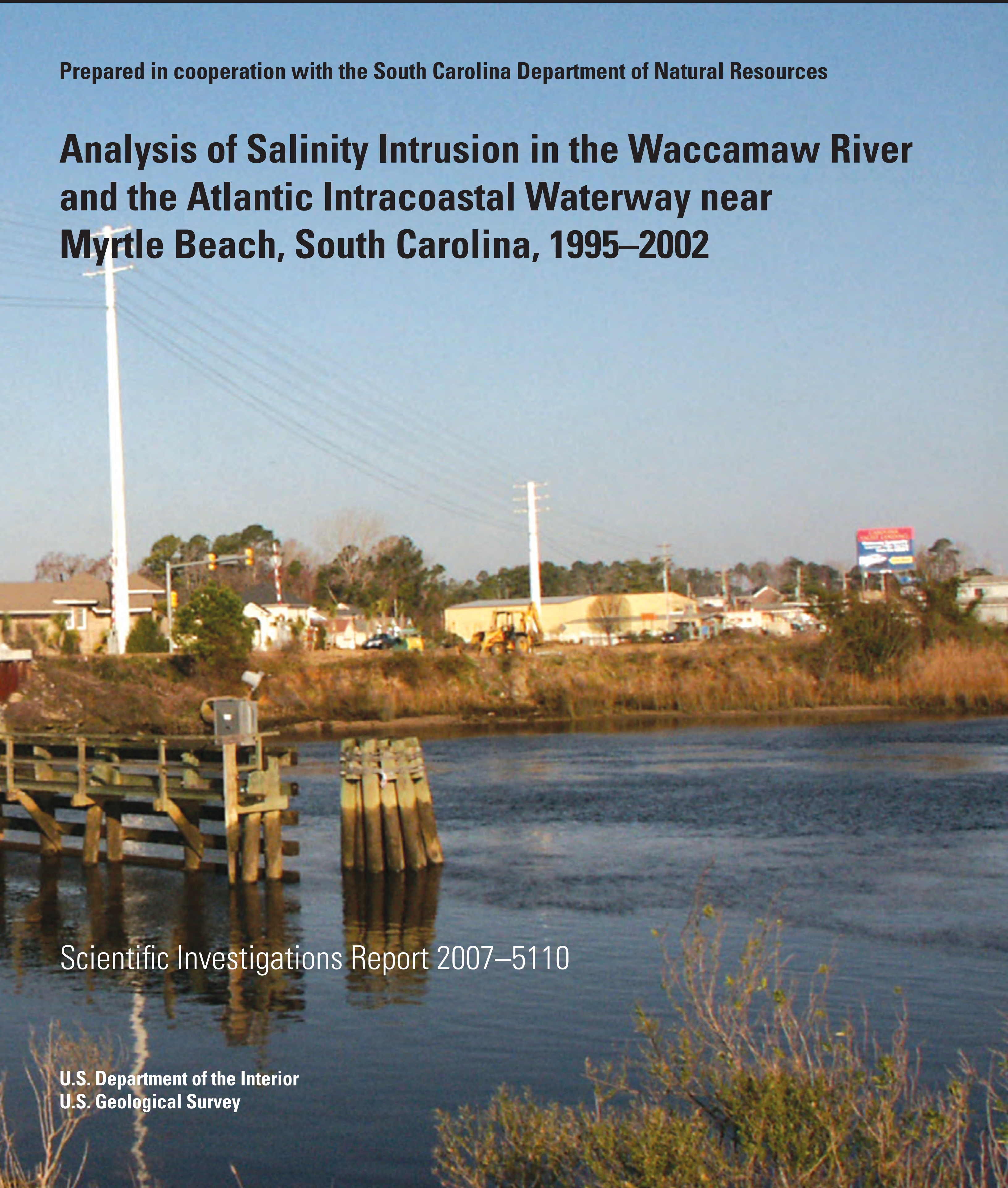
USGS
science for a changing world

Prepared in cooperation with the South Carolina Department of Natural Resources

Analysis of Salinity Intrusion in the Waccamaw River and the Atlantic Intracoastal Waterway near Myrtle Beach, South Carolina, 1995–2002



Scientific Investigations Report 2007–5110

U.S. Department of the Interior
U.S. Geological Survey

Cover. **Atlantic Intracoastal Waterway at S.C. Highway 9 at Nixons Crossroad, South Carolina (station 02110777)** *Photograph by John W. Erbland, U.S. Geological Survey, Conway Field Office*

Analysis of Salinity Intrusion in the Waccamaw River and Atlantic Intracoastal Waterway near Myrtle Beach, South Carolina, 1995–2002

By Paul A. Conrads and Edwin A. Roehl, Jr.

Prepared in cooperation with the South Carolina Department of Natural Resources

Scientific Investigations Report 2007–5110

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
Mark D. Myers, Director

U.S. Geological Survey, Reston, Virginia: 2007

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Suggested citation:
Conrads, P.A., and Roehl, E.A., Jr., 2007, Analysis of salinity intrusion in the Waccamaw River and Atlantic Intracoastal Waterway near Myrtle Beach, South Carolina, 1995–2002: U.S. Geological Survey Scientific Investigations Report 2007–5110, 41 p., 2 apps. (available online at http://pubs.water.usgs.gov/sir2007-5110)

Contents

Figures

Tables

Conversion Factors and Datums

Multiply	By	To obtain
	Length	
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
acre	0.004047	square kilometer (km^2)
square mile (mi^2)	2.590	square kilometer (km^2)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

Vertical coordinate information is referenced to the National Geodetic Vertical Datum of 1929 (NGVD 1929).

Horizontal coordinate (latitude and longitude) information is referenced to the North American Datum of 1983 (NAD 83).

Specific conductance is given in microsiemens per centimeter at 25 degrees Celsius (µS/cm at 25 °C).

Acronyms and Abbreviations Used in the Report

AI	artificial intelligence
AIW	Atlantic Intracoastal Waterway
ANN	artificial neural network
BEP	back error propagation
CRADA	Cooperative Research and Development Agreement
DO	dissolved oxygen
DSS	decision support system
FERC	Federal Energy Regulatory Commission
GUI	graphical user interface
MLP	multilayer perceptron
MSM	marsh succession model
M2M	model-to-marsh model
N.C.	North Carolina
OLS	ordinary least squares
PME	percent model error
PRISM	Pee Dee River and Atlantic Intracoastal Waterway Salinity Intrusion Model
psu	practical salinity units
Q	streamflow
RMSE	root mean square error
R^2	coefficient of determination
S.C.	South Carolina
SCDHEC	South Carolina Department of Health and Environmental Control
SCDNR	South Carolina Department of Natural Resources
SSR	state-space reconstruction
TMDL	total maximum daily load
USGS	U.S. Geological Survey
WL	water level
XWL	tide range
2D	two dimensional
3D	three dimensional
3DM	three-dimensional hydrodynamic model
3DVis	three-dimensional visualization

Analysis of Salinity Intrusion in the Waccamaw River and Atlantic Intracoastal Waterway near Myrtle Beach, South Carolina, 1995–2002

By Paul A. Conrads and Edwin A. Roehl, Jr.[1]

Abstract

Six reservoirs in North Carolina discharge into the Pee Dee River, which flows 160 miles through South Carolina to the coastal communities near Myrtle Beach, South Carolina. During the Southeast's record-breaking drought from 1998 to 2003, salinity intrusions inundated a coastal municipal freshwater intake, limiting water supplies. To evaluate the effects of regulated flows of the Pee Dee River on salinity intrusion in the Waccamaw River and Atlantic Intracoastal Waterway, the South Carolina Department of Natural Resources and a consortium of stakeholders entered into a cooperative agreement with the U.S. Geological Survey to apply data-mining techniques to the long-term time series to analyze and simulate salinity dynamics near the freshwater intakes along the Grand Strand of South Carolina. Salinity intrusion in tidal rivers results from the interaction of three principal forces—streamflow, mean tidal water levels, and tidal range. To analyze, model, and simulate hydrodynamic behaviors at critical coastal gages, data-mining techniques were applied to over 20 years of hourly streamflow, coastal water-quality, and water-level data. Artificial neural network models were trained to learn the variable interactions that cause salinity intrusions. Streamflow data from the 18,300-square-mile basin were input to the model as time-delayed variables and accumulated tributary inflows. Tidal inputs to the models were obtained by decomposing tidal water-level data into a "periodic" signal of tidal range and a "chaotic" signal of mean water levels. The artificial neural network models were able to convincingly reproduce historical behaviors and generate alternative scenarios of interest.

To make the models directly available to all stakeholders along the Pee Dee and Waccamaw Rivers and Atlantic Intracoastal Waterway, an easy-to-use decision support system (DSS) was developed as a spreadsheet application that integrates the historical database, artificial neural network models, model controls, streaming graphics, and model output. An additional feature is a built-in optimizer that dynamically calculates the amount of flow needed to suppress salinity intrusions as tidal ranges and water levels vary over days and months. This DSS greatly reduced the number of long-term simulations needed for stakeholders to determine the minimum flow required to adequately protect the freshwater intakes.

Introduction

The Pee Dee and Waccamaw Rivers and Atlantic Intracoastal Waterway (AIW), as with many major estuarine systems, meet many local and regional water-resources needs. The tidal portions of these systems provide water supply to the growing coastal communities along the Grand Strand of South Carolina (S.C.), provide assimilative capacity for municipal dischargers, and provide navigation along the AIW (fig. 1). With increases in industrial and residential development along the South Carolina coast and in the Pee Dee River basin, there are competing, and often conflicting, interests in the water resources of the Pee Dee River. As part of the Federal Energy Regulatory Commission (FERC) re-licensing of six reservoirs in North Carolina (N.C.), the ecological and hydrological effects of the controlled releases from the dams to the Pee Dee River are being evaluated, including effects on salinity intrusion along the Grand Strand of South Carolina.

The U.S. Geological Survey (USGS), in cooperation with the South Carolina Department of Natural Resources (SCDNR), the Pee Dee River Coalitions, Progress Energy, and Alcoa Power, initiated a study to: (1) develop empirical models to predict specific conductance at selected coastal gaging stations; (2) develop a spreadsheet application that integrates historical databases, empirical specific-conductance models, optimization routines, model controls, and streaming graphics; and (3) develop a three-dimensional visualization routine that will spatially extrapolate the model results over the salinity reaches of the system.

[1] Advanced Data Mining, LLC, Greenville, South Carolina.

Figure 1. Study area including the Pee Dee and Waccamaw River basins and Atlantic Intracoastal Waterway in South Carolina.

In order to meet the objects of this study (and previous studies), the USGS entered into a Cooperative Research and Development Agreement (CRADA) with Advanced Data Mining in 2002 to collaborate on applying data mining and artificial neural network (ANN) models to water-resources investigations. The emerging field of data mining addresses the issue of extracting information from large databases (Weiss and Indurkhya, 1998). Data mining is a powerful tool for converting large databases into knowledge for use in solving problems that are otherwise imponderable because of the large numbers of explanatory variables or poorly understood process physics. Data-mining methods come from different technical fields such as signal processing, statistics, artificial intelligence, and advanced visualization. It employs methods for maximizing the information content of data, determining which variables have the strongest correlations to the problems of interest, and developing models that predict future outcomes. This knowledge encompasses both understanding of cause-effect relations and predicting the consequences of alternative actions. Data mining is used extensively in financial services, banking, advertising, manufacturing, and e-commerce to classify the behaviors of organizations and individuals and to predict future outcomes.

Purpose and Scope

This report presents the results of a study that analyzes salinity intrusion due to changing streamflow and tidal water-level conditions. This report documents the development of the **P**ee Dee **R**iver and Atlantic **I**ntracoastal Waterway **S**alinity Intrusion **M**odel Decision Support System (PRISM DSS) including the examples of applying the PRISM DSS to the Waccamaw River and AIW to evaluate salinity intrusions along the coast.

An important part of the USGS mission is to provide scientific information for the effective water-resources management of the Nation. To assess the quantity and quality of the Nation's surface water, the USGS collects hydrologic and water-quality data from rivers, lakes, and estuaries using standardized methods, and maintains the data from these stations in a national database. Often these databases are underutilized and under-interpreted for addressing contemporary hydrologic issues. The techniques presented in this report demonstrate how valuable information can be extracted from existing USGS databases to assist local, State, and Federal agencies. The application of data-mining techniques, including ANN models, to the Pee Dee and Waccamaw Rivers and AIW demonstrates how empirical models of complex hydrologic systems can be developed, disparate databases and models can be integrated, and study results can be easily disseminated to meet the needs of a broad range of end users.

Description of Study Area

The Pee Dee River basin, including the Waccamaw River tributary, supplies freshwater inflows to the Grand Strand of South Carolina, an area of rapidly growing coastal communities from Little River Inlet to the north to Winyah Bay to the south (fig. 1). The headwaters of the Pee Dee River are in the Blue Ridge Province of North Carolina and Virginia and drain 6,800 square miles (mi^2) of North Carolina above Blewett Falls Lake (fig. 2) before flowing through South Carolina to the Atlantic Ocean (Seaber and others, 1994). Above the confluence with the Uwharrie River, the stream is known as the Yadkin River, and below as the Pee Dee River, or Great Pee Dee River (U.S. Geological Survey, 1986). The Pee Dee River flows through seven impoundments in North Carolina. The first reservoir is the W. Kerr Scott Lake west of Wilkesboro, N.C. Downstream, a chain of five reservoirs impounds 50 miles of the river consisting of High Rock Lake, Tuckertown Reservoir, Badin Lake, Falls Lake, and Lake Tillery (North Carolina Department of Environment and Natural Resources, 2001). The seventh is Blewett Falls Lake located approximately 15 miles upstream from the South Carolina State line (table 1).

Table 1. Lake name, surface area, and owners of seven North Carolina Reservoirs on the Yadkin-Pee Dee River, North Carolina (from North Carolina Department of Environment and Natural Resources, Division of Water Quality, 1998).

Lake	Surface area, in acres	Owner
W. Kerr Scott Lake	1,480	U.S. Army Corps of Engineers
High Rock Lake	12,200	Alcoa Power Generation, Inc.
Tuckertown Reservoir	2,550	Alcoa Power Generation, Inc.
Badin Lake	5,350	Alcoa Power Generation, Inc.
Falls Lake	203	Alcoa Power Generation, Inc.
Lake Tillery	5,260	Carolina Power and Light Co.
Blewett Falls Lake	2,570	Carolina Power and Light Co.

Figure 2. The Yadkin-Pee Dee River basin in North and South Carolina.

The Pee Dee River below Blewett Falls Reservoir drains approximately 11,700 mi^2 and has five major tributaries: the Little Pee Dee, Lynches, Black, Waccamaw, and Sampit Rivers (Seaber and others, 1994; figs. 1 and 2). The Little Pee Dee, Lynches, and Black River tributaries are unregulated and undeveloped, and drain rural areas. Downstream from U.S. Highway 701, the Pee Dee River branches successively into Bull, Thoroughfare, and Schooner Creeks (fig. 3). These three creeks eventually flow into the Waccamaw River and Winyah Bay. The majority of the freshwater flow to the AIW from the Pee Dee River basin is carried by Bull Creek to the Waccamaw River.

The Waccamaw River originates in North Carolina and enters the AIW about 10 miles north of the mouth of Bull Creek. Prior to the 1930s, the Waccamaw River flowed to the south toward Winyah Bay. In the 1930s, the U.S. Army Corps of Engineers constructed a canal to form the waterway from Enterprise Landing to the Little River Inlet, which altered the flow of the Waccamaw River north toward Little River Inlet along the AIW. A large portion of the flow from the Waccamaw River flows north through the AIW to the Atlantic Ocean through Little River Inlet (Drewes and Conrads, 1995). The Waccamaw River drains extensive cypress and hardwood swamps (U.S. Geological Survey, 1986). There are seven USGS gaging stations on the Pee Dee River and its major tributaries (table 2).

The reach of the AIW from just south of Little River Inlet to just north of Hagley Landing provides freshwater for the coastal communities of the Grand Strand (fig. 3). In the 1980s, major water purveyors changed from ground-water to surface-water sources to avoid taste and odor problems associated with the ground-water supplies (Carswell and others, 1988). Three municipal surface-water intakes are in the tidal, freshwater portions of the AIW, Waccamaw River, and Bull

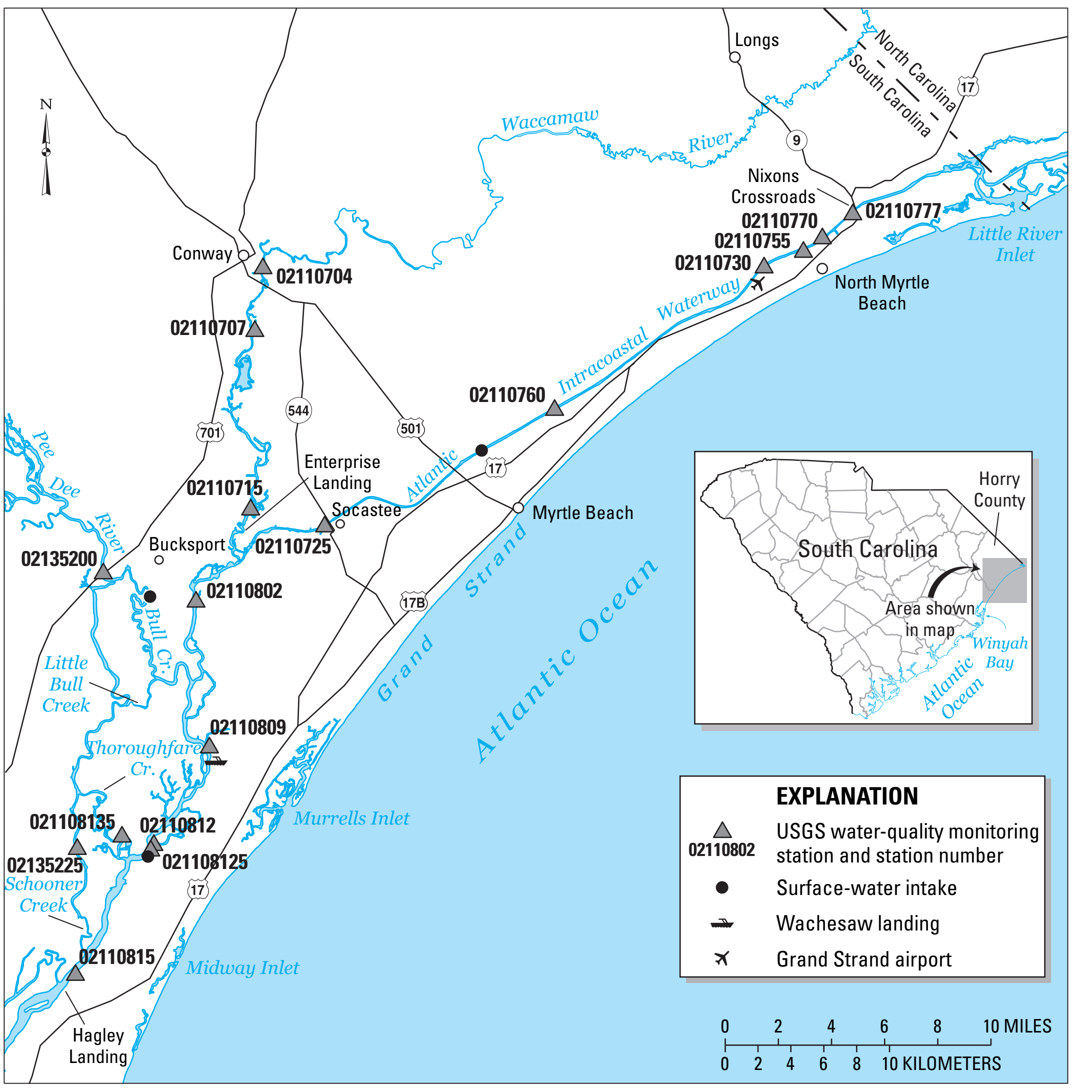


Figure 3. Specific-conductance gages and surface-water intakes on the Waccamaw River and Atlantic Intracoastal Waterway in the Grand Strand of South Carolina.

Table 2. Station name, station number, period of record, drainage area, and river mile for U.S. Geological Survey gaging stations on the Pee Dee River and its major tributaries, North and South Carolina.

[mi², square mile]

Station name	Station number	Period of record	Drainage area, in mi²	River mile
Waccamaw River near Longs, SC	02110500	1950 to 2007	1,110	85.4[a]
Pee Dee River near Rockingham, NC	02129000	1906 to 2007	6,863	192[a]
Pee Dee River near Bennettsville, SC	02130561	1990 to 2007	7,600	153.0[a]
Pee Dee River at Pee Dee, SC	02131000	1938 to 2007	8,830	100.2[b]
Lynches River at Effingham, SC	02132000	1929 to 2007	1, 030	43.4[b]
Little Pee Dee River at Galivants Ferry, SC	02135000	1942 to 2007	2,790	41.7[b]
Black River at Kingstree, SC	02136000	1929 to 2007	1,252	86.7[b]

[a] River mile measured from confluence with Winyah Bay.

[b] River mile measured from confluence with the Pee Dee River.

Creek (fig. 3). During the drought from 1998 to 2002, salinity intrusion forced a municipal intake to close temporarily until increased streamflow and changing meteorological conditions moved the freshwater-saltwater interface downstream from the intake.

Previous Studies

Numerous investigations have been conducted to address water-availability issues in the Grand Strand and use of data-mining techniques to study salinity dynamics in estuarine systems. Carswell and others (1988) investigated the freshwater supply potential of the AIW as an alternative to ground-water sources. Using statistical analysis and mechanistic models, it was determined that the AIW could provide a reliable supply of freshwater. A salinity warning system also was developed that uses real-time gages to trigger alerts when specific-conductance values exceed specified thresholds. Drewes and Conrads (1995) determined the assimilative capacity of the Waccamaw River and AIW using dynamic one-dimension flow and water-quality models. The models were used by the South Carolina Department of Health and Environmental Control (SCDHEC) for determining the total maximum daily load (TMDL) for dissolved oxygen (South Carolina Department of Health and Environmental Control, 1998).

Roehl and others (2000) used data-mining techniques, including ANN models, to simulate the response of the freshwater-saltwater interfaces in the Cooper River, South Carolina, to changing reservoir releases. Conrads and others (2006) developed a DSS to integrate hydrodynamic and ecological models being used to evaluate a potential deepening of the Savannah Harbor. 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 ANN models. The ANN models simulated riverine and marsh water levels and salinity in the vicinity of the Savannah National Wildlife Refuge for the full range of 11½ years of data from riverine and marsh gaging networks. The 3DM, MSM, and M2M were integrated in a DSS for use by various regulatory and scientific stakeholders.

