CASE STUDY: ANALYSIS AND FORECASTING OF SALINITY IN APALACHICOLA BAY, FLORIDA, USING BOX-JENKINS ARIMA MODELS

By Hongbing Sun[1] and Manfred Koch[2]

ABSTRACT: The Apalachicola Bay is one of the most productive estuaries in Florida. Variations of salinity directly influence the productivity of the aquatic habitats. Physical conditions that affect the salinity include tidal elevations, wind and current velocities, precipitation, and the discharge of the Apalachicola River. In the present paper, cross-correlation techniques, autoregressive integrated moving average (ARIMA), and dynamic regression transfer models using the Box-Jenkins methodology are employed to analyze the time series data. The rational distributed lag transfer functions between hourly variations of tidal water levels and salinity allow forecasting of short-term fluctuations in the salinity, whereas multivariate correlation analyses of daily salinity with river discharge, wind stresses, water levels and currents, and precipitation shed light on the important control variables. Several conclusions with regard to the hydrodynamics and water quality of the bay can be drawn from identification of auto- and cross correlations and the appropriate ARIMA models. Fluctuations of tidal water levels result only in short-term periodic variations in salinity, with a linear transfer function that has a lag-two as the highest coefficient. The cross-correlation analysis shows that the Apalachicola River, being the major fresh-water source of the bay, strongly affects the currents and salinity in the bay area over the long term. Though regional precipitation controls the amount of fresh-water inflow, either through river discharge or groundwater seepage, its effect on the daily variations in salinity is statistically insignificant. In contrast, the effect of daily wind stress is significant. Salinity is positively correlated with western currents in the bay because most of the oceanic flow enters the bay from the east. A lag between the daily discharge and salinity indicates that up to a week is required for the peak of the inflow fresh water to flush through the exit of the bay.

INTRODUCTION

The Apalachicola Bay is located in northwest Florida, adjacent to the Gulf of Mexico (Fig. 1), and is one of the most productive estuaries both in Florida and the entire Northern Hemisphere. It yields 90% of the oysters consumed in Florida and 10% of those consumed in the United States (Johnson 1993). Variations in salinity directly control seafood production, as well as the overall ecological system of the bay, including peripheral marsh zones that serve as important fish nursery areas. The main factors controlling the variations in salinity are the discharge of the Apalachicola River, water level, regional precipitation, strength of winds and gulf currents, and to a lesser extent, surface runoff and groundwater discharge. The circulation of the saline water and nutrients is driven by tides and winds and by the Apalachicola river flow. Fresh-water currents are generally stronger at the river entrance, the bay's east entrance, and at the exits of the estuary. Bottom water is more salty than the surface water of the bay because of the density difference between salty and fresh water. The critical salinity concentrations that arise from these mixing processes create a unique habitat environment for oysters to grow in the Apalachicola Bay. Minor seasonal variations in the above control variables are often sufficient to temporarily destabilize the balance of the bay ecosystem in a detrimental manner (Johnson 1993) for reasons that are not yet understood. These variations may result in significant decreases in the oyster harvest.

This paper quantifies some of the interacting controlling variables, such as the water levels, currents and winds, river discharge, and regional precipitation, that may affect both the short- and long-term variations in the salinity of the bay. For the study of the short-term variations in the observed data (which are mainly due to tidal effects), techniques of univariate time series analysis were used and autoregressive integrated moving average (ARIMA) models constructed employing the Box-Jenkins method (Box and Jenkins 1976). Unlike the spectral tidal analysis carried out by Koch and Sun (1999) for the bay, which solely allowed quantification of the major tidal constituents throughout the estuary, forecasting of salinity variations also can be performed by means of the univariate short-term ARIMA models presented in this paper. The advantage of this time series modeling technique (which is used in the time domain) over ordinary spectral analysis (used in the frequency domain) [Priestley (1981); Hippel and McLeod (1994) for comprehensive discussion of the two complementary techniques] is particularly evident when considering the variations in the control variables and attempting to forecast the short-term behavior of salinity in the bay.

This multivariate analysis of variations of salinity was attempted to provide guidance for hydroengineering projects relating to water control and to aid in the study of the fresh-water needs, ecology, and sea food production in and around Apalachicola Bay. Understanding the physical dynamics of the latter therefore satisfies not only a goal of fundamental estuarine science, but should ultimately result in ecological and economic benefits to the bay region.

DATA AND TIDAL ANALYSIS

Daily precipitation [Fig. 2(a)] and discharges of the Apalachicola River at the Sumatra gauge [Fig. 2(b)] were obtained from the Florida Climate Center and the U.S. Geological Survey (USGS), respectively. Water elevations, wind speed, current velocity, and salinity were collected at multiple stations by the Northwest Florida Water Management District at half-hour intervals from April 1993 to August 1994 and also are shown as averaged daily values in Figs. 2(c–e) and Fig. 3, respectively. Salinity records from a total of 20 stations were analyzed and short gaps in the data series were filled using cubic splines. Data records with gaps of more than a week

[1]Dept. of Geological and Marine Sci., Rider Univ., Lawrenceville, NJ 08648.

[2]Dept. of Geohydraulics and Engrg. Hydro., Univ. of Kassel, Kurt-Wolters Strasse 3, 34109 Kassel, Germany.

Note. Discussion open until February 1, 2002. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on July 28, 1995; revised April 27, 2001. This paper is part of the ***Journal of Hydraulic Engineering***, Vol. 127, No. 9, September, 2001. ©ASCE, ISSN 0733-9429/01/0009-0718–0727/$8.00 + $.50 per page. Paper No. 22329.

