

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Process-Based Statistical Models Predict Dynamic Estuarine Salinity

*Christina L. Durham, David B. Eggleston and Amy J. Nail*

## Abstract

Climate change is increasing variation in freshwater input and the intensity of this variation in estuarine systems throughout the world. Estuarine salinity responds to dynamic meteorological and hydrological processes with important consequences to physical features, such as vertical stratification, as well as living resources, such as the distribution, abundance and diversity of species. We developed and evaluated two space-time statistical models to predict bottom salinity in Pamlico Sound, NC: (i) *process* and (ii) *time* models. Both models used 20-years of observed salinity and contained a deterministic component designed to represent four key processes that affect salinity: (1) recent and long-term fresh water influx (FWI) from four rivers, (2) mixing with the ocean through inlets, (3) hurricane incidence, and (4) interactions among these variables. Freshwater discharge and distance from an inlet to the Atlantic Ocean explained the most variance in dynamic salinity. The final *process* model explained 89% of spatiotemporal variability in salinity in a withheld dataset, whereas the final *time* model explained 87% of the variability within the same withheld data set. This study provides a methodological template for modeling salinity and other normally-distributed abiotic variables in this lagoonal estuary.

**Keywords:** estuaries, space-time model, spatial covariance, freshwater inflow, process-based model, salinity

## 1. Introduction

Estuarine salinity responds to dynamic meteorological and hydrological processes [1] with important consequences to physical features, such as vertical stratification, as well as living resources, such as the distribution, abundance and diversity of species [2–5]. For example, relatively low mixing and subsequent salinity stratification can lead to hypoxia in areas where organically-rich sediments are not adequately re-oxygenated, causing emigration of mobile fauna and degradation of ecosystem functions [5–9]. Rapid salinity changes, such as those associated with large rainfall events or tropical cyclones, can cause death of postlarval stages that are sensitive to unusually low salinities [10], and mass seaward migration and subsequent hyper-aggregation of mobile, commercially important species that can result in (1) shifts of juveniles from primary nursery areas protected from trawling to secondary non-nursery areas vulnerable to fishing pressure [11], (2) overharvest of adults due to increases in fishery catchability [12],

or (3) bias fishery-independent surveys that leads to over-inflated population abundance estimates [12]. Thus, the need to accurately predict the spatiotemporal dynamics of salinity is unprecedented. The specific goals of this study were to: (1) evaluate several statistical models to hindcast and forecast salinity in the second largest estuary and largest lagoonal estuary in the United States—Pamlico Sound, North Carolina, USA, and (2) assess salinity observations, predictions, and standard errors under five hydrologic scenarios characteristic of historic and future climate changes.

Pamlico Sound (PS) is a relatively shallow estuary with a mean depth of 4 m and a maximum depth of 7 m. PS circulation is dominated by wind-driven currents and freshwater input [13, 14]. Seasonal cyclonic storms are also an important climatological component of the PS system. Since 1996, over three tropical storms or hurricanes have passed within 300 km of the North Carolina coast per year [10]. Given the important role that salinity plays in the abiotic and biotic system components of estuaries, and the likelihood that global climate change will increase the frequency of extreme weather events (e.g., floods, droughts, hurricanes—[9, 15, 16]), there is a critical need for models that can accurately forecast spatiotemporal variation in salinity (e.g., [17]). A recent review by Iglesias et al. [17] highlights the strengths of applying numerical modeling tools to characterize morpho-hydrodynamic processes in estuarine and coastal systems. *Numerical methods* can include a large variety of models and techniques, such as finite element, finite difference, finite volume, or Eulerian-Lagrangian models (e.g., [17–19]). Complex, three-dimensional numerical models used for simulation and forecasting of dynamic estuarine salinity can require significant effort and computation time that is beyond the capabilities of many local management agencies. Local management agencies sometimes require a quick turnaround time for long-term simulations or short-term forecasts of estuarine salinity conditions, which could be produced using location-specific statistical models. Therefore, the goals of this study were to (1) develop and evaluate two types of *statistical models* of bottom salinity in PS, and (2) apply the best models to produce sound-wide retrospective maps of bottom salinity based on observational data. Bottom (as opposed to surface) salinity was chosen as the variable of interest because it characterizes habitats of mobile demersal species that are important members of benthic food webs, and that are the targets of valuable commercial and recreational fisheries. Hereafter, the term ‘salinity’ will always refer to bottom salinity unless otherwise noted.

### 1.1 Statistical models to predict dynamic salinity

Producing retrospective salinity maps based on observational data does not require a statistical model based on hydrological mechanisms that affect salinity; it is possible to perform individual spatial interpolations for each time period of interest using an ordinary kriging model or a universal kriging model with a simple spatial trend. Predicting salinity under a hypothetical set of conditions, however, does require a model that can ‘learn’ about hydrological mechanisms based on retrospective data (e.g., [20, 21]). Thus, the more comprehensive goal of this study was to produce retrospective maps of salinity by developing a space-time statistical model in which the mean function represents the hydrological mechanisms that affect salinity, and a spatial covariance function makes up the difference between the observed salinity data and the mean function’s salinity prediction.

To create such a model, we constructed explanatory variables that accounted for the effect of riverine freshwater inflow (FWI), distance to inlet sources of oceanic saltwater, and hurricane incidence on salinities at different locations in PS. We used

a forward-selection process to choose which of these variables to keep in the model. Standard errors based on the covariance function allowed for assessment of strengths and weaknesses of the representation of the hydrology in the mean function. Since an additional goal of this study was to provide a template for researchers to build process-based models of normally-distributed estuarine variables, we considered only models that could be fit using procedures in the SAS<sup>®</sup> software package, yet can be adopted to R-statistical software.

Other process-based models of PS salinity in the literature—all of which are differential-equation-based deterministic models—provided important insights into how different variables influenced spatiotemporal salinity variation in PS ([22, 23], and others). However, these models ultimately lacked the spatial resolution and/or coverage of the entire area of interest of this study, and none quantified uncertainty at every space-time prediction location. For example, Xu et al. [24] predicted surface and bottom salinity, and temperature at 30-second intervals over a spatial grid with varying cell size (200–800 m<sup>2</sup>) in the Pamlico River Estuary (PRE), a PS tributary, using a customized extension of the Environmental Fluid Dynamics Code [25] to incorporate FWI from major tributary rivers, as well as tide and wind effects on circulation. Although this model incorporated environmental variation and produced salinity predictions suitable to assess long-term space-time trends, the PRE makes up only 18% of the area of PS. Predicting salinity across the entire PS using this model would require spatial domain expansion and re-parameterization, and such extensions are not planned (J. Lin, NC State University, pers. comm. on behalf of Xu et al. [24]).

Though we are unaware of researchers that have constructed space-time statistical models of salinity in PS, there are examples of applying statistical models for spatial prediction of salinity in other estuaries. For example, Rathbun [26] used independent multiple linear regression models with spatially-correlated errors to predict salinity and dissolved oxygen (DO) in Charleston Harbor, SC over a two-week time period in 1988 as a function of spatial coordinates and distance to the estuary mouth. Chehata et al. [27] performed three-dimensional spatial interpolation of salinity and DO measurements in Chesapeake Bay. Qiu and Wan [20] developed a salinity model based on time series analyses of salinity data for the Caloosahatchee River Estuary, Florida, USA. The structure of their model consisted of an autoregressive term representing the system persistence and an exogenous term accounting for physical drivers including freshwater inflow, rainfall, and tidal water surface elevation that cause salinity to vary. The model was calibrated and validated using up to 20 years of measured data collected they found that the time series model offers comparable or superior performance compared with its 3-D, numerical counterpart. This model has been used as a tool for water resources management projects relating to ecosystem restoration and water control in south Florida [20]. Similarly, Ross et al. [21] examined the response of salinity in the Delaware Estuary, USA to climatic variations using statistical models and long-term (1950-present) records of salinity from the U.S. Geological Survey and the Haskin Shellfish Research Laboratory. The statistical models included non-parametric terms and were robust against auto-correlated and heteroscedastic errors. After using the models to adjust for the influence of streamflow and seasonal effects on salinity, several locations in the estuary showed significant upward trends in salinity. Insignificant trends are found at locations that are normally upstream of the salt front. The models indicate a positive correlation between rising sea levels and increasing residual salinity, with salinity rising from 2.5 to 4.4 psu per meter of sea-level rise. The results suggest that continued sea-level rise in the future will cause salinity to increase regardless of any variation in fresh water influx [21]. Urquhart



et al. [28] present the results of multiple statistical models that predicted daily, gridded surface salinity at 1 km resolution across Chesapeake Bay, USA as a function of surface reflectance estimates of salinity from the NASA Moderate Resolution Imaging Spectroradiometer (MODIS), onboard the Aqua platform satellite. Eight statistical methods were tested, and sea surface salinity was accurately predicted via remote sensed products with an accuracy that was more than sufficient for many physical and ecological applications [28].

None of these previous studies, however, attempted to explicitly represent the hydrological processes by which fresh and saltwater mixing affects estuarine salinity. In this paper, we describe the development of candidate explanatory variables to represent mechanisms affecting PS salinity and how that development led to consideration of two fundamentally different mean functions. We then describe the forward selection process by which candidate variables were chosen to be retained in the models, and how candidate covariance functions were selected to pair with each mean function. Next, we examined maps of salinity observations, predictions, and standard errors under five hydrologic scenarios, analyzed these results, and provided overall implications of the findings.