Approach

The variability of salinity in the Waccamaw River and AIW is a result of many factors including streamflow of the Pee Dee River basin and tidal conditions of Little River Inlet and Winyah Bay. In order to simulate the dynamic response of salinity, empirical models were developed to predict specific conductance[2] for gages near Little River Inlet and Hagley Landing for changing streamflow and tidal conditions. For the Pee Dee and Waccamaw Rivers and AIW, extensive continuous data sets of streamflow, tidal water level, and specific conductance are available. Empirical specific-conductance models were developed directly from the data using data-mining techniques and ANN models.

The application of data-mining techniques to salinity intrusion was undertaken in four phases: (1) evaluating the suitability of the long-term (20 years) USGS specific-conductance data; (2) simulating the salinity intrusion at four gage locations at the north end of the system in the AIW near the Little River Inlet and at five gage locations at the south end of the system in the Waccamaw River upstream from Winyah Bay (fig. 1); (3) analyzing the causes of the large

[2] Specific conductance is the property often used to compute salinity.

salinity intrusions into the river systems; and (4) developing a DSS that integrates historical databases, model controls, and model output into a spreadsheet application with a graphical user interface (GUI) that allows a user to simulate scenarios of interest.

Acknowledgments

The complexity of this study required interagency cooperation in addition to individual contributions. The authors thank Danny Johnson of the SCDNR for his coordination of the project and Phil Lucas of Progress Energy for hosting bimonthly project status meetings and a workshop. The authors also thank Marty Barfield of the Pee Dee River Coalition, Gene Ellis of Alcoa, Bud Badr of the SCDNR, Larry Turner of the SCDHEC, and Eric Horner and Thomas Fransen of the North Carolina Department of Environment and Natural Resources for their participation and contributions to the bimonthly meetings and workshop.

Data-Collection Networks

Many resource agencies have collected data in the Pee Dee River basin and AIW estuary, including the USGS, National Oceanic and Atmospheric Administration, the SCDHEC, and local colleges and universities. For this study, data from three data-collection networks were used to build, train, and test the specific-conductance ANN models. One network is the long-term streamflow network in the Pee Dee and Waccamaw River basin upstream from the tidal influence (fig. 1). Data from the streamflow network originate as early as 1906 (table 2). Over 50 years (1950 to 2007) of concurrent data are available for five stations on the principal tributaries—the Waccamaw, Lynches, Little Pee Dee, and Black Rivers.

The second network is the coastal network of specific-conductance gages in the Grand Strand (fig. 3; table 3). The coastal network does not have the temporal continuity of the

Table 3. Station name, station number, parameters, and period of record for U.S. Geological Survey tidal gaging stations along the Grand Strand of South Carolina.

[S.C., South Carolina; WL, water level; Q, flow; SC, specific conductance; AIW, Atlantic Intracoastal Waterway. Artificial neural network models developed for stations are in **bold** text]

Station name	Station number	Name used in this report	Parameters	Period of record
Waccamaw River at Conway Marina at Conway, S.C.	02110704	Conway Marina	WL, Q, SC	1991–2007
Waccamaw River at Pitch Landing, S.C.	02110707		WL, SC	1986–1989
Waccamaw River at Peachtree Landing, S.C.	02110715		SC	1990–1991
AIW at S.C. Highway 544 at Socastee, S.C.	02110725	Highway 544	WL, SC	1986–1992
AIW at Vereens Marina at North Myrtle Beach, S.C.	02110730		WL, SC	1983–1991
AIW at Briarcliffe Acres, S.C.	**02110755**		SC	1983–2007
AIW at Myrtlewood Golf Course, S.C.	**02110760**		SC	1986–1989, 1994–2007
AIW at Grand Strand Airport at North Myrtle Beach, S.C.	**02110770**		SC	1987–2007
AIW at S.C. Highway 9 at Nixons Crossroads, S.C.	**02110777**	Highway 9	WL, Q, SC	1986–2007
Waccamaw River at Bucksport, S.C.	02110802		WL, Q, SC	1983–1995
Waccamaw River at Wachesaw Landing, S.C.	**02110809**		SC	1986–1989, 2002–2007
Waccamaw River at Mt. Rena Landing near Murrells Inlet, S.C.	**02110812**		SC	1986–1989
Waccamaw River near Pawleys Island, S.C.	**021108125**	Pawleys Island	WL, SC	2002–2007
Thoroughfare Creek at Berlin near Pawleys Island, S.C.	**021108135**		WL, SC	1989
Waccamaw River at Hagley Landing, S.C.	**02110815**	Hagley Landing	WL, SC	1986–2007
Waccamaw River at U.S. Highway 17 at Georgetown, S.C.[a]	02110850		SC	1985–1989
Pee Dee River at U.S. Highway 701, S.C.	02135200	Highway 701	SC	1986–1994
Pee Dee River at Arundel Plantation near Jackson, S.C.	02135225		WL, SC	1989
Winyah Bay at Mouth near Georgetown, S.C.[a]	02136390		SC	1986–1989

[a] Station not shown in figure 3.

streamflow network. Gages often were installed to support special investigations and discontinued upon completion of the particular study. The coastal network does provide 15-minute data for extended periods (3 to 12 years) for 19 stations over the last 20 years. During the past 20 years of data collection, the coastal network has measured various extreme meteorological conditions, including large rainfalls in a 24-hour period, the passing of major hurricanes offshore, and drought conditions. In addition to data from the USGS networks, wind speed and direction data were obtained from the Southeast Climate Center from its Charleston Harbor gage.

Characterization of Streamflow, Water Level, and Specific Conductance

Estuarine systems are complex systems that are constantly responding to changing hydrologic, tidal, and meteorological conditions. Dyer (1997) stated that the challenge of studying estuaries is "… that river flow, tidal range, and sediment distribution are continually changing and this is exacerbated by the continually changing weather influences. Consequently, some estuaries may never really be steady-state systems; they may be trying to reach a balance they never achieve." The estuarine portions of the AIW and Waccamaw Rivers are constantly integrating the changing streamflow of the Pee Dee River basin, changing tidal conditions of the Atlantic Ocean, and changing meteorological conditions, including wind direction and speed, rainfall, low- and high-pressure systems, and hurricanes. The following sections describe data preparation (including calculated variables and signal processing) and characterize the streamflow and tidal water levels and how these affect the salinity intrusion in the rivers.

Calculated Variables, Data Preparation, and Signal Processing

Tidal systems are highly dynamic and exhibit complex behaviors that evolve over multiple time scales. 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 more influential 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. For the Pee Dee River study, nested moving-window averages of 25 and 13 hours were used to remove the high-frequency tidal cycle. Removing the semi-diurnal tidal frequency 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. 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.

One variable was computed from the field measurements of the physical properties—tide range (XWL). Tidal dynamics are a dominant force for estuarine systems, and the tide range is a significant variable for determining the lunar phase of the tide and flushing dynamics of coastal rivers. Tidal range is calculated from water level (or gage height) and is defined as the water level at high tide minus the water level at low tide for each semi-diurnal tidal cycle. Examples of filtered time series signal and tidal ranges are provided in subsequent sections of the report.

Characterization of Streamflow

Streamflow in the Pee Dee River in South Carolina is regulated by releases from Blewett Falls Lake near Rockingham, N.C. Duration hydrographs based on 78 years of data are shown in figure 4. Daily duration graphs characterize the state of a stream with respect to time. The plotted percentiles are best explained by an example. Suppose 78 years of daily value flow data exist for a station and the 75th-percentile flow is 10,000 cubic feet per second (ft^3/s) for a particular day of the year, say January 3. This means that 75 percent of all flows that occurred on January 3 of each of the 78 years of data were equal to or less than 10,000 ft^3/s. It also is assumed that flows between the 0- and 10th-percentiles occur during very dry hydrologic conditions, and likewise, it is assumed that flows between the 90th- and 100th-percentiles occur during very wet hydrologic conditions. It is assumed that flows between the 25th- and 75th-percentiles occur during normal hydrologic conditions. Flow at station 02129000, Pee Dee River at Rockingham, N.C., ranges from a minimum of less than 500 ft^3/s during periods of low flow to greater than 60,000 ft^3/s or more during periods of high flows (fig. 4). Seasonally, the highest flows occur in late winter and early spring (February through March), and the lowest flows occur in late summer and early fall (July through October).

The large oscillation in the percentile flow traces, especially in the 95th-percentile flows, is a result of the regulated flow and large variability in releases from the Blewett Falls Lake. The Little Pee Dee River, which is the largest tributary to the Pee Dee River, and the other tributaries are unregulated, and the percentile flows do not have the large oscillations of the regulated streams. The regulated flows decrease the relative distribution of the low and medium percentile flows as compared to unregulated streams. For the Little Pee Dee River, there is a larger distribution between the 5th- and 75th- percentile flows as compared to the same percentiles for the Pee Dee River (fig. 4).

Although reservoir regulation changes the natural flow regime over the short term (hours to days), generally, reservoir operation reflects longer term (weeks to months)

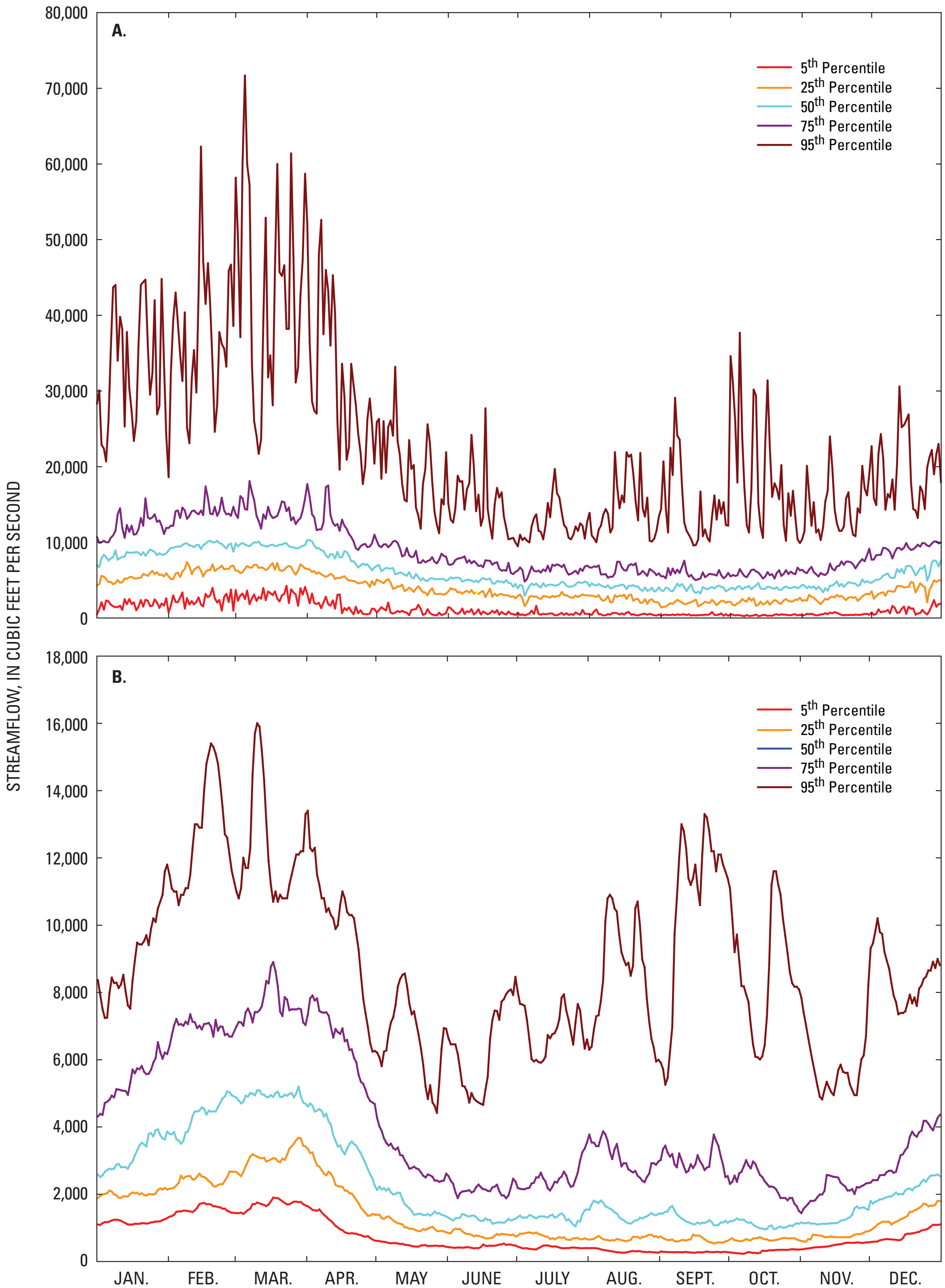


Figure 4. Duration hydrographs for (A) station 02129000, Pee Dee River near Rockingham, North Carolina, and (B) station 02135000, Little Pee Dee River at Galivants Ferry, South Carolina. Percentile flows for the Pee Dee River near Rockingham are based on streamflow data from 1928 to 2005. Percentile flows for the Little Pee Dee River at Galivants Ferry are based on streamflow data from 1942 to 2005.

meteorological conditions similar to unregulated streams. Two years of hourly streamflow conditions are shown in figure 5 for the Pee Dee River at Pee Dee (station 02131000) and for the unregulated tributaries of the Lynches, Little Pee Dee, and Black Rivers (stations 02132000, 02135000, 02136000, respectively). A 7-day moving window average was applied to the streamflow for the Pee Dee River and shows a similar response to low- and high-flow conditions of the unregulated streams (fig. 5B).

Characterization of Water Level

The AIW and Waccamaw River experience semi-diurnal tides of two high tides and two low tides in a 24.8-hour period. The semi-diurnal tides exhibit periodic cycles of high- and low-tide ranges (water-level difference between high and low tide) on a 14-day cycle. Spring tides are periods of increased tide range during the time of full and new moons. Neap tides are periods of decreased tide range around the time of waxing and waning moons. The mean and spring tidal ranges for Little River Inlet are 4.41 and 5.07 feet (ft), respectively, and for Winyah Bay are 4.60 and 5.40 ft, respectively (table 4: National Oceanic and Atmospheric Administration, 2005). As the tidal wave propagates upstream, the tidal range decreases with the increased freshwater flow of the Pee Dee and Waccamaw Rivers and energy losses due to the decrease in channel geometry. In the AIW, the mean tide range decreases to 1.97 ft at the S.C. Highway 544 Bridge at Socastee (table 4). In the Waccamaw River, the mean tide range decreases to 1.24 ft at Conway. An approximate 3- to 4-hour lag of the tide exists from Nixons Crossroad at S.C. Highway 9 (station 02110777) to S.C. Highway 544 (station 02110725) on the AIW, and an approximate 5-hour lag exists between Hagley Landing (station 02110815) and Conway Marina (02110704) on the Waccamaw River.

The water levels at four USGS stations on the AIW and the Pee Dee and Waccamaw Rivers for a 17-day period during October 1998 are shown in figure 6A. The spring-tide period, characterized by a large amplitude in tidal range, occurred around October 7; the neap-tide period, characterized by a relatively small amplitude in tidal range, occurred around October 14. During periods of medium and high streamflow, the tidal signals at Waccamaw River at Conway Marina and Pee Dee River at U.S. Highway 701 are overwhelmed by the river flows. Water levels for the same four stations on the AIW and Pee Dee and Waccamaw Rivers were compared for a 45-day period during January and February 1999 (fig. 6B). Streamflow at Pee Dee River at Pee Dee, S.C. (station 02131000) peaked at 22,400 ft^3/s on January 29 and

Table 4. Mean tide range, spring tide range, and mean tide levels for locations on the Atlantic Intracoastal Waterway and Waccamaw River, South Carolina.

[Data from National Oceanic and Atmospheric Administration, 2005]

Location (fig. 3)	Mean tide range, in feet	Spring tide range, in feet	Mean tide level, in feet[a]
Atlantic Intracoastal Waterway			
Little River	4.41	5.07	2.35
Nixons Crossroads	4.00	4.56	2.11
Myrtle Beach Airport	2.88	3.34	1.60
North Myrtle Beach	1.85	2.15	1.08
Socastee Bridge (S.C. Highway 544)	1.97	2.29	1.09
Winyah Bay and Waccamaw River			
Winyah Bay Entrance	4.60	5.40	2.50
Waccamaw River Entrance	3.60	4.18	1.96
Hagley Landing	3.50	4.06	1.88
Thoroughfare Creek	3.33	3.86	1.84
Wachesaw Landing	2.74	3.18	1.53
Bull Creek Entrance	2.46	2.85	1.38
Bucksport	2.22	2.58	1.27
Enterprise Landing	2.00	2.40	1.10
Conway	1.24	1.44	0.76

[a] The arithmetic mean of high and low tides. The height of the mean tide level is listed relative to the mean lower low water datum.

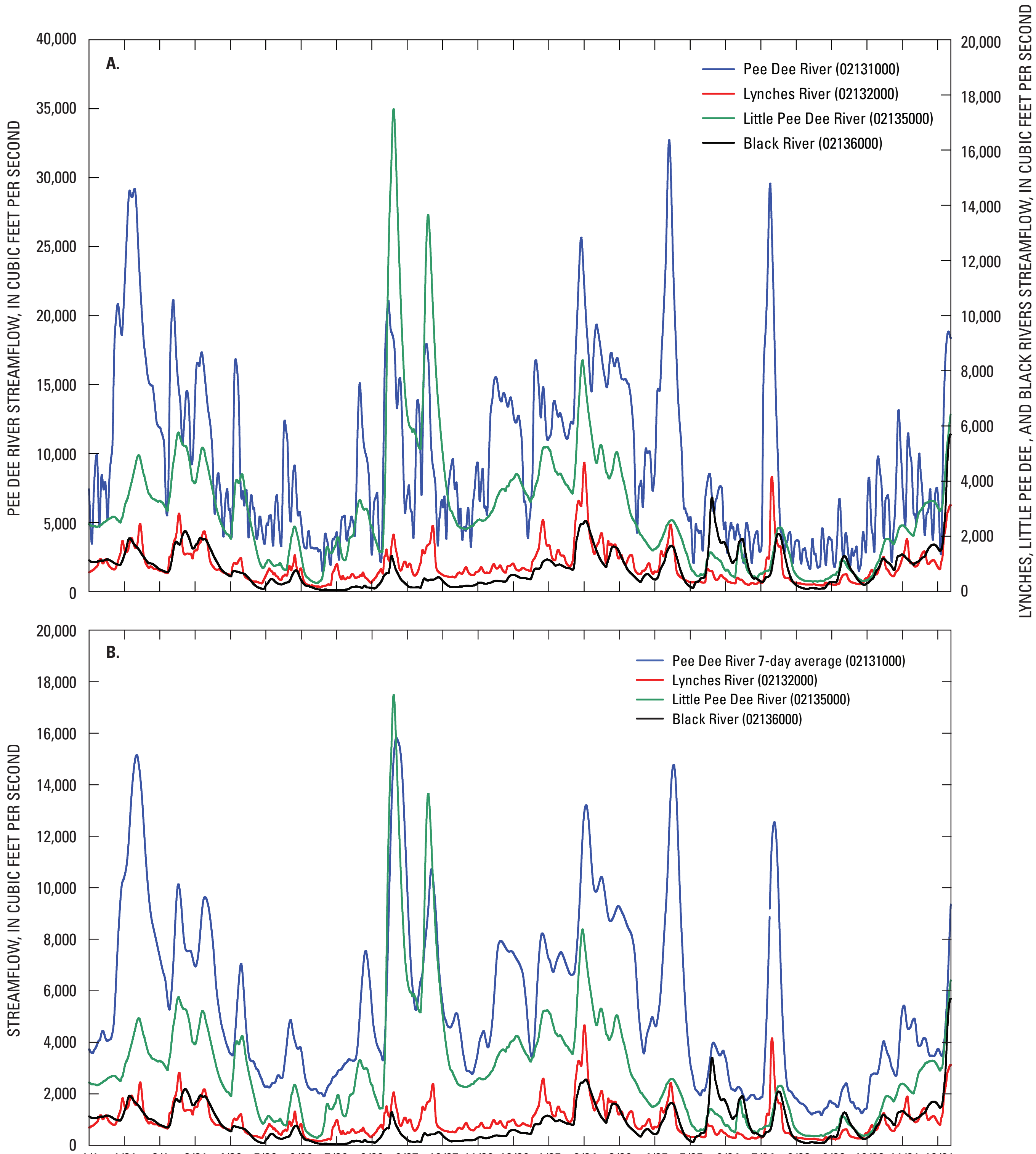


Figure 5. (A) Hourly streamflow for the Pee Dee, Lynches, Little Pee Dee, and Black Rivers and (B) 7-day average flow for the Pee Dee River and hourly streamflow for the Lynches, Little Pee Dee, and Black Rivers for the period January 1, 1996, to December 31, 1997.

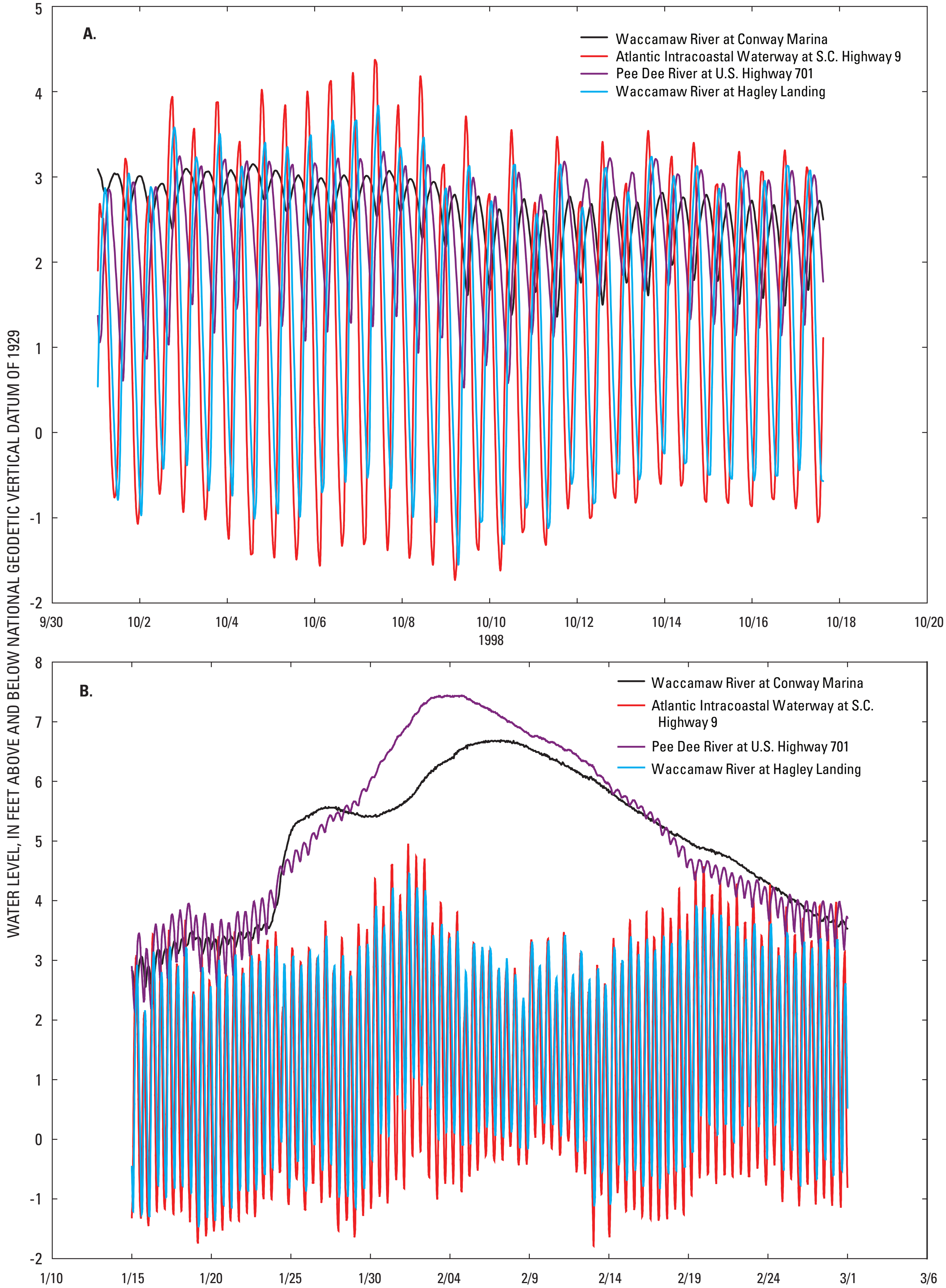


Figure 6. Hourly water levels at four gaging stations on the Pee Dee and Waccamaw Rivers and the Atlantic Intracoastal Waterway for (A) October 1 to October 17, 1998, and (B) January 15 to March 1, 1999.