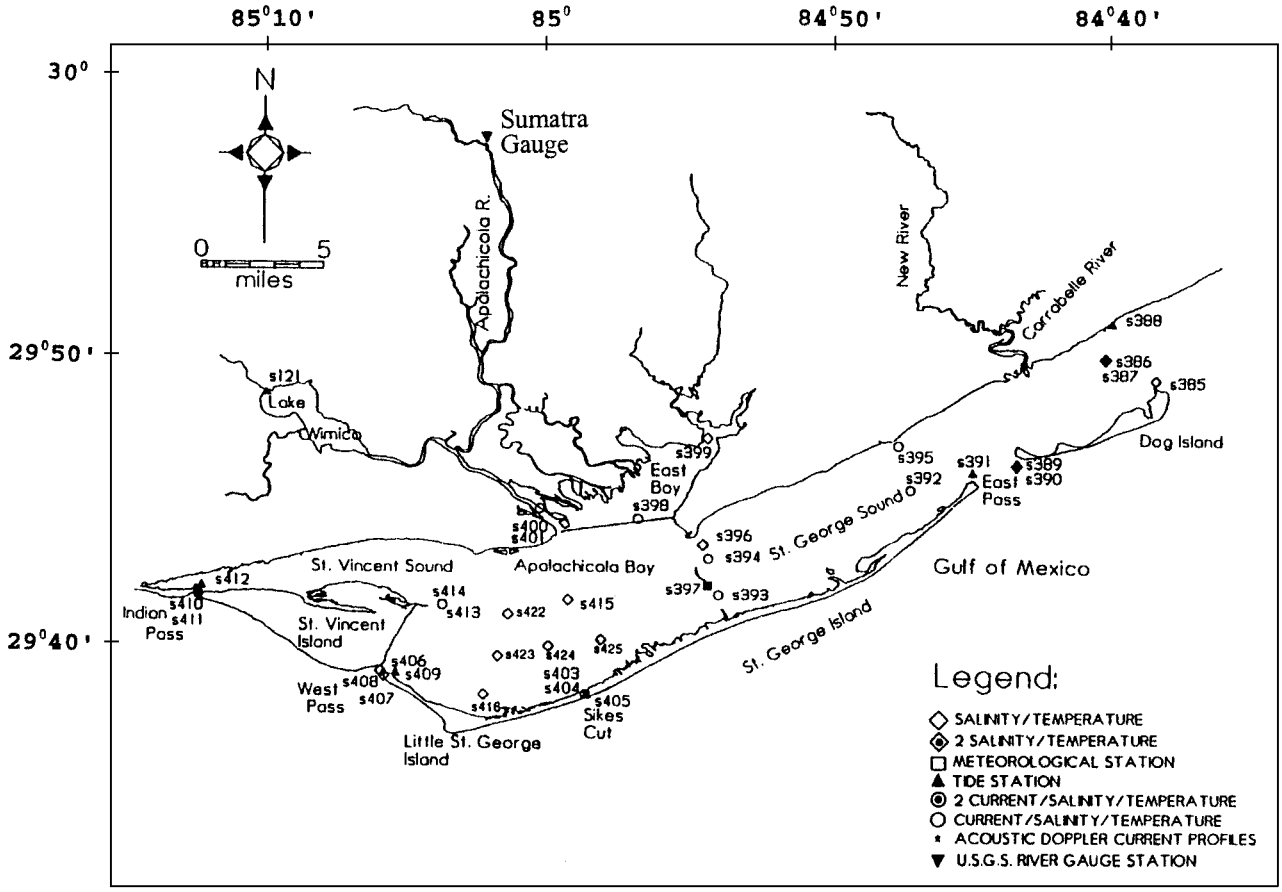

FIG. 1. Study area of Apalachicola Bay Showing Location of Current Meters, Salinity Meters, Tidal Stations, Wind Meter, and River Gauge

were eliminated entirely. Missing tidal records were generated by means of a least-squares harmonic analysis that included 25 major tidal constituents. A detailed analysis of the amplitudes and epochs of tidal water levels and currents for the Apalachicola Bay was presented in Koch and Sun (1999). Tidal heights are of mixed types that change from semidiurnal at the bay entrance in the east to diurnal in the west. Although this cannot be readily seen in Fig. 2(c) because of the low time resolution used in the plotting, the fortnightly amplification of the tidal elevations during the two phases of the new and full moon, respectively, are easy to observe. Tidal currents in the bay, on the other hand, are more of semidiurnal type. Tidal ranges gradually decrease from the east to the west side of the bay, but tidal currents are strongest in the narrow exits of the bay, as they should be; see Koch and Sun (1999) for details.

HYDRODYNAMICS OF SALINE WATER CIRCULATION IN BAY

Annual average daily inflow of the Apalachicola River is approximately 650 m^3/s (USGS 1994). With an estimated 64 km shoreline on the bay's landward side and an average annual precipitation of 150 cm, the estimated contribution from surface runoff and direct groundwater discharge to the bay is at least two magnitudes less than the freshwater discharge of the Apalachicola River from hydrograph separation analysis. Saline water generally originates from the east side of the bay and East Pass under normal conditions. Exits in the southwest area of the bay comprise Indian Pass, West Pass, and a man-made small canal, Sikes Cut (Fig. 1). River flow enters the bay approximately perpendicularly to the current flow, creating a substantial mixing zone between the fresh and saline water. Figs. 3(a–b) show salinity variations for groups of stations in the east and west sides of the bay, respectively. Because the entrance to the Apalachicola River is in the northwest segment of the bay area, the west and north sides of the bay are less saline than the east and south sides. The west and north sides of the bay also have larger seasonal fluctuations because of larger seasonal changes in river discharge and surface runoff as a consequence of strong seasonal variations in precipitation.

The velocities of the currents in the bay area vary from 0 m/s to more than 1 m/s, and the flow directions change from predominantly southwest at high tides to northeast at low tides. The central part of the bay, around stations s422, s423, and s416, has relatively weak currents (Fig. 1). Comparing the current distribution, the currents are stronger at the river entrance and the eastward ocean entrance [Fig. 2(e)]. This distribution pattern can also be verified by the sedimentation pattern observed at the bottom of the bay (Brooks 1973; Jones et al. 1992). Most of the bay bottom around stations s422 and s416 is covered by clay, with other bottom sections consisting of sporadic sandy sediments (Brooks 1973; Jones et al. 1992). Generally, the presence of sandy sediments implies a stronger current activity within the area than the presence of clay sediments.

Because of vertical density differences, a salinity stratification with depth exists (Fig. 3). Dense saline water stays at the bottom unless the waters are well mixed. Thus, the salinity fluctuations in the surface water generally are stronger than those in the bottom water. The salinity mixing is driven by current, tidal, and wind actions. The flow dynamics in the bay are governed mainly by momentum conservation, mass continuity, and the Reynolds equations describe the stress-strain constitutive relationships (Blumberg and Herring 1987). Using the appropriate boundary initial and forcing conditions, preliminary hydrodynamic simulations were carried out by Jones et al. (1994) that were able to explain some of the spatial variations in the salinity distribution, both horizontally and

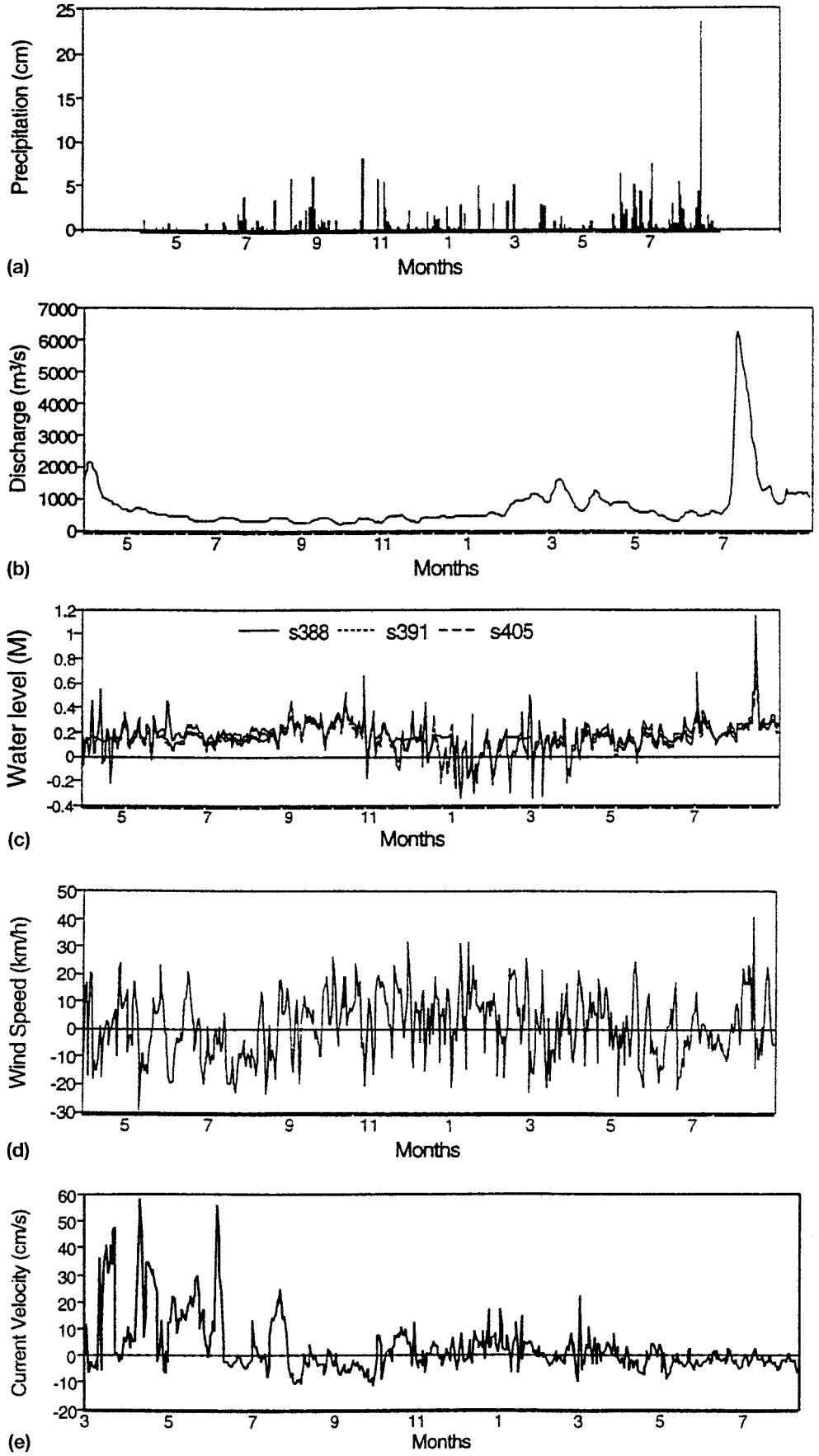

FIG. 2. Average Daily Time Series Data for Five Major Variables That Affected Salinity Variations from 4/1/1993 to 8/31/1994: (a) Precipitation; (b) Apalachicola River Discharge; (c) Water Levels for Three Selected Stations; (d) Wind Speed Projected onto SW–NE 63° Major Axis of Bay (Positive in NE, Negative in SW Direction); (e) Current Velocity for Station s392 Projected onto SW–NE 63° Major Axis