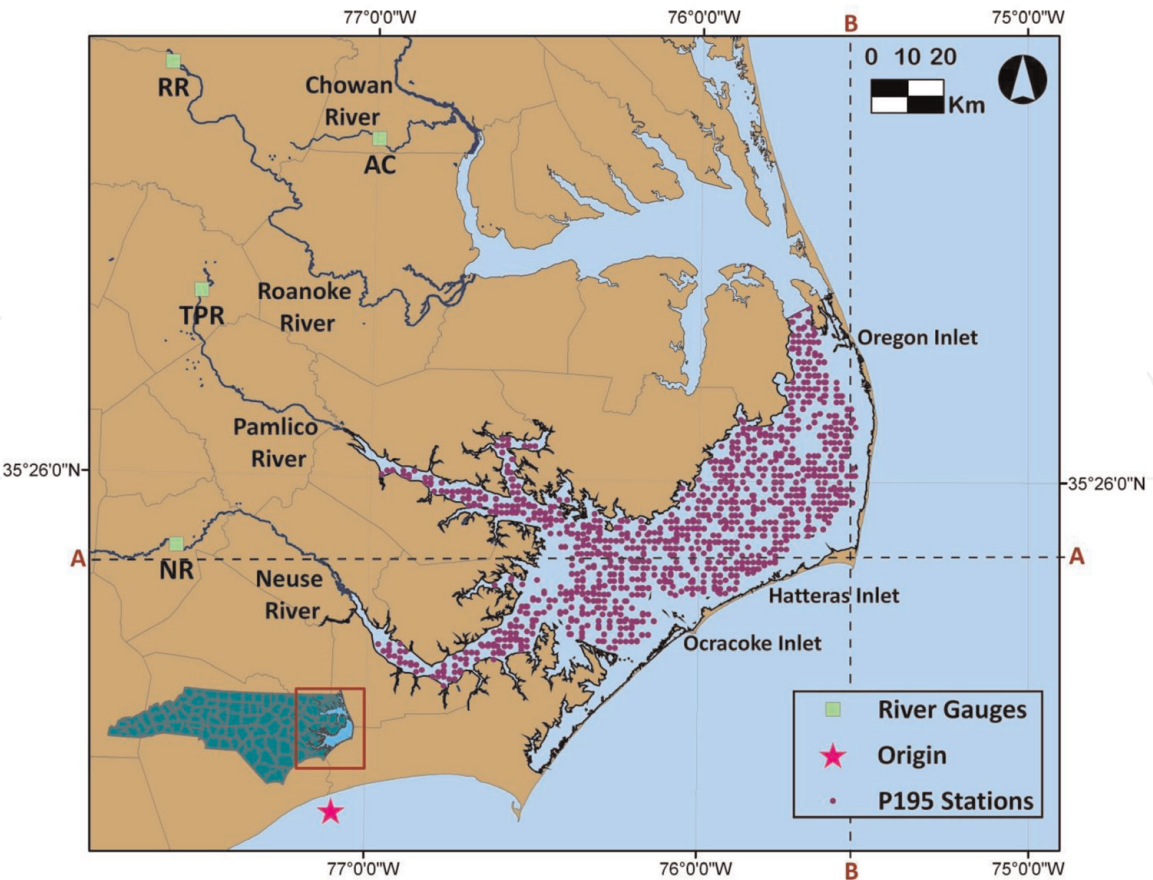
## 2. Methods and results

### 2.1 Data and notation

We used bottom salinity values measured by the North Carolina Division of Marine Fisheries (NC DMF) Pamlico Sound Trawl Survey Program 195 (*the survey*) every June and September from 1987 to 2006. The survey is conducted only in June and September each year. Designed to assess species abundance at depths over 2 m, the survey uses a weighted stratified random sampling design. For each time period, coordinates of stations are randomly generated within each of seven water body strata, with more stations allocated to larger strata, for a total of 54 stations per time period. Hereafter, we denote with  $S$  the spatial domain that includes all points sampled within the seven strata mentioned over the entire 1987–2006 temporal domain. **Figure 1** shows the geographic location of each sampling station in  $S$ . Salinity was measured using a YSI-85 multi-function meter at the beginning of each trawl and recorded along with depth and spatial reference coordinates. All spatial coordinates used in this analysis were converted from decimal degrees to northings and eastings in nautical miles (nmi) from a reference point (the origin in **Figure 1**) located southwest of  $S$  at  $34.6^\circ\text{N}$ ,  $-77.1^\circ\text{W}$ . Salinity is always reported using the Practical Salinity Scale.

The temporal domain contains  $T = 40$  time periods, or month/year combinations, indexed by the subscript  $t$ , so that  $t = 1, \dots, T$ . A time period is approximately 2.5 weeks long, the time it takes to sample all stations. Since locations of the 54 stations sampled in each time period differed slightly, and since some data were missing in each time period, let  $n_t$  represent the number of sites in time period  $t$ . Site refers to a specific spatial location nested within a particular time period and is indexed using the subscript  $i$  where  $i = 1, \dots, n_t$ . The dataset included  $N = 2100$  total observations of salinity, where  $N = \sum_{t=1}^T n_t$ . Denoted with  $sal_{it}$  observed salinity at site  $i$  in time period  $t$ .

The fresh water influx (FWI) data represented watersheds of the Neuse, Pamlico, Roanoke, and Chowan rivers, which comprise 80% of the land draining into PS [29]. FWI observations were average daily river discharge rates collected by one



**Figure 1.** Pamlico Sound, NC and the Chowan, Roanoke, Pamlico, and Neuse Rivers. Green squares show the four river gauge stations used in this study. Purple dots indicate all P195 trawl survey sample stations for the 1987–2006 time domain. The pink star indicates the reference point from which northings and eastings were calculated. As referenced in Section 3.5, Parallel A is located at 35° 16' N latitude and meridian B is at 75° 42' W longitude.

US Geological Survey (USGS) gauge station per tributary (**Figure 1**): Neuse River (NR) station 02089500 in Kinston; Tar-Pamlico River (TPR) station 02083500 in Tarboro; Roanoke River (RR) station 02080500 in Roanoke Rapids; and Ahoskie Creek (AC) station 02053500 in Ahoskie, which gauges Chowan River inflow. Discharge rates in ft<sup>3</sup>/s for every day during the time domain (7305 days) were downloaded from the USGS Water Resources website for the state of North Carolina (USGS 2009) and were converted to m<sup>3</sup>/s. For each river, the gauge chosen was the furthest downstream gauge that recorded data over the entire temporal domain.

## 2.2 Candidate explanatory variables

The creation of explanatory variables reflects the modeling context—the objectives, the geographical features of the spatial domain, and the space-time coverage and resolution of the data—but the general thought process can be modified by other researchers in a different context. We index the term *it* as any variable that varies in both space and time, and with *t* any variable that varies over time but is constant over *S* within a time period.

## 2.3 Freshwater influx indices

Sixty-one days is the average freshwater residence time of the four major rivers flowing into PS [30–32], accounting for the temporal lag between the upriver

gauging of freshwater and the delivery of that water to S. Therefore, we defined the long-term metric  $2moFWI_{r_t}$  for river  $r$  and time period  $t$  where  $r = 1, \dots, 4$ , and  $t = 1, \dots, T = 40$  as the average daily discharge rate in the 61 days prior to  $m_t$ , the first day of the survey in time period  $t$ . Because Ramus et al. [33] calculated a seven-day residence time for the Neuse and Pamlico Rivers after Hurricanes Dennis and Floyd deposited 1 m of rainfall in eastern NC less than 2 weeks before the September 1999 survey, we defined the short-term metric  $1wkFWI_{r_t}$ , by averaging daily discharge rates in the 7 days prior to  $m_t$ .

Since freshwater from river  $r$  in time period  $t$  should have more of an effect on  $sal_{it}$  the closer site  $i$  is to the river, a unique measure of the influence of  $1wkFWI_{r_t}$  and  $2moFWI_{r_t}$  for each site was formed by dividing each by  $dist_{r_{it}}$ ,  $r = 1, \dots, 4$ , the distance separating the gauge on river  $r$  from site  $i$  within time period  $t$ :

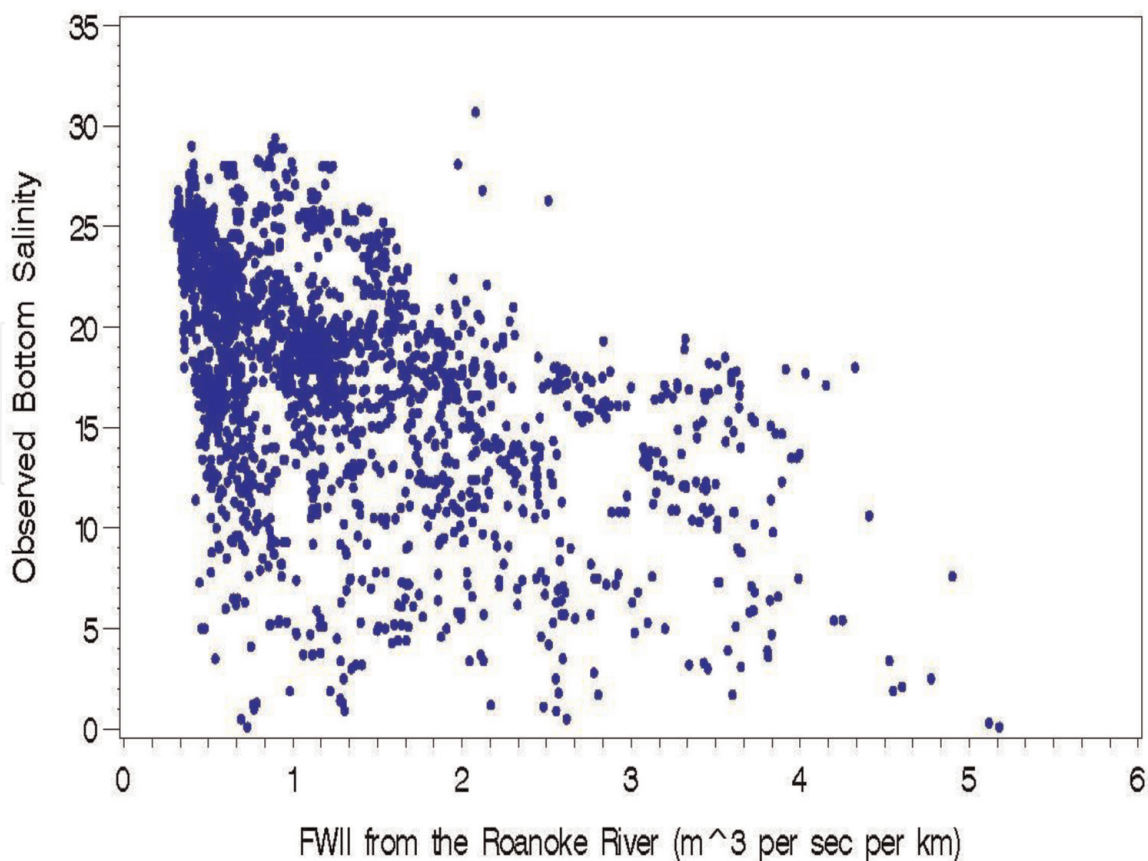
$$1wkFWII_{r_{it}} = \frac{1wkFWI_{r_t}}{dist_{r_{it}}}, \text{ and } 2moFWII_{r_{it}} = \frac{2moFWI_{r_t}}{dist_{r_{it}}} \quad (1)$$

The coordinates of each gauge station were used to calculate distance because the gauge was the location of the  $1wkFWI_{r_t}$  and  $2moFWI_{r_t}$  observations. Like all distances in this study,  $dist_{r_{it}}$  represents distance “as the crow flies” as opposed to water-path distance. Though the superiority of using water-path distance when modeling water-quality variables in stream and estuarine systems seems intuitive, results from studies that compare these two distance metrics are inconclusive. For example, Gardner et al. [34] found more accurate predictions of stream temperatures when models incorporated water-path distance, but only when this distance was modified and weighted by stream order. Peterson and Urquhart [35] predicted various nutrient concentrations in 17 Maryland rivers and concluded that using water-path distance works well when modeling certain nutrients, but not others, and that the crow-flies distance appeared to be the most suitable distance measure overall. Comparing the accuracy of predictions of water quality parameters generated from two different multiple linear regression models containing the explanatory variable “distance to inlet mouth”, Little et al. [36] found that predictions from models using water-path distance were no more accurate than those from models using crow-flies distance. None of these studies demonstrated marked predictive improvement using water-path distance, therefore we used crow-flies distance from each of the four river gauges to each of 2100 sample stations and over 6000 prediction locations.

The plot in **Figure 2** of  $sal_{it}$  against Roanoke River  $2moFWII_{r_{it}}$  typifies the relationships between salinity and each of the eight  $1wkFWII_{r_{it}}$  and  $2moFWII_{r_{it}}$  variables. Larger values of the metric are associated with smaller values of salinity, but groups of observations have different slopes. Closer examination revealed that the different groups corresponded to different time periods. We attempted to account for the different slopes in two ways, first by considering the 28 pair-wise interactions among the  $1wkFWII_{r_{it}}$  and  $2moFWII_{r_{it}}$  and second by considering 39 time-period indicator variables defined as

$$\begin{aligned} timeper\_1_t &= \begin{cases} 1 & \text{if } t = 1 \\ 0 & \text{otherwise} \end{cases}, \dots, timeper\_t_t = \begin{cases} 1 & \text{if } t = t \\ 0 & \text{otherwise} \end{cases}, \dots, timeper\_39_t \\ &= \begin{cases} 1 & \text{if } t = 39 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$





**Figure 2.** Observed bottom salinity (psu) vs. the Roanoke River two-month relative freshwater influx index ( $2moFWII_{x_{it}}$ ) in  $m^3 s^{-1} km^{-1}$  from 1987 to 2006. Groups of values within the same time period exhibit relationships with different slopes.

