7th International Conference on Hydroinformatics
HIC 2006, Nice, FRANCE

INTEGRATING 3D HYDRODYNAMIC TRANSPORT AND ECOLOGICAL PLANT MODELS OF THE SAVANNAH RIVER ESTUARY USING ARTIFICIAL NEURAL NETWORK MODELS

RUBY C. DAAMEN
Advanced Data Mining, 3620 Pelham Road, PMB 351
Greenville, SC 29615, United States

EDWIN A. ROEHL
Advanced Data Mining, 3620 Pelham Road, PMB 351
Greenville, SC 29615, United States

PAUL A. CONRADS
U.S. Geological Survey, 720 Gracern Road, Suite 129
Columbia, SC 29210, United States

WILEY M. KITCHENS
U.S. Geological Survey, P.O Box 10485
Gainesville, FL 32611, United States

The Savannah Harbor is one of the busiest ports on the East Coast of the USA. The harbor is located downstream from 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 maintaining 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. A three-dimensional hydrodynamic model (3DM) and a "marsh succession model" (MSM) were developed by different scientific teams to evaluate the environmental impacts of the harbor deepening. 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 (DSS) for use by various regulatory and scientific stakeholders.

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 maintaining water quality and predicting the potential impacts of a proposed harbor deepening on the estuary and tidal wetlands. 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 changes. The MSM predicted 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 these databases had not been compiled into a usable form. Variables having particular relevance include those describing bathymetry, meteorology, discharge (Q), water level (WL), specific conductance (SC), water temperature (WT), and dissolved oxygen concentration (DO). Most of the databases were composed of time series that varied by variable type, periods of record, measurement frequency, location, and reliability. Scientists 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, USA. The type of ANN used in these cases was the multi-layered perceptron (MLP) described by Jensen [5], which is a multivariate, non-linear regression method based on machine learning. In a side-by-side comparison, Conrads and Roehl [2] found that ANN models had prediction errors 60-82% lower than those of a state-of-the-practice mechanistic model when predicting the effect of WT, SC, and DO on Charleston's Cooper River. In a regulatory application, Conrads and others [3] describe an ANN-based model for the permitting of three wastewater treatment plants that discharge into the Beaufort River estuary.. Permits were issued only 35 months after the program's development began, as compared to 10 or more years for similar modeling projects in Myrtle Beach and Charleston. The shortened time was due to demonstrably better prediction accuracy, and packaging of the model and databases as a decision support system (DSS), which made it easy for decision makers to use the models directly

MODELING

A modeling approach similar to the one developed for the Beaufort River study was used for the Savannah River Estuary. The Beaufort model incorporated 118 separate ANN *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 variables tend to move together; predicting point-source impacts at each of seven real-time stream 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* composed 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 available data.

For the Savannah River Estuary Study, linking the riverine predictions of the 3DM to the MSM required the development of another *super-model*, called the "model-to-marsh" or M2M model. 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. Similar to the Beaufort River super-model, cascading *sub-models* in the M2M are used for decorrelating variables, predicting river impacts, and predicting tidal marsh impacts.

Linking the 3DM to the MSM is accomplished by using predicted differences in WL and SC values for the river generated by the 3DM as input to the M2M. Using the predicted difference for the river, the M2M predicts the change in WL and SC in the tidal marshes. These predictions are then used by the MSM to predict changes in the plant communities in the tidal marshes.

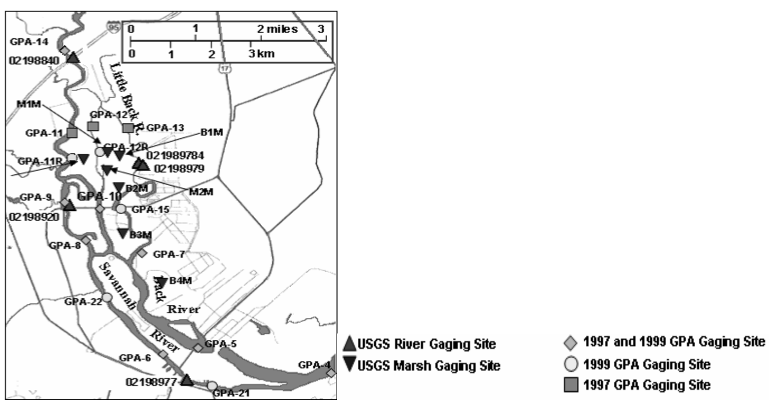


Figure 1. Gaging sites of the U.S. Geological Survey (USGS) and Georgia Ports Authority (GPA) in the Savannah River Estuary.