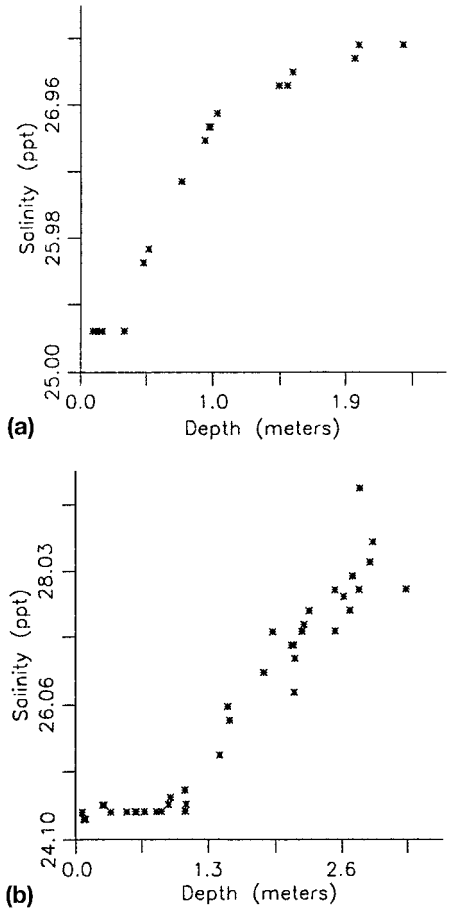

FIG. 3. Vertical Salinity Profiles in 6/1993: (a) Station s389; (b) Station s395

vertically. For the surface layer, results of such a calibration exercise are shown in Fig. 4. The above figures give one a sense of the scale of mixing between the saline water from the Gulf and the fresh water from the Apalachicola River. These critical salinity concentrations create the unique habitat environment for oysters to grow in the Apalachicola Bay (Johnson 1993).

BOX-JENKINS ARIMA AND LINEAR TRANSFER FUNCTION MODELS FOR SHORT-TERM FLUCTUATIONS OF SALINITY, WATER LEVELS, AND CURRENTS

Univariate time series analysis using Box-Jenkins ARIMA models (Box and Jenkins 1976) is a major tool in hydrology and has been used extensively, mainly for the prediction of such surface water processes as precipitation and streamflow events (Govindasamy 1991; Lettenmaier and Wood 1993; Maidment 1993; Salas 1993). On the other hand, in contrast to economic forecasting, the use of multivariate dynamic regression or transfer models between different variables has been somewhat more limited in hydrological applications (McCuen and Snyder 1986).

Univariate Box-Jenkins ARIMA Models: Methodology and Comparison with Spectral Method

A general nonseasonal/seasonal ARIMA (p, d, q) $(P, D, Q)_s$ model with nonseasonal parameters p, d, q, seasonal parameters P, D, Q, and seasonality S can be written as

$$\varphi_p(B^S)\varphi_p(B)\nabla_S^D\nabla^d z_t = C + \theta_Q(B^S)\theta_q(B)a_t \quad (1)$$

where $\varphi_p(B)$ and $\varphi_P(B^S)$ = series of the nonseasonal and seasonal autoregressive (AR) components of order p and P of the time series z_t, respectively, and $\theta_q(B)$ and $\theta_Q(B^S)$ = moving average (MA) components of the random shock a_t. For example, $\varphi_p(B)$ can be written as $\varphi_p(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_P B^P)$, and similar expressions hold the other three series. The backshift operator B is defined as $B^k z_t = z_{t-k}$. The nonseasonal and seasonal difference operators ∇^d and ∇_S^D, defined as $\nabla^d = (1 - B)^d$ and $\nabla_S^D = (1 - B^S)^D$, respectively, are used to make the time series z_t stationary, by either nonseasonal or, in the presence of seasonality, seasonal differencing with lag times S. Finally, C is a constant term of regression. For a more detailed description of the terminology, see Bowerman and O'Connell (1987) or Pankraz (1991).

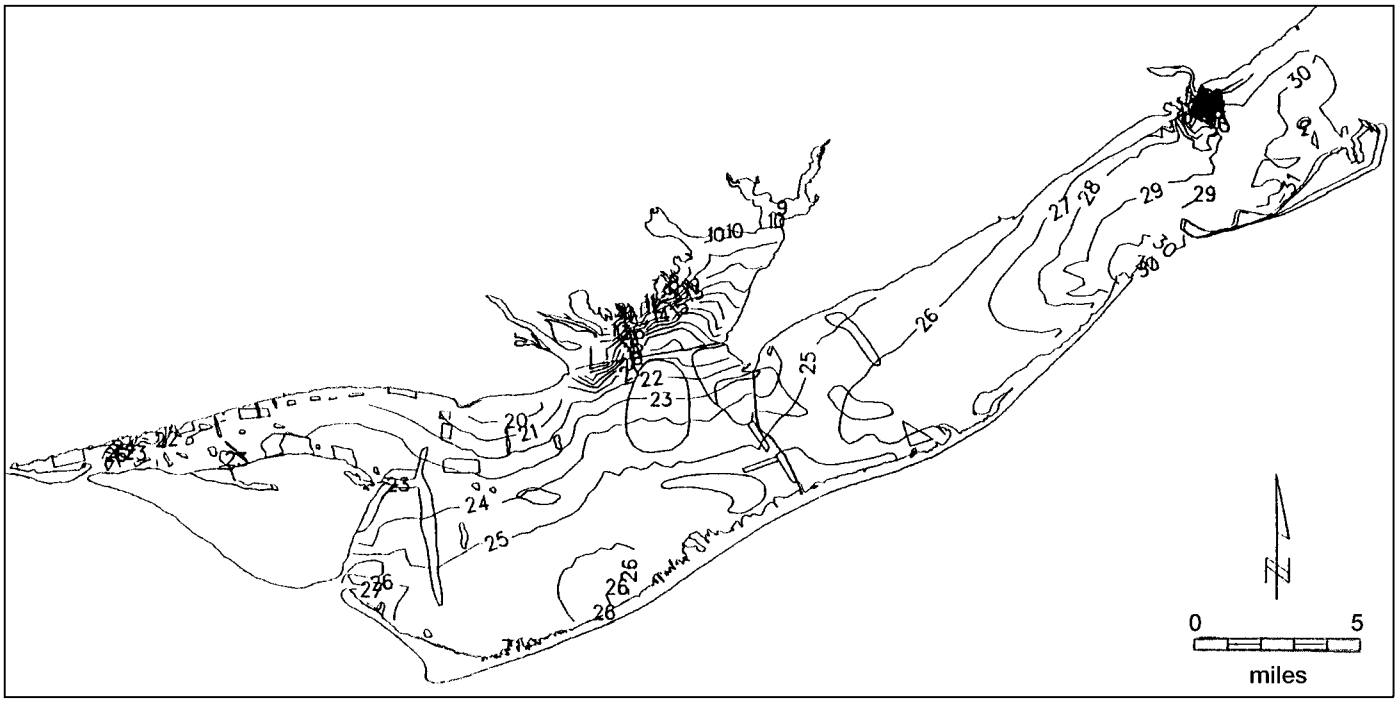

FIG. 4. Distribution of Salinity in Surface Layer of Bay as Predicted by Calibrated Hydrocirculation Model

As prescribed by the Box-Jenkins formalism, the buildup of the general ARIMA model (1) for a time series requires three stages: (1) identification; (2) diagnostic checking; and (3) estimation. The first two stages involve computation of the total and partial autocorrelations $\rho_T(\tau) = E[z(t)\cdot z(t+\tau)]/E[z^2(t)]$ (E = expectation operator; τ = lag time), and $\rho_P(\tau)$, respectively. A graphical check of the latter to determine the positions of spikes and their general decay pattern for various lag times z_{t-k}, ($k = 1, \ldots, p$); and applying various t tests of statistical significance to ascertain which lag coefficients do or do not contribute to the tentative ARIMA model. The model is then estimated in the third stage by maximum likelihood least-squares regression. The various steps of the Box-Jenkins formalism are implemented computationally in the SAS statistical package (SAS 1993, 1999), which has been used throughout this study.