(A fortieth indicator variable was not used because it would create a non-full-rank design matrix, and the effect for the fortieth time period can be derived using the intercept.) This latter consideration led to the creation of two distinct mean function models: the *process* and *time* models. The first has process variables only, and the second has process variables in addition to the time-period indicator variables to address the possibility that salinity is affected by some aspect of physical phenomena that is not accounted for by any other variable in the model.

## 2.4 Saltwater mixing and tidal signal

Although salinity on the inner-continental shelf of the U.S. Southeast Atlantic coast exhibits some spatial variability near PS [37], we follow Xie et al. [38] and assume constant open ocean salinity. This assumption allows for modeling the effect of ocean water mixing as a function of only the distance to inlet, as opposed to distance interacting with the salinity of the ocean water, from each spatial location in the sound to each of the major PS inlets: Oregon, Hatteras, and Ocracoke. Exploratory analyses reveal that models using a single variable (distance to the nearest inlet) rather than three variables (distances to each of the three inlets), explains the same amount of variability in salinity when other explanatory variables are also included. Therefore, we consider for inclusion in subsequent models the variable *closest\_inlet\_dist<sub>it</sub>*, defined to be the distance separating site *i*, sampled in time period *t*, from the center of the most proximate inlet.



## 2.5 Wind speed and direction

A prevailing wind field that is north/northeast from March to August and south/southwest from September to February is the primary driver of currents in PS [39]. Thus, wind speed and direction were incorporated into the modeling process using the categorical variable  $month_t$ , where

$$month_t = \begin{cases} 1 & \text{if } t \text{ is in Sept} \\ 0 & \text{if } t \text{ is in June} \end{cases}$$

is used to examine the effects of seasonal wind patterns on the spatial distribution of salinity.

## 2.6 Evaporation and direct precipitation

Holding other factors constant, sound-wide salinity in time periods that experience more evaporation of water from the surface of PS would likely be higher than those in time periods that experienced less evaporation, but no evaporation data were available for the space-time domain of interest. Salinity in time periods for which there was more direct precipitation into S should be lower than those in lower-precipitation time periods, however precipitation data were only available at two weather stations on the edges of PS from which information about individual spatial locations within PS would be difficult to infer. Giese et al. [40] found that direct precipitation constitutes only 8% of mean PS freshwater input, thus the signal from riverine FWI should dominate in explaining salinity variability. Therefore, we did not include evaporation or direct precipitation variables in our models.

## 2.7 Spatial coordinates

Estuarine salinity varies over space such that functions of spatial coordinates might explain variability in salinity not accounted for by the other variables. Scatterplots of salinity versus easting and northing suggested that salinity is quadratic in the former and cubic in the latter. The quadratic function of easting can be explained by examining a west-to-east path through PS along the 35° 16' N parallel (A in **Figure 1**): salinity should initially increase, reach a maximum at the saltwater plume near Ocracoke and Hatteras Inlets, and decrease again on the other side of the plume in the waters on the western shore of Hatteras Island near Buxton, NC. The cubic function of northing is best described by examining a north-to-south path along longitude of 75° 42' W (B in **Figure 1**), where salinity should increase traveling south from Albemarle Sound, reach a local maximum near Oregon Inlet, decrease continuing past the saltwater inlet plume, and increase again as the Hatteras Inlet saltwater plume is reached. Thus,  $easting_{it}$ ,  $easting_{it}^2$ ,  $northing_{it}$ ,  $northing_{it}^2$ , and the interactions  $northing_{it} * easting_{it}$ ,  $northing_{it}^2 * easting_{it}$ ,  $northing_{it} * easting_{it}^2$ , and  $northing_{it}^2 * easting_{it}^2$  are considered as explanatory variables. All coordinates are centered before they are squared or cubed by subtracting the mean over all observations.

## 2.8 Hurricanes

Hurricanes can rapidly introduce large volumes of freshwater to estuaries via riverine influx, push large volumes of saltwater in through inlets via storm surge,

and alter circulation patterns through abrupt changes in wind speed and direction [7, 10]. Hurricanes can also open new inlets to PS, which can alter current flow and increase saltwater intrusion [41]. The variable  $1wkFWII_{rit}$  should capture variability in salinity due to hurricane-produced FWI. Three additional variables may account for non-FWI-related variability in salinity due to hurricane passage. These variables are unique to a given time period  $t$  but are constant over all sites  $i$  within  $t$ . The continuous variable  $inverse\_days\_survey_t$  is the reciprocal of the number of days between the most recent hurricane and  $m_t$ , except when there is no hurricane within the 61 days, and then it takes the value zero. The categorical variable  $category_t$  equals the category of the most recent hurricane rated on the Stafford-Simpson scale (1, ..., 5), but if no hurricane made landfall in the 61 days prior to  $m_t$ , it takes the value zero. Finally, the discrete variable  $num\_storms_t$  equals the number of hurricanes making landfall in NC in the 61 days prior to  $m_t$ .

## 2.9 Variable selection

Section 3 identifies 46 candidate explanatory variables for the process model mean function:  $1wkFWII_{rit}$  and  $2moFWII_{rit}$  (8), plus selected pair-wise interactions (explained below) (24); spatial coordinates, their powers, and specified interactions (9);  $closest\_inlet\_dist_{it}$ ;  $month_t$ ; and hurricane variables  $inverse\_days\_survey_t$ ,  $category_t$ , and  $num\_storms_t$ . For the time model, there were an additional 39 time period indicator variables. Some variables—in either model—may be redundant. There is overlap among the hurricane variables, and spatial coordinates may not be necessary if other variables explain more variability in salinity. The set of variables included in the final model(s) should balance goodness-of-fit with parsimony. We first describe the variable-selection process for the process model, then for the time model.

## 2.10 Process model

The results of eight separate ordinary least squares linear regression models of salinity make up the rows **Table 1**. The first five consist of an intercept and a single explanatory variable:  $closest\_inlet\_dist_{it}$ ,  $category_t$ ,  $inverse\_days\_survey_t$ ,  $num\_storms_t$ , and  $month_t$ . The sixth and seventh contain an intercept plus, respectively, the sets of four short and long-term freshwater influx indices  $\{1wkFWII_{rit}, r = 1, \dots, 4\}$ , and  $\{2moFWII_{rit}, r = 1, \dots, 4\}$ . We treated the short and long-term sets of indices as groups assuming that if an index evaluated for one river is meaningful, then it is also meaningful for other rivers. We discuss the eighth row in Section 4.2.

Adjusted  $R^2$  is a modification of  $R^2$  that penalizes the number of explanatory variables. While  $R^2$  increases as more variables are added to a model, adjusted  $R^2$  increases only if the added variable decreases the error sum of squares enough to offset the loss in error degrees of freedom.

The model with the long-term freshwater influx indices had the largest adjusted  $R^2$  at 0.38, followed by the model with the distance from the nearest inlet (0.34), and the model with the short-term FWI indices (0.27). None of the other four models explained more than 5% of the variability in salinity. We chose the model with the long-term freshwater influx indices as the base upon which to build the mean function.

To this base model we added the variable  $closest\_inlet\_dist_{it}$  since the model containing this variable had the second-best performance, thus beginning a forward-selection process. Each time we added a variable or set of variables to the model, we kept it in the model if the new adjusted  $R^2$  exceeded the old. Variables

Explanatory variable or set of explanatory variables	Adj R <sup>2</sup>
<i>closest_inlet_dist<sub>it</sub></i>	0.34
<i>category<sub>t</sub></i>	0.049
<i>inverse_days_survey<sub>t</sub></i>	0.035
<i>num_storms<sub>t</sub></i>	0.029
<i>month<sub>t</sub></i>	0.015
<i>1wkFWII_r<sub>it</sub></i> , <i>r</i> = 1, ..., 4	0.27
<i>2moFWII_r<sub>it</sub></i> , <i>r</i> = 1, ..., 4	0.38
<i>timeper_τ<sub>t</sub></i> , <i>τ</i> = 1, ..., 39	0.41

**Table 1.** Adjusted R<sup>2</sup> for the eight initial linear regression models. All regressions include an intercept plus the variables listed.

from the seven initial models were then added in order of decreasing adjusted R<sup>2</sup>. Following this procedure, the mean trend model grew to contain 10 variables —{*2moFWII\_r<sub>it</sub>*, *r* = 1, ..., 4}, *closest\_inlet\_dist<sub>it</sub>*, {*1wkFWII\_r<sub>it</sub>*, *r* = 1, ..., 4}, and *inverse\_days\_survey<sub>t</sub>*—with adjusted R<sup>2</sup> 0.57.

Because the effect of FWI from one river on a given location in PS could change based on the FWI from another river during the same time period, we evaluated the addition of the 6 pair-wise interactions among the four *1wkFWII\_r<sub>it</sub>*, the 6 pair-wise interactions among the four *2moFWII\_r<sub>it</sub>*, and the twelve interactions between the *1wkFWII\_r<sub>it</sub>* and the *2moFWII\_r<sub>it</sub>*, excluding interactions of one river's *1wkFWII\_r<sub>it</sub>* with its own *2moFWII\_r<sub>it</sub>*. Despite a decrease in error degrees of freedom by 24, adjusted R<sup>2</sup> was 0.66, so the set was retained.

Spatial coordinate variables were evaluated last in groups according to their polynomial order, with squared and cubic terms added before interactions. We considered these variables last because we wanted to include them only if they explained additional variability in the response after more interpretable variables were included. We determined that including all variables except *northing<sub>it</sub><sup>2</sup> \* easting<sub>it</sub><sup>2</sup>* increased the adjusted R<sup>2</sup>. The final process model mean function thus had an adjusted R<sup>2</sup> of 0.73 and included the following: {*2moFWII\_r<sub>it</sub>*, *r* = 1, ..., 4}; *closest\_inlet\_dist<sub>it</sub>*; {*1wkFWII\_r<sub>it</sub>*, *r* = 1, ..., 4}; {*1wkFWII\_r<sub>it</sub>* \* *1wkFWII\_q<sub>it</sub>*, *r* ≠ *q*}; {*2moFWII\_r<sub>it</sub>* \* *2moFWII\_q<sub>it</sub>*, *r* ≠ *q*}; {*1wkFWII\_r<sub>it</sub>* \* *2moFWII\_q<sub>it</sub>*, *r* ≠ *q*}; *inverse\_days\_survey<sub>t</sub>*; *easting<sub>it</sub>*, *easting<sub>it</sub><sup>2</sup>*, *northing<sub>it</sub>*, *northing<sub>it</sub><sup>2</sup>*, and interactions *northing<sub>it</sub>* \* *easting<sub>it</sub>*, *northing<sub>it</sub><sup>2</sup>* \* *easting<sub>it</sub>*, and *northing<sub>it</sub><sup>2</sup>* \* *easting<sub>it</sub><sup>2</sup>*.