Waccamaw River near Longs, S.C. (station 02110500) peaked at 7,280 ft^3/s on February 7 (not shown on fig. 6B). For the Pee Dee River at U.S. Highway 701, the tidal signal is minimized at water levels greater than 6 ft. For the Waccamaw River at Conway Marina, the tidal signal is overwhelmed when water levels are greater than 4 ft above National Geodetic Vertical Datum of 1929 (NGVD 1929).

A graph of the tidal range at the AIW at S.C. Highway 9 for the period July 1994 through June 1996 clearly shows the 14- and 28-day spring-neap tidal cycles along with seasonal and semi-annual cycles (fig. 7). For example, a high spring tide (tidal range greater than 4.5 ft) is followed 14 days later by a low spring tide (tidal range less than 4.5 ft). A similar 28-day pattern is apparent in the neap tides where a low neap tide (tidal range less than 3.5 ft) is followed 14 days later by a high neap tide (tidal range greater than 3.5 ft). The biggest differences in spring and neap tides occur in the spring (March and April) and the fall (October and November) of the year. Minimum differences between the spring and neap tides occur in the summer (June and July) and in the winter (December and January) of the year.

Characterization of Specific Conductance

The location of the saltwater-freshwater interface is a balance between upstream river flows and downstream tidal forcing (fig. 8). During periods of high streamflow, it is difficult for salinity to intrude upstream, and the saltwater-freshwater interface is moved downstream toward the ocean. During periods of low streamflow, salinity is able to intrude upstream, and the saltwater-freshwater interface is moved upstream by tidal forcing—either by an increase in mean water levels or a change in tidal range, or a combination of the two. Historically, streamflow on the Pee Dee River has ranged between 500 and 60,000 ft^3/s. Salinity in the Waccamaw River

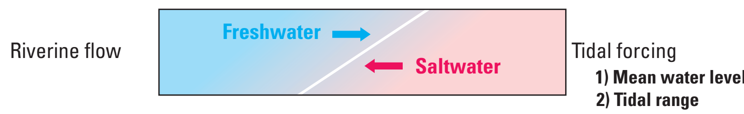


Figure 8. Conceptual model of the location of the freshwater-saltwater interface.

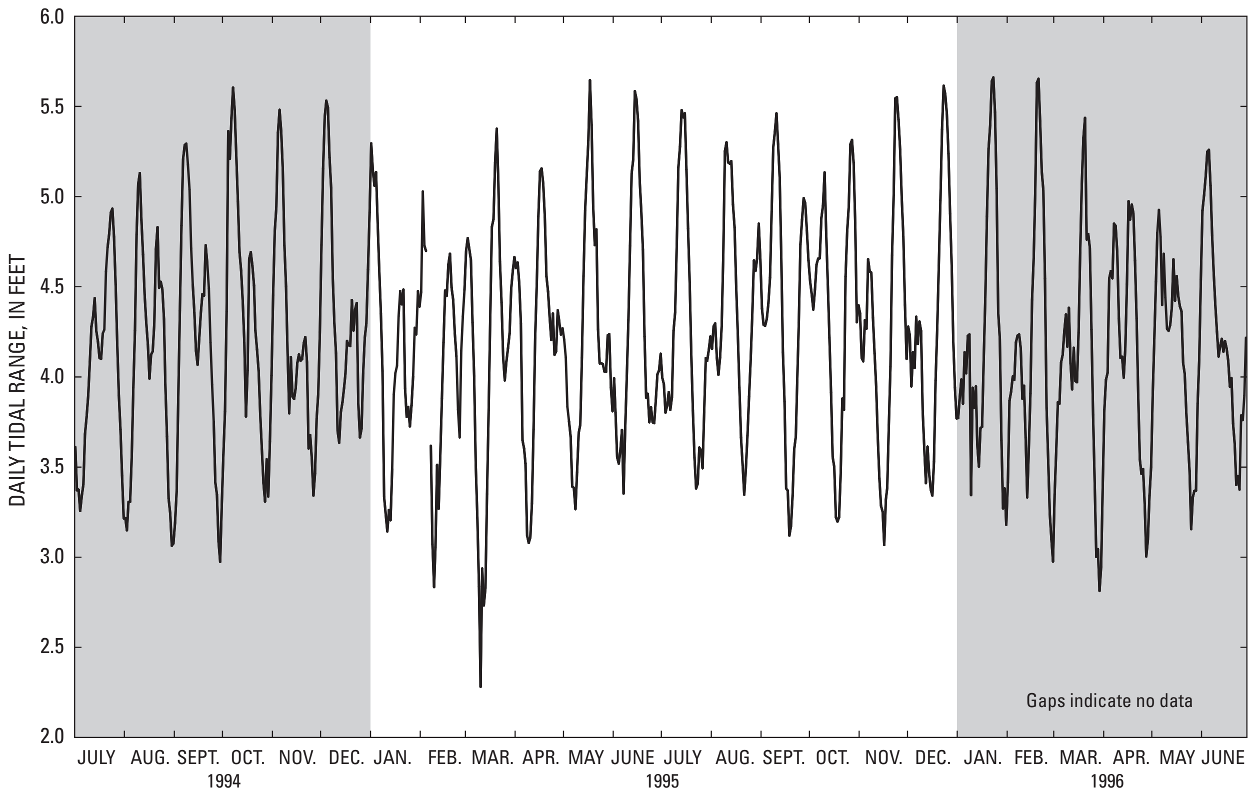


Figure 7. Daily tidal range at Atlantic Intracoastal Waterway at South Carolina Highway 9 (station 02110777) for the period July 1994 through June 1996.

and AIW is constantly responding to changing streamflow and tidal conditions. The daily mean specific conductance for the Waccamaw River at Hagley Landing (station 02110815) and the daily mean streamflow for the Pee Dee River at Pee Dee, S.C. (station 02131000) for the October 1995 to February 2003 period are shown in figure 9. The period includes the full range of flows for the system from the high flows of the El Niño in 1998 to the low flows of the extended drought from 1998 to 2002. During periods of medium and high flows (streamflow greater than 5,000 ft^3/s), the specific conductance is generally low. During periods of low flow (streamflow less than 5,000 ft^3/s), specific-conductance values increase, representing periods of salinity intrusion. During the low-flow periods prior to the high flows measured during the El Niño of 1998, salinity intrusion with specific-conductance values of 10,000 to 15,000 microsiemens per centimeter at 25 degrees Celsius (μS/cm at 25 °C) were not uncommon. After the high flow of 1998 and during the extended drought, flows were even lower and remained lower for extended periods. This resulted in greater salinity with daily mean values often exceeding 15,000 μS/cm, with occasional intrusions around 25,000 μS/cm. A specific conductance of 25,000 μS/cm is equal to a salinity of 15.2 practical salinity units (psu).

Modeling Specific Conductance

Simulating salinity for estuarine systems typically is done using dynamic deterministic models that incorporate the mathematical descriptions of the physics of coastal hydrodynamics. These one-, two-, or three-dimensional models often require extensive data collection and are time consuming to apply to complex coastal systems with satisfactory results. Conrads and Roehl (2005) assert that in estuaries, mechanistic model calibration is "…particularly difficult due to low watershed gradients, poorly defined drainage areas, tidal complexities, and a lack of understanding of watershed and marsh processes." Although mechanistic models have been the state of the practice for regulatory evaluations of anthropogenic effects on estuarine systems, developments in the field of advanced statistics, machine learning, and data mining offer opportunities to develop empirical ANN models that are often more accurate. Conrads and Roehl (1999) compared the application of a deterministic model and an ANN model to simulate dissolved-oxygen (DO) concentrations for 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

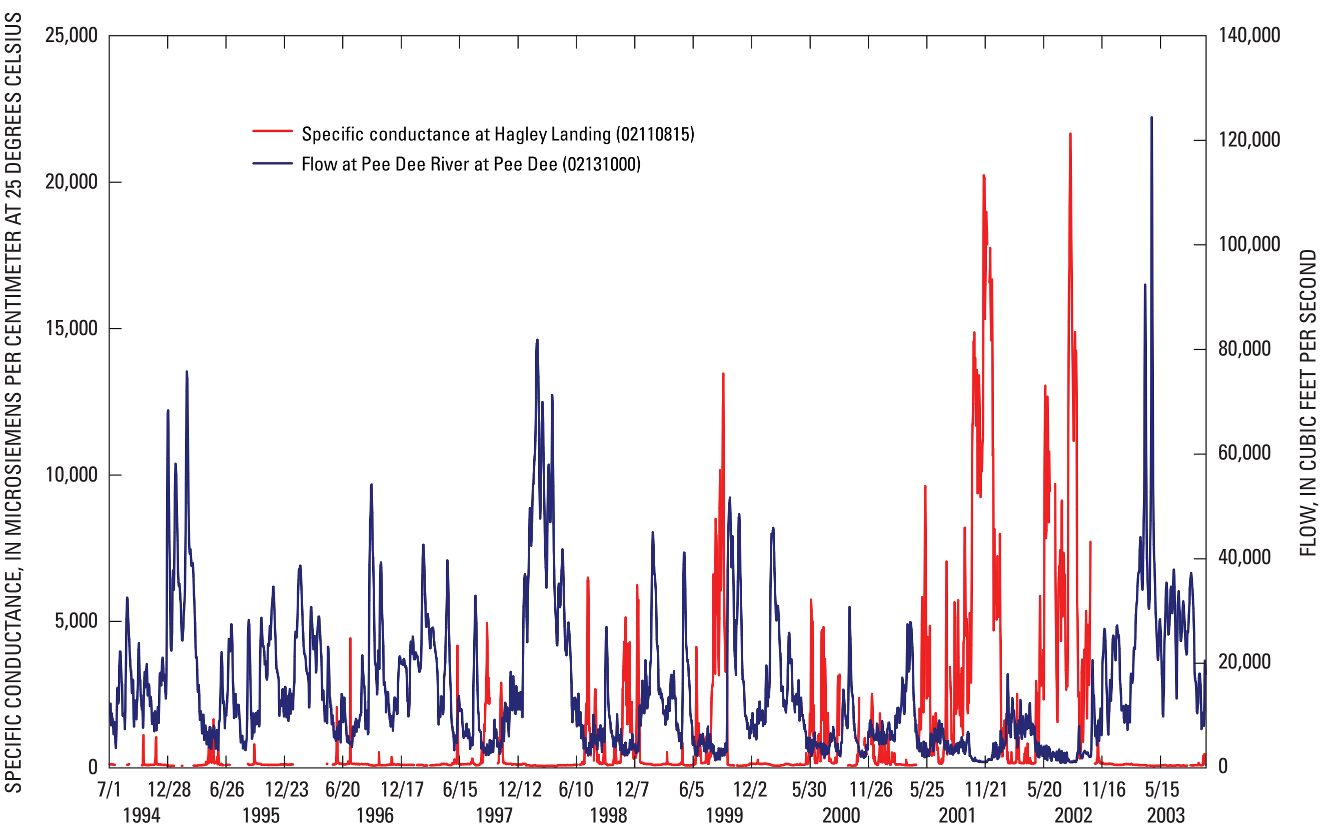


Figure 9. Daily specific conductance at Waccamaw River at Hagley Landing, South Carolina, and streamflow at Pee Dee River at Pee Dee, South Carolina, for the period October 1, 1995, to September 30, 2003.

amounts of data, the incorporation of optimization routines, and model dissemination in spreadsheet applications. With the real-time gaging network on the Waccamaw River and AIW and the availability of large databases of hydrologic and water-quality data, the SCDNR realized an opportunity existed to develop an empirical model using data-mining techniques, including ANNs, to simulate salinity intrusion in the Waccamaw River and AIW.

The emerging field of data mining addresses the issue of extracting information from large databases. Data mining is composed 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 relations from example cases and data. For numerical data, commonly used methods include ANNs, 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 of South Carolina and Georgia (Conrads and others, 2006; Conrads and others, 2003; Conrads, Roehl, and Cook, 2002; Conrads, Roehl, and Martello, 2002; Roehl and others, 2000; Roehl and Conrads, 1999; Conrads and Roehl, 1999) and stream temperatures in western Oregon (Risley and others, 2003). These studies have demonstrated the ability of data mining to predict water level, water temperature, dissolved oxygen, and specific conductance and to assess the effects of reservoir releases and point and nonpoint sources on receiving streams.

The ultimate goal of this study is to produce an effective model to predict salinity intrusion in the freshwater portions of the Waccamaw River and AIW for a given set of streamflow, water-level, and tidal-range conditions. The approach taken uses all available streamflow, water-level, and specific-conductance measurements from the individual gages since the establishment of the coastal gaging network in 1983. The modeling approach uses correlation functions that were synthesized directly from data to predict how the change in specific conductance at each gage location is affected by streamflow and tidal conditions over time.

Signal Decomposition, Correlation Analysis, and State-Space Reconstruction

The behavior, or dynamics, in a natural system results from interactions between multiple physical forces. For example, the specific conductance at a fixed location is subject to daily, seasonal, and annual streamflow conditions and semi-diurnal, fortnightly, seasonal, and annual tidal water-level conditions. For the application of the ANN models to the Waccamaw River and AIW, data-mining methods were applied to maximize the information content in raw data while diminishing the influence of poor or missing measurements. Methods include digital filtering using multiple moving window averages (described previously), time derivatives, time delays, and running averages. Signals, or time series, manifest three types of behavior: periodic, noise, or chaotic. Periodic behavior is perfectly predictable. Examples of periodic behavior are the diurnal sunlight 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, and is unpredictable. Chaotic behavior is neither totally periodic nor noise, and always has a physical cause. Weather provides an example of chaotic behavior. Chaotic behavior is somewhat predictable, especially for small timeframes and prediction horizons.

Signal Decomposition

Signal decomposition involves splitting a signal into sub-signals, called "components," that are independently attributable to different physical forces. To analyze and model these time series, the periodic and chaotic components of the signals need to be separated. As previously discussed, digital filtering can separate out the chaotic component in the tidal water-level, or gage height, time series. Computation of the tide range time series from the water-level time series separates out the periodic components of the water-level time series. Digital filtering also can diminish the effect of noise in a signal to improve the amount of useful information that it contains. An example of 60 days of hourly gage-height data and filtered gage-height data for the AIW at S.C. Highway 9 is shown in figure 10. The filtered gage heights represent the chaotic component of the tidal gage-height signal.

The use of time derivatives is a common analytical method for analysis of the dynamics of a system. Time derivatives were computed from the measured, computed, and filtered variables on the Waccamaw River and the AIW to further understand the dynamics of the system. The 1-day derivative of the filtered water-level time series for a 60-day period was plotted with the original time series and the filtered data (fig. 10). The 1-day derivatives show the rate of change of the chaotic component of the water-level time series. For the 60-day period, the daily change in filtered gage height ranged from –0.5 ft to 0.5 ft.

Often time delays exist 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 simulate a dynamic system. For the Waccamaw River and the AIW, there are time delays between the measured streamflow at the Pee Dee

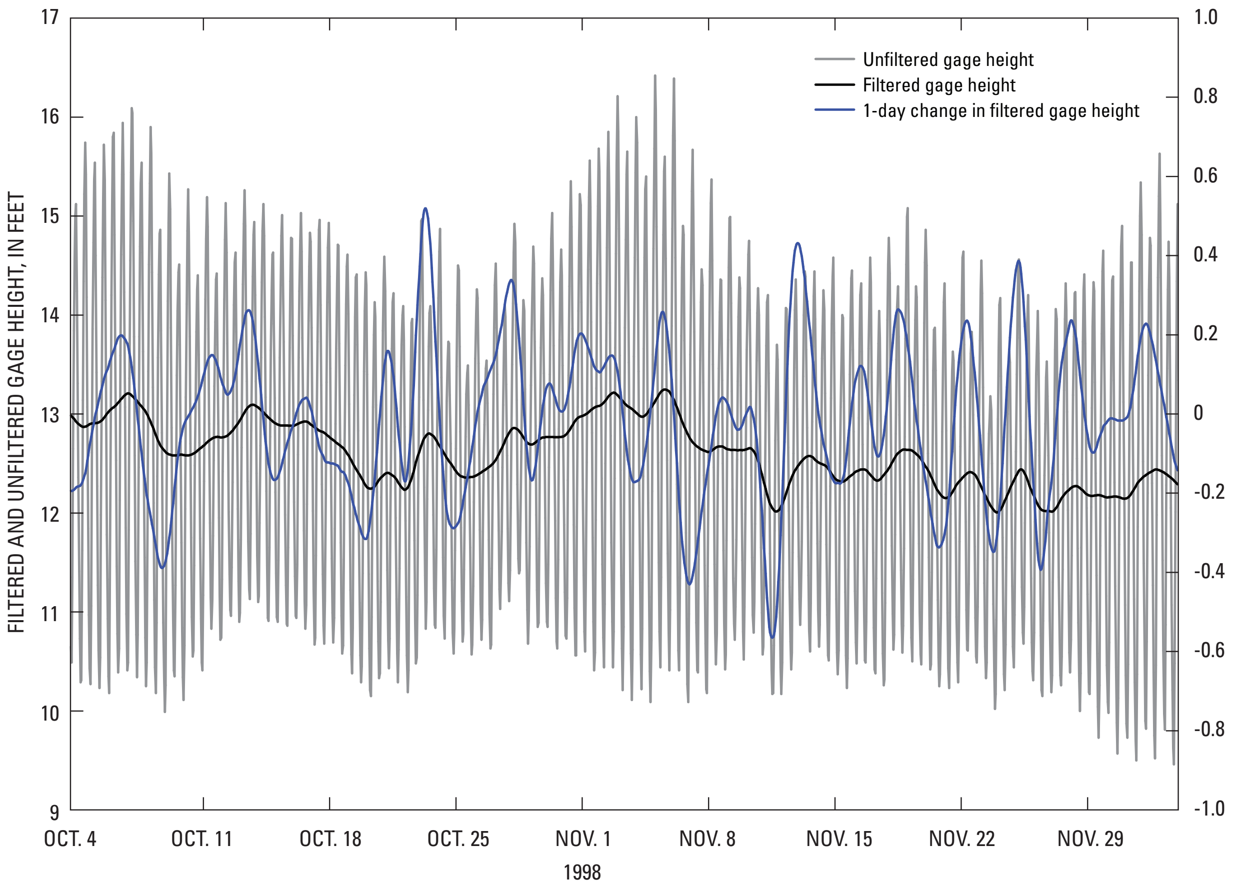


Figure 10. Hourly gage heights at Atlantic Intracoastal Waterway at South Carolina Highway 9 (station 02110777), filtered gage heights, and 1-day change in filtered gage heights for the 60-day period October 4 to December 4, 1998.

River at Pee Dee, S.C. (station 02131000), and the response in specific conductance at the gages in the Waccamaw River and the AIW. Time delays also exist among measured streamflow values for the four tributaries to the Pee Dee River. Time delays from when the flow enters the system to when the river responds to the flow were determined for each gage.

Correlation Analysis

The relations between the many variables and their various components are ascertained through correlation analyses to provide deeper understanding of system dynamics. Sensitivity analysis quantifies the relations between a dependent variable of interest and causal variables. For example, salinity intrusion is dependent on streamflow and tides. Computing sensitivities requires defining the relations among variables through modeling.

The computer systematically correlates factors that most influence parameters of interest (for example, specific conductance) to determine combinations of controlled and uncontrolled variables (for example, streamflow and tidal conditions). Correlation methods based on statistics and ANNs are applied in combination. Promising results found by the computer are validated by comparing them to known patterns of behavior.

State-Space Reconstruction

Chaos Theory provides a conceptual framework called “state-space reconstruction” (SSR) for representing dynamic relations. Data collected at a point in time can be organized as a vector of measurements; for example, element one of the vector might be the water level, element two the streamflow, and so on. Engineers will say that a process evolves from one state to another in time, and that a vector of measurements, also referred to as a “state vector,” represents the process state at the moment the measurements were taken. A sequence of state vectors represents a “state history.” Mathematicians will say that the state vector is a point in a “state space” having a number of dimensions equal to the number of elements in the vector. For example, eight vector elements equates to eight dimensions. Empirical modeling is the fitting of a multidimensional surface to the points arrayed in state space.

Chaos Theory proposes that a process can be optimally represented (reconstructed) by a collection of state vectors Y(t) using an optimal number of measurements, equal to “local dimension” d_L, that are spaced in time by integer multiples of an optimal time delay τ_d (Abarbanel, 1996)[3].

[3]In Chaos Theory, d_L and τ_d are called “dynamical invariants” and are analogous to the amplitude, frequency, and phase angle of periodic time series.