Initially, the time domain ARIMA method does not appear to provide a real advantage over the spectral method used commonly in time series analysis. In the spectral method the time series $z(t)$ is analyzed through its Fourier representation $z(t) = A_0/2 + \Sigma[A_n \cos(n\omega_0 t) + B_n \sin(n\omega_0 t)]$ (with ω_0 = fundamental frequency; n = integer), with the coefficients A_n and B_n, calculated by classical Fourier analysis (Priestley 1981). Moreover, since the Fourier transform of the autocorrelation function $\mathcal{F}\{\rho_T(\tau)\}$ is equal to the power spectrum $h(\omega) = |F(\omega)|^2$ with $F(\omega) = \mathcal{F}\{z(t)\}$, the coefficients C_i of the discrete spectrum of $h(\omega = \omega_0 n)$ are given by $C_n = \{\Sigma[A_n^2 + B_n^2]\}^{1/2}$. Furthermore, for a purely harmonic process such as the tidal heights (Koch and Sun 1999), $z(t) = \Sigma a_n \cos(w_n t + \varphi_n)$ (with a_n = amplitude; w_n = angular frequency or tidal velocity; φ_n = phase), it is easy to show (Priestley 1981; Hippel and McLeod 1994) that autocorrelation $\rho_T(\tau)$ is itself harmonic and can be written in the form $\rho_T(\tau) = (1/\Sigma a_n^2)\cdot\Sigma a_n^2 \cos(w_n\tau)$. Therefore, the power spectrum $h(\omega) = \mathcal{F}\{\rho_T(\tau)\}$ is given by $h(\omega) = (1/2\Sigma a_n^2)\cdot\Sigma a_n^2[\delta(w + w_n) + \delta(w - w_n)]$, (with δ = Dirac delta function); that is, it has spikes at the corresponding dominant frequencies w_n of the original process. This means that no more information about the time series $z(t)$ can be gained from looking either at its autocorrelation $\rho_T(\tau)$ or at its power spectrum $h(\omega)$. The advantage of the ARIMA time domain technique arises when building a prediction or forecasting time series model, as is of particular interest in the present application.

Auto and Cross correlations of Salinity, Water Levels, and Current Velocities

As mentioned, in the first step of the ARIMA process, auto- and cross correlations are computed for the short-term fluctuations of hourly, semidiurnal, and diurnal variations of the salinity, tidal water levels, and flow currents. As shown by the autocorrelations [Figs. 5(a–b)], the major periodicities of the water levels and the salinity are of mostly semidiurnal nature. This is also corroborated by the independent spectral analysis of tidal water levels and currents of Koch and Sun (1999), indicating again the formal equivalency of time series and spectral analyses.

The cross correlations of the salinity versus both the water levels and current velocities for different stations projected onto the principal Southwest-Northeast (SW–NE) elongation of the bay [Figs. 6(a–b)] provide information on flushing and fresh/saltwater mixing processes that occur in the bay. Thus, the statistically significant cross correlations between the salinity at the mouth of the Apalachicola River (station s398) and the water levels at the opening between St. George Island and Dog Island (station s391) [Fig. 6(a)] can be taken as evidence of tidal mixing of the fresher river water with saline water that enters the bay in the northeast during a high tide. On the other hand, although the water levels and current velocities cross correlate significantly [see Sun and Koch (1966) for details], the cross correlations between the salinity at station s385 and the current velocities at station s392 are statistically insignificant [Fig. 6(b)]. This might be due to the fact that both stations are located within the bay proper and are thus experiencing fewer relevant short-term salinity fluctuations.

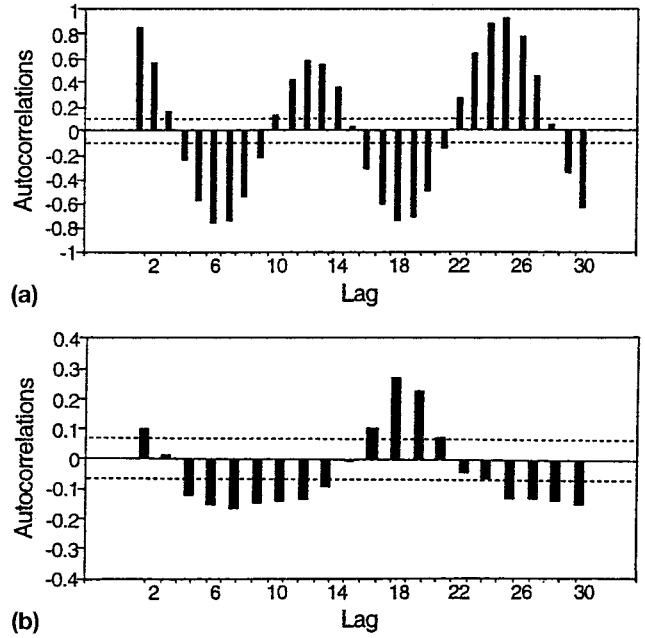

FIG. 5. Autocorrelations of Hourly Tidal Water Levels and Salinity for Two Stations: (a) Water Level at s391; (b) Salinity at s389. Horizontal Lines Delineate $\pm 2\sigma$ (95%) Confidence Interval of Correlation Coefficients (Lag Units Are in Hours in All Plots with Lags)

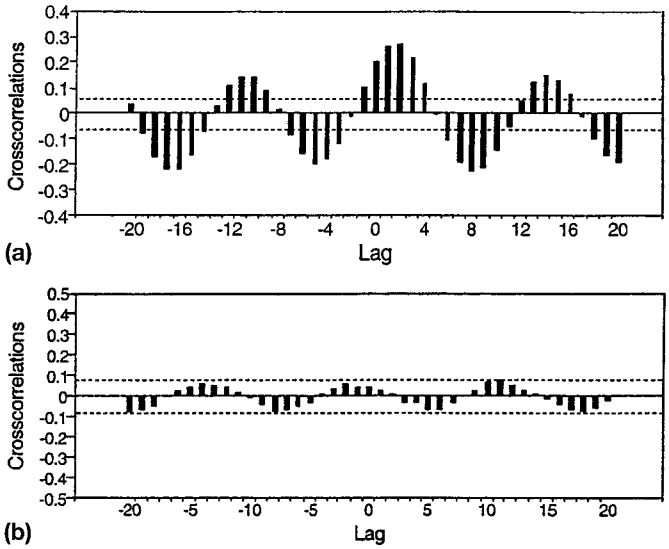

FIG. 6. Cross Correlations of Hourly Salinity Variation versus Tidal Elevation and Current Velocity: (a) Salinity at s398 versus Water Level at s391; (b) Salinity at s385 versus Current at s392. Horizontal Lines Delineate $\pm 2\sigma$ Confidence Interval of Correlation Coefficients

ESTIMATION OF ARIMA MODELS FOR HOURLY SALINITY

Based on the results of (1) identification of the leading lag coefficients in the autocorrelations of the relevant time series and (2) diagnostic checking (based on statistical tests), (3) estimation of the statistically significant ARIMA φ coefficients in (1), can be endeavored. Computationally this is achieved in

the SAS package by means of a maximum-likelihood least-squares procedure.