2.11 Time model

To build the time model, we followed the same procedure described above, selecting for the base of the mean function a set of time period indicator variables because a linear regression of *sal<sub>it</sub>* on these variables had an adjusted R<sup>2</sup> of 0.41 (Table 1). (Note that such a model is equivalent to fitting an ANOVA model using the time periods as groups.) Again, we added other sets of explanatory variables in order of decreasing adjusted R<sup>2</sup>. Before evaluating interactions, the mean trend time model had an adjusted R<sup>2</sup> of 0.78 and contained 48 variables: {*timeper\_τ<sub>t</sub>*, *τ* = 1, ..., 39}, {*2moFWII\_r<sub>it</sub>*, *r* = 1, ..., 4}, *closest\_inlet\_dist<sub>it</sub>*, and {*1wkFWII\_r<sub>it</sub>*, *r* = 1, ..., 4}. When interactions among the {*timeper\_τ<sub>t</sub>*, *τ* = 1, ..., 39} and the {*2moFWII\_r<sub>it</sub>*, *r* = 1, ..., 4} were added, the

model was not full rank (not all columns in the design matrix were linearly independent). Because we created this second model to evaluate these interactions, we removed the  $\{1wkFWII_{rit}, r = 1, \dots, 4\}$ , the most recent variable addition, to include them. This new model, including the interactions, became the base since its adjusted  $R^2$  (0.89) was larger than that of the previous mean trend time model (0.78). After investigating spatial coordinate variables, the final mean trend time model (below) had an adjusted  $R^2$  of 0.91 and included 204 variables:  $\{timeper_{\tau_t}, \tau = 1, \dots, 39\}$ ,  $\{2moFWII_{rit}, r = 1, \dots, 4\}$ ,  $closest\_inlet\_dist_{it}$ ,  $\{timeper_{\tau_t} * 2moFWII_{rit}; \tau = 1, \dots, 39; r = 1, \dots, 4\}$ ,  $easting_{it}$ ,  $easting_{it}^2$ ,  $northing_{it}$ , and  $northing_{it}^2$ . To avoid confusion later, note that the adjusted  $R^2$  of 0.73 for the process model and 0.91 for the time model were based on fitting each model to the full dataset. In the next section, we report  $R^2$  (not adjusted  $R^2$ ) based on a cross-validation dataset.

## 2.12 Modeling spatially correlated error

The variable selection analyses above used ordinary least squares (OLS) regression to model salinity as a function of explanatory variables. That model can be written as

$$sal_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \dots + \beta_P x_{Pit} + \varepsilon_{it}, \quad t = 1, \dots, 40, \quad i = 1, \dots, n_t \quad (2)$$

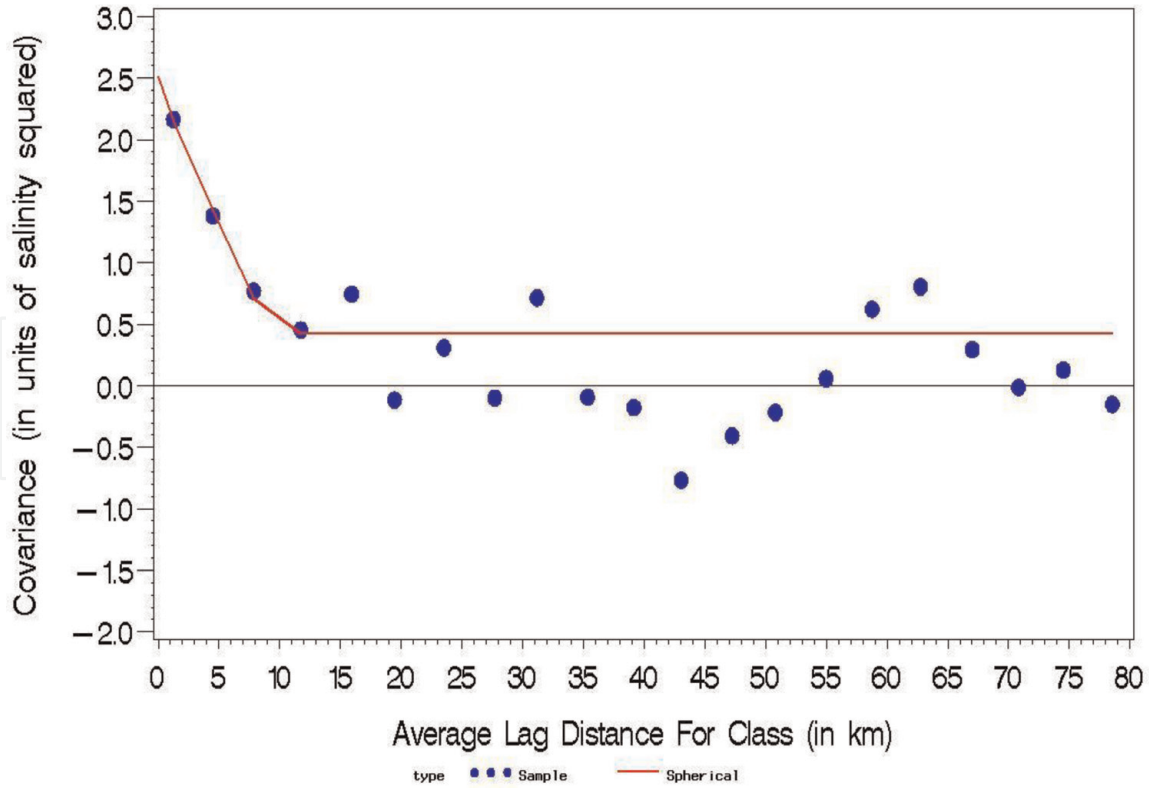
where  $x_{pit}$  represents the value of the  $p^{th}$  explanatory variable at space-time location  $it$ , for  $p = 1, \dots, P$ , where  $P$  is the total number of explanatory variables.  $\beta_0, \beta_1, \dots, \beta_P$  represent the intercept and regression coefficients, and deviations from the mean trend  $\varepsilon_{it}$  are assumed to be independent and identically distributed  $\varepsilon_{it} \sim N(0, \sigma^2)$  with mean 0 and variance  $\sigma^2$ . The model can be equivalently written as  $sal_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \varepsilon_{it}$ , where  $\mathbf{x}_{it}$  is the  $(P + 1) \times 1$  vector containing the values of the explanatory variables at space-time location  $it$ , and  $\boldsymbol{\beta}$  represents the  $(P + 1) \times 1$  vector of regression coefficients. The same model written in matrix form is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (3)$$

where bold print indicates vectors so that  $\mathbf{Y}$ ,  $\boldsymbol{\varepsilon}$ , and  $\mathbf{0}$  are  $N \times 1$  vectors containing, respectively, all observations of salinity in the space-time domain, all deviations from the mean function, and all zeros.  $\mathbf{X}$  is the  $N \times (P + 1)$  design matrix whose rows represent space-time locations and whose columns contain the values of the explanatory variables (with a column of ones for the intercept), and  $\mathbf{I}$  is the  $N \times N$  identity matrix. Since a histogram of salinity observations is somewhat symmetric and bell-shaped, use of the normal distribution is justified.

Rarely, however, does the assumption of independent and identically distributed errors hold for observations of natural phenomena associated with locations in space and time. While it is intuitive that values of salinity located close together in space should be similar, it is also generally the case that the deviations from the mean function of observations located close together are similar. That similarity is referred to as spatial covariance, and the spatial covariance between deviations from the mean trend at two locations within the same time period can be modeled as a function of the distance separating them. Including in the overall model both a deterministic mean function and a spatial covariance function allowed predictions of salinity at locations where there were no observations.





**Figure 3.** Sample covariogram for June 1994 calculated from process model residuals (blue dots). The solid red line illustrates a spherical covariance function fit to the covariogram. Covariance is in units of salinity squared.

Valid covariance functions ensure that the covariance matrix will be positive definite, which, in turn, ensures that variances will be non-negative. Each covariance function has a shape defined by a range parameter, a partial sill, and sometimes a nugget effect. Appendix Table A1 gives formulas for determining spatial covariance according to the exponential, Gaussian, and spherical covariance functions, each with and without a nugget effect. **Figure 3** shows an example of the spherical covariance function—the solid red line—fit to a sample covariogram—the blue dots—of deviations from the process model for June 1994. The range parameter— $\theta$  for the exponential and Gaussian covariance functions,  $\rho$  for spherical—is related to the distance that must separate two sites before their deviations are independent, where independence corresponds to a covariance of zero or virtually zero. In **Figure 3**, the range is approximately 10 km. In the absence of a nugget effect, the partial sill  $\sigma^2$  is the value of the covariance at distance zero—that is, it is the variance of deviations from the mean—and in **Figure 3** this value is approximately 2.5 squared units of salinity. In the presence of the nugget  $\sigma_n^2$ , there is a discontinuity in the covariance function at distance zero, so that the intercept is slightly greater than the limit of the smooth part of the function as distance approaches zero. In this case, the variance of the deviations is equal to the sum of the partial sill and nugget:  $\sigma^2 + \sigma_n^2$ . It may be the case that variance is higher when values of deviations are higher. Since covariance parameters represent physical quantities that may change over time, we used the capabilities of SAS<sup>®</sup> Proc Mixed to allow a different partial sill and range parameter for each time period.

Model (3), modified to include spatial correlation, becomes

$$Y = X\beta + \epsilon, \quad \epsilon \sim N(\mathbf{0}, \Sigma), \quad (4)$$

where  $\Sigma$  represents the  $N \times N$  block-diagonal covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \Sigma_T \end{bmatrix},$$

where zero matrices for off-diagonal elements indicate that deviations in one time period are not correlated with those in another. We make this assumption partially due to the long time span separating June and September, but also because no SAS<sup>®</sup> procedure has the capacity to model such space-time correlation while at the same time allowing every time period to have different spatial covariance parameters and allowing a mean function to be fit. Diagonal elements  $\Sigma_1, \Sigma_2, \dots, \Sigma_t, \dots, \Sigma_T$  are individual spatial covariance matrices for each time period with dimensions  $n_t \times n_t$ , and elements  $[\Sigma_t]_{ij} = \text{Cov}\{\varepsilon_{it}, \varepsilon_{jt}\}$  representing the spatial covariance between sites  $i$  and  $j$  in time period  $t$ .

Understanding how predictions of salinity and prediction standard errors are generated from this model will make the results and analysis in Sections 6 and 7 easier to understand. To predict salinity at space-time locations where it is not observed, the following results are needed. Superscripts differentiate between locations where salinity is observed and unobserved. Model (4), represents observations of salinity (by virtue of the dimensions of the vectors and matrices), but we model salinity observations and unobserved values of salinity at other space-time locations using a similar model, the joint distribution of unobserved and observed salinity, given by

$$\begin{pmatrix} \mathbf{Y}^o \\ \mathbf{Y}^u \end{pmatrix} \sim N \left\{ \begin{pmatrix} X^o \boldsymbol{\beta} \\ X^u \boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \Sigma^o & \Sigma^{ou} \\ \Sigma^{uo} & \Sigma^u \end{pmatrix} \right\}. \quad (5)$$

Here,  $\mathbf{Y}^o$  represents the  $N \times 1$  vector of salinity observations, and, letting  $N^u$  represent the number of space-time locations at which we want to predict salinity,  $\mathbf{Y}^u$  represents the  $N^u \times 1$  vector of unknown values of salinity at these locations. All the symbols in (5) have the same meaning as in (4), except for the distinction between observed and unobserved locations. The  $N \times N^u$  matrix  $\Sigma^{ou}$  contains the cross-covariance between observed and unobserved locations. Thus,

$$\Sigma^{ou} = \begin{bmatrix} \Sigma_1^{ou} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_2^{ou} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \Sigma_T^{ou} \end{bmatrix},$$

and  $\Sigma^{uo} = (\Sigma^{ou})^T$ . The elements  $[\Sigma_t^{ou}]_{ij} = \text{Cov}\{\varepsilon_{it}^o, \varepsilon_{jt}^u\}$  also come from the spatial covariance function.