For a multivariate process of k independent variables, Y(t) is expressed as:

$$Y(t) = \{[x_1(t), x_1(t-\tau_{d1}),..., x_1(t-(d_{L1}-1)\tau_{d1})],..., [x_i(t), x_i(t-\tau_{di}),..., x_i(t-(d_{Li}-1)\tau_{di})],..., [x_k(t), x_k(t-\tau_{dk}),..., x_k(t-(d_{Lk}-1)\tau_{dk})]\} \quad (1)$$

i = 1 to k, where each $x(t,\tau_{di})$ represents a different dimension in state space and, therefore, a different element in a state vector. Values of d_{Li} and τ_{di} are estimated analytically or experimentally from the data. The mathematical formulations for models are derived from those for state vectors. A dependent variable of interest, y(t), can be predicted from prior measurements (also known as forecasting) of k independent variables (Roehl and others, 2000) and expressed as:

$$y(t) = F\{[x_1(t-\tau_{p1}), x_1(t-\tau_{p1}-\tau_{d1}),..., x_1(t-\tau_{p1}-(d_{M1}-1)\tau_{d1})],....,[x_i(t-\tau_{pi}), x_i(t-\tau_{pi}-\tau_{di}),..., x_i(t-\tau_{pi}-(d_{Mi}-1)\tau_{di})], ...,[x_k(t-\tau_{pk}), x_k(t-\tau_{pk}-\tau_{dk}),..., x_k(t-\tau_{pk}-(d_{Mk}-1)\tau_{dk})]\} \quad (2)$$

i = 1 to k, where F is an empirical function such as an ANN, each $x_k(t,\tau_{pi},\tau_{di})$ is a different input to F, and τ_{pi} is yet another time delay. For each variable, τ_{pi} is either: constrained to the time delay at which an input variable becomes uncorrelated to all other inputs, but can still provide useful information about y(t); constrained to the time delay of the most recent available measurement of x_k; or the time delay at which an input variable is most highly correlated to y(t). Here, the state-space local dimension d_{Lk} of equation 1 is replaced with an empirically determined model input variable dimension d_{Mk}, less than or equal to d_{Lk}. The empirically derived input variable dimension parameter d_{Mk} tends to decrease with the number of independent variables.

Limitations of the Historical Data Sets

As with any modeling effort, empirical or deterministic, the reliability of the model is dependent on the quality of the data and range of measured conditions used for training or calibrating the model. The available period of record for the data-collection networks can limit the range of streamflow, water-level, tide-range, and salinity conditions that the ANN model can accurately simulate. As noted previously, substantial changes in the salinity response of the system can occur due to a small change in streamflow (fig. 9). Although data are available from the coastal network for the early 1980s, the data were not always of a sufficient quality to use for developing empirical models. Environmental monitoring technology has changed substantially over the last 20 years. One of the most substantial changes for monitoring in estuarine systems has been improvement in the clocks in the recording equipment. For monitoring estuarine environments, it is essential that the correct time of the passing of the tidal wave, and associated physical properties, is recorded. Timing errors are analogous to physically moving the gage upstream or downstream. The timer used for paper-punch recorders in the 1980s and early 1990s often drifted from the true time. The USGS South Carolina Water Science Center started using satellite telemetry in the mid-1980s and instrumented all gages in the coastal network of the Grand Strand in the mid-1990s. The clocks associated with the satellite telemetry are much more accurate, and the gage has a limited “window” (less than 15 seconds) to transmit the recorded data to a satellite. As a result, timing errors are nearly eliminated from the real-time data.

The timing errors of the early data often are not apparent upon inspection of a time series. Prediction errors (differences between measured and simulated values) of preliminary specific-conductance models showed larger errors using the data prior to 1995 than the data after 1995. It was decided to use the data after 1995 for the development of the ANN models. The exceptions were two stations, Waccamaw River near Mt. Rena Landing near Murrells Inlet, S.C. (station 02110812) and Thoroughfare Creek at Berlin near Pawleys Island, S.C. (station 021108135), which were discontinued in 1989. A large range in hydrologic conditions occurred after 1995, including high-flow conditions with the El Niño of 1998, the extended drought from 1998 to 2002, and many hurricanes and tropical storms.

Artificial Neural Networks

Models generally fall into one of two categories: deterministic (or mechanistic) or empirical. Deterministic models are created from first-principles equations, whereas empirical modeling adapts generalized mathematical functions to fit a line or surface through data from two or more variables. The most common empirical approach is ordinary least squares (OLS), which relates variables using straight lines, planes, or hyper-planes, whether the actual relations are linear or not. Calibrating either type of model attempts to optimally synthesize a line or surface through the observed data. Calibrating models is difficult when data have substantial measurement error or are incomplete, or when the variables for which data are available provide only a partial explanation of the causes of variability. The principal advantages that empirical models have over deterministic models are that they can be developed much faster and are more accurate when the modeled systems are well characterized by data. Empirical models, however, are prone to problems when poorly applied. Overfitting and multicollinearity caused by correlated input variables can lead to invalid mappings between input and output variables (Roehl and others, 2003).

An ANN model is a flexible mathematical structure capable of describing complex nonlinear relations between input and output data sets. The structure of ANN models is loosely based on the biological nervous system (Hinton, 1992). Although numerous types of ANNs exist, the most commonly used type of ANN is the multilayer perceptron (MLP) (Rosenblatt, 1958). As shown in figure 11, 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 neurons, 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.

Typically, linear transfer functions are used to simply scale input values from the input layer to the hidden layer and generally fall within the range that corresponds to the most linear part of the s-shaped sigmoid transfer functions used from the hidden layer to the output layer (fig. 11). Each connection has a "weight," wi, 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 composed 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). Jensen (1994) describes the details of the MLP ANN, the type of ANN used in this study. MLP ANNs can synthesize functions to fit high-dimension, nonlinear multivariate data. Devine and others (2003) and Conrads and Roehl (2005) describe their use of MLP ANN in multiple applications to model and control combined manmade and natural systems, including disinfection byproduct formation, industrial air emissions monitoring, and surface-water systems affected by point and nonpoint-source pollution.

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 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 ranges are well distributed throughout the state space of interest,
- The input variables selected by the modeler share "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 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.

Subdividing a complex modeling problem into subproblems and then addressing each is an effective means to achieving the best possible results. A collection of sub-models whose calculations are coordinated by a computer program constitutes a "super model." For the Pee Dee River study, daily and hourly ANN models (submodels) were developed for specific conductance at gages proximal to salinity intrusion. These sub-models were then incorporated into a "super-model" application that integrates the model controls, model database, and model outputs. The "super model" for the project is the **P**ee Dee **R**iver and Atlantic **I**ntracoastal Waterway **S**alinity Intrusion **M**odel (PRISM) DSS described later in the report. The ANN models and plots described and shown in this report were developed using the iQuest™ data-mining software[4] (Version

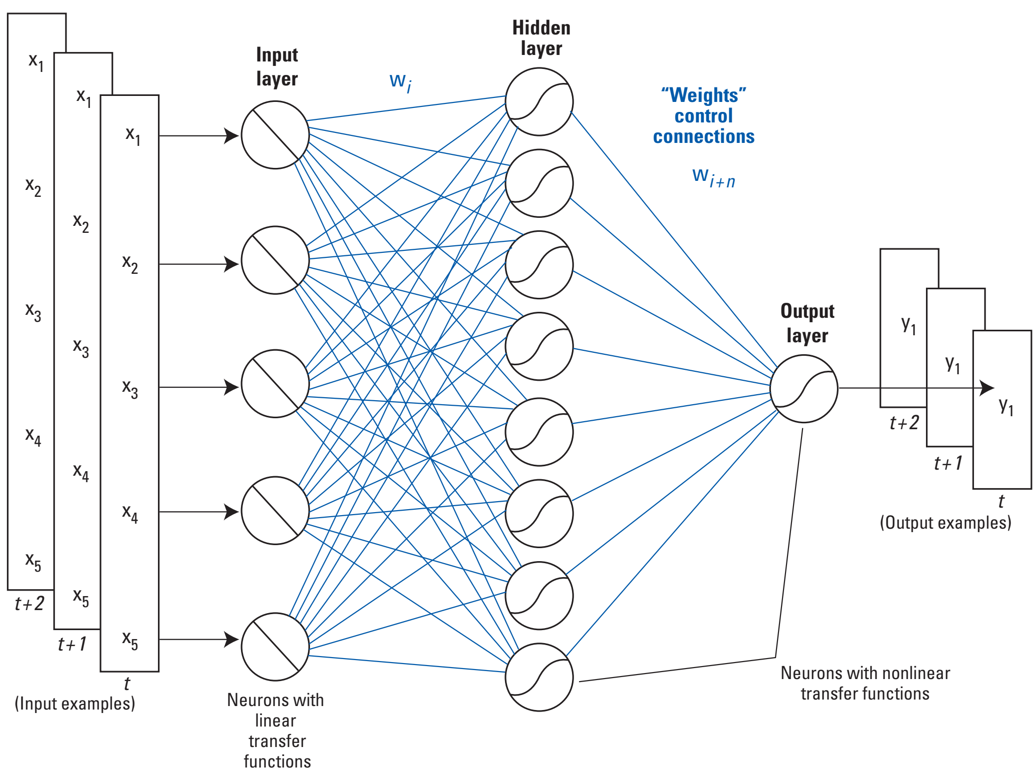


Figure 11. Multilayer perceptron artificial neural network architecture.

2.03C DM Rev31). The ANN models were deployed in the DSS using the Visual Basic run-time library of the iQuest R/T™ software.

Statistical Measures of Prediction Accuracy

The coefficient of determination (R^2), mean error (ME), root mean square error (RMSE), and percent model error (PME) were computed for the training and testing data sets for each model and are listed in table 5. Model accuracy usually is reported in terms of R^2 and commonly is 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 provide about the behavior of another variable?" For example, in the first context, an $R^2 = 0.6$ might be disappointing, whereas in the latter, it is merely an accounting of how much information is shared by the variables being used. The developers believe that specific-conductance models are unusually more accurate relative to one-dimensional, two-dimensional, and three-dimensional finite-difference models developed for comparably complex estuaries and tidal marsh systems.

The ME and RMSE statistics provide a measure of the prediction accuracy of the ANN models. The ME is a measure of the bias of model predictions—whether the model over or underpredicts the measured data. The ME is presented as the adjustment to the simulated values to equal the measured values; therefore, a negative ME indicates an over-simulation by the model, and a positive ME indicates an underprediction by the ANN model. Mean errors near zero may be misleading because negative and positive discrepancies in the simulations can cancel each other. Root mean square error addresses the limitations of ME by computing the magnitude, rather than the direction (sign) of the discrepancies. The units of the ME and RMSE statistics are the same as the simulated variable of the model.

The minimum and maximum specific-conductance values of the measured output are listed in table 5. The accuracy of the models, as given by RMSE, should be evaluated with respect to the range of the output variable. A model may have a low RMSE, but if the range of the output variable is small, the model may only be accurate for a small range of conditions, and the model error may be a relatively large percentage of the model response. Likewise, a model may have a large RMSE, but if the range of the output variable is large, the model error may be a relatively small percentage of the total model response. The PME was computed by dividing the RMSE by the range of the measured data. Generally, between the training and testing data sets, the specific-conductance models for the north end have R^2 between 0.62 and 0.92 and PME between 1.9 and 9.3 percent and the specific-conductance models for the south end have R^2 between 0.69 to 0.96 and PME between 1.8 and 7.3 percent.

[4] The iQuest™ software is exclusively distributed by Advanced Data Mining, LLC, 3620 Pelham Road, PMB 351, Greenville, SC 29615-5044, phone: (864) 201-8679, email: info@advdatamining.com, URL: *http://www.advdmi.com* .

Development of Specific-Conductance Models

The following sections describe the development of the specific-conductance models for a site. The specific conductance at nine gages was modeled. Four of the nine gages are in the north end of the system near Little River Inlet and are referred to as the "North End models" (stations 02110755, 02110760, 02110770, and 02110777). Five of the nine gages are in the south end of the system near Hagley Landing and Winyah Bay and are referred to as the "South End models" (stations 02110809, 02110812, 021108125, 021108135, 02110815). The specific-conductance models for each gage were developed in two stages. The first stage simulated the daily average specific conductance to capture the long-term dynamics of the system. The second stage simulated the higher frequency hourly specific conductance, using the predicted daily specific conductance as a carrier signal. The ANN model architecture for each gage and the use of model predictions from the "daily" model to the "hourly" model is shown in figure 12.

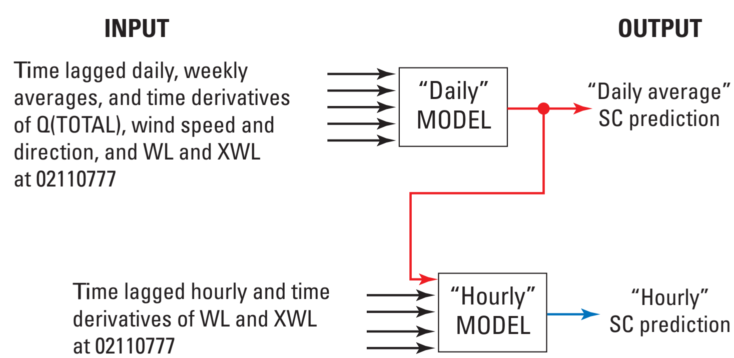


Figure 12. Artificial neural network model architecture for specific-conductance modeling at each gage.

The daily and hourly models use three general types of input signals, or time series: streamflow signal(s), water-level signal(s), and tidal range signal(s). The signals may be the daily values and(or) time derivatives of the signals. The streamflow signal is the total streamflow entering the system at station 02131000, the Pee Dee River at Pee Dee, S.C. The water-level and tidal range signals were computed from the water-level time series from station 02110777, AIW at S.C. Highway 9. The water-level data at this station are the least correlated to the streamflow time series, which makes it the

Table 5. Summary statistics for models used in the study.

[min, minimum; SC, specific conductance; max, maximum; n, number of vectors; R^2, coefficient of determination; SSE, sum of square error; ME, mean error; RMSE, root mean square error; PME, percent model error; --, no data]

Model name	Gage number	Output variable	Number of hidden neurons	Range of output variable		TRAINING DATA SET					
				Min, SC	Max, SC	n	R^2	SSE, SC	ME, SC	RMSE, SC	PME
North End Models											
psc755a-final	02110755	SC110755A	4	71	6,218	1,453	0.82	66,308,706	0.7	214	3.5
psc755h-final	02110755	SC110755	6	56	18,900	10,842	0.75	3,900,800,000	17.7	600	3.2
psc760a-final	02110760	CSC110760A	4	51	263	1,306	0.74	453,763	-1.8	19	8.8
psc760h-final	02110760	SC110760	2	55	288	7,173	0.82	3,056,943	-0.8	21	8.9
psc770a-final	02110770	SC110770A	4	64	15,956	10,758	0.88	3,666,200,000	-0.0	584	3.7
psc770h-final	02110770	SC110770	3	20	31,100	4,433	0.92	11,360,000,000	-0.2	1,601	5.2
psc777a-final	02110777	SC110777A	4	80	36,800	3,211	0.88	16,045,000,000	237.6	2,236	6.1
psc777h-final	02110777	SC110777	4	20	49,000	8,178	0.90	110,940,000,000	70.9	3,684	7.5
South End Models											
psc809a-final[a]	02110809	SC110809A	1	121	2,811	3,235	0.96	22,453,556	1.2	83	3.1
psc809h-final[a]	02110809	SC110809	2	120	7,490	3,301	0.92	75,387,428	10.0	151	2.1
psc812a-final[a]	02110812	SC110812A	2	51	1,429	25,252	0.69	64,404,895	-4.4	51	3.7
psc812h-final[a]	02110812	SC110812	3	50	6,100	2,406	0.75	116,340,000	0.3	220	3.6
psc8125a-final[a]	021108125	SC1108125A	1	68	12,962	12,173	0.90	3,033,900,000	11.1	499	3.9
psc8125h-final[a]	021108125	SC1108125	2	67	18,700	2,097	0.90	2,285,800,000	82.0	1,045	5.6
psc8135a-final[a]	021108135	SC1108135A	2	77	220	1,849	0.85	200,109	0.2	10	7.3
psc8135h-final[a]	021108135	SC1108135	2	70	880	1,897	0.87	385,496	0.6	14	1.8
psc815a-final	02110815	SC110815A	2	115	21,751	5,690	0.87	8,399,100,000	-3.6	1,215	5.6
psc815h-final	02110815	SC110815	3	60	27,050	5,580	0.91	19,844,000,000	-7.7	1,886	7.0

Table 5. Summary statistics for models used in the study. — Continued

[min, minimum; SC, specific conductance; max, maximum; n, number of vectors; R^2, coefficient of determination; SSE, sum of square error; ME, mean error; RMSE, root mean square error; PME, percent model error; --, no data]

Model name	Gage number	Output variable	Number of hidden neurons	Range of output variable		TESTING DATA SET					
				Min, SC	Max, SC	n	R^2	SSE, SC	ME, SC	RMSE, SC	PME
North End Models											
psc755a-final	02110755	SC110755A	4	62	4,390	1,220	0.62	30,726,952	-8.2	159	3.7
psc755h-final	02110755	SC110755	6	50	18,900	56,704	0.69	6,955,900,000	19.4	350	1.9
psc760a-final	02110760	CSC110760A	4	51	263	1,366	0.73	516,567	-2.1	19	9.2
psc760h-final	02110760	SC110760	2	54	291	11,158	0.80	5,038,214	-0.6	21	9.0
psc770a-final	02110770	SC110770A	4	62	15,968	57,150	0.87	26,268,000,000	-32.3	678	4.3
psc770h-final	02110770	SC110770	3	0	32,900	57,358	0.84	80,423,000,000	15.1	1,184	3.6
psc777a-final	02110777	SC110777A	4	80	36,813	56,673	0.82	668,210,000,000	-171.8	3,434	9.3
psc777h-final	02110777	SC110777	4	20	51,600	57,215	0.83	993,160,000,000	-7.1	4,166	8.1
South End Models											
psc809a-final[a]	02110809	SC110809A	1	--	--	--	--	--	--	--	--
psc809h-final[a]	02110809	SC110809	2	--	--	--	--	--	--	--	--
psc812a-final[a]	02110812	SC110812A	2	--	--	--	--	--	--	--	--
psc812h-final[a]	02110812	SC110812	3	--	--	--	--	--	--	--	--
psc8125a-final[a]	021108125	SC1108125A	1	--	--	--	--	--	--	--	--
psc8125h-final[a]	021108125	SC1108125	2	--	--	--	--	--	--	--	--
psc8135a-final[a]	021108135	SC1108135A	2	--	--	--	--	--	--	--	--
psc8135h-final[a]	021108135	SC1108135	2	--	--	--	--	--	--	--	--
psc815a-final	02110815	SC110815A	2	115	21,758	34,091	0.87	62,053,000,000	-60.4	1,349	6.2
psc815h-final	02110815	SC110815	3	50	28,100	54,865	0.85	96,869,000,000	-15.3	1,329	4.7

[a] Due to limitation of available data, only training data were used.

best independent water-level and tide range data to use as an explanatory variable.

The procedure for developing the North End and South End models follows a similar approach. A description of one of the models each for the North and South End is given below. One example is given for each group of models and is followed by a general description of the model performance for all the stations. Model summaries and variable descriptions are presented in table 6. Variables used in the ANN models are listed in Appendix 1.

Table 6. Model name and input and output variables for models used in the study.

Model	Input variable	Output variable
North End Models		
psc755a-final	IQTOTA3	SC110755A
	IQTOTA7L3	
	CWL110777A	
	CWL110777AD2	
	CWL110777AD2L2	
	ICXWL110777A	
	ICXWL110777AD2	
	ICXWL110777AD2L2	
	ICXWL110777AD2L4	
	ICXWL110777AD2L6	
psc755h-final	PSC755ALD18H	SC110755
	CWL110777	
	CWL110777D3	
	CWL110777D3(003)	
	CWL110777D3(006)	
	CWL110777D3(009)	
	ICXWL110777	
	ICXWL110777D3	
	ICXWL110777D3(003)	
	ICXWL110777D6(006)	
psc760a-final	ICXWL110777A28	CSC110760A
	IQTOTA7L2	
	IQTOTA14L9	
	IQTOTA14L23	
	IQTOTA14L37	
	CWL110777AL2	
psc760h-final	CWL110777	SC110760
	CWL110777D3	
	PSC760ALD18H	
psc770a-final	IQTOTA2	SC110770A
	IQTOTA3L2	
	CWL110777A	
	CWL110777AD2	
	CWL110777AD3L2	
	ICXWL110777A	
	ICXWL110777AD2	
	ICXWL110777AD2L2	
	ICXWL110777AD2L4	
psc770h-final	CWL110777	SC110770
	CWL110777D3	
	CWL110777D3(003)	
	CWL110777D3(006)	
	CWL110777D3(009)	
	ICXWL110777	
	ICXWL110777D3	
	ICXWL110777D3(003)	
	PSC770ALD18H	
psc777a-final	IQTOTA2	SC110777A
	IQTOTA3L2	
	CWL110777A	
	CWL110777AD2	
	CWL110777AD3L2	
	ICXWL110777A	
	ICXWL110777AD2	
psc777h-final	CWL110777	SC110777
	CWL110777D3	
	CWL110777D3(003)	
	CWL110777D3(006)	
	CWL110777D3(009)	
	ICXWL110777	
	ICXWL110777D3	
	ICXWL110777D3(003)	
	ICXWL110777D6(006)	
	PSC777ALD18H	
South End Models		
psc809a-final	PSC815A	SC110809A
psc809h-final	PSC809ALD18H	SC110809
	CWL110777	
	CWL110777D3	
	PSC809ALD18HD2	
	CWL110777D3(003)	

Table 6. Model name and input and output variables for models used in the study. — Continued

Model	Input variable	Output variable
psc812a-final	PSC815A	SC110812A
	CWL110777AL2	
	CWL110777AD2L2	
	ICXWL110777AL2	
psc812h-final	PSC812ALD18H	SC110812
	CWL110777	
	CWL110777D3	
	CWL110777D3(003)	
	ICXWL110777	
psc8125a-final	CWL110777AD2	SC1108125A
	ICXWL110777A	
	PSC815A	
	PSC815AD2	
psc8125h-final	PSC8125ALD18H	SC1108125
	CWL110777	
	CWL110777D3	
	CWL110777D3(003)	
	ICXWL110777D3	
psc8135a-final	PSC815A	SC1108135A
	PSC815AD2	
	CWL110777A	
psc8135h-final	CWL110777	SC1108135
	CWL110777D3	
	PSC8135ALD18H	
	PSC8135ALD18HD2	
psc815a-final	IQTOTA3	SC110815A
	IQTOTA7L3	
	CWL110777AL1	
	CWL110777AD3L1	
	ICXWL110777AL1	
	SPEEDA4	
	DIRECTA4	
psc815h-final	PSC815ALD18H	SC110815
	CWL110777	
	CWL110777D3	
	CWL110777D3(003)	
	ICXWL110777	
	ICXWL110777D3	
	ICXWL110777D3(003)	

In developing ANN models, 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 test data sets, but the most common is random selection of a specified percentage of the total population of measurements. For data sets with numerous salinity intrusion events, the data sets were randomly bifurcated into training and testing data sets. Percentages of the data set used for testing ranged from 5 to 54 percent. For the large data sets, typically with over 50,000 data points, a zone-average, or box, filter of the data was used to separate the data into training and testing data sets. Using the zone-average filter, all the data are used in the test data set, and a small selected sample of the data is used for the training data set. The filter separates the data set into a user-specified number of zones or boxes and determines the input vectors with the highest information content, and reserves these vectors for the training data set. The percentage of training and testing data depended on the length of the period of record and the range of hydrologic conditions in the data set. Typically, the zone averaging filter uses a small percentage of the data (less than 15 percent) for the training data set.