As a representative example, the seasonal ARIMA model obtained for the hourly salinity time series through the above three steps at station s389 is

$$\nabla \ln S_t = \frac{1 + 0.122B^{12}}{1 - 0.146B} a_t \tag{2}$$

The left side of (2) indicates that, prior to the model buildup, the natural logarithm ln of the original hourly salinity time series has been taken (a Box-Cox transformation), and the first time difference $\ln S_t - \ln S_{t-1}$ (denoted by the operator $\nabla = 1 - B$) has been used. These two operations were intended to modify the data to produce a stationary data series such that the mean, variance, and autocorrelation of the data series are stabilized. The coefficients for the backshift operators in the numerator and denominator were initially obtained through the above two procedures and calibrated during examination of the normal, partial, and inverse autocorrelations of the residues (Pankraz 1991).

The highest AR term in the denominator of (2) indicates that the salinity values are persistent in themselves for up to one lag time (hour). The lag 12 for the MA term in the numerator is due to the seasonality of the salinity time series with a period of 12 h, which, of course, is a consequence of the major semidiurnal tidal cycle of that period (Koch and Sun 1999). Similar forms for the ARIMA models for salinity were obtained for most of the other stations in the bay, though with lag-term coefficients that vary from one station location to another.

Transfer Function Models between Salinity and Tidal Water Heights

As discussed earlier, the salinity in the bay itself is not an independent control variable but is mainly dependent upon the hydrodynamic circulation and ensuing mixing processes. For the short-term intervals considered in this section, tidal effects may therefore have the strongest influence on salinity. Though this has already transpired from the special seasonal form of the MA term in the ARIMA model (2) for salinity and from the cross correlations between the tidal water heights and salinity in Fig. 6(a), a more quantitative analysis can be performed by the estimation of linear transfer functions (LTF), or dynamic regression models, between these two variables (Pankraz 1991).

A linear transfer function model between the dependent variable S_t (salinity) and the independent variable H_t (tidal heights) can be written in the general form:

$$S_t = C + v(B)H_t \tag{3}$$

where $v(B)$ = rational distributed lag transfer or impulse function given more specifically as $v(B) = \beta w(B)B^b/\delta(B)$, with $w(B)$ and $\delta(B)$ polynomial delay functions of the form $w(B) = (1 + w_1B + w_2B^2 + w_3B^3 + \cdots)$ and $\delta(B) = (1 + \delta_1B + \delta_2B^2 + \delta_3B^3 + \cdots)$, where B^b specifies the ''dead time'' of lag b. Thus b = delay time before S_t responds to H_t at all; $w(B)$ then denotes the further-delayed response of S_t to H_t and $\delta(B)$ the AR response of S_t to itself; and C and β = constants and N_t = stochastic disturbance. The latter can itself be described as a general seasonal ARIMA process (1), with an MA random shock component a_t (Pankraz 1991).

Because of the particular circulation hydrodynamics in the Apalachicola Bay, where most of the tidal flow enters the bay at the East Pass isthmus between St. George and Dog Island, linear transfer functions between the salinity at stations within the bay and tidal water levels at a station close to the East Pass are of specific interest. For this purpose, the water heights at tidal station s391 (see Fig. 1 for location) have been used as a reference in most of the analyses.

Fig. 7 shows SAS-calculated impulse response weights $v(B)$ of the LTF between the salinity at site s385 in the eastern section of the bay and water levels at station s391 at the East Pass isthmus as a function of the lag-time coefficient B. The impulse response weights $v(B)$ for this and other stations exhibit generally a compound exponential decay; however, spikes will reappear at the major seasonal lag cycles. The signs of the $v(B)$ coefficients provide some qualitative information about how changes in water heights will affect salinity values. For the example shown, the weights $v(B)$ are significant up to a maximum lag time of 1 h, and this is somewhat indicative of the time it takes before water levels at the East Pass isthmus affect salinity in the very east of the bay.

From the analysis of the prewhitened input X_t (water heights) and output Y_t (salinity) series, tentative LTF models were identified and fine-tuned after diagnostic checking for various bay stations. The prewhitening process is a process of filtering off the autocorrelated signals in the data series itself, such that after the filtering, only an uncorrelated residual series is left. Again, for reasons stated earlier, a Box-Cox natural logarithm transformation and a first-order time difference of the salinity series was taken prior to estimation. As a representative example, the best LTF model between the salinity at site s385 and water levels at station s391 was computed as

$$\nabla \ln S_t = \frac{0.08 - 0.16B + 0.07B^2}{1 - 0.125B + 0.54B^2} H_t + \frac{1 - 0.8B^2 - 0.12B^5 + 0.12B^{11}}{1 + 0.16B - 0.58B^2} a_t \tag{4}$$

The highest AR terms in the two denominators of (4) indicate that salinity levels are persistent in themselves for up to two lag times (in hours). In the numerator of the AR term the twice-as-high value of the B coefficient (0.16) compared

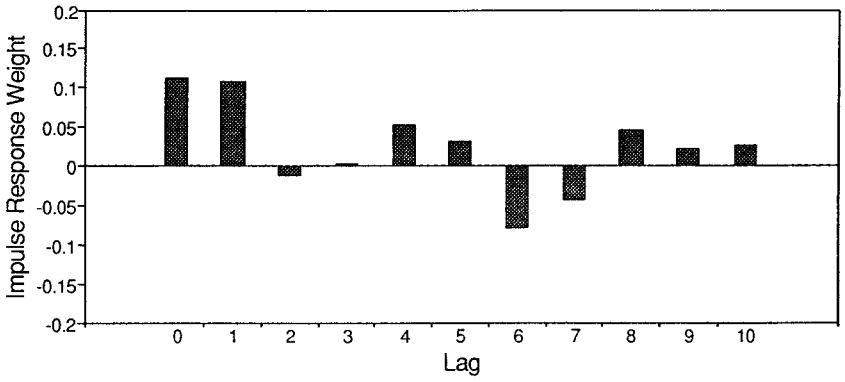

FIG. 7. Impulse Response Weights of Linear Transfer Function between Salinity at Site s385 and Water Levels at Station s391 (Lag Unit Is in Hours)

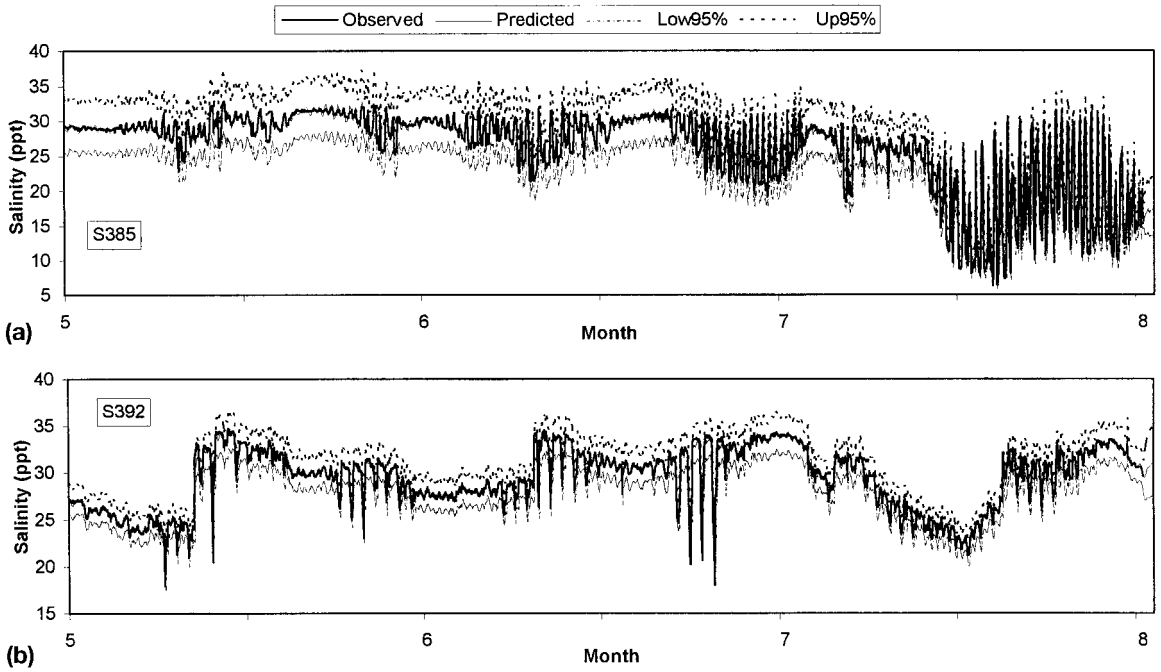

FIG. 8. One-Step-ahead Predicted Hourly Salinity Using LTF Model (4) at Stations s385 (a) and s392 (b) Based on Observed Tidal Height at Station s391 for May 1994 through August 1994