Let  $\boldsymbol{\psi}$  represent the vector that contains all spatial covariance parameters for every time period—either 80 or 81 parameters depending on whether a nugget effect is used. Standard normal distribution theory gives the distribution of unobserved salinity  $\mathbf{Y}^u$  conditioned on knowing the values of observations  $\mathbf{Y}^o$  and all of the parameter values:

$$\mathbf{Y}^u | \mathbf{Y}^o, \boldsymbol{\beta}, \boldsymbol{\psi} \sim N \left\{ X^u \boldsymbol{\beta} + \Sigma^{uo} (\Sigma^o)^{-1} (\mathbf{Y}^o - X^o \boldsymbol{\beta}), \Sigma^u - \Sigma^{uo} (\Sigma^o)^{-1} \Sigma^{ou} \right\}. \quad (6)$$

The pipe symbol ( $|$ ) means “given” or “conditioned on knowing the values of” the terms following the pipe symbol. The terms before the comma represent the mean of the multivariate normal distribution, which is used for the salinity prediction, and the terms after the comma represent the variance-covariance matrix, which is used for prediction standard errors. Salinity predictions are the sum of the mean trend,  $X^u\beta$ , and the spatial interpolation of observation deviations from the mean trend,  $\Sigma^{uo}(\Sigma^o)^{-1}(Y^o - X^o\beta)$ . If the deviations  $(Y^o - X^o\beta)$  are large for a given time period, then the partial sill  $\sigma_t^2$  will be large for that time period, so that diagonal elements of  $\Sigma^o$  and  $\Sigma^u$  will be large. For a given location, the prediction standard error is the diagonal element of the matrix  $\Sigma^u - \Sigma^{uo}(\Sigma^o)^{-1}\Sigma^{ou}$ . If the diagonal elements of  $\Sigma^o$  and  $\Sigma^u$  are large, then the diagonal elements of  $(\Sigma^o)^{-1}$  are small, and the prediction standard error is a large number minus a small number. That is, the prediction standard error will be high for time periods in which observation deviations from the mean function are large. When observation deviations from the mean trend are small, the reverse is true, and prediction standard errors tend to be low for that time period.

The salinity predictor

$$X^u\beta + \Sigma^{uo}(\Sigma^o)^{-1}(Y^o - X^o\beta) \quad (7)$$

is an *exact predictor*: the prediction of salinity at a site where there is an observation will exactly equal the observation. For this reason, to determine which spatial covariance function to use, we randomly selected 10% of the observations to withhold as a cross-validation dataset, the *test dataset*; the remaining 90% we term the *base dataset*. For every combination of the two mean functions—process and time—and the six spatial covariance functions in Appendix Table A1, we fit model (4) to the base dataset, and predicted salinity values at the space–time locations of the test dataset using the results given in (5) and (6). When the model predicted salinity to be less than zero, we set the prediction equal to zero before calculating the following statistics. Predictions of negative values could be avoided using a truncated normal distribution, but SAS<sup>®</sup> Proc Mixed does not permit specification of this distribution. The root mean squared error (RMSE) of predictions—with the same units as salinity—are given in **Table 2**, along with the slope, intercept, and coefficient of determination ( $R^2$ ) from a regression of actual salinity values in the test dataset on predictions of them. If predictions were perfect, this regression would have slope equal to one, intercept equal to zero, and  $R^2$  equal to 1.

Salinity predictions are better when a spatial covariance function is combined with either mean function. For example, of the time models, the exponential covariance function with a nugget produced predictions with the lowest RMSE (2.1), slope closest to one (0.92), and intercept closest to zero (1.55). Comparing process models, the exponential and spherical, each with and without a nugget, performed equally well, and better than the time models. To select the best model from this group of four, we examined statistics based on how well the model fit the base dataset. The model with an exponential covariance function with a nugget had the lowest AIC (7580.0) and BIC (7711.7) and was thus chosen as the final model. It explained 89% of variability in the test dataset and generated predictions with RMSE 2.0.

Next, we fit this model using the full dataset, and produced retrospective maps of salinity predictions and standard errors at evenly spaced 1 nmi (1.85 km) increments for each time period. Forty-two salinity predictions—less than 0.1% of the total number of predictions—were negative and set to zero.

	Model type	−2 log likelihood	AIC	BIC	RMSE (psu)	Slope/ $\beta_1$	Intercept/ $\beta_0$	R <sup>2</sup>
Process	IID	9935.9	9937.9	9943.5	2.9	0.98	0.84	0.74
	Exponential	7430.7	7584.7	7714.7	2.0	0.95	1.03	0.89
	Exponential + $\sigma_n^2$ *	7424.0	7580.0	7711.7	2.0	0.96	0.96	0.89
	Gaussian	8198.0	8356.0	8489.5	2.3	0.94	1.37	0.84
	Gaussian + $\sigma_n^2$ *	7532.0	7686.0	7816.0	2.1	0.94	1.15	0.87
	Spherical	7570.0	7722.0	7850.4	2.0	0.95	1.07	0.88
	Spherical + $\sigma_n^2$ *	7571.6	7727.6	7859.3	2.0	0.96	0.93	0.89
Time	IID	7077.5	7079.5	7084.9	2.6	0.83*	3.47*	0.83
	Exponential	Infinite						
	Exponential + $\sigma_n^2$	6217.1	6367.1	6493.7	2.1	0.92*	1.55*	0.87
	Gaussian	6281.0	6433.0	6561.3	2.2	0.90*	1.98*	0.86
	Gaussian + $\sigma_n^2$ *	6214.0	6366.0	6494.4	2.2	0.91*	1.90*	0.86
	Spherical	6199.6	6315.6	6479.9	2.2	0.91*	1.86*	0.86
	Spherical + $\sigma_n^2$	6201.3	6357.3	6489.1	2.2	0.91*	1.86*	0.86

Process and time mean functions with no spatial covariance (IID) and with each of six covariance functions were used. The symbol “ $\sigma_n^2$ ” indicates that a nugget was included. Stars (\*) indicate rejection of the appropriate null hypothesis at the  $\alpha = 0.05$  level of significance:  $H_{01}$ :  $\sigma_n^2 = 0$ ;  $H_{02}$ :  $\beta_1 = 1$ ;  $H_{03}$ :  $\beta_0 = 0$ . The exponential plus nugget process model is highlighted as it was chosen as the best model of PS salinity for our modeling context.

**Table 2.**  
Summary statistics comparing salinity observations in the test dataset to predictions based on fitting models to the base dataset.

2.13 Examining freshwater influx scenarios

To examine variations in the spatial distribution of salinity under drought, average, and flood conditions, we classified freshwater influx from each river within each time period ( $1wkFWI_{r_t}$  and  $2moFWI_{r_t}$ ) as LOW if it fell below the 25th percentile of observed FWI across all time periods, MODERATE if it fell between the 25th and 75th percentiles, HIGH if it fell between the 75th and 95th percentiles, and FLOOD if it fell above the 95th percentile. Next, we classified one-week and two-month FWI for the entire time period as LOW (or HIGH) if at least two rivers exhibited low (or high) inflow, MODERATE if at least three rivers exhibited moderate inflow, and FLOOD if at least one river exhibited extremely high (>95th percentile) inflow. These classifications are mutually exclusive, though some of the 40 time periods did not fall into any of them. The first two columns of **Table 3** list the 16 combinations of classifications, and the third column shows the classification and salinity rank for each time period. Time periods were ranked 1–40 by mean predicted salinity (1 = highest mean salinity; 40 = lowest).

*Moderate-to-moderate FWI.* June 2005 (**Figure 4**) experienced moderate FWI in both the 2 months and 1 week prior to the survey in PS with predicted salinity ranked 37th—the lowest of the moderate-to-moderate time periods. Legend colors for model predictions in the left pane and observations in the upper right pane of **Figure 4** (as well as **Figure 5A** and **B**, **6A** and **B**) are based on percentiles of the distribution of observed salinity across all time periods: minimum to 5%; 5–10%; 10–25%; 25–50%; 50–75%; 75–90%; 90–95%; and 95% to maximum. From the left pane of **Figure 4**, predicted salinity in June 2005 increased moving east across PS,



<i>2mo_FWI<sub>rt</sub></i>	<i>1wk_FWI<sub>rt</sub></i>	Time periods and mean predicted salinity rank (mmyy, r)
Flood	Flood	(0603, 40), (0999*, 30)
	High	none
	Moderate	(0687, 28), (0689, 27)
	Low	None
High	Flood	(0903*, 39), (0690, 29)
	High	(0904*, 32)
	Moderate	(0698, 38), (0693, 36), (0697, 35)
	Low	None
Moderate	Flood	(0996*, 33)
	High	(0696, 26), (0900, 24)
	Moderate	(0605, 37), (0989, 31), (0601, 26), (0600, 22), (0604, 21), (0688, 16), (0990, 13), (0692, 10)
	Low	(0694, 17)
Low	Flood	(0987, 18)
	High	(0695, 6)
	Moderate	(0905, 20)
	Low	(0997, 15), (0699, 12), (0901, 8), (0902, 7), (0993, 5), (0602, 4), (0988, 3), (0994, 1)

Only time periods that fit each scenario as defined in Section 6 are listed; the remaining 7 time periods were not classified. Boldfaced time periods are examined in Section 6. Stars (\*) indicate time periods in which hurricanes occurred within the 61 days prior to the survey.

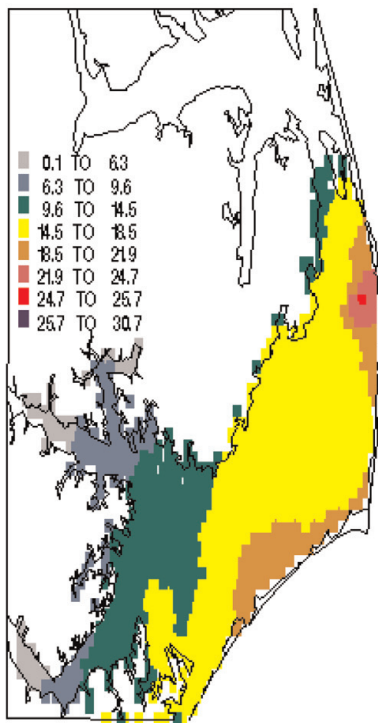
**Table 3.**  
Sixteen combinations of *2mo\_* and *1wk\_FWI<sub>rt</sub>* classifications; time periods that exhibit each set of conditions; and mean predicted salinity rank (1 = highest).

reaching a maximum just south of Oregon Inlet. We note the same east-west salinity gradient when comparing this pane to the June 2005 map of observed salinities (top right pane), indicating that prediction maps typically mirror trends seen in observation maps. The area of highest predicted salinity corresponds to a lone purple observation of 26.5 just south of Oregon Inlet (**Figure 4**). Plumes of relatively higher salinity are evident in the vicinity of all three ocean inlets (**Figure 4**).