Short data sets often include only a few salinity intrusion events. For a behaviorally complex system, such as the salinity dynamics in the AIW and Waccamaw River, it was deemed too risky to set aside data for independent testing of ANN performance. To do so would prevent the ANN models from learning from data representing unique and possibly important behavioral states. For these stations, all the data, from the average salinity conditions to the few salinity intrusion events, were used in the model training data set. The training of ANN models fits nonlinear surfaces through the data. The "flexibility" of the surfaces is determined by the modeler, typically through the number of hidden neurons and the training parameter rates. Highly nonlinear, or highly "flexible," surfaces tend to over-fit the data and produce models with erroneous results. Sparse or noisy data are prone to over-fitting if surface fits are made overly complex. The complex surfaces may fit the limited data but are not representative of behavior of the system.

To mitigate the extrapolation and sparseness issues, the ANN models were conservatively trained using a method called "stop training" to both fit the data and extrapolate in a minimally nonlinear, 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. The data-mining software 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 a progressively nonlinear fit. This inflection point was used to trigger stop training.

North End Models

The daily specific-conductance model (psc770a-final) for AIW at Grand Strand Airport (station 02110770) uses streamflow, water-level, and tidal-range inputs (table 6). The streamflow inputs are 2-day moving window average flow (IQTOTA2) and 3-day moving window average flow lagged 2 days (IQTOTA3L2). The water-level data inputs are daily water levels from the AIW at S.C. Highway 9 (CWL110777A), the time derivative of the 2-day change in water level (CWL110777AD2), and the 3-day change in water level lagged 2 days (CWL110777AD3L2). The tidal-range inputs are daily tidal range (ICXWL110777A), the time derivatives of the 2-day change in tidal range (ICXWL110777AD2), and the 2-day change in tidal range lagged 2 and 4 days (ICXWL110777AD2L2 and ICXWL-110777AD2L4, respectively). For testing and training the daily model, 67,908 data values were available, and about 16 percent (10,758 data points) were used for training. The coefficients of determination, R^2, for the training and testing were 0.88 and 0.87, respectively (table 5). The daily model used four hidden-layer neurons.

The hourly specific-conductance model (psc770h-final) uses the simulated daily specific conductance from the daily specific-conductance model and water-level and tidal inputs from the AIW at S.C. Highway 9 gage. The simulated daily specific-conductance input, lagged 18 hours (PSC770ALD18H), captures the long-term movement of the specific conductance that is characterized by the streamflow and the chaotic components of the water-level signal. Water-level and tidal range data and time derivatives and time delays of these inputs are used to capture the semidiurnal tidal signal. The water-level inputs include hourly water levels at the AIW at S.C. Highway 9 (CWL110777), the 3-hour change in water level (CWL110777D3), and the 3-hour change in water level lagged 3, 6, and 9 hours (CWL110777D3(003), CWL110777D3(006), and CWL-110777D3(009), respectively). Tidal-range inputs include the hourly tidal range (ICXWL110777), the 3-hour change in tidal range (ICXWL110777D3), and the 3-hour change in tidal range lagged 3 hours (ICXWL110777D3(003)). For testing and training the hourly model, 61,791 data values were available, and about 7 percent (4,433 data points) were used for training. The coefficients of determination, R^2, for the training and testing were 0.92 and 0.84, respectively (table 5). The daily model used three hidden-layer neurons.

The measured and simulated daily specific-conductance values for the four North End specific-conductance models are shown in figure 13. The daily models are able to simulate the timing and range of the salinity intrusion event well. The measured and simulated hourly specific-conductance values for the four North End models are shown in figure 14.

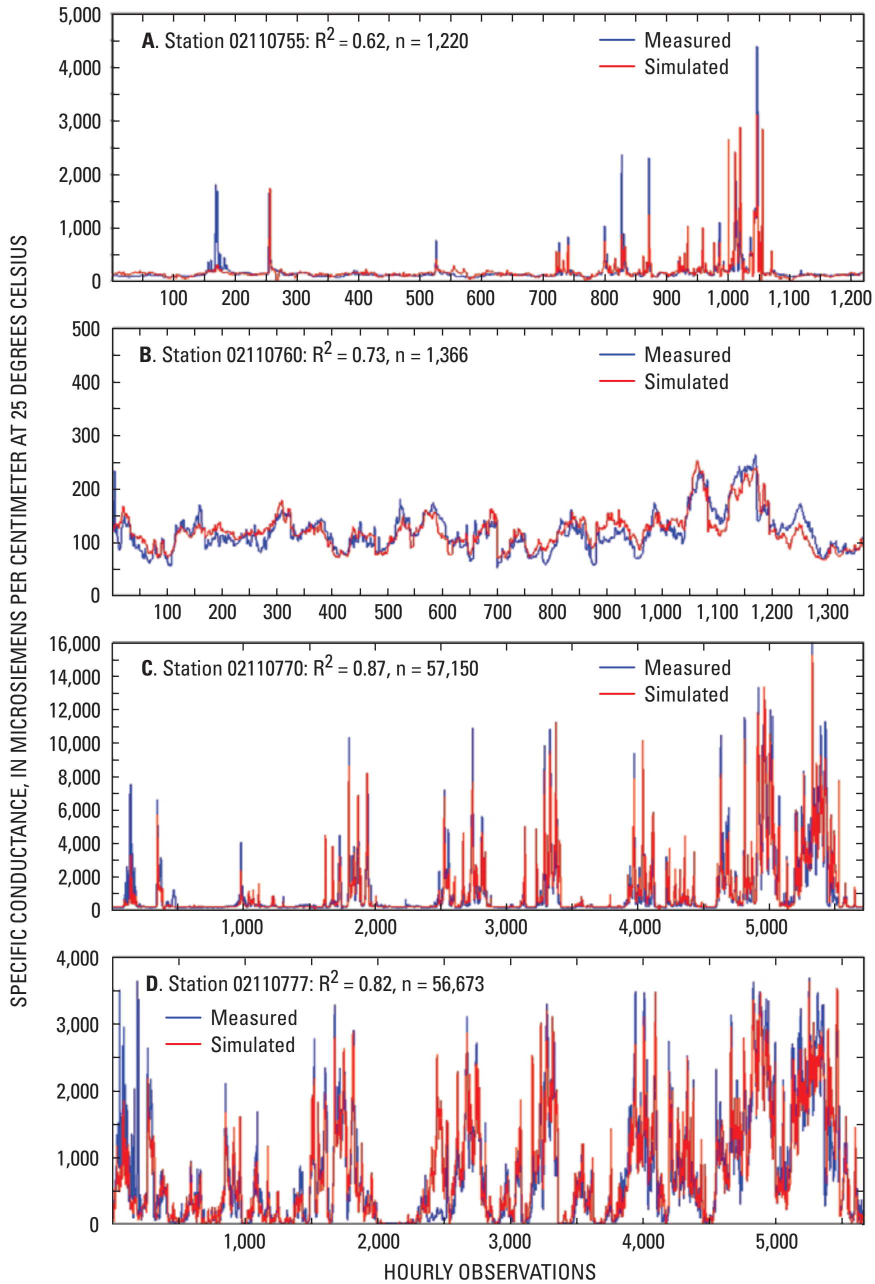


Figure 13. Measured and simulated specific-conductance values from the daily model for four U.S. Geological Survey gages on the Atlantic Intracoastal Waterway. Data were bifurcated into training and testing data sets. Graphs show hourly observations from the testing data set along with the coefficient of determination (R^2) and number (n) of data points.

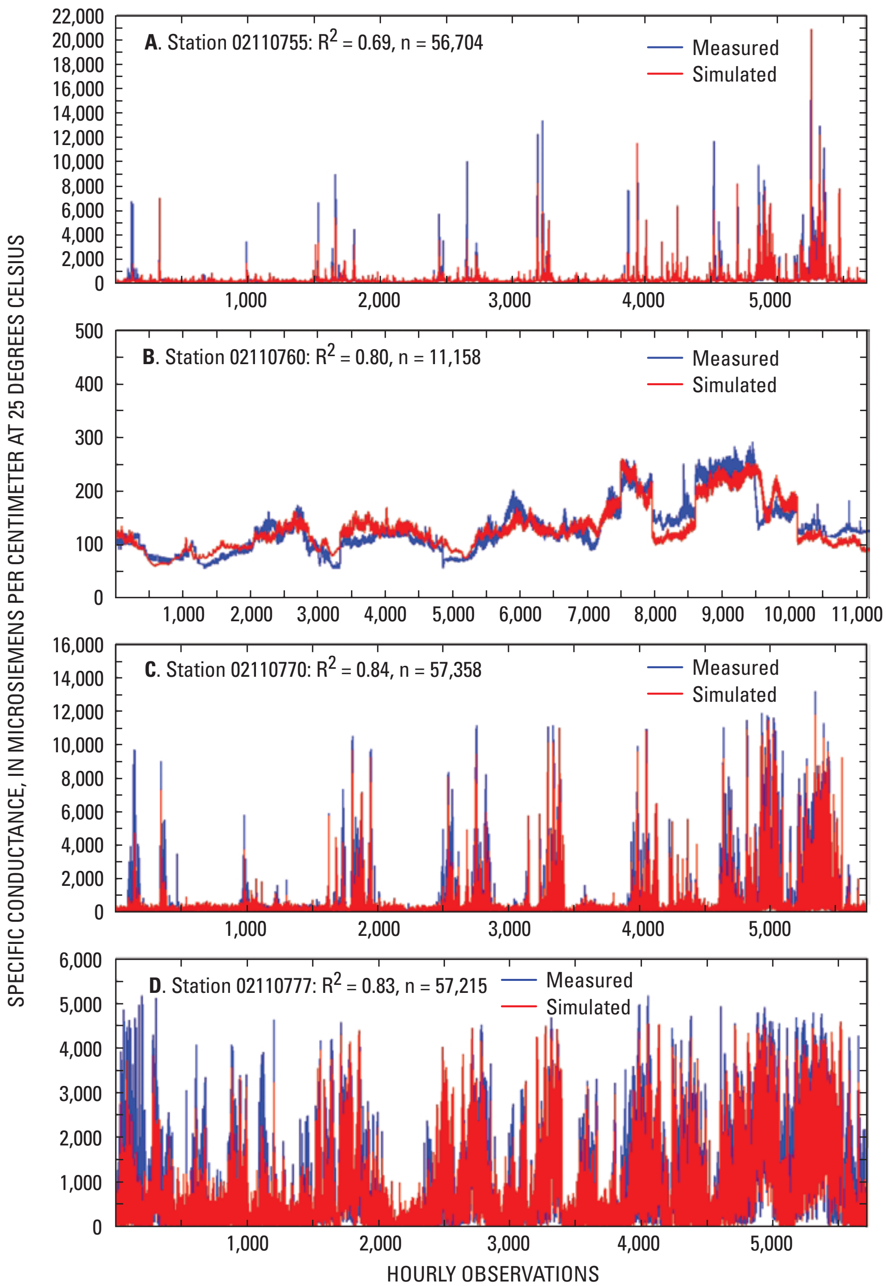


Figure 14. Measured and simulated specific-conductance values from the hourly model for four U.S. Geological Survey gages on the Atlantic Intracoastal Waterway. Data were bifurcated into training and testing data sets. Graphs show hourly observations from the testing data set along with the coefficient of determination (R^2) and number (n) of data points.

South End Models

For four of the five stations for the South End Models (02110815, 021108135, 021108125, 02110812, and 02110809), limited data were available for training and testing the ANN models (table 3). Waccamaw River at Hagley Landing (station 02110815) has data from 1986 to 2007, but the other four stations have limited data. Some stations were installed in 2002 near the end of the recent drought. To develop good models for these sites with limited data, the predicted daily specific-conductance values from the Waccamaw River at Hagley Landing model were used for input values to four of the South End models (021108135, 021108125, 02110812, and 02110809).

The daily specific-conductance model (psc815a-final) for Waccamaw River at Hagley Landing uses streamflow, water-level, and tidal range inputs (table 6). The streamflow inputs are the 3-day moving window average flow (IQTOTA3) and the 7-day moving window average flow lagged 3 days (IQTOTA7L3). The water-level data inputs, which are daily water levels from the AIW at S.C. Highway 9 lagged 1 day (CWL110777AL1) and the time derivative of the 3-day change in water level lagged 1 day (CWL110777AD3L1). The tidal range input is daily tidal range lagged 1 day (ICXWL110777AL1). The daily specific-conductance model also uses the 4-day moving window average wind speed (SPEEDA4) and wind direction (DIRECTA4). The zone-average filter separated the data into training and testing data sets. All the data are in the testing data set (34,091 data points), and 5,690 data points (17 percent) are used for training the model. The coefficient of determination, R^2, for both the training and testing, was 0.87 (table 5). The daily model used two hidden-layer neurons.

The hourly specific-conductance model (psc815h-final) uses the simulated daily specific conductance from the daily specific-conductance model and water-level and tidal inputs from the AIW at S.C. Highway 9 gage. The simulated daily specific-conductance input, lagged 18 hours (PSC815ALD18H), captures the long-term movement of the specific conductance that is characterized by the streamflow and the chaotic components of the water-level signal. Water-level and tidal range data and time derivatives and time delays of these inputs are used to capture the semidiurnal tidal signal. The water-level inputs include hourly water level at the AIW at S.C. Highway 9 (CWL110777), the 3-hour change in water level (CWL110777D3), and the 3-hour change in water level lagged 3 hours (CWL110777D3(003)). The tidal range inputs include hourly tidal range at AIW at S.C. Highway 9 (ICXWL110777), the 3-hour change in water level

(ICXWL110777D3), and the 3-hour change in water level lagged 3 hours (ICXWL110777D3(003)). The zone-average filter separated the data into training and testing data sets. All the data are in the testing data set (54,865 data points), and 5,580 data points (10 percent) are used for training the model. The coefficients of determination, R^2, for the training and testing were 0.91 and 0.85, respectively (table 5). The measured and simulated daily and hourly specific-conductance values for the Waccamaw River at Hagley Landing models are shown in figure 15. The daily models are able to simulate the timing and range of the salinity intrusion event well.

The daily specific-conductance model (psc8125a-final) for the Pawleys Island gage uses water-level and tidal range inputs from AIW at S.C. Highway 9 and simulated specific conductance from the Waccamaw River at Hagley Landing model for inputs (table 6). The water-level data input is the 2-day change in daily water level (CWL110777AD2). The tidal range input is the daily tidal range (ICXWL110777A). Simulated specific-conductance inputs for station 02110815 include the simulated daily values (PSC815A) and the 2-day change in daily specific-conductance values (PSC815AD2). As discussed previously, due to the limited data and the limited number of salinity-intrusion events, the data were not bifurcated into training and testing data sets. For training the daily model, 12,173 data values were available. The coefficient of determination, R^2, for the training data set was 0.90 (table 5). The daily model used one hidden-layer neuron.

The hourly specific-conductance model (psc8125h-final) uses the simulated daily specific conductance from the daily specific-conductance model and water-level and tidal inputs from the AIW at S.C. Highway 9 gage. The simulated daily specific-conductance input, lagged 18 hours (PSC815ALD18H), captures the long-term movement of the specific conductance that is characterized by the streamflow and the chaotic components of the water-level signal. Water-level and tidal-range data and time derivatives and time delays of these inputs are used to capture the semidiurnal tidal signal. The water-level inputs include hourly water level at AIW at S.C. Highway 9 (CWL110777), the 3-hour change in water level (CWL110777D3), and the 3-hour change in water level lagged 3 hours (CWL110777D3(003)). The tidal-range input is the 3-hour change in hourly tidal range (ICXWL110777D3). For training the daily model, 2,097 data values were available. The coefficient of determination, R^2, for the training data sets was 0.90 (table 5). The hourly model used two hidden-layer neurons. The measured and simulated daily and hourly specific-conductance values for the four models using simulated inputs from the Waccamaw River at Hagley Landing model are shown in figures 16 and 17, respectively. The daily models are able to simulate the timing and range of the salinity intrusion event well.

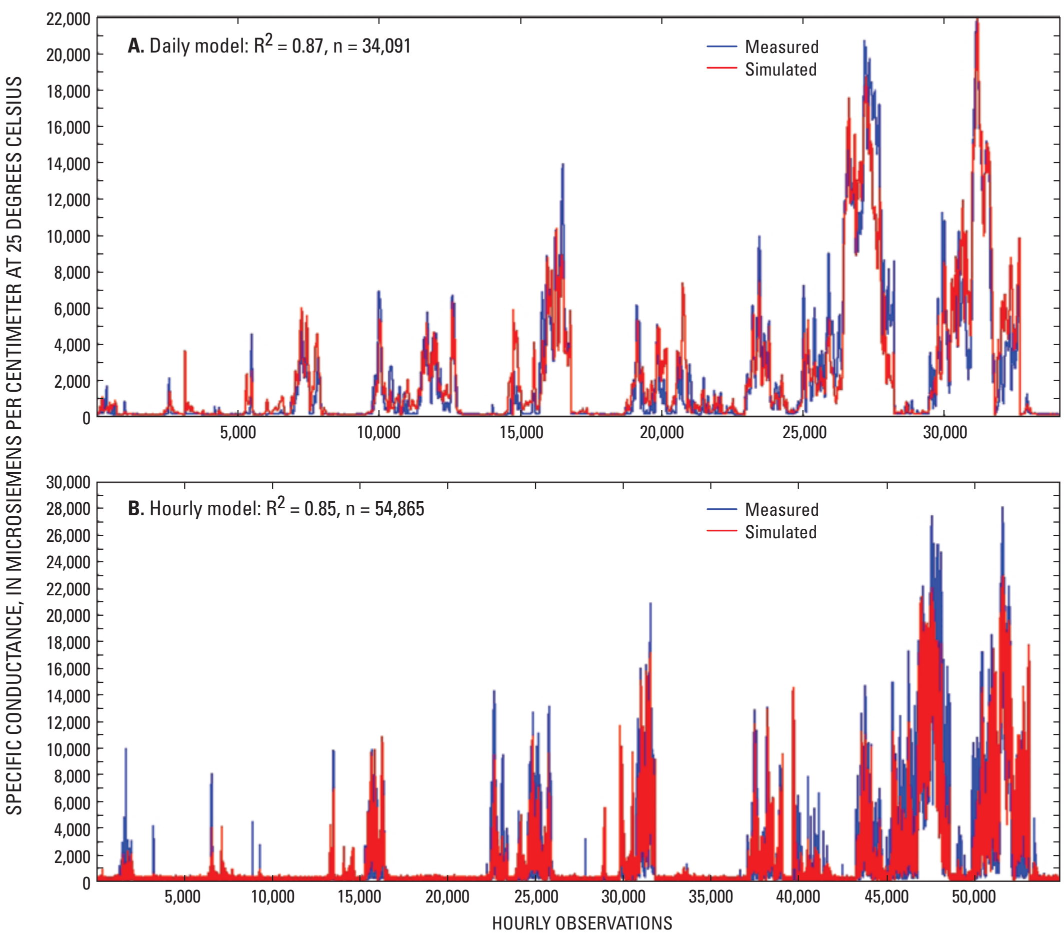


Figure 15. Measured and simulated specific-conductance values for Waccamaw River at Hagley Landing for the (A) daily model and (B) hourly model. Data were bifurcated into training and testing data sets. Graphs show hourly observations from the testing data set along with the coefficient of determination (R^2) and number (n) of data points.

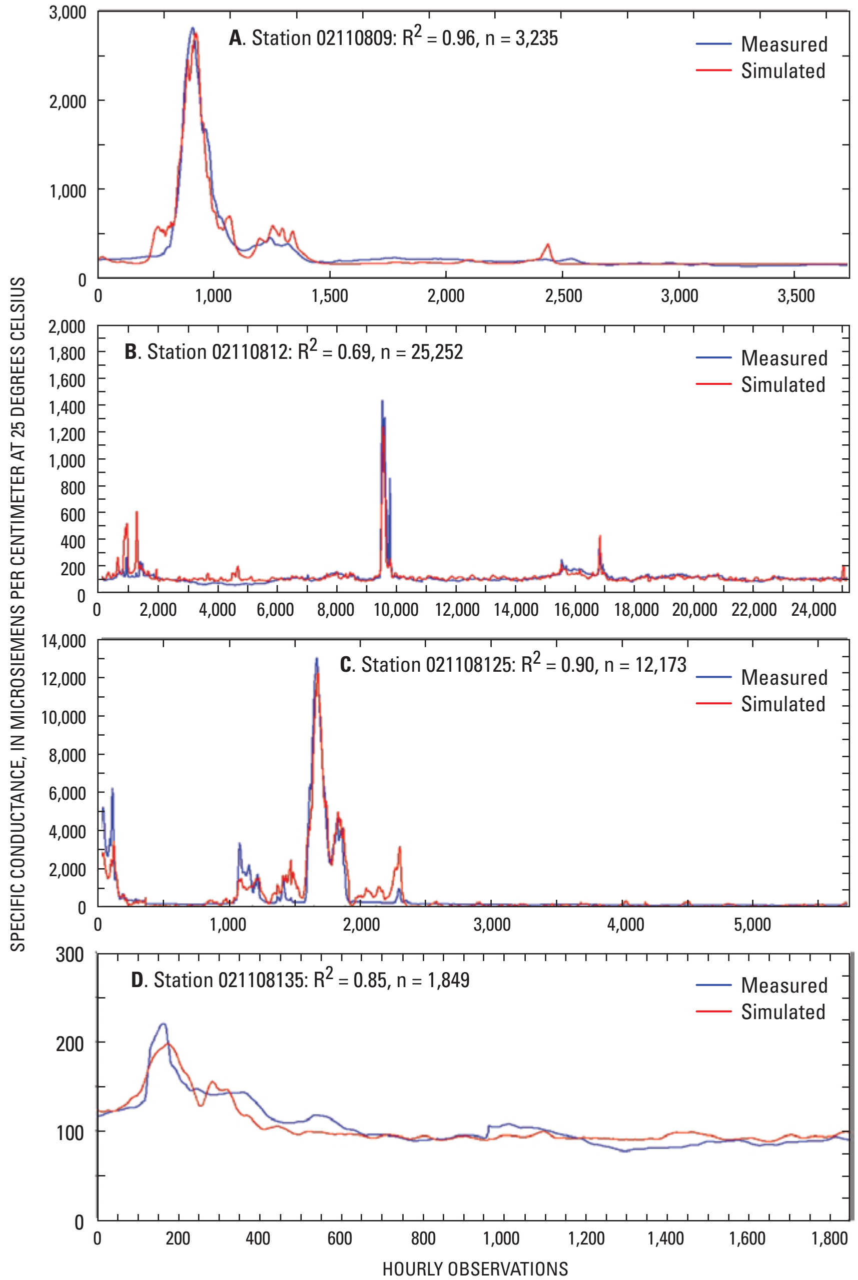


Figure 16. Measured and simulated specific-conductance values from the daily model for four U.S. Geological Survey gages on the Waccamaw River and Thoroughfare Creek. Graphs show hourly observations from the training data set along with the coefficient of determination (R^2) and number (n) of data points.