with the B^2 coefficient (0,07) supports quantitatively the conclusions drawn from the impulse response weights. Finally, the lag 11 and 5 for the MA terms approximate the periods of three major tidal semidiurnal constituents M_2, S_2; N_2; and M_4, M_6, respectively. Similar forms for the LTF were obtained for most of the other salinity stations in the bay that have a complete data record.

One-Step-Ahead Forecast of Salinity

Knowing the dynamic regression model (3) between H_t and S_t, an explicit forecast model for S_t as a function of earlier values for both S_{t-k} and H_{t-k} can be formulated. This can be considered as a fourth and final step of the Box-Jenkins technique, which is forecasting. A special case of forecasting is the one-step-ahead forecast, whereby all observed variables S_t and H_t up to time t are used to predict $\hat{S}_{t+1}$ at time $t + 1$. Fig. 8 shows the results for the forecast salinity at stations s392 and s385 [the LTF for s385 is given in (4)] based on water levels at station s391. One observes a close match between the forecast and observed salinity data series. It should be noted that forecasts with longer lead times have been performed, but their confidence intervals (as shown by the 95% confidence lines in Fig. 8) soon become unreasonably wide. This is due to the fact that the LTF model (4) includes only lag-time coefficients B of up to order two, so that values of $\hat{S}_{t+l}$ for lead times greater than 2 h must be estimated from previously forecast values $\hat{S}_{t+l-1}$ and $\hat{S}_{t+l-2}$.

CORRELATION ANALYSIS OF LONG-TERM VARIATIONS OF SALINITY

One would expect long-term variations in the salinity of the bay to be controlled by (1) discharge of the Apalachicola River; (2) wind stress; (3) water levels; (4) direction and speed of the bay currents; and (5) amount of direct precipitation. To analyze these effects quantitatively, cross correlations of the daily averages of the salinity with these control variables were computed.

Apalachicola River Discharge

Discharges of the Apalachicola River comprise the major source of the fresh water for the bay, with an annual average

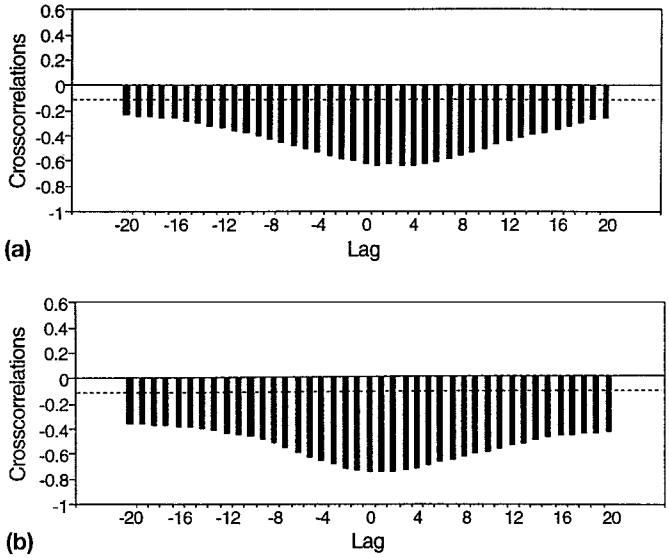

FIG. 9. Cross Correlations of Average Daily Salinity Values at Stations s385 (a) and s398 (b) versus Apalachicola River Daily Stream Discharge

inflow of 650 m^3/s. Figs. 9(a and b) show that for two stations at different locations away from the mouth of the river, distributions of the salinity, unsurprisingly, are strongly and negative correlated with the discharge. Moreover, the negative correlation coefficients are stronger in the SW portion of the bay, with peaks at shorter lag times (station s385 close to the river mouth) than in the NE section (station s398). The peak lag time is indicative of the time it takes to mix or to flush the fresh water out of the river.

The peak cross-correlation coefficients of the river discharge versus salinity are plotted in Fig. 10. The lag time (in days) is marked in parentheses after the peak coefficient, and it can approximate the mixing or flushing time of the river discharge. The longest lag times are distributed around stations s422 and s416, which are located in the southeast (SE) cul-de-sac of the bay, where relatively stagnant conditions prevail. The lag time for the arriving peak at these sites averages six days, which means that the estimated mixing or flushing time is approxi-

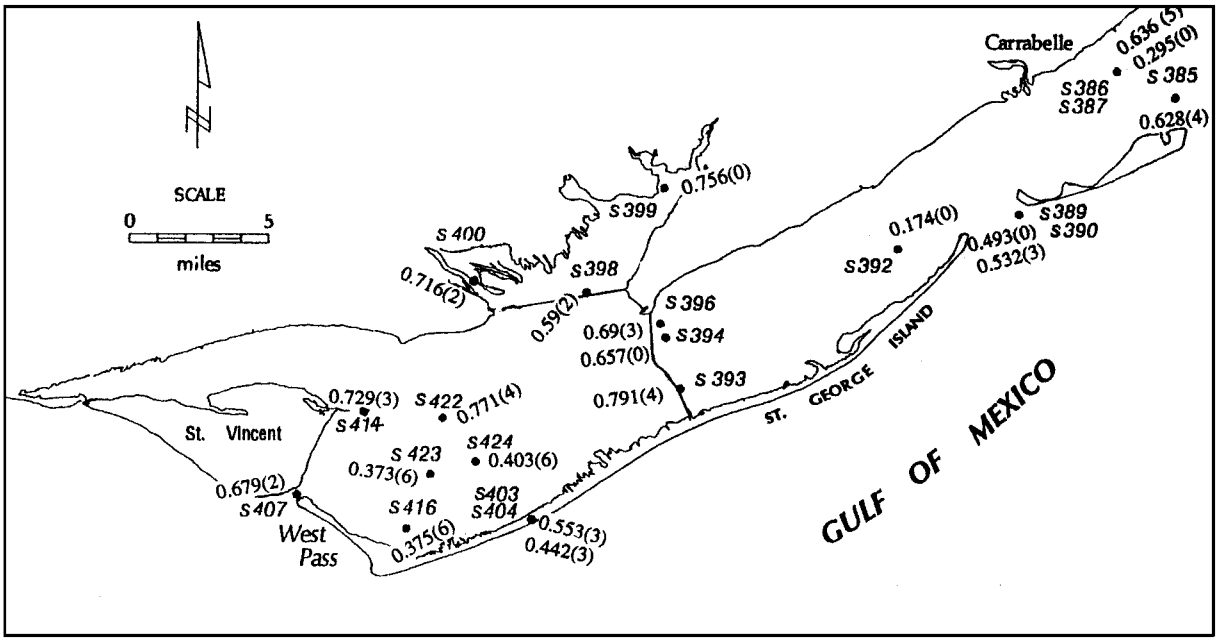

FIG. 10. Peak Cross Correlation Coefficient Map of Daily Stream Discharge at Sumatra Stream Gauge versus Average Daily Salinity Variations. Numbers in Parentheses Are Lag Numbers (Delay Times in Hours)

mately a week after the river discharge has entered the bay. The correlation coefficients are generally higher for the surficial than for the bottom salinities [compare station s386 (surface) with s387 (bottom) and s403 (surface) with s404 (bottom)]. This is due to the fact that surface water is more mobile than bottom water. The strong negative cross-correlation coefficients also indicate that the highest fresh-water levels of the bay are in phase with flooding events, whereas the highest saltwater levels of the bay occur during drought conditions.