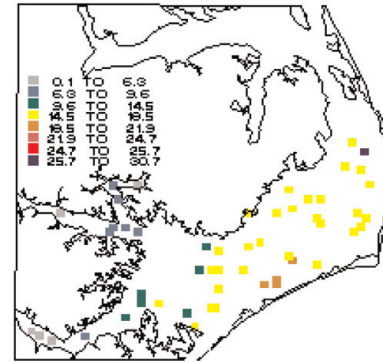
The lower right pane of **Figure 4** (as well as **Figure 5A and B, 6A and B**) displays prediction standard errors (SE) with the same units as salinity. The same eight percentile groups classify colors on the SE legend, here based on the distribution of prediction standard errors across all time periods. The transition from low SE at sample sites to higher SE moving away from sample sites reflects the fact that the exact predictor (6) reproduces observations, so confidence intervals closer to sample sites are narrower than those further away.

This spatial trend in SEs is further illustrated by comparing locations of high SE in the same time period, which are also consistent over time. High SEs occur between the mouths of the Neuse and Pamlico Rivers and along a margin of varying width following the outline of the Outer Banks, areas within which sampling does not occur (**Figure 1**). We note here that because SEs increase as distance from sample site increases, we chose to generate only interpolated (and not extrapolated)

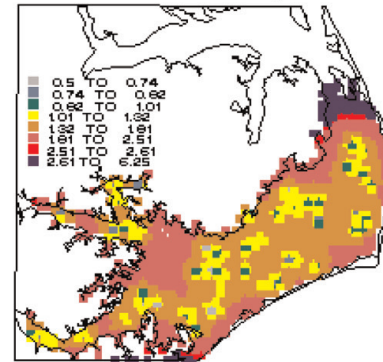
**Predicted P195 Bottom Salinity June 2005**



**Observed P195 Bottom Salinity June 2005**



**SE of P195 Bottom Salinity June 2005**



**Figure 4.**

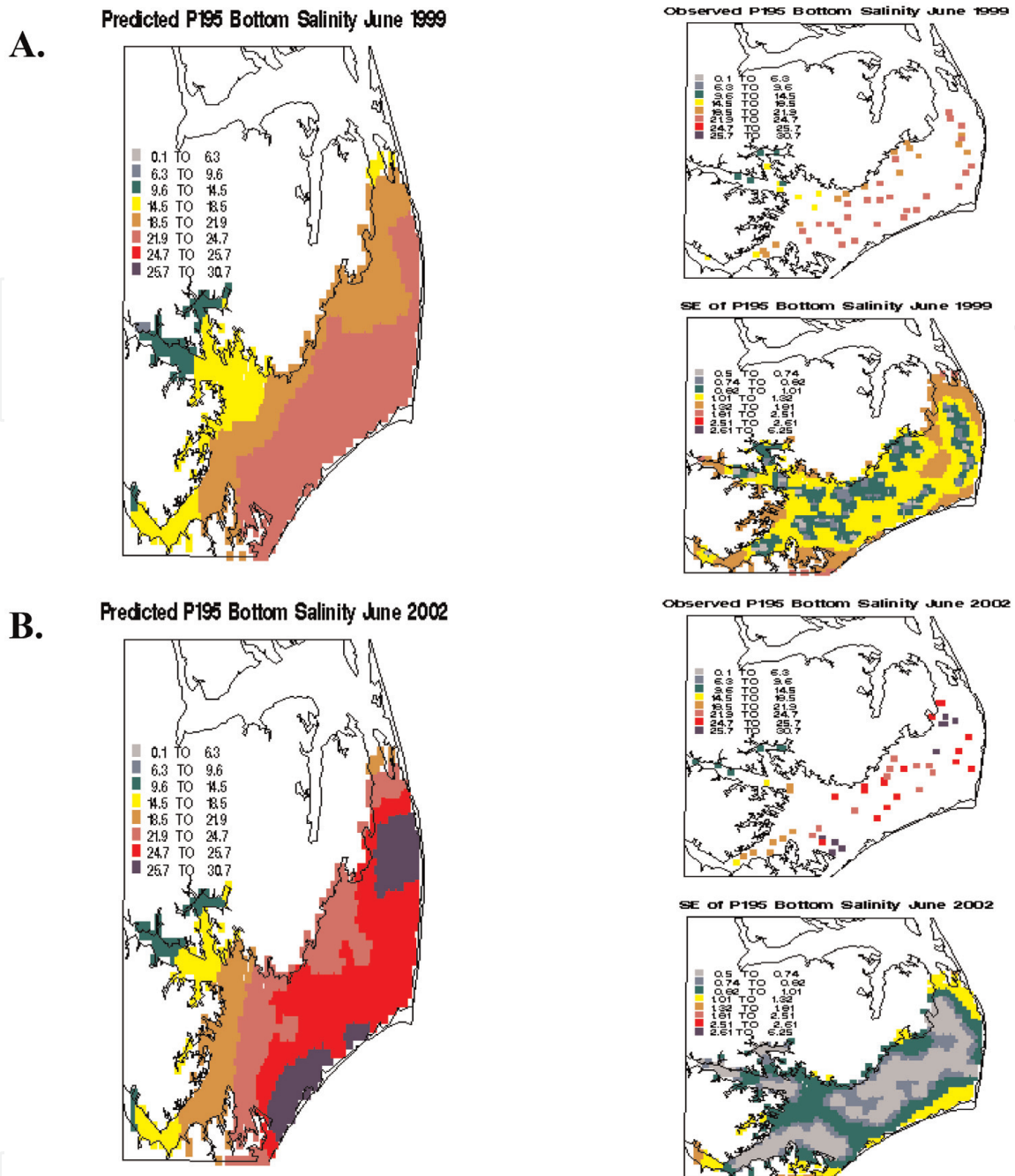
Salinity model predictions (left), prediction standard errors (bottom right), and P195 survey observations (top right) for June 2005, classified as moderate-to-moderate FWI.

salinity predictions. In June 2005, as in all other time periods, predictions were generated only for locations within S, which does not extend either to Albemarle Sound or to the heads of the Neuse and Pamlico Rivers (Figure 4).

*Low to low FWI in early and late-stage drought.* June 1999 (Figure 5A) and June 2002 (Figure 5B)—which mark early and late stages of North Carolina's 1998–2002 drought [42]—experienced low long- and short-term FWI with predicted salinity ranking 12th and 4th, respectively. At every point in PS, predicted salinity in these two time periods was higher than in June 2005, and predicted salinity was much higher in June 2002 than June 1999, though both have similar values for  $1wkFWI_{r_t}$  and  $2moFWI_{r_t}$  variables from three of the four tributary rivers. The difference may be due to (1) the fact that in the fourth river, the Roanoke,  $1wkFWI_{r_t}$  and  $2moFWI_{r_t}$  in June 1999 were twice their values in June 2002, or (2) that by June 2002, NC had been experiencing drought conditions for 4 years (186 weeks) as opposed to less than one (30 weeks) and that this cumulative FWI deficit became more pronounced over time.

Though June 2002 salinity observations have a larger mean and greater variability, the majority of prediction standard errors are less than 1.01. In June 1999, however, SEs fell between 1.01 and 1.81 at all prediction locations except those that were very close to observations. This result shows that the conditions affecting salinity in PS were better represented by the mean function in June 2002 than they were in June 1999.

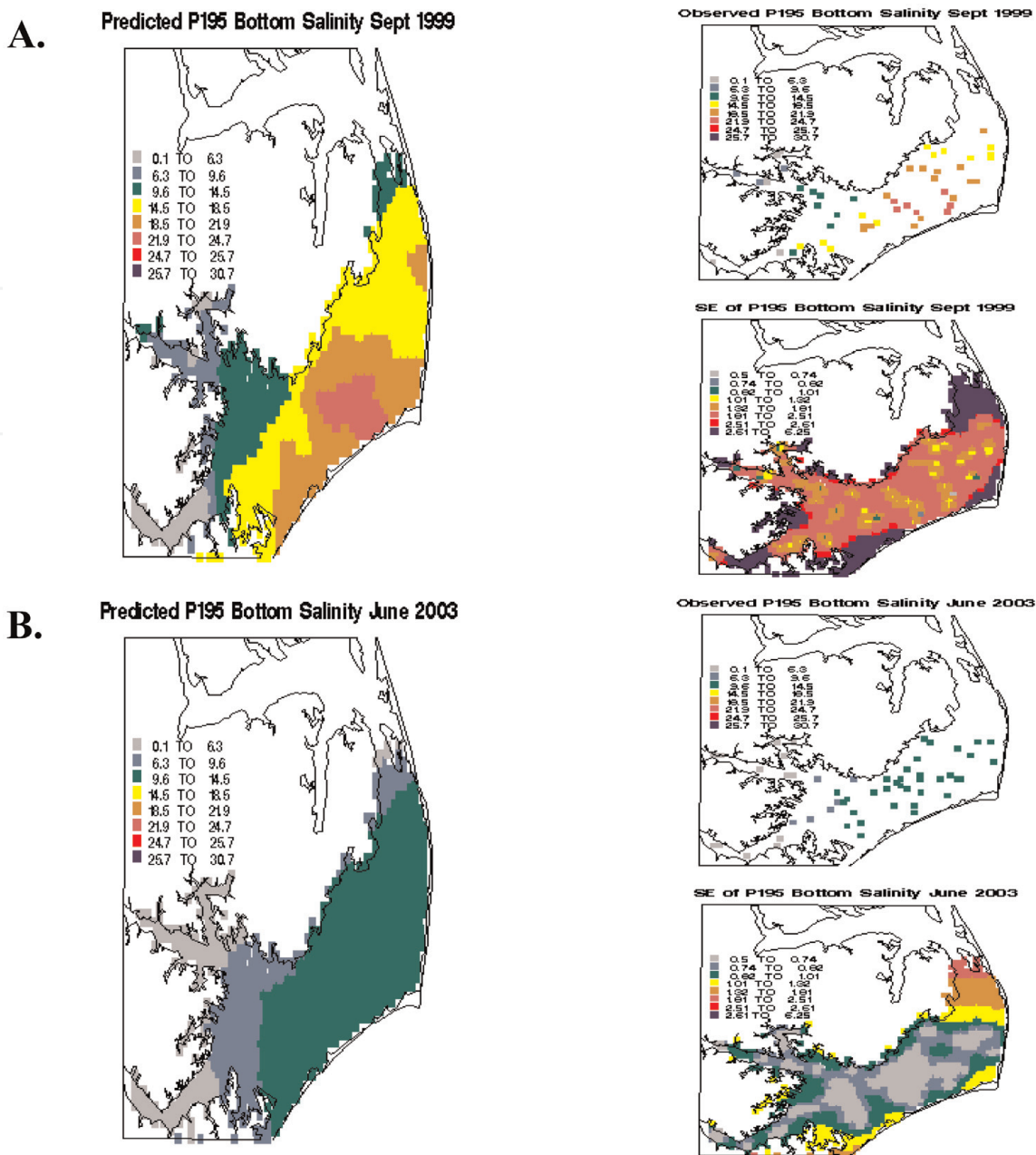
*Flood to flood FWI—with and without hurricanes.* FWI was extremely high in September 1999 (Figure 5A) as a result of the 500-year floods produced by Hurricanes Dennis and Floyd that occurred 24 and 12 days before the survey, respectively. In June 2003 (Figure 5B), extremely high FWI was due to an eight-month period of above-average precipitation totals prior to the survey. Though these are



**Figure 5.** Salinity model predictions (left), prediction standard errors (bottom right), and P195 survey observations (top right) for June 1999 (A) and June 2002 (B), both classified as low-to-low FWI.

the only two time periods categorized as flood-to-flood, predicted salinity in September 1999 ranks a surprisingly high 30th, while in June 2003 it ranks 40th. Observed and predicted salinity for these two time periods are lower than those in the low-FWI time periods of June 1999 and 2002, but in September 1999, salinity was higher at most prediction locations, and more variable, than in moderate-FWI June 2005. Water at locations near the two southerly inlets to PS was more saline in September 1999 than in these same locations during moderate-FWI of June 2005 likely due to storm surge-generated inlet plumes. Salinity at locations near the Neuse and Tar-Pamlico Rivers was similar to that in June 2005. Standard errors were lower sound-wide in June 2003 than in September 1999. SEs in September 1999 were highest sound-wide relative to the other four time periods examined (Figures 5 and 6).