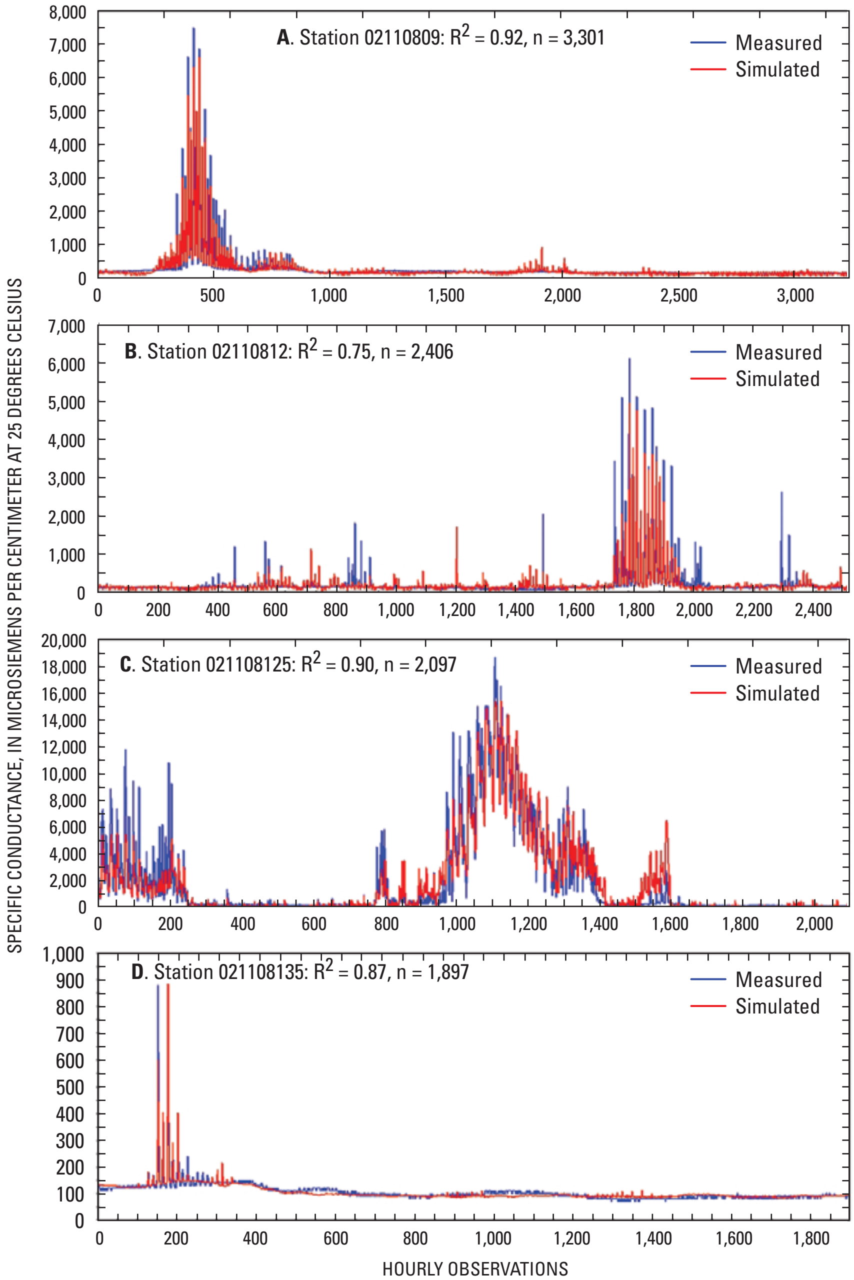


Figure 17. Measured and simulated specific-conductance values from the hourly model for four U.S. Geological Survey gages on the Waccamaw River and Thoroughfare Creek. Graphs show hourly observations from the training data set along with the coefficient of determination (R^2) and number (n) of data points.

Analysis of Estuary Dynamics Using Three-Dimensional Response Surfaces

Salinity intrusions typically occur when low streamflows are accompanied by high mean water levels or changes in tidal range (fig. 8). Two-dimensional (2D) plots show the interaction between one explanatory variable (flow, water level, or tide range) and a response variable (specific conductance), but often the interaction of two explanatory variables is difficult to discern. Figure 18 shows 2D plots of the specific conductance at Waccamaw River at Hagley Landing, streamflow at the Pee Dee River at Pee Dee, S.C., and gage height and tidal range at the AIW at S.C. Highway 9 for calendar years 2001 and 2002. The specific-conductance response to changing flow conditions (fig. 18A) is easily discerned, but the responses to changing gage heights (fig. 18B) and tidal ranges (fig. 18C) are difficult to discern from visual inspection of the time-series plots because of the lack of a distinct pattern in the water-level

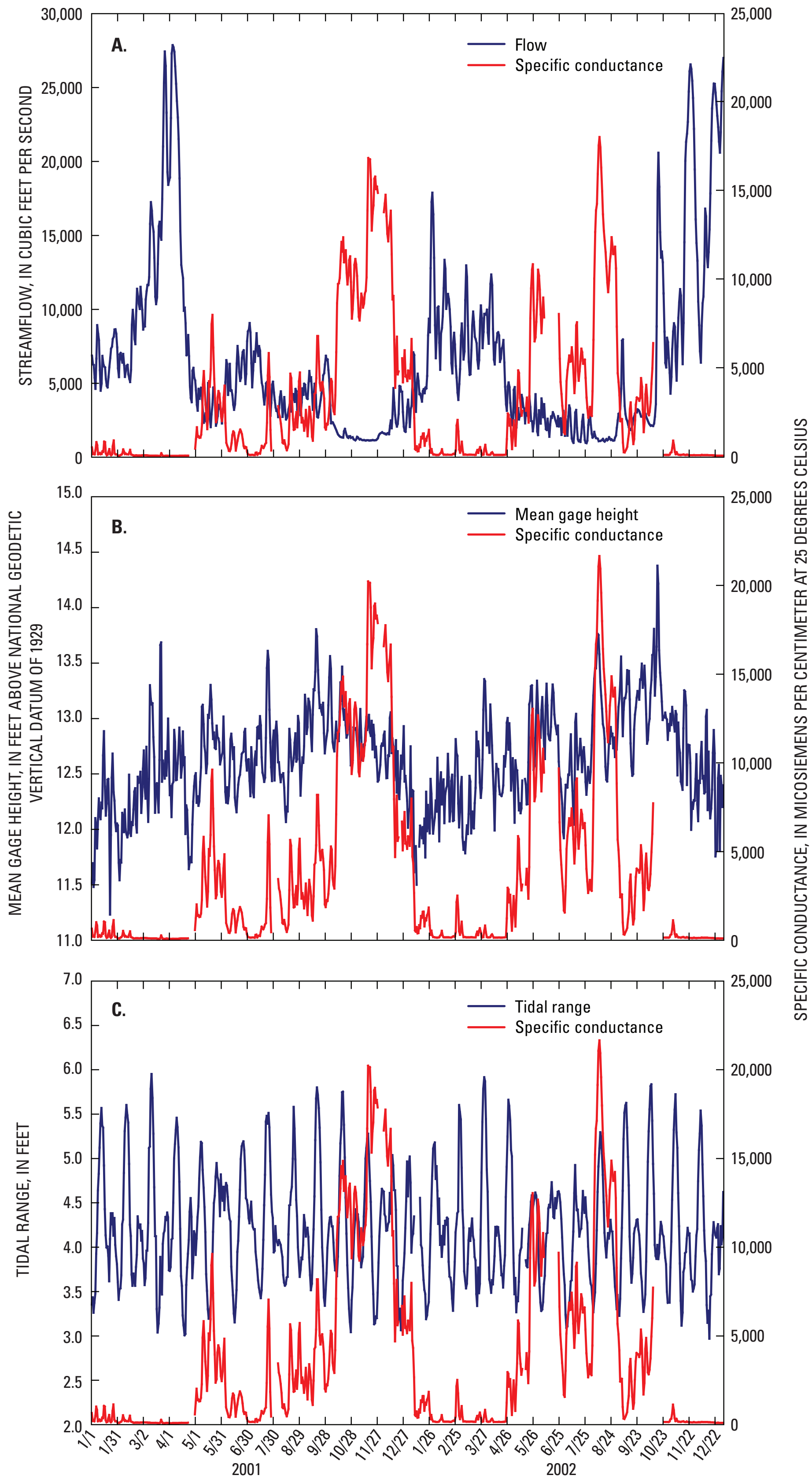


Figure 18. Daily specific conductance at Waccamaw River at Hagley Landing and (A) flow at Pee Dee River at Pee Dee, South Carolina, and (B) gage height and (C) tidal range at the Atlantic Intracoastal Waterway at South Carolina Highway 9 from January 2001 to December 2002.

time series. In the tidal-range time series, there appears to be some correlation with timing of the spring tides, but the largest salinity intrusion events do not occur during the highest spring tides.

The specific-conductance ANN models can be used to examine salinity intrusion in the Waccamaw River and AIW and to determine the hydrologic and tidal conditions that cause large salinity intrusions. Three-dimensional (3D) response surfaces can be generated to plot two explanatory variables (among water level, tidal range, and streamflow) with a response variable (specific conductance). The data for the response surface is computed by the ANN model across the full ranges of the displayed input variables, while the "unshown" inputs (all the models have more than two inputs with the exception of psc809a-final) are set to a constant value, such as a historical minimum, mean, or maximum value. The response surface is a manifestation of the historical interaction of the parameter at the site and, therefore, a valuable tool for understanding the physics of a system and for comparing how variable interactions differ during various hydrologic and tidal conditions.

The causes of the large salinity intrusion events at the Waccamaw River at Hagley Landing gage can be understood by comparing a series of 3D response surfaces showing the interaction of flow, water level, and tide range (figs. 19 and 20). The specific-conductance response to total daily inflow to the system and water level for neap and spring tide conditions are shown in the two response surfaces in figure 19. The response surface on the right shows the gage response under spring tide conditions. The neap tide response surface shows a slight increase in specific-conductance response for low-flow conditions and high water levels as compared to the spring tide response surface. The response surfaces show salinity intrusion occurs during low-flow and high water-level conditions. Also shown is a small increase in the maximum intrusion during neap tides (maximum approximately 10,000 µS/cm) as compared to spring tides (maximum approximately 8,000 µS/cm).

Alternatively, the salinity intrusion during high water levels can be seen by plotting total inflow and tide ranges for low and high water conditions (fig. 20). The low water response surface shows little specific conductance response through all streamflow and tidal conditions. The high water 3D response surface shows large specific-conductance response during low streamflow and all tide. The low water 3D response surfaces show that for all flow conditions and tidal ranges during low water level, the specific-conductance response is negligible (fig. 20). The high water 3D response surface shows the interaction of total inflow and tidal range during high water conditions and shows that the large intrusion events occur during the convergence of low flow and high water conditions for all tidal ranges (fig. 20).

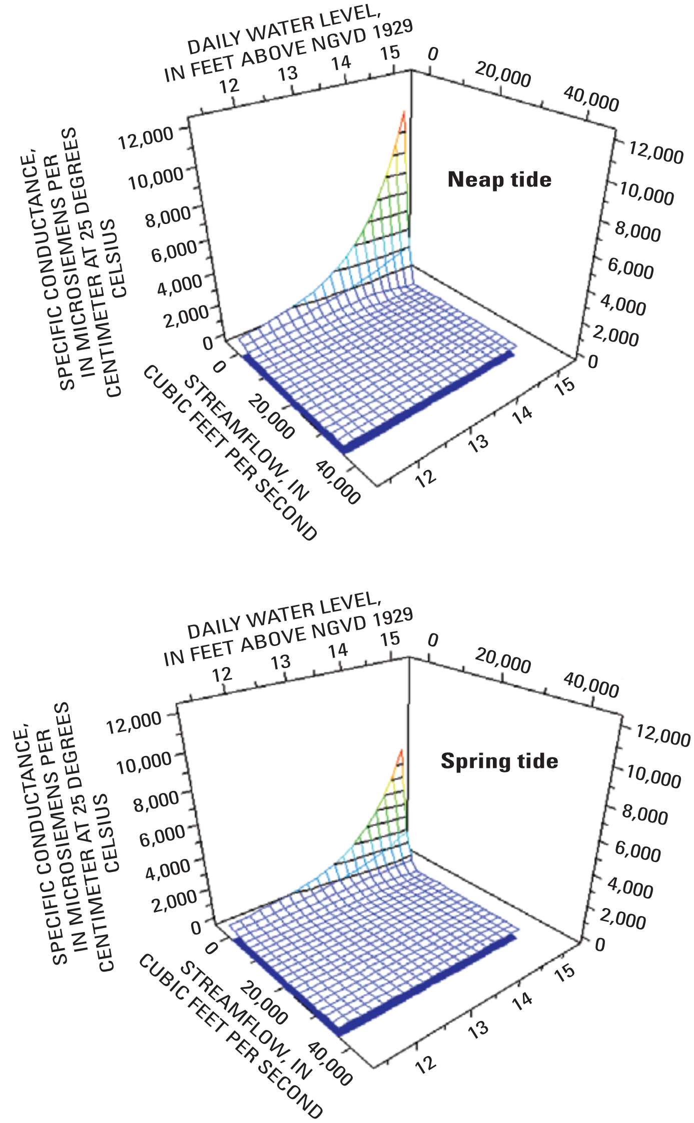


Figure 19. The interaction of daily water level and 3-day moving window average streamflow on daily specific conductance for Waccamaw River at Hagley Landing for neap tide and spring tide conditions.

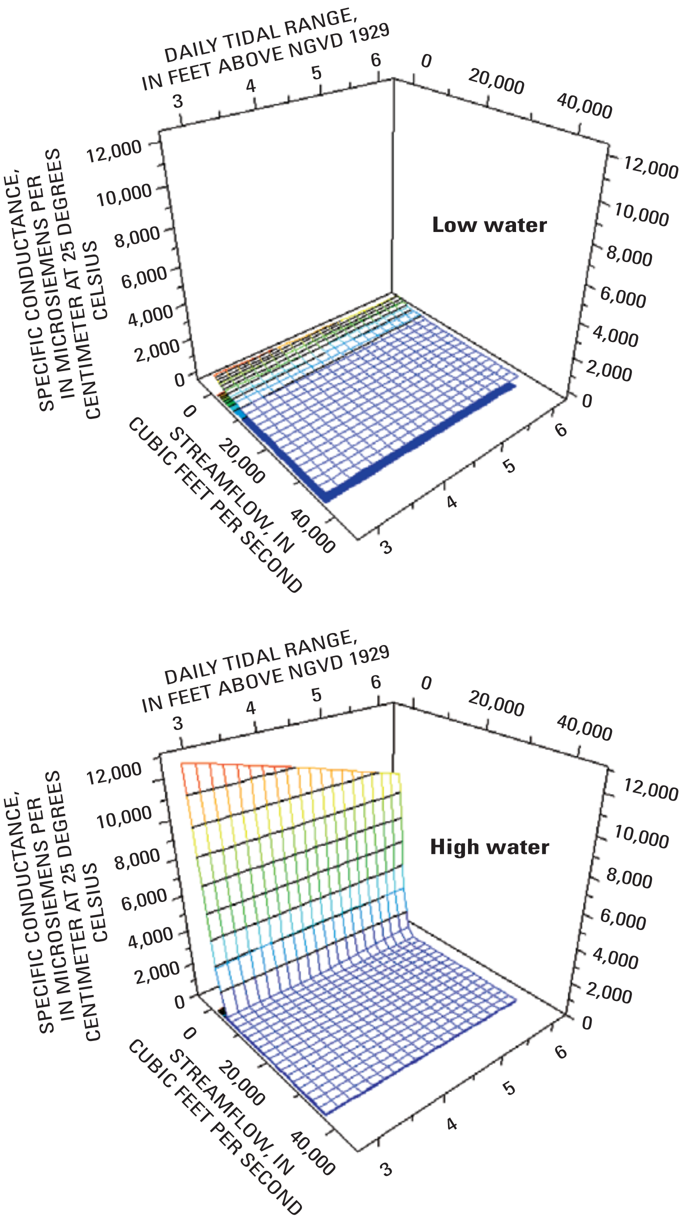


Figure 20. The interaction of daily tidal range and 3-day moving window average streamflow on daily specific conductance for Waccamaw River at Hagley Landing for low-water and high-water conditions. (Note: The small z-axis (vertical scale) was used for the response surface shown in figures 19 and 20.)

Development of the Decision Support System

Natural-resource managers and stakeholders face difficult challenges when managing interactions between natural and manmade systems. Even though the collective interests and computer skills of the community of managers, scientists, and other stakeholders are quite varied, equal access to the scientific knowledge is needed for them to make the best possible decisions. Dutta and others (1997) define decision support systems (DSSs) as, "systems 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." Environmental resource managers commonly use complex mathematical (mechanistic) models based on first principle physical equations to evaluate options for using the resource without damage. While there appears to be no strict criteria that distinguish a DSS from other types of programs, Dutta and others (1997) suggest that artificial intelligence (AI) is a characteristic of more advanced DSSs: "With the help of AI techniques DSSs have incorporated the heuristic models of decision makers and provided increasingly richer support for decision making. Artificial intelligence systems also have benefited from DSS research as they have scaled down their goal from replacing to supporting decision makers."

The authors of this report have previously developed two DSSs in South Carolina and Georgia to support the permitting of three water reclamation facilities that discharge into South Carolina's Beaufort River estuary (Conrads, Roehl, and Martello, 2002; Conrads and others, 2003; Conrads and Roehl, 2005) and to evaluate the environmental effects of a proposed deepening of the Savannah Harbor (Conrads and others, 2006). These DSSs are spreadsheet applications that provided predictive models with real-time databases for ANN model simulation, graphical user interfaces, and displays of results. Additional features include optimizers, integrations with other models and software tools, and color contouring of simulation output data. These features make the DSSs easily distributable and immediately usable by all water-resource managers.

The development of a DSS for the Pee Dee and Waccamaw Rivers and AIW required a number of steps (described previously), including: (1) merging all the data into a single comprehensive database; (2) developing specific-conductance ANN sub-models; and (3) developing a Microsoft Office Excel™ application that integrates the new database, ANN sub-models, and optimization routine into a single package that is easy to use and disseminate.

Architecture

The basic architectural elements of the PRISM DSS are shown in figure 21. The DSS reads and writes files for the various run-time options that can be selected by the user through the system's GUI. A historical database, containing 7½ years of hydrodynamic data, is read into the simulator along with the ANN sub-models at the start of a simulation. By using GUI controls, the user can evaluate alternative flow scenarios. The outputs generated by the simulator are written to files for post processing in Microsoft Office Excel™ or

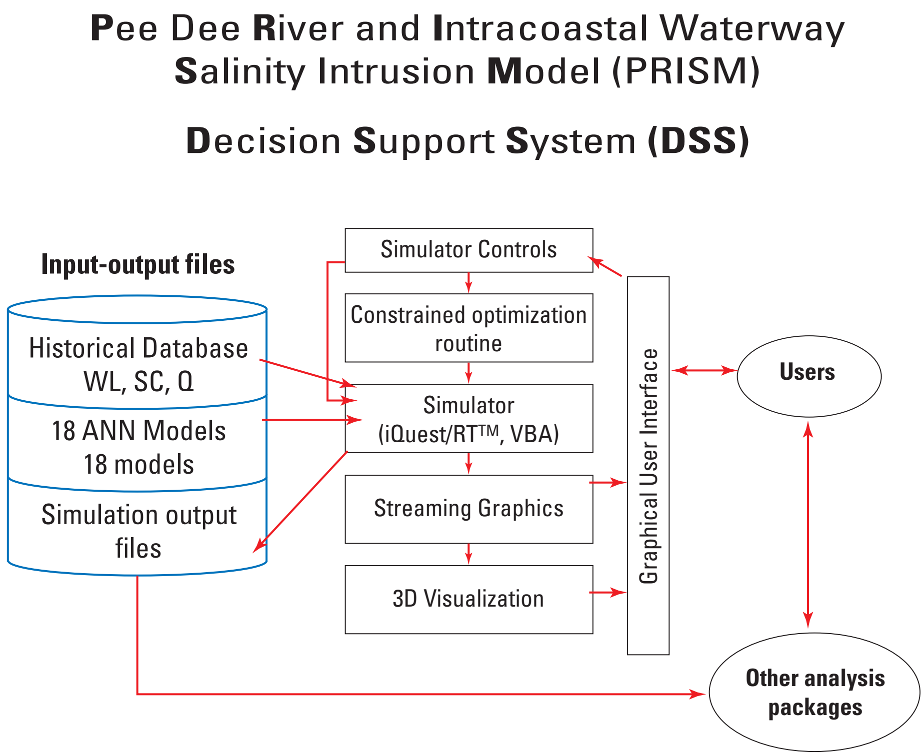


Figure 21. Architecture of the PRISM decision support system.

other analysis software packages. The DSS also provides streaming graphics for each gage during simulations and 3D visualization of the specific-conductance response for the North End and South End models.

Historical Databases

Review of the measured data in the historical databases was necessary for quality control due to a variety of problems previously discussed, including erroneous and missing values and phase shifts. The resulting database comprises 7½ years of hourly data (approximately 66,000 time stamps) for 29 variables. A summary of the historical databases stored in the DSS is described below.

- Pee Dee River and tributary streamflow and the AIW at S.C. Highway 9 water levels—7½ years of water-level measurements in the AIW near Little River Inlet and river and major tributary flows measured by the USGS.
- Riverine water level and specific conductance—7½ years of hourly measurements collected in the AIW (four north end gages) and Waccamaw River (five south end gages) by the USGS.
- Wind speed and direction—7½ years of hourly wind-speed and wind-direction data collected by the National Weather Service.