Wind Stress

Wind stress is a major driving force for current velocities in the bay area and should therefore have an effect on the mixing and flushing times of the seawater that flows into the bay from the east. Because of its shallow depth (averaging 2.7 m), wind stress also changes the water levels (tidal ranges) of the bay. In the cross-correlation analysis, the speeds and directions of the average daily winds are projected onto the elongated SW–NE axis of the bay, such that winds in the NE directions are associated with positive signs and those in SW directions with negative signs.

Figs. 11(a–b) show the cross-correlation coefficients for two stations, and the map of Fig. 12 plots their maximum magnitudes and the corresponding lag times. Both the significance and the sign of these correlation coefficients vary with station locations. For s385 [Fig. 11(a)] and s389 at the far east site of the bay, positive coefficients demonstrate that salinities are

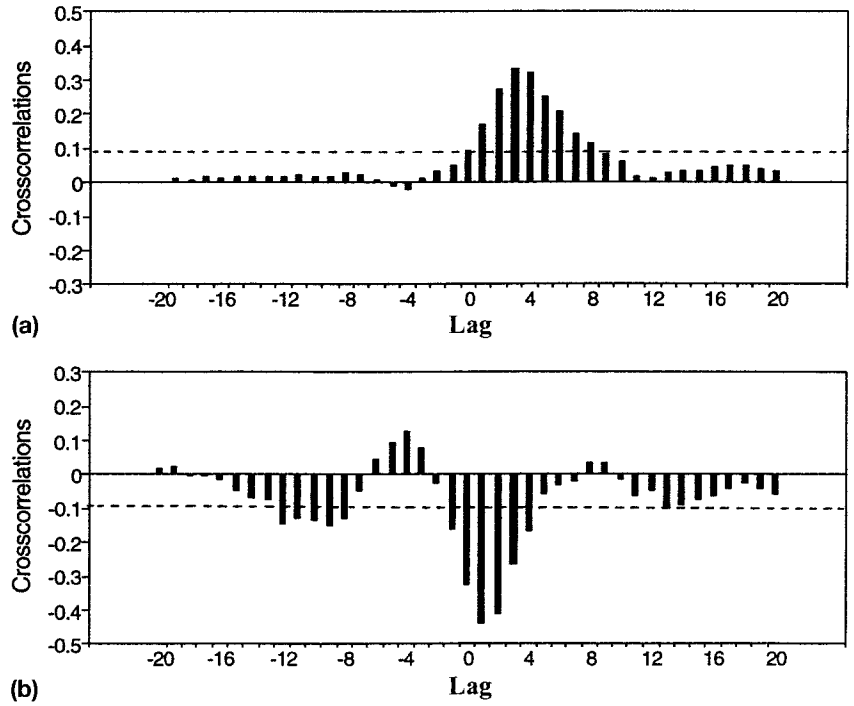

FIG. 11. Cross Correlations of Hourly Salinity Variation for Stations s385 (a) and s416 (b) versus Average Daily Wind Speed Projected in NE 63° Direction

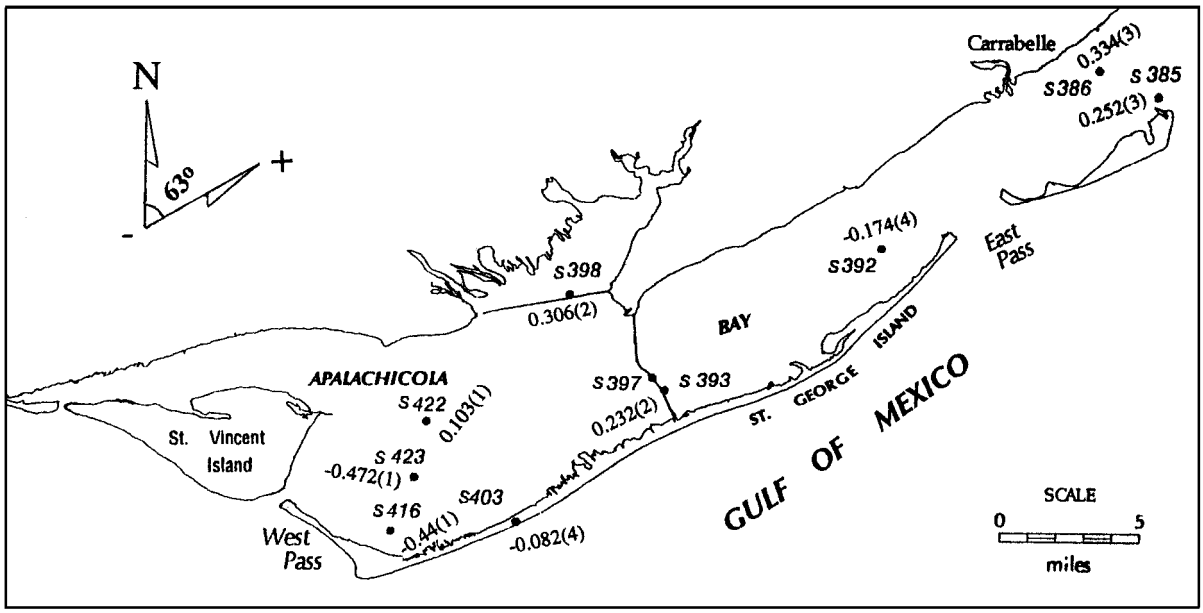

FIG. 12. Map of Maximum Cross-Correlation Coefficients of Average Daily Salinity Variations versus Average Daily Wind Speeds Projected onto NE 63° Direction. Corresponding Lag Times (Hours) in Parentheses

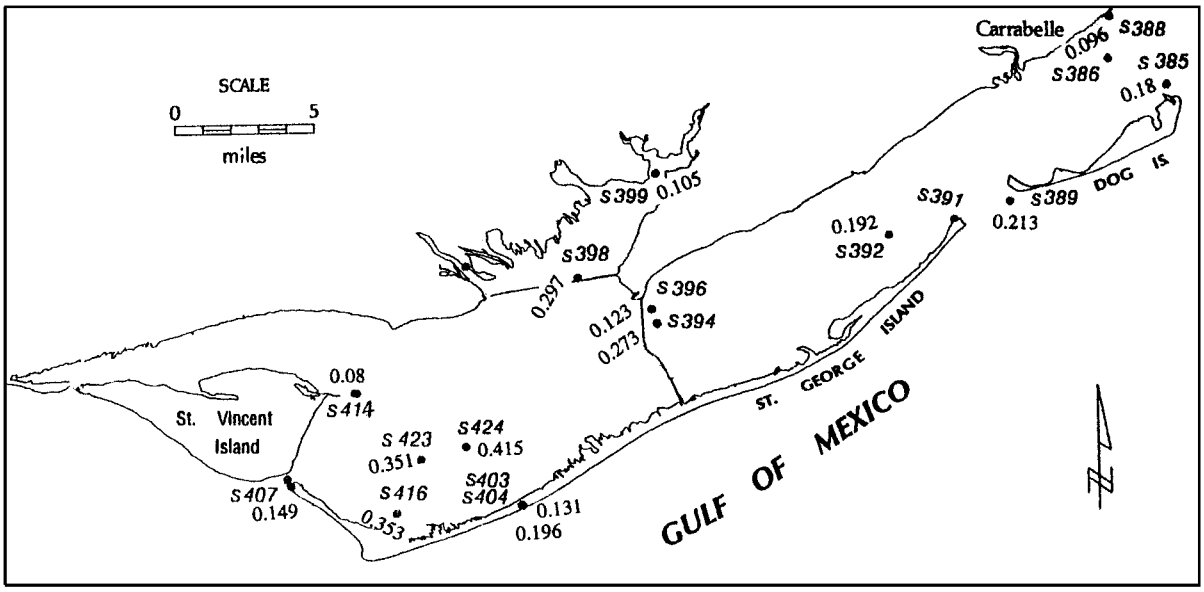

FIG. 13. Map of Maximum Cross-Correlation Coefficients of Daily Salinity Variations versus Water Levels at Station s391

generally higher when the wind is blowing toward the northeast due to the inflow from East Pass (Fig. 1). In contrast, for stations in the center of the bay, such as s416 [Fig. 11(b)], s423, and s403, the negative cross-correlation coefficients are evidence that the salinity is generally higher when the wind is blowing toward the southwest, bringing in more saline water toward the West Pass exit. The salinity-wind cross-correlation coefficient is also positive at the mouth of the Apalachicola River (site s398), which corroborates the basic hydrodynamic fact that for wind in the northeast direction, saline bay water is driven into the mouth of the river, therefore partly reversing the freshwater dilution.

