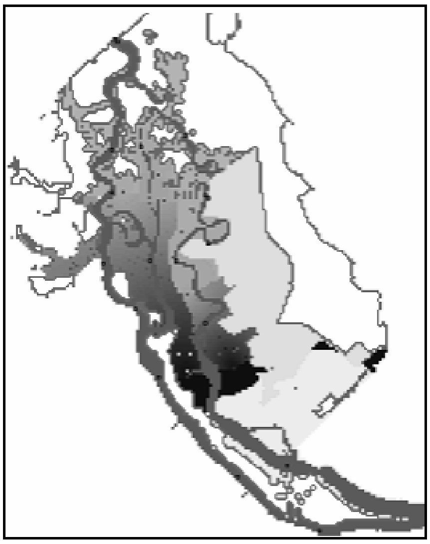

Figure 10. 2D contour map of SC on a grid of the study area.

4. CONCLUSIONS

Years of hydrodynamic and plant data, a traditional three-dimensional hydrodynamic model (3DM), and a marsh succession model (MSM) were generated by teams of scientists to understand and describe how the Savannah River and surrounding wetlands would be affected by a proposed harbor deepening. The total investment represented several million dollars. It was found to be necessary for yet another model to be created that would defensibly link the riverine predictions of the 3DM to the observed behaviors in the marshes. This link was provided by a non-traditional "model to marsh" (M2M) model, which was developed using data mining techniques, including signal decomposition and artificial neural networks (ANN). The M2M's 127 ANN sub-models were trained to accurately represent the chaotic and periodic interactions of river flows, tides, water-levels, and salinities manifest in 11½ years of time series collected from 56 gaging sites. The data included a record drought and two El Nino events, assuring the M2M's representativeness for a broad range of conditions. The M2M's packaging as a user-friendly decision support system has greatly facilitated its use by the stakeholder community.

5. REFERENCES

Abarbanel, H.D.I., *Analysis of Observed Chaotic Data*, Springer-Verlag New York, Inc., New York, 4-12, 1996.

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.B. 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.

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

Roehl, E.A. and T. Murray, Forecasting water user demand at Beaufort-Jasper Water and Sewer Authority, South Carolina Environmental Conference, Myrtle Beach, March 2005.