Model Simulation Control, Optimization Routine, Streaming Graphics, and Three-Dimensional Visualization Program

The simulator in the PRISM DSS integrates the historical database with the 18 ANN models. The date/time controls on the user control panel (fig. 22) are used to adjust start and end dates and time-step size for a simulation. The simulator allows the user to run “what-if?” simulations by varying the streamflow from its historical values. Two types of inputs to a mode are: (1) controllable variables, such as the regulated flows from the North Carolina reservoirs, and (2) uncontrollable state variables, such as the tidal water levels and tidal range. To evaluate alternative courses of action, the controllable inputs can be manually manipulated by the user, while the uncontrollable and constantly changing variables representing water level, tidal range, and unregulated freshwater flows are set to their historical values. The user has four simulation input variable options:

- percentage of historical streamflow to the system,
- user-set streamflow to a constant value,
- user-defined hydrograph, and
- optimized streamflow to meet specific conductance threshold.

Explanations of how to use each of the options in the PRISM DSS can be found in the user’s manual in appendix 2.

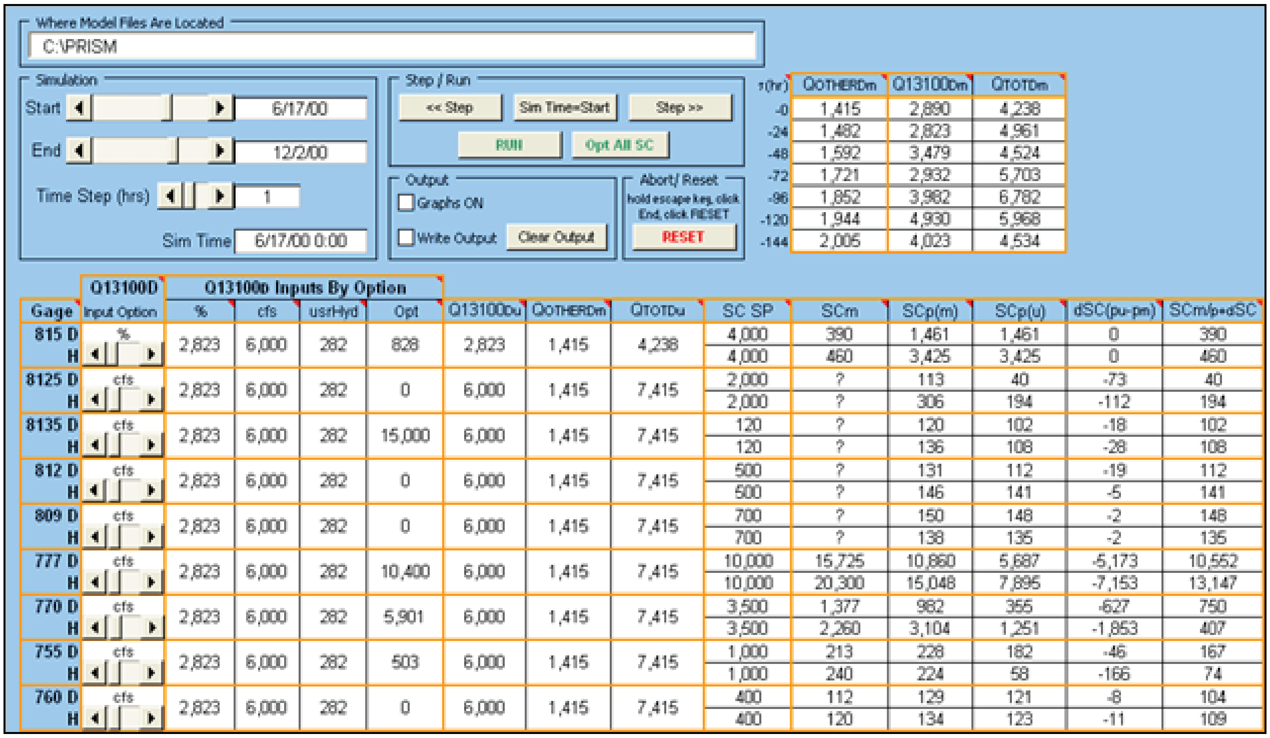


Figure 22. Simulator controls used to set parameters and run a simulation in the PRISM decision support system.

An alternative to manually setting flow inputs is to use another program called a "constrained optimizer" (fig. 23). An optimizer can be coupled to a model to have it automatically predict a specified output, called a setpoint, as state variables change over time. For each time step, state variable values are input to the model, and the optimizer iteratively modulates the controllable variable until the model achieves the setpoint value within a specified tolerance. Limits (constraints) can be placed on the range of values that the controllable variables can have. In PRISM, an optimizer based on the Secant Algorithm described by Burden and others (1978) can be used with the models for each gage. The optimizer can greatly reduce the number of simulations needed to determine the flow required for the Pee Dee River at Pee Dee, S.C., to control the specific conductance at a gage with widely varying water levels, tidal ranges, and unregulated flows.

For each gage, a worksheet displays a streaming graphics while a simulation is running for any four simulated variables selected by the user (fig. 24). The graphs display simulation input data (flow, water level, tide range, and wind speed and direction) and output data, including the historical measured data, simulated historical conditions (to show model accuracy), and the simulated output using streamflow set by the user using the GUI controls or an input file.

The PRISM's three-dimensional visualization (3DVis) worksheet provides graphical specific-conductance profiles at the south and north ends of the study area. It is designed to visualize and animate periods of special interest. Data and the controls for operating the 3DVis worksheet are on the left side of the 3DVis worksheet (fig. 25). The data are a subset of that on the "Run" worksheet (appendix 2) and are provided for reference while using the 3DVis worksheet. Two plots for the north end of the study area are shown in figure 25. The left plot shows the specific-conductance profile representing the actual historical data (when available), and the right plot shows the profile predicted by the gage model pairs using the user-specified flow condition. Note that the predicted

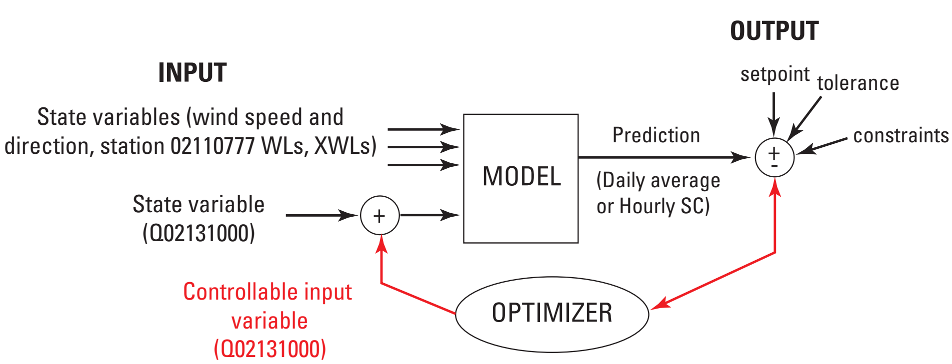


Figure 23. Optimization routine used to modulate a controllable input variable to predict an output variable in the PRISM decision support system.

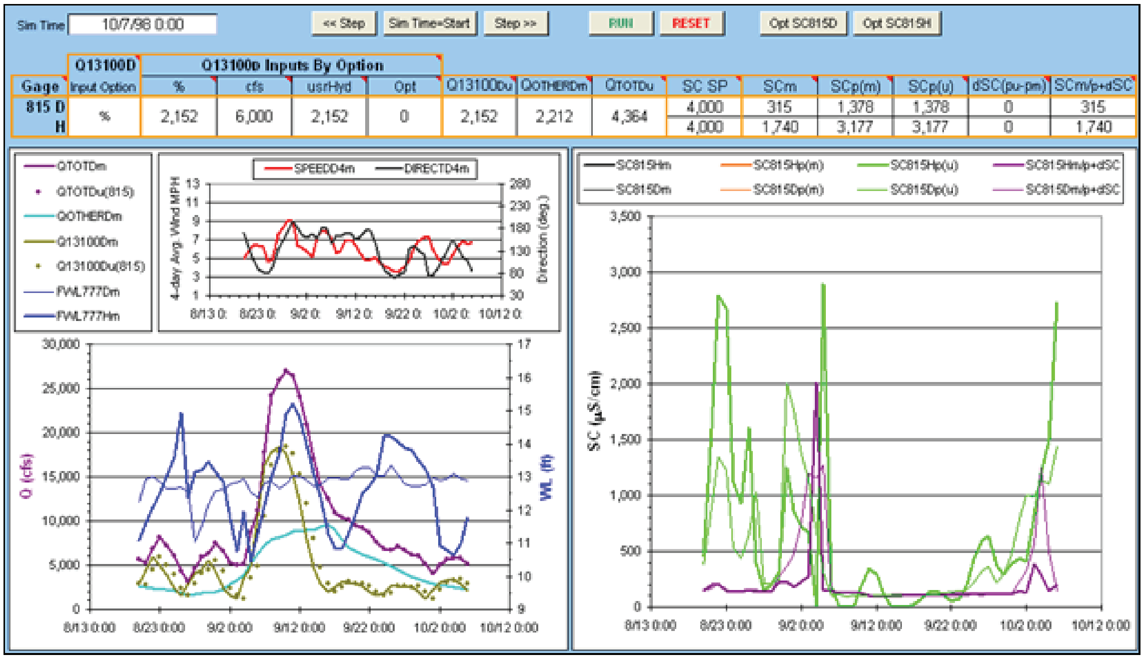
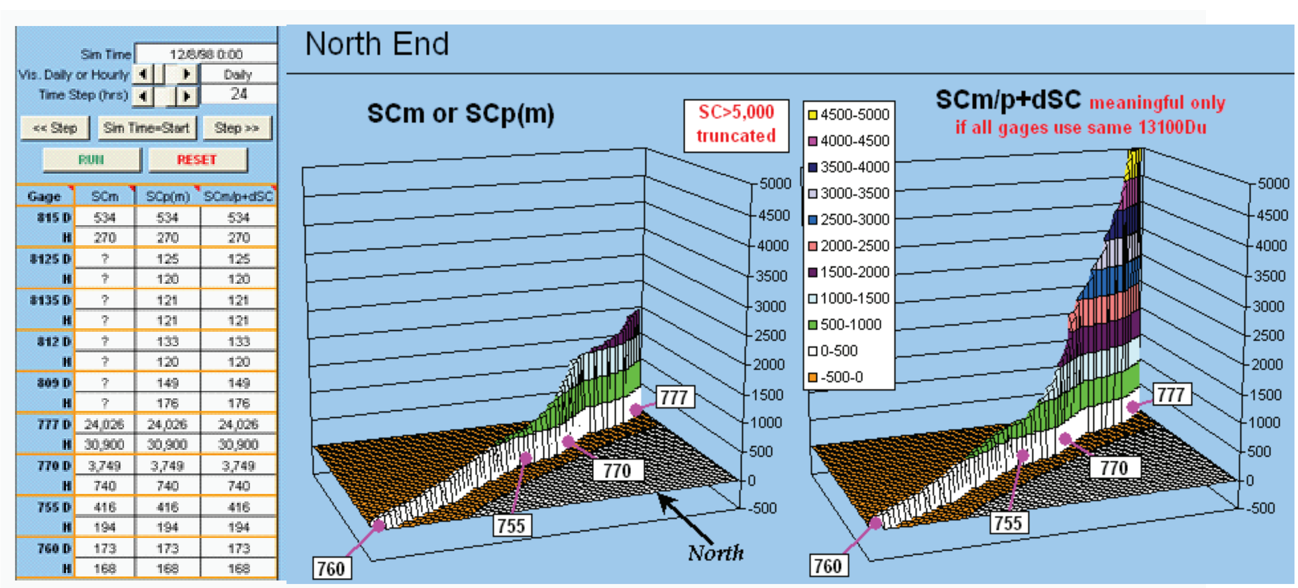


Figure 24. Streaming graphics displayed during simulation in the PRISM decision support system.



Note:

1) Gage numbers (first column on left) indicate last three or four digits of the eight- or nine-digit station numbers.

2) Pee Dee flow input (131000Du) is set at 80 percent of the historical streamflow.

Figure 25. Three-dimensional visualization (3DVis) worksheet showing specific-conductance intrusion at the north end of the study area for a simulation scenario.

profiles are only meaningful if the models are set up to use exactly the same input flow condition. For example, the plots shown in figure 25 were created using all models with the Pee Dee River flows set at 80 percent of the actual historical flows.

Application of the PRISM Decision Support System

The development of the ANN models and the DSS application for the Pee Dee and Waccamaw Rivers and AIW provides water-resource managers from State and local agencies a tool to evaluate salinity dynamics in the vicinity of municipal water intakes for various regulated flow conditions. Prior to the final determination of the required regulated flow to protect municipal intakes, several issues concerning protection of coastal water intakes must be addressed. What level of protection is required? What is the maximum specific-conductance value that is acceptable? Is an intake required to stay online 100 percent of the time, or are limited intrusion events acceptable? What regulated flow volumes from the North Carolina reservoirs are realistic? The PRISM DSS allows users to simulate various regulated flow conditions and analyze the specific-conductance response at nine tidal gages along the Grand Strand and assist in understanding the complex interaction of streamflow on the Pee Dee River and salinity dynamics of the AIW and Waccamaw River.

It is instructive to analyze riverine and estuarine systems under extreme conditions. Often the critical dynamics of a system manifest themselves during these periods rather than during more common hydrologic conditions. The 5-year drought (1998–2002) in South Carolina provides an opportunity to analyze salinity dynamics and hydrologic conditions during the worst extended drought on record. To evaluate the hydrologic conditions of the Pee Dee River for a particular year, an actual daily streamflow hydrograph can be compared to a duration flow hydrograph. During the last year of the drought, 2002, the streamflow recorded at the Pee Dee River at Pee Dee, S.C., gage was mostly between the 5^{th} percentile and the 25^{th} percentile flows for January to April (fig. 26). During the summer, May to September, flows were near the 5^{th} percentile and below the 5^{th} percentile for extended periods. The end of the drought is clearly seen with the increase in flows in October to between the 25^{th} and 75^{th} percentiles.

The period from November 2001 to October 2002 provides a 12-month period when the system experienced extreme low-flow and average-flow conditions. The salinity intrusion during this 12-month period ranges from moderate intrusions at the Waccamaw River at Hagley Landing gage (station 02110815) of 5,000 to 10,000 µS/cm, to large intrusions of greater than 10,000 µS/cm (fig. 18). Two gages of

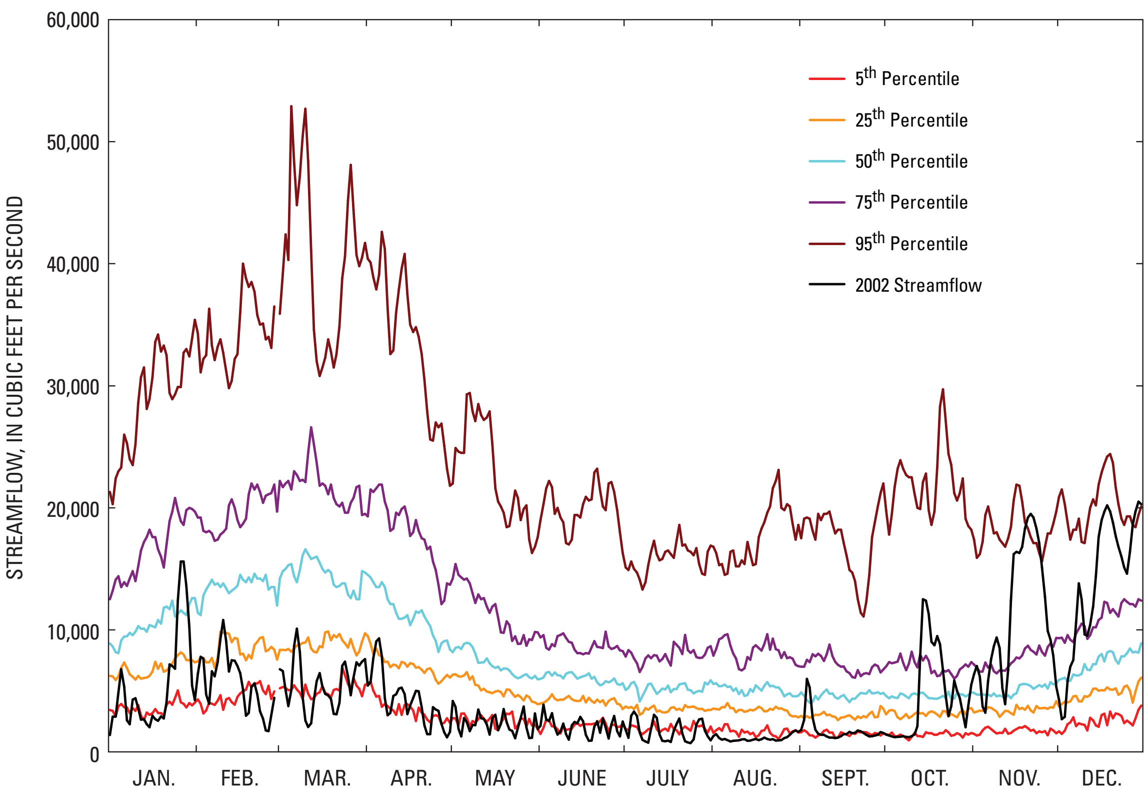


Figure 26. Duration hydrographs for the Pee Dee River near Pee Dee, South Carolina (station 02131000), and daily hydrograph for the calendar year 2002. Percentile flows are based on streamflow data from 1939 to 2006.

particular interest during the summer of 2002 were the Waccamaw River near Pawleys Island gage (station 021108125) just downstream from a municipal intake and the Waccamaw River at Hagley Landing gage (station 02110815). During the summer, large salinity intrusions forced the municipal intake near the Pawleys Island gage to shut down. The following sections describe the application of the PRISM DSS to three hydrologic scenarios—constant regulated flow, minimum regulated flow, and variable flow to maintain a specific-conductance threshold (optimized regulated flow). Results from these applications are evaluated either at the Pawleys Island gage or at the Hagley Landing gage. The results from these scenarios are intended to demonstrate the utility of the PRISM DSS and are not intended to be interpreted as a regulatory application of the DSS.

Constant Regulated Flow

To evaluate the effect of low-flow conditions on the specific-conductance response in the system, the PRISM was set up to simulate constant regulated streamflows of 1,000 and 3,000 ft^3/s for the November 1, 2001, to October 31, 2002, period. The user-set flow condition adjusts only the regulated flows into the system. The unregulated flows are not affected by the user setting and are set at their historical values. The daily specific-conductance response at the Pawleys Island gage to the two regulated flow conditions is shown in figure 27. With the exception of the period from November 1 to December 10, 2001, the constant regulated flow of 1,000 ft^3/s substantially reduces the total inflow to the Waccamaw River (black and green traces, respectively, fig. 27A). The flow greater than 1,000 ft^3/s is the flow contribution from the unregulated tributaries. From December 2001 through April 2002, the unregulated flow to the system was as much as 3,000 ft^3/s. Through the late spring and summer months (the most severe period of the 5-year drought), the flows from the tributaries were greatly reduced, and the total inflow to the system was largely from the regulated flows. As a result, the specific-conductance response to the constant regulated flow at the Pawleys Island gage increases, especially during the period of May through July 2002 (fig. 27A). Intrusion events between 2,000 and 3,000 µS/cm for measured inflow conditions increase to 4,000 to 7,000 µS/cm for the user-set inflow condition during this period. The specific conductance response seen during the period of large salinity intrusions of mid-November 2001 and early August 2002 (fig. 27A) are generally unaffected by the user-set inflow condition because the measured total inflow to the system is approximately equal to the user-set inflow condition.

The constant regulated flow of 3,000 ft^3/s substantially reduces the salinity intrusion at the Pawleys Island gage (fig. 27B). The measured intrusion events of 2,000 to 3,000 µS/cm between May and July 2002 are reduced to 500 to 1,000 µS/cm with the increased regulated flows. The high specific-conductance intrusion in August 2002 of greater than 20,000 µS/cm is reduced to less than 1,000 µS/cm.

Minimum Regulated Flow

A simulation input option in the PRISM DSS is a user-defined hydrograph. With this option, an hourly flow hydrograph is created outside of the PRISM DSS and loaded into the application. A simulation period is selected by the user, and the PRISM DSS uses the user-defined hydrograph and tidal conditions for the simulation period as inputs. Two scenarios were simulated using this option. Two hydrographs for the November 2001 through October 2002 period were created by setting the minimum flow to 1,500 ft^3/s and 2,500 ft^3/s, respectively. The daily specific-conductance response to the two flow conditions at the Pawleys Island gage is shown in figure 28. For the minimum flow of 1,500 ft^3/s (fig. 28A), inflow to the system is increased slightly (by approximately 800 ft^3/s) in November 2001 and during periods in July and August 2002 (less than 500 ft^3/s increase). As shown in figure 28A, the larger salinity intrusions greater than 2,500 µS/cm are reduced, while the intrusions less than 2,500 µS/cm (July 2002) are generally unaffected by the minimum regulated flow of 1,500 ft^3/s. Increasing the minimum flow to 2,500 ft^3/s (fig. 28B) reduces the vast majority of salinity intrusions to less than 1,500 µS/cm (fig. 28B).

Optimized Regulated Flow

Another option in the PRISM DSS to evaluate specific-conductance response in the Waccamaw River and AIW is to use the optimization routine in the application. To use the optimizer, the user specifies the set-point for the maximum specific conductance at a gage. During the simulation, the computer determines the required flow to meet the prescribed specific-conductance threshold. The optimizer uses the secant method algorithm that has the advantage of being easy to implement (as compared to other root-finding algorithms like the Newton method), but has the disadvantage of the possibility of not converging on a solution. To keep the search routine from becoming numerically unstable and interrupting a simulation period, an upper limit of 15,000 ft^3/s is coded into the optimizer. If the flow value of the solution estimate exceeds the upper limit, the search ends, a flow value of 15,000 ft^3/s is returned, and the simulation continues on to the next day. Although this upper limit may produce some noise in the time series of optimized flows, the optimization routine is a powerful alternative approach to evaluate regulated flows for a particular specific-conductance level.

The DSS was set up to evaluate the additional flow necessary to prevent specific conductance values from exceeding 5,000 µS/cm at Hagley Landing for the period of record available in the application, July 14, 1995, to December 31, 2002. The results of the scenario are shown in figure 29. To limit the specific conductance to 5,000 µS/cm, the regulated flows were modulated less than 10 percent of the simulation days. Flow augmentation ranged from 165 to 8,720 ft^3/s.