**Figure 6.** Salinity model predictions (left), prediction standard errors (bottom right), and P195 survey observations (top right) for September 1999 (A) and June 2003 (B), both classified as flood-to-flood FWI.

### 3. Discussion

Because water exchange between lagoonal estuaries and the open ocean can be relatively restricted, there is a relatively high potential in systems like PS for changes in precipitation patterns and storm frequencies associated with global climate change to result in changes in salinity patterns and subsequent ecosystem alterations. Changes in precipitation will affect the amount and timing of river flow, which will impact nutrient cycling, estuarine flushing rates, and salinity. Increased storm activity may open new inlets, which would alter current flow, increase tidal action, and allow a greater influx of seawater that carries with it both different chemical signals and mobile species. Salinity is therefore a practical estuarine characteristic to use to study the impacts of these changes, as both effects mentioned above include enhanced water exchange that impacts overall estuarine salinity content [43, 44].



We developed and evaluated two statistical models, using the best model to hindcast salinity in PS. The process mean function combined with the exponential covariance with a nugget explained 89% of the variability in a test dataset with a RMSE of 2.0 and produced relatively accurate retrospective salinity maps under a wide range of freshwater influx and system-state scenarios. Much of this accuracy was due to allowing the range and partial sill parameters of the spatial covariance to be time-period specific. We then examined variations in the spatial distribution of salinity under varying freshwater influx (FWI) conditions such as drought, average FWI, and flood conditions, and identified the following patterns. In years with moderate FWI, the salinity gradient increased from west to east in PS as expected, and was highest adjacent to the major inlets, with highest salinities near Oregon Inlet. In years with low FWI indicative of drought conditions, the overall mean and variance in salinity increased in PS. In years with floods, salinities displayed a high degree of spatial variation, with salinities being lower near the tributaries as expected, yet also displaying occasional sharp increases in salinity near inlets due to influx of ocean water into PS via the major inlets.

### 3.1 Improvements to model predictions

For retrospective prediction purposes, model improvements could focus on improvements to the mean trend, the covariance, or both, and such improvements could be evaluated using the test dataset. A reasonable goal might be to increase  $R^2$  to 0.93 or to reduce RMSE to 1.5. Improvements for the purpose of prospective prediction of salinity under hypothetical, unobserved conditions, a situation in which spatial covariance among observation deviations cannot be used, would entail improving the mean function exclusively. Locations and time periods with high SEs highlight conditions not well-represented by the current mean function. A reasonable goal here would be to produce a model for which all values of SE fall beneath the current median (1.32).

*Mean function.* The mean function alone explained over two-thirds of the variability in salinity in both process and time models. While this is a noteworthy accomplishment, there remains room for multiple improvements. High SE values in **Figure 5A** show that the mean function is unable to capture the interaction between high FWI in September 1999 and hurricane storm surges. One hurricane explanatory variable, *inverse\_days\_survey<sub>t</sub>*, remained in the final process model. Its parameter estimate was positive, reflecting that strong hurricane winds push more saltwater into PS through inlets than would enter under typical seasonal wind conditions, but alone it explained only 4% of salinity variability in the full dataset. The *inverse\_days\_survey<sub>t</sub>* variable did not differentiate between a year in which a single hurricane passed within 12 days of the survey and a year in which such a hurricane followed another that passed 12 days earlier. A future effort might attempt to account for cumulative build-up of storm surge on observed PS salinities.

Though *closest\_inlet\_dist<sub>it</sub>* alone explained a third of the variability in salinity over all time periods, variability in inlet-plume size across **Figures 3–5** suggests that this distance metric should be modified based on wind speed and direction, using more finely resolved wind information than the *month<sub>t</sub>* variable. Devising a way to use the *u* and *v* components of wind to interact with *closest\_inlet\_dist<sub>it</sub>* could allow both the size and the direction of the inlet plume to vary such that east-to-west winds create different plume sizes and shapes than winds from the southeast-to-northwest. Considerable exploratory analysis would be needed to determine what pre-survey time lag should be considered to affect observed survey salinities.

Differences in both salinity values and SE estimates between early-stage drought during June 1999 and late-stage drought during June 2002 suggest accounting for

effects of FWI over a longer duration than 61 days. Doing so might explain differences in salinity patterns seen in time periods with similar one-week and two-month FWI conditions. Molina [45] calculated an 11 month mean residence time for freshwater in PS. We could incorporate this effect by adding a third freshwater influx index to the mean function or by adding an autoregressive component to the model so that salinity in a given time period was a function of mean salinity in the previous time period. The first option would be tedious from a data-manipulation standpoint, but much easier from a mathematical model-fitting standpoint, because SAS<sup>®</sup> Proc Mixed could still be used. The second option necessitates a change in the covariance function, as we can no longer assume that salinity deviations from the mean function at a given space-time point were independent in time. This second option would also require specialized hand-written code, as no current SAS<sup>®</sup> Proc allows such a dynamic space-time model to be fit.

Differences in salinity patterns between June 1999 and June 2002, our two low-to-low FWI time periods, could be attributed to differences in FWI from the Roanoke River, one of the two northern rivers whose connection to PS is indirect. This observation warrants further investigation into the calculation of the FWII indices; namely, an investigation of water-path distance as a possible substitute for crow-flies distance between river gauges and sites in PS. Although we did not find a study that demonstrated marked predictive improvement using water-path distance under all circumstances ([36, 46], and others), it would be interesting in future work to compare differences in PS salinity predictions using both distance methods. Recall that Gardner et al. [34] noted more accurate predictions of stream temperatures when models incorporated water-path distance, but only when this distance was further modified and weighted by stream order. It might be the case that water-path distance out-performs crow-flies distance in predicting estuarine salinity when care is taken to make all explanatory variables as meaningful as possible. Development of an automated procedure for calculating water-path distances similar to the one used in [47] would make such an investigation more practically feasible.

*Covariance function.* Two mutually-exclusive improvements to the covariance function, as implemented in SAS<sup>®</sup> Proc Mixed, could be investigated: using either the Matern covariance function or an anisotropic covariance function to achieve greater flexibility in each time period. The Matern covariance function has a smoothing parameter in addition to partial sill and range parameters. When the smoothing parameter takes the value of 0.5, the Matern covariance function is the same as the exponential covariance function—as the smoothness parameter approaches infinity, the covariance function approaches the Gaussian covariance function. Using the Matern covariance function is thus equivalent to allowing a third parameter to determine which two-parameter covariance function is appropriate, as opposed to using the same two-parameter covariance function for every time period. The computational cost of this flexibility is high—in a similar model with only four separate groups of covariance parameters, compared to the 40 groups in this paper—co-author Amy Nail experienced computation time of 2 h (versus a 2 min run time using the two-parameter exponential covariance function here). The added computational burden is due to the complex nature of the Matern covariance function and to the necessity of estimating one additional covariance parameter per time period (for a total of 40 additional parameters).

Another way to achieve flexibility while still specifying a single covariance function for every time period, would be to allow an anisotropic covariance function. Geometric anisotropy allows for different range parameters in different directions. For example, if the water current in PS were flowing directly north-to-south, two points separated by a north-to-south vector might have more similar values of salinity than would two points separated by a west-to-east vector of the same length.

Fortunately, the parameterization of a geometric anisotropic covariance function is such that if anisotropy were unnecessary, the parameters would take values that effectively result in an isotropic covariance function. The cost of this added flexibility is the need to estimate two additional covariance parameters per time period, for a total of 80 additional parameters. Computation time might be less here than for Matern, since anisotropic covariance functional forms are less complex.

4. Conclusions

We created a statistical model combining a process mean function with an exponential spatial covariance function with a nugget to predict salinity in a lagoonal estuary. This model can generate predictions of bottom salinity for Pamlico Sound, NC that are more spatially-resolute than any previous bottom salinity predictions encountered in the literature for this system. The salinity maps produced using the model are useful for researchers to build an intuitive understanding of salinity dynamics under PS conditions covered by these 40 time periods. Salinity predictions can also be used to inform future analyses including, but not limited to, the examination of historical distribution patterns of estuarine species relative to salinity variability and the prediction of salinity changes under various global climate change scenarios.

Acknowledgements

We thank the North Carolina Division of Marine Fisheries and the United States Geological Survey for providing datasets used in this study. We also thank editor A. Manning for helpful comments that improved the manuscript. Funding for this project was provided by the Environmental Defense Fund (Program Manager Pam Baker), North Carolina Coastal Recreational Fishing License Program (Grant No. 2010-H-004), North Carolina Sea Grant (R12-HCE-2) and the National Science Foundation (OCE-1155609) to D. Eggleston. A. Nail was supported as a VIGRE Postdoctoral Fellow by NSF grant DMS 0354189.

Appendix

See Table A1.

Name of covariance function	$\text{Cov}\{\epsilon_{ti}, \epsilon_{tj}\} =$	
	With nugget effect	Without nugget effect
Exponential	$\sigma_{n_t}^2 I(d_{ij} = 0) + \sigma_t^2 \exp\left(\frac{-d_{ij}}{\theta_t}\right)$	$\sigma_t^2 \exp\left(\frac{-d_{ij}}{\theta_t}\right)$
Gaussian	$\sigma_{n_t}^2 I(d_{ij} = 0) + \sigma_t^2 \exp\left(\frac{-d_{ij}^2}{\theta_t^2}\right)$	$\sigma_t^2 \exp\left(\frac{-d_{ij}^2}{\theta_t^2}\right)$
Spherical	$\sigma_{n_t}^2 I(d_{ij} = 0) + \sigma_t^2 \left[1 - \left(\frac{3d_{ij}}{2\rho_t}\right) + \left(\frac{d_{ij}^3}{2\rho_t^3}\right)\right] I(d_{ij} \leq \rho_t)$	$\sigma_t^2 \left[1 - \left(\frac{3d_{ij}}{2\rho_t}\right) + \left(\frac{d_{ij}^3}{2\rho_t^3}\right)\right] I(d_{ij} \leq \rho_t)$
Note: For all models, $\sigma_{n_t}^2, \sigma_t^2 > 0$ , and $\theta_t \geq 0$ , $-\infty < \rho < \infty$ , and $d_{ij}$ is the distance separating sites $i$ and $j$ . Note: $I(\text{statement}) = 1$ if statement is true and 0 otherwise.		

Table A1.  
Formulas for the three spatial covariance functions used in this analysis.