Historical Databases

The locations of the real-time gaging sites are 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 composed 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 3 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 collected from seven stations (2000-2005).
- GPA marsh WL and SC – 19 months of half hourly SC and WL data collected from 10 stations.

Much of the field data was collected during a record-setting 4½-year drought, raising concerns that the data was not representative of "normal" hydrodynamic conditions. Figure 2 shows that the record-low river flows during the drought led to unprecedented seawater 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 Niño events when flows were substantially above average, and presumably periods of normal conditions.

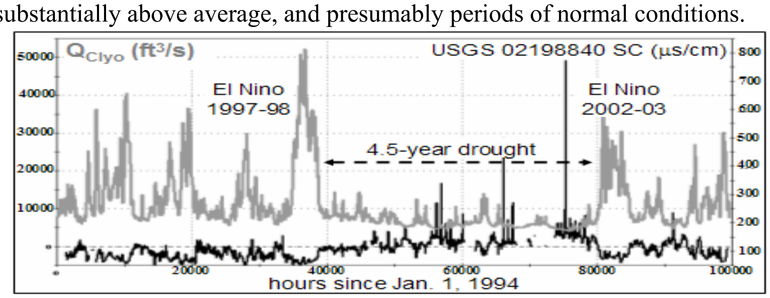


Figure 2. 11½ years of hourly Q_{Clyo} and SC at USGS 02198840, which was the farthest inland riverine gage. The SC spikes at center right occur at 28-day intervals, and are coincident with peaking of the tidal range during the lowest flows of the drought.

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 dynamic behaviors that underlie periodic and chaotic signals are different. Multiple periodic signals are superpositions of individual periodic signals that are represented by three constants — phase, amplitude, and frequency. Abarbanel [1] describes how chaotic univariate systems can be optimally represented by *dynamical invariants* — characteristic *time delays* and *dimensions*. Roehl and othersl [6] describe an ANN model that predicted the salt-front location in the Cooper River, 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 2 weeks, so that when input to an ANN with multiple time delays, flow histories of up to 44 days were represented.

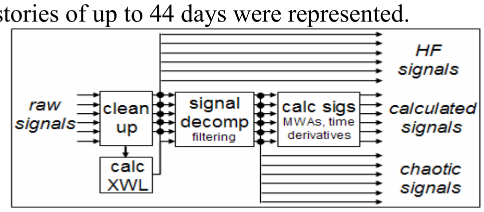


Figure 3. Signal processing and decomposition.

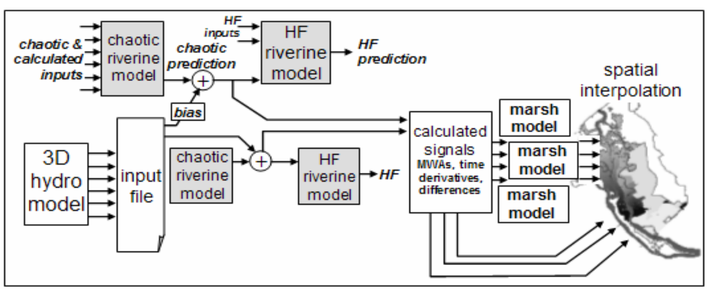


Figure 4. Data flow through the super-model decomposition. Separate sub-models were used for each WL and SC prediction.

M2M

The M2M super-model was composed 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 gaging sites 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 gaging sites.

The chaotic predictions at the main channel sites were then transformed into calculated signals to decorrelate them and to represent dynamic 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 freshwater 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 site and at a nearby marsh site. The R^2 of the SC predictions at most of the gaging sites reanged between 0.8 and 0.9. The R^2 of the WL predictions generally were greater than 0.9.