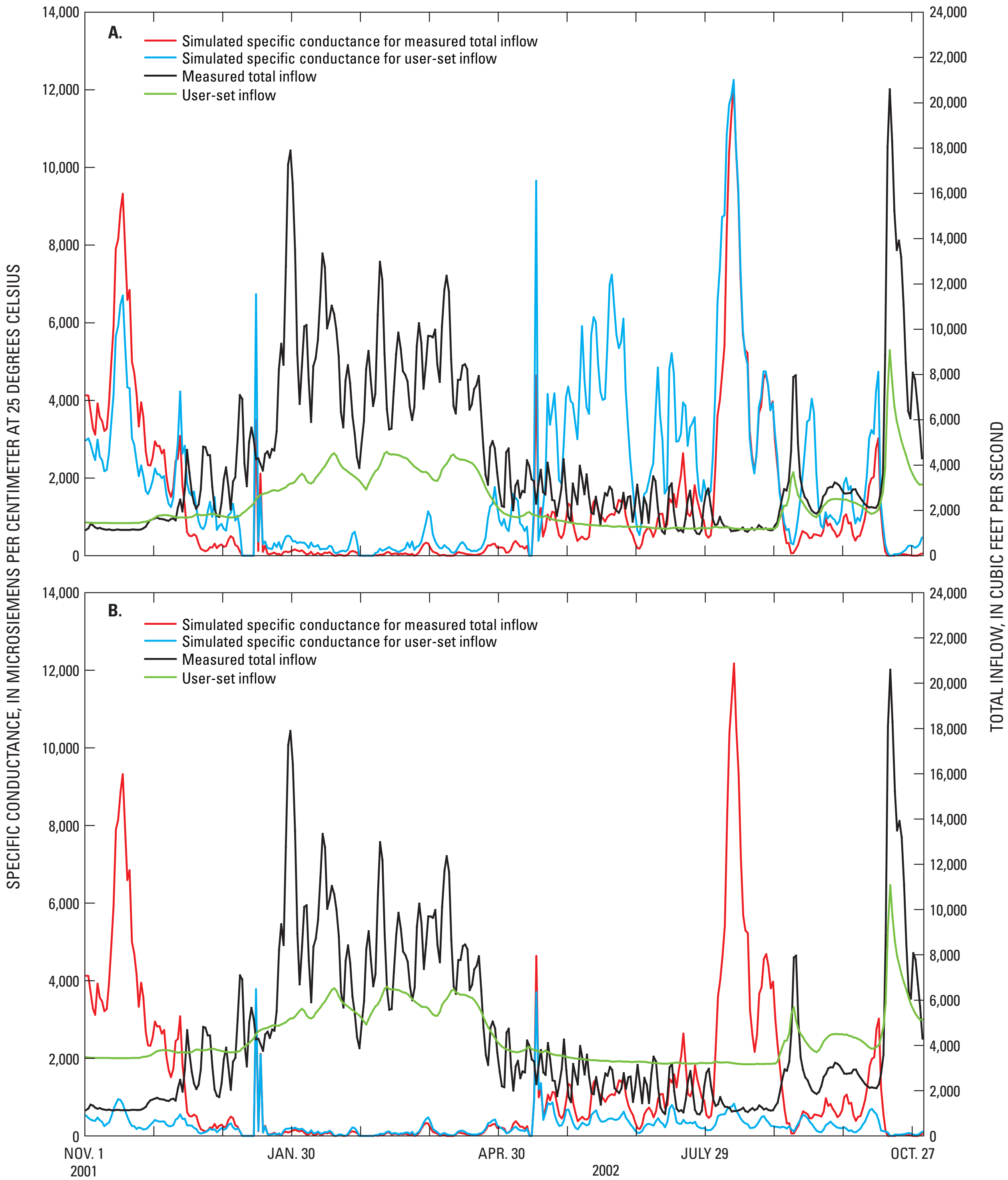


Figure 27. Simulated specific conductance for actual and user-set flow conditions at the Waccamaw River near Pawleys Island, South Carolina, (station 021108125) for the period November 1, 2001, to October 31, 2002, for constant regulated streamflows of (A) 1,000 ft^3/s and (B) 3,000 ft^3/s. User-set inflow includes user-defined constant regulated flow and historical unregulated flow.

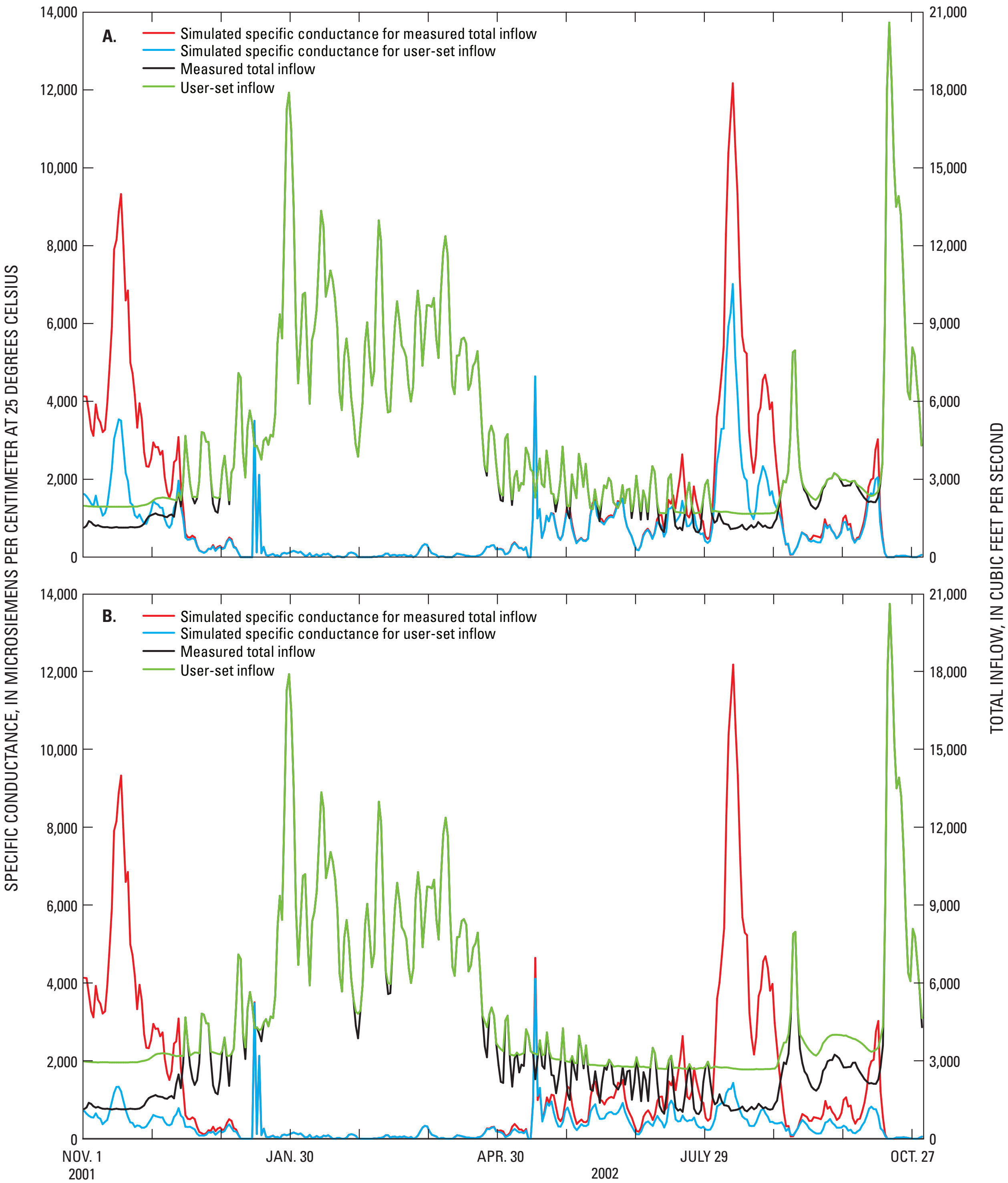


Figure 28. Simulated specific conductance for actual and user-set flow conditions at the Waccamaw River near Pawleys Island, South Carolina, (station 021108125) for the period November 1, 2001, to October 31, 2002, for minimum streamflows of (A) 1,500 ft^3/s and (B) 2,500 ft^3/s. User-set inflow includes user-defined constant regulated flow and historical unregulated flow.

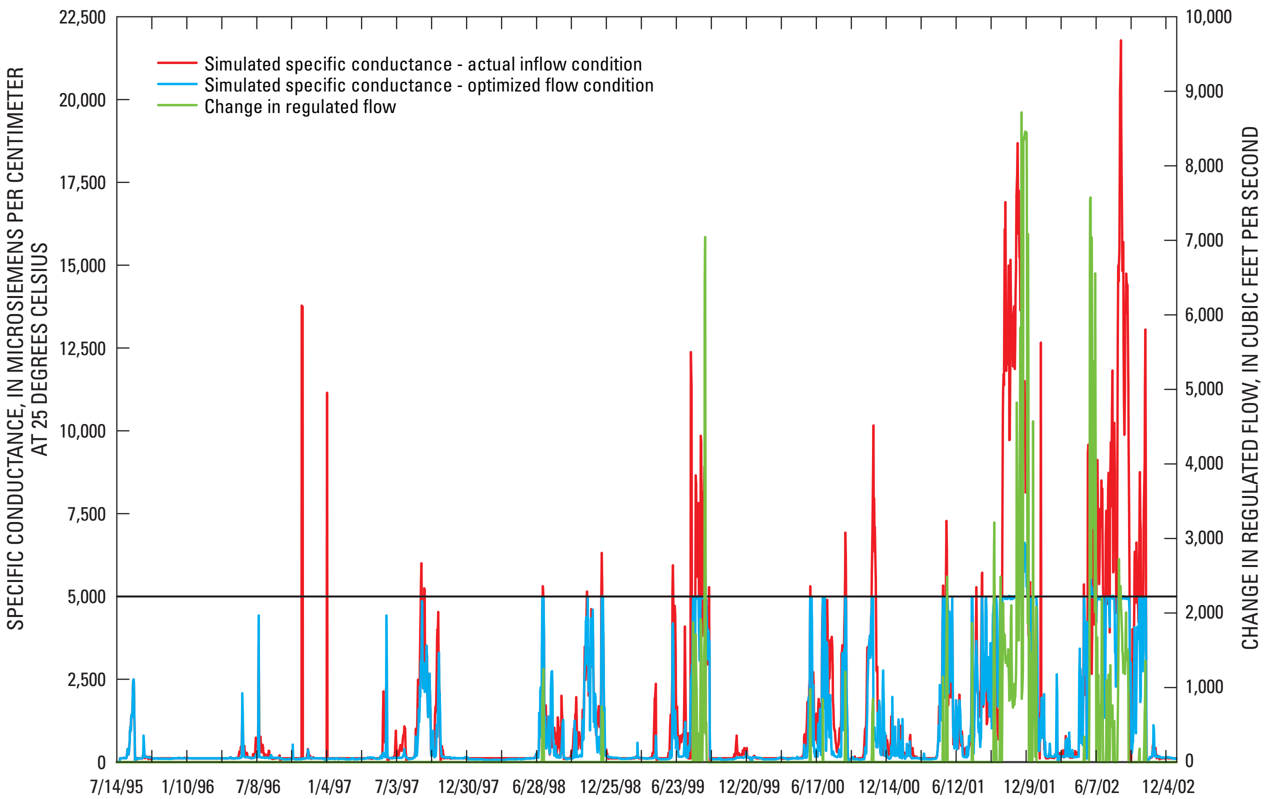


Figure 29. Simulated specific conductance for actual and optimized inflow conditions at Waccamaw River at Hagley Landing for the period July 14, 1995, to December 31, 2002. Maximum specific conductance was set at 5,000 microsiemens per centimeter (black line). The optimized change in inflow also is shown.

Summary

To evaluate the effects of regulated flows of the Pee Dee River on salinity intrusion in the Waccamaw River and Atlantic Intracoastal Waterway, the South Carolina Department of Natural Resources and a consortium of stakeholders entered into a cooperative agreement with the U.S. Geological Survey to apply data-mining techniques to the long-term time series to analyze and simulate salinity dynamics near the freshwater intakes along the Grand Strand of South Carolina. The specific-conductance artificial neural network models, historical database, model simulation controls, streaming graphics, optimization routine, and model output were integrated into a decision support system (DSS) named the **P**ee Dee **R**iver and Atlantic **I**ntracoastal Waterway **S**alinity Intrusion **M**odel (PRISM) DSS. The PRISM DSS allows the user to manipulate the streamflow inputs to the system. Four options are available: percentage of historical streamflow, constant streamflow, user-defined hydrograph, and optimized regulated flows. Output from the PRISM DSS includes tabular time series of measured data, predictions of the measured data, predictions of the user-specified conditions, and differences in simulated and user-specified values. A three-dimensional (3D) visualization routine also was developed that displays longitudinal specific-conductance conditions. The visualization routine uses predictions at the gaging station locations and interpolates values between stations. The PRISM DSS is a spreadsheet application that facilitates the dissemination and utility of the DSS.

The empirical artificial neural network models were developed using data-mining techniques. Data mining is a powerful tool for converting large databases into knowledge to solve complex problems resulting from large numbers of explanatory variables or poorly understood process physics. For the application of the artificial neural network models to the Pee Dee and Waccamaw Rivers and Atlantic Intracoastal Waterway, data-mining methods were applied to maximize the information content in the raw data. Signal processing techniques include signal decomposition, digital filtering, time derivatives, time delays, and running averages. Signal inputs to the artificial neural network models used “state-space reconstruction” for representing dynamic relations within the system. Inputs to the specific-conductance artificial neural network models include time series, or signals, of streamflow, tidal water level, and tidal range. For a complex tidal system like the Waccamaw River and AIW, the statistical accuracy of the models and predictive capability are satisfactory. Generally, the river specific-conductance models have coefficient of determination (R^2) values ranging from 0.62 to 0.96.

The PRISM DSS was run using three user-specified regulated flow input options. One input option simulated the 12-month period at the end of the recent 5-year drought using constant regulated flows of 1,000 cubic feet per second (ft^3/s) and 3,000 ft^3/s. Using the constant 1,000 ft^3/s regulated flow condition, salinity intrusion increased at the Pawleys Island gage on the Waccamaw River. For the 3,000 ft^3/s regulated flow condition, all the salinity intrusions were reduced to below 1,000 microsiemens per centimeter (µS/cm). The second input option demonstrated the use of a user-defined hydrograph to simulate minimum regulated flow to the system. For this input option, two hydrographs were computed that set the minimum flow at 1,500 ft^3/s and 2,500 ft^3/s. The minimum flow of 2,500 ft^3/s reduced the majority of salinity intrusions to less than 1,500 µS/cm at the Pawleys Island gage, whereas a minimum regulated flow of 1,500 ft^3/s reduced the larger intrusion events but generally had limited effect on intrusion events of less than 2,500 µS/cm. The third regulated flow option was the use of the optimizer to set the regulated flow to control the maximum allowable specific conductance at the Hagley Landing gage. With this option, the maximum allowable specific conductance for the Hagley Landing gage was set at 5,000 µS/cm. Using a secant method optimization routine, the PRISM DSS computed the necessary flow to prevent specific conductance intrusions greater than 5,000 µS/cm. For this scenario, the complete period of record available in the PRISM DSS database was used (July 14, 1995, to December 31, 2002). Regulated flows had to be augmented for approximately 10 percent of the time. Daily flows were increased between 165 ft^3/s and 8,720 ft^3/s to prevent the specific conductance from reaching 5,000 µS/cm at Hagley Landing.

References

Abarbanel, H.D.I., 1996, Analysis of observed chaotic data: New York, Springer-Verlag New York, p. 4–12 and p. 39.

Burden, R.L., Faires, J.D., and Reynolds, A.C., 1978, Numerical analysis: Boston, Prindle, Weber, and Schmidt, p. 40.

Carswell, W.J., Sanders, C.L., Jr., and Johnson, D.M., 1988, Freshwater supply potential of the Atlantic Intracoastal Waterway near Myrtle Beach, South Carolina: U.S. Geological Survey Water-Resources Investigations Report 88–4066, 45 p.

Conrads, P.A., and Roehl, E.A., Jr., 1999, Comparing physics-based and neural network models for predicting salinity, water temperature, and dissolved oxygen concentration in a complex tidally affected river basin, *in* Proceeding of the South Carolina Environmental Conference, Myrtle Beach, March 1999, 6 p.

Conrads, P.A., and Roehl, E.A., Jr., 2005, Integration of data-mining techniques with mechanistic models to determine the impacts of non-point source loading on dissolved oxygen in tidal waters, *in* Proceeding of the South Carolina Environmental Conference, Myrtle Beach, March 2005, 6 p.

Conrads, P.A., Roehl, E.A., Jr., and Cook, J.B., 2002, Estimation of tidal marsh loading effects in a complex estuary, *in* Proceeding of the American Water Resources Association Annual Conference, New Orleans, May 2002, 6 p.

Conrads, P.A., Roehl, E.A., Jr., Daamen, R.C., and Kitchens, W.M., 2006, Simulation of water levels and salinity in the rivers and tidal marshes in the vicinity of the Savannah National Wildlife Refuge, Coastal South Carolina and Georgia: U.S. Geological Survey Scientific Investigations Report 2006–5187, 134 p.

Conrads, P.A., Roehl, E.A., Jr., and Martello, W.P., 2002, Estimating point-source impacts on the Beaufort River using neural network models, *in* Proceeding of the American Water Resources Association Annual Conference, New Orleans, May 2002, 6 p.

Conrads, P.A., Roehl, E.A., Jr., and Martello, W.P., 2003, Development of an empirical model of a complex, tidally affected river using artificial neural networks, *in* Proceeding of the Water Environment Federation TMDL 2003 Specialty Conference, Chicago, IL, November 2003, 30 p.

Devine, T.W., Roehl, E.A., Jr., and Busby, J.B., 2003, Virtual sensors—Cost effective monitoring, *in* Proceeding of the Air and Waste Management Association Annual Conference: San Diego, CA, June 2003.

Drewes, P.A., and Conrads, P.A., 1995, Assimilative capacity of the Waccamaw River and the AtlanticIntracoastal Waterway near Myrtle Beach, South Carolina, 1989–92: U.S. Geological Survey Water-Resources Investigations Report 95–4111, 58 p.

Dutta, Soumitra, Wierenga, Berend, and Dalebout, Arco, 1997, Case-based reasoning systems—From automation to decision-aiding and stimulation: IEEE Transactions on Knowledge and Data Engineering, v. 9, no. 6, p. 911–912.

Dyer, K.R., 1997, Estuaries—A physical introduction (2d ed.): New York, John Wiley and Sons, Ltd, 195 p.

Hinton, G.E., 1992, How neural networks learn from experience: Scientific American, v. 267, no. 3, p.144–151.

Jensen, B.A., 1994, Expert systems—Neural networks: Instrument Engineers' Handbook (3d ed.), Chilton, Radnor, PA.

National Oceanic and Atmospheric Administration, 2005, Tide station locations and ranges; accessed December 15, 2006, at *http://tidesandcurrents.noaa.gov/tides05/tab2ec3a.html#64*

North Carolina Department of Environment and Natural Resources, Division of Water Quality, 1998, Yadkin-Pee Dee River basinwide water quality management plan, Chapter 2—General basin description, 35 p.; accessed May 1, 2007, at *http://h2o.enr.state.nc.us/basinwide/yadkin/yadch2.doc*

North Carolina Department of Environment and Natural Resources, Division of Water Resources, 2001, North Carolina State Water Plan Basin 18, Yadkin-Pee Dee, 7 p.; accessed May 1, 2007, at *http://www.ncwater.org/Reports_and_Publications/swsp/swsp_jan2001/final_pdfs/B18_YadkinPeeDee.pdf*

Risley, J.C., Roehl, E.A., Jr., and Conrads, P.A., 2003, Estimating water temperatures in small streams in Western Oregon using neural network models: U.S. Geological Survey Water-Resources Investigations Report 02–4218, 67 p.

Roehl, E.A. Jr., and Conrads, P.A., 1999, Near real-time control for matching wastewater discharges to the assimilative capacity of a complex, tidally affected river basin, *in* Proceeding of the 1999 South Carolina Environmental Conference, Myrtle Beach, March 15–16, 1999, 6 p.

Roehl, E.A., Jr., and Conrads, P.A., 2006, Hydrologic modeling using multivariate state space reconstruction, *in* Gourbesville, Philippe, Cunge, Jean, Guinot, Vincent, and Liong, Shie-Yui, eds., Proceedings of Hydroinformatics 2006, v. 1, p. 765–772.

Roehl, Ed, Jr., Conrads, Paul, and Cook, J.B., 2003, Discussion of "Using complex permittivity and artificial neural networks for contaminant prediction:" Journal of Environmental Engineering, v. 129, p. 1069–1071.

Roehl, E.A., Jr., Conrads, P.A., and Daamen, R.C., 2006, Features of advanced decision support systems for environmental studies, management, and regulation, *in* Proceedings of the International Environmental Modeling and Software Society Summit, Burlington, VT, July 2006, 8 p.

Roehl, E.A., Jr., Conrads, P.A., and Roehl, T.A.S., 2000, Real-time control of the salt front in a complex tidally affected river basin, Smart Engineering System Design, v. 10, *in* Proceedings of the Artificial Neural Networks in Engineering Conference: New York, ASME Press, p. 947–954.

Rosenblatt, F., 1958, The perceptron—A probabilistic model for information storage and organization in the brain: Psychological Review, v. 65, no. 6, p. 386–408.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J., 1986, Learning internal representations by error propagation, *in* Rumelhart, D.E., McClelland, J.L., and the PDP Research Group, Parallel distributed processing—Explorations in the microstructure of cognition, v 1: Cambridge, MA, MIT Press, p. 318–362.

Seaber, P.R., Kapinos, F.P., and Knapp, G.L., 1994, Hydrologic unit maps: U.S. Geological Survey Water-Supply Paper 2294, 63 p.

South Carolina Department of Health and Environmental Control, 1998, Total maximum daily load determination for the Waccamaw River and the Atlantic Intracoastal Water Way near Myrtle Beach, SC—State of South Carolina Administrative Record TMDL Submittal for the Waccamaw River and Atlantic Intracoastal Water Biochemical Oxygen Demand, 64 p.

U.S. Geological Survey, 1986, National water summary 1985—Hydrologic events and surface-water resources: U.S. Geological Survey Water-Supply Paper 2300, 506 p.

Weiss, S.M., and Indurkhya, Nitin, 1998, Predictive data mining—A practical guide: San Francisco, Morgan Kaufmann Publishers, Inc., p. 1.

Prepared by:
USGS Enterprise Publishing Network
Raleigh Publishing Service Center
3916 Sunset Ridge Road
Raleigh, NC 27607

For additional information regarding this publication, contact:
Director
USGS South Carolina Water Science Center
720 Gracern Road, Suite 129
Columbia, SC 29210
phone: 1-803-750-6100
email: dc_sc@usgs.gov

Or visit the South Carolina Water Science Center website at:
http://sc.water.usgs.gov

Conrads and Roehl—**Salinity Intrusion in the Waccamaw River and Intracoastal Waterway near Myrtle Beach, SC—Scientific Investigations Report 2007–5110**

Printed on recycled paper