IntechOpen

## Author details

Christina L. Durham<sup>1</sup>, David B. Eggleston<sup>1\*</sup> and Amy J. Nail<sup>2</sup>

<sup>1</sup> Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC, United States

<sup>2</sup> Department of Statistics, North Carolina State University, Raleigh, NC, United States

\*Address all correspondence to: [eggleston@ncsu.edu](mailto:eggleston@ncsu.edu)

## IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 



## References

- [1] Cloern JE, Nichols FH. Time scales and mechanisms of estuarine variability, a synthesis from studies of San Francisco Bay. *Hydrobiologia*. 1985;**129**: 229
- [2] Anderson AM, Davis WJ, Lynch MP, Schubel JR. Effects of Hurricane Agnes on the Environment and Organisms of Chesapeake Bay. Baltimore, MD: The Chesapeake Bay Research Council, Johns Hopkins University; 1973
- [3] Bell GW, Eggleston DB. Species-specific avoidance responses by blue crabs and fish to chronic and episodic hypoxia. *Marine Biology*. 2005;**146**: 761-770
- [4] Mallin MA, Corbett CA. How hurricane attributes determine the extent of environmental effects: Multiple hurricanes and different coastal systems. *Estuaries and Coasts*. 2006;**29**:1046-1061
- [5] Bell GW, Eggleston DB, Wolcott TG. Behavioral responses of free-ranging blue crabs to episodic hypoxia. I. Movement. *Marine Ecology Progress Series*. 2003;**259**:215-225
- [6] Baird D, Christian RR, Peterson CH, Johnson GA. Consequences of hypoxia on estuarine ecosystem function: Energy diversion from consumers to microbes. *Ecological Applications*. 2004;**14**:805-822
- [7] Burkholder J, Eggleston D, Glasgow H, Brownie C, Reed R, Melia G, et al. Comparative impacts of major hurricanes on the Neuse River and Western Pamlico Sound ecosystems. *Proceedings of the National Academy of Science*. 2004;**101**: 9291-9296
- [8] Paerl HW, Pinckney JL, Fear JM, Peierls BL. Ecosystem responses to internal and watershed organic matter loading: Consequences for hypoxia in the eutrophying Neuse River Estuary, North Carolina, USA. *Marine Ecology Progress Series*. 1998;**166**:17-25
- [9] Paerl HW, Hall NS, Hounshell AG, Luettich RA, Rossignol KL, Osburn CL, et al. Recent increase in catastrophic tropical cyclone flooding in coastal North Carolina, USA: Long-term observations suggest a regime shift. *Scientific Reports*. 2019;**9**:10620. DOI: 10.1038/s41598-019-46928-9
- [10] Eggleston DB, Reyns NB, Etherington LL, Plaia G, Xie L. Tropical storm and environmental forcing on regional blue crab settlement. *Fisheries Oceanography*. 2010;**19**(2):89-106
- [11] Searcy S, Eggleston DB, Hare J. Environmental influences on the relationship between juvenile and larval growth for Atlantic croaker, *Micropogonias undulatus*. *Marine Ecology Progress Series*. 2007;**349**:81-88
- [12] Eggleston DB, Johnson E, Hightower J. Population Dynamics and Stock Assessment of the Blue Crab in North Carolina. Final Report for Contracts 99-FEG-10 and 00-FEG-11 to the North Carolina Fishery Resource Grant Program, NC Sea Grant, and the NC Department of Environmental Health and Natural Resources. Division of Marine Fisheries; 2004. p. 252
- [13] Lin G. A numerical model of the hydrodynamics of the Albemarle-Pamlico-Croatan Sounds system, North Carolina [M.S. thesis]. Raleigh, NC: North Carolina State University; 1992. 118 pp
- [14] Pietrafesa LJ, Janowitz GS. Physical oceanographic processes affecting larval transport around and through North Carolina inlets. *American Fisheries Society Symposium*. 1988;**3**:34-50

- [15] Bender MA, Knutson TA, Tuleya RE, Sirutis JJ, Vecchi GA, Garner ST, et al. Modeled impact of anthropogenic warming on the frequency of intense Atlantic hurricanes. *Science*. 2010;**327**:454-458
- [16] Federov AV, Brierley CM, Emanuel K. Tropical cyclones and permanent El Nino in the early Pliocene epoch. *Nature*. 2010;**463**: 1066-1070
- [17] Iglesias I, Avilez-Valente P, Pinho JL, Bio A, Vieira JM, Bastos L, et al. Numerical modeling tools applied to estuarine and coastal hydrodynamics: A user perspective. In: *Coastal and Marine Environments—Physical Processes and Numerical Modelling*. Rijeka: InTechOpen; 2019. pp. 1-20. DOI: 10.5772/intechopen.85521
- [18] Haase A, Eggleston D, Luettich R, Weaver R, Puckett B. Estuarine circulation and predicted oyster larval dispersal among a network of reserves. *Estuarine, Coastal & Shelf Science*. 2012;**101**:33-43
- [19] Puckett BJ, Eggleston DB, Kerr PC, Luettich R. Larval dispersal and population connectivity among a network of marine reserves. *Fisheries Oceanography*. 2014;**23**(4):342-361
- [20] Qiu C, Wan Y. Time series modeling and prediction of salinity in the Caloosahatchee River Estuary. *Water Resources Research*. 2013;**49**: 5804-5816. DOI: 10.1002/wrcr.20415
- [21] Ross AC, Najjar RG, Li M, Mann ME, Ford SE, Katz B. Sea-level rise and other influences on decadal-scale salinity variability in a coastal plain estuary. *Estuarine, Coastal and Shelf Science*. 2015;**157**:79-92
- [22] Lin J, Xie L, Pietrafesa LJ, Ramus JS, Paerl HW. Water quality gradients across Albemarle-Pamlico estuarine system: Seasonal variations and model applications. *Journal of Coastal Research*. 2007;**23**:213-229
- [23] Xia M, Xie L, Pietrafesa L. Modeling of the cape fear river estuary plume. *Estuaries and Coasts*. 2007;**30**: 698-709
- [24] Xu H, Lin J, Wang D. Numerical study on salinity stratification in the Pamlico River Estuary. *Estuarine, Coastal and Shelf Science*. 2008;**80**: 74-84
- [25] Hamrick JM. User's manual for the environmental fluid dynamics computer code. In: *Special Report in Applied Marine Science and Ocean Engineering No. 331*. Virginia: Virginia Institute of Marine Science/School of Marine Science, The College of William and Mary; 1996
- [26] Rathbun SL. Spatial modeling in irregularly shaped regions: Kriging estuaries. *Environmetrics*. 1998;**9**: 109-129
- [27] Chehata M, Jasinski D, Monteith MC, Samuels WB. Mapping three-dimensional water-quality data in the Chesapeake Bay using Geostatistics 1. *Journal of the American Water Resources Association*. 2007;**43**:813-828
- [28] Urquhart EA, Zaitchik BF, Hoffman MJ, Guikema SD, Geiger EF. Remotely sensed estimates of surface salinity in the Chesapeake Bay: A statistical approach. *Remote Sensing of the Environment*. 2012;**123**:522-531
- [29] Bales JD. Effects of hurricane Floyd inland flooding, September–October 1999, on tributaries to Pamlico Sound, North Carolina. *Estuaries and Coasts*. 2003;**26**:1319-1328
- [30] Bales JD, Robbins JC. Simulation of Hydrodynamics and Solute Transport in the Pamlico River Estuary, North Carolina. US Geological Survey, Raleigh,

- NC (USGS Open-file Rep. No. 94-454); 1995
- [31] Lilly JP. The Roanoke River and Albemarle sound. In: Jones FB, Phelps SB, editors. Washington County, NC: A Tapestry. Winston-Salem, NC: Jonsten Printing Company; 1998
- [32] Paerl HW, Bales JD, Ausley LW, Buzzelli CP, Crowder LB, Eby LA, et al. Ecosystem impacts of three sequential hurricanes (Dennis, Floyd, and Irene) on the United States' largest lagoonal estuary, Pamlico Sound, NC. *Proceedings of the National Academy of Sciences*. 2001;**98**:5655-5660
- [33] Ramus J, Eby LA, McClellan CM, Crowder LB. Phytoplankton forcing by a record freshwater discharge event into a large lagoonal estuary. *Estuaries and Coasts*. 2003;**26**:1344-1352
- [34] Gardner B, Sullivan PJ, Lembo AJ. Predicting stream temperatures: Geostatistical model comparison using alternative distance metrics. *Canadian Journal of Fisheries and Aquatic Sciences*. 2003;**60**:344-351
- [35] Peterson EE, Urquhart NS. Predicting water quality impaired stream segments using landscape-scale data and a regional geostatistical model: A case study in Maryland. *Environmental Monitoring and Assessment*. 2006;**121**:613-636
- [36] Little LS, Edwards D, Porter DE. Kriging in estuaries: As the crow flies, or as the fish swims? *Journal of Experimental Marine Biology and Ecology*. 1997;**213**:1
- [37] Pietrafesa LJ, Morrison JM, McCann MP, Churchill J, Bohm E, Houghton RW. Water mass linkages between the Middle and South Atlantic Bights. *Deep-Sea Research*. 1994;**41**: 365-389
- [38] Xie L, Pietrafesa LJ. Systemwide modeling of wind and density driven circulation in Croatan-Albemarle-Pamlico estuary system part I: Model configuration and testing. *Journal of Coastal Research*. 1999;**15**:1163-1177
- [39] Pietrafesa LJ, Janowitz GS. Final Report on the Albemarle Pamlico Coupling Study. Raleigh, NC: NC Albemarle-Pamlico Estuarine Study; 1991. 223 pp
- [40] Giese GL, Wilder HB, Parker GG. Hydrology of major estuaries and sounds of North Carolina. US Geological Survey Water-Supply Paper. 1979
- [41] Paerl HW, Valdes LM, Peierls BL, Weaver RS, Gallo T, Joyner AR, et al. Ecological effects of a recent rise in Atlantic hurricane activity on North Carolina's Pamlico Sound System: Putting Hurricane Isabel in perspective. In: Sellner KG, Chesapeake Research Consortium, editors. Hurricane Isabel in Perspective. Edgewater, MD: CRC Publication 05-160; 2005
- [42] Weaver JC. The drought of 1998–2002 in North Carolina—Precipitation and hydrologic conditions. Report U.S. Geological Survey Scientific Investigations Report 2005-5053. 2005. 88 pp
- [43] Brinson MM, Bradshaw HD, Jones MN. Transitions in forested wetlands along gradients of salinity and hydroperiod. *Journal of the Elisha Mitchell Scientific Society*. 1985;**101**: 76-94
- [44] Corbett DR, Vance D, Letrick E, Mallinson D, Culver S. Decadal-scale sediment dynamics and environmental change in the Albemarle Estuarine System, North Carolina. *Estuarine, Coastal and Shelf Science*. 2007;**71**: 717-729
- [45] Molina JR. Estuarine exchange model of the pamlico and albemarle sounds [M.S. thesis]. Raleigh, NC: North Carolina State University; 2002. 56 pp

[46] Peterson EE, Merton AA, Theobald DM, Urquhart NS. Patterns of spatial autocorrelation in stream water chemistry. *Environmental Monitoring and Assessment*. 2006;**121**: 569-594

[47] Jensen OP, Christman MC, Miller TJ. Landscape-based geostatistics: A case study of the distribution of blue crab in Chesapeake Bay. *Environmetrics*. 2006; **17**:605-621