Roehl and others [7] describe the use of 3D response surfaces to visualize the functional forms of 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 that represent the behaviors at a riverine site and a nearby marsh site. While the behavior at the riverine gaging site is highly non-linear with respect to freshwater flows and tides, the marsh response to the riverine SC is relatively linear. This indicates the reasonableness of using ANNs trained with backwater 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.

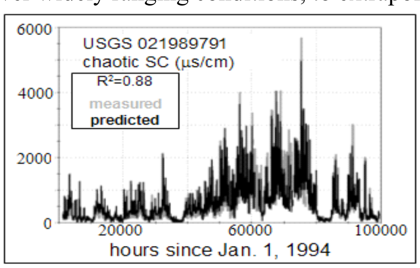


Figure 5. Measured and predicted chaotic riverine SC. Increased SC at center right occurred during the drought.

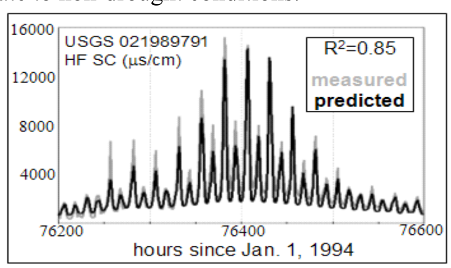


Figure 6. Measured and predicted HF riverine SC with 16.6 days shown during the drought.

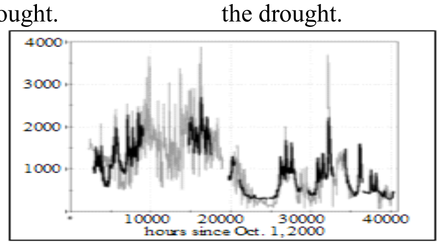
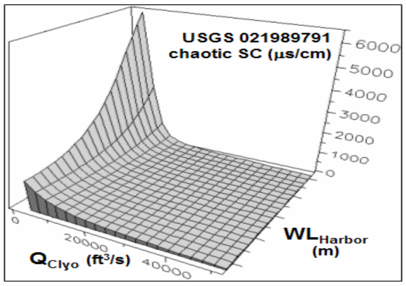


Figure 7. Measured and predicted marsh SC. Gaps mark missing input data. Marsh parameters are difficult to monitor for extended periods because of the physical instability of gaging sites.



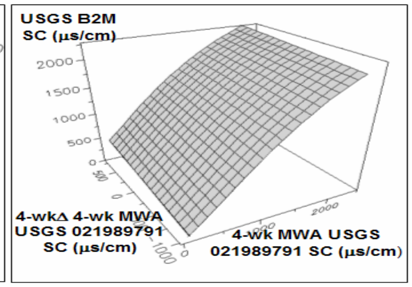


Figure 8. 3D response-surface generated with a chaotic model of SC. The spikes in Figure 5 occur at low Q_{Clyo} and high WL_{Harbor}.

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.

Simulation and Decision Support

Daamen and Roehl [4] describe how 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 200,000 records of half-hourly measurements, for running long-term simulations. The DSS 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 that manifest in such a long historical record.

3DM Integration

Figure 4 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.

MSM Integration

Figure 4 shows that riverine and marsh predictions at gaged sites 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 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 SCs 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.

CONCLUSIONS

The M2M leverages and integrates millions of dollars of field data collection and modeling performed over more than a decade by several scientific organizations. A divide-and-conquer super-model solution, enabled by signal decomposition and accurate ANN sub-models, allowed a large amount of disparate data and intermediate works to be optimally used in their entirety. The packaging of the super-model and data in a DSS makes the scientific products immediately accessible and useful to all stakeholders.

REFERENCES

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

[2] 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.

[3] 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.

[4] Daamen, R.C., and E.A. Roehl, Integrating multiple databases and estuary models into a comprehensive software tool for regulatory support, South Carolina Environmental Conference, Myrtle Beach, March 2005.

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

[6] Roehl, E.A., P.A. Conrads, and T.A. Roehl, Real-time control of the salt front in a complex, tidally affected river basin, Proceedings of the Artificial Neural Networks in Engineering Conference, St. Louis, 947-954, 2000

[7] Roehl, E.A., P.A. Conrads, and J.B. Cook., Discussion of using complex permittivity and artificial neural networks for contaminant prediction, *J. Environmental Engineering*, 1069-1071, November 2003.