Water Levels

Because of the daily averaging process used here, diurnal and semidiurnal tidal periodic variations in the water elevations in the bay are eliminated. Instead, the effects or salinity of longer-term seasonal fluctuations of the water levels, which arise from persistent weather fronts and storms, are investigated. Fig. 13 shows that mostly positive correlation coefficients exist between these two variables; that is, high water levels correlate with high salinity and vice versa. Again, differences are observed for various station locations in the bay, with station s424 in the center of the bay experiencing a stronger effect than s386 at the east entrance.

Current Velocities

Similar to the water levels, the long-term, daily-averaged current velocities (projected onto the NE axis of the bay) are affected mainly by persistent seasonal weather patterns and storm events. However, there is also a strong influence of the river discharge on the NE current velocities that extends far into the NE section of the bay, as illustrated in Fig. 14(a) by the significant positive correlation between these two variables for station s392. This may partly explain the significant negative cross correlations observed between the NE current velocity and the salinity at that station (and for most of the other sites in the bay) in Fig. 14(b).

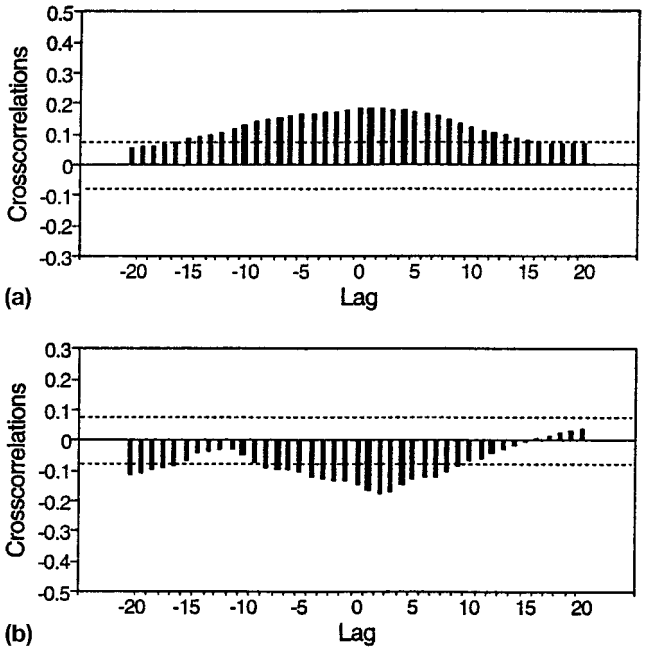

FIG. 14. Cross Correlations between Current Velocities (Projected onto Principal NE 63° Direction) and River Discharge (a) and between Current Velocities and Salinity (b) for Station s392

CONCLUSIONS

Through the application of the two techniques of univariate Box-Jenkins ARIMA time series analysis and cross-correlation analysis, it has been possible to identify the most relevant hydrological parameters that affect variations in the salinity of the Apalachicola Bay. Particularly, the time-series cross correlation made it possible to statistically define the interaction of different parameters that affect the salinity change in Apalachicola Bay. The strength of the coefficients and the lag times in the application provided help one understand the hydrodynamic circulation of the water body through the approach of data analysis. We believe the same type of approach can be applied in the study of other estuaries as well. In practical application, since the balance of the bay ecosystem and, especially, the quality of the oyster habitat appear to be controlled to a large extent by the salinity of the bay, the results of the present study should provide help to water management agencies regarding the proper amount of fresh water inflow to the bay area. For example, based on the strength of the cross-correlation coefficient and lag time between the salinity and the discharge amount of the Apalachicola River, the amount and schedule of the fresh-water inflow modification can be managed in the upper stream reservoirs such that oyster production and other usage of fresh water all could be optimized. The linear transfer function model of 1 and 2 h predictions of the water salinity change based on river discharge ahead of the discharge modification can also come to play in that case, such that the impact of salinity change in the bay can be predicted ahead of time.

REFERENCES

Blumberg, A. F., and Herring, J. J. (1987). ''Circulation modeling using orthogonal curvilinear coordinates.'' *Three-dimensional models of marine and estuarine dynamics*, J. C. J. Nihoul and B. M. Jamart, eds., Elsevier, Amsterdam, 55–58.

Bowerman, B. L., and O'Connell, R. T. (1987). *Time series forecasting*, Duxbury Press, Boston.

Box, G. E. P., and Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*, Holden-Day, San Francisco.

Brooks, H. (1973). ''Geological oceanography.'' *A summary of knowledge of the Eastern Gulf of Mexico*, Florida Institute of Oceanography, St. Petersburg, Fla.

Govindasamy, R. (1991). ''Univariate Box-Jenkins forecasts of water discharge in the Missouri River.'' *Int. J. Water Resour. Devel.*, 3, 168–177.

Hippel, K. W., and McLeod, A. I. (1994). *Time series modeling of water resources and environmental systems*, Elsevier, Amsterdam.

Johnson, V. (1993). ''Apalachicola Bay: Endangered estuary.'' *Florida Water*, 2, 14–20.

Jones, W. K., Galperin, B., Wu, T. S., and Weisberg, R. H. (1994). ''Preliminary circulation simulations in Apalachicola Bay, FL.'' *Water Resour. Spec. Rep.*, Northwest Florida Water Management District, Havana, Fla.

Koch, M., and Sun, H. (1999). *Tidal and non-tidal characteristics of water levels and flow in the Apalachicola Bay, Florida*, C. A. Brebbia and P. Anagnostpolous, eds., WIT Press, Computational Mechanics Publications, Southampton, U.K., 357–366.

Lettenmaier, D. P., and Wood, E. F. (1993). ''Hydrological forecasting.'' *Handbook of hydrology*, D. R. Maidment, ed., McGraw-Hill, New York.

Maidment, D. R. (1993). *Handbook of hydrology*, McGraw-Hill, New York.

McCuen, R. M., and Snyder, W. M. (1986). *Hydrological modeling, statistical methods and applications*, Prentice-Hall, Englewood Cliffs, N.J.

Pankraz, A. (1991). *Forecasting with dynamic regression models*, Wiley-Interscience, New York.

Priestley, M. B. (1981). *Spectral analysis and time series*, Academic Press, Orlando, Fla.

Salas, J. D. (1993). ''Analysis and modeling of hydrological time series.'' *Handbook of hydrology*, D. R. Maidment, ed., McGraw-Hill, New York.

SAS Institute Inc. (1993). *SAS/ETS user's guide*, Cary, N.C.

SAS Institute Inc. (1999). *SAS/ETS user's guide*, Cary, N.C.

Sun, H., and Koch, M. (1996). ''Time series analysis of water quality parameters in an estuary using Box-Jenkins ARIMA models and cross correlation techniques.'' *Computational methods in water resources*, Vol. XI, Computational Mechanics Publications, Boston, 230–239.

U.S. Geological Survey (USGS). (1994). *Water resources data Florida, water year 1993–1994*. Tallahassee, Fla.

NOTATION

The following symbols are used in this paper:

A_n and B_n = coefficients of Fourier function;
a_t = random shock at time t;
ARIMA (p, d, q) $(P, D, Q)_s$ = autoregressive integrated moving average. (p, d, q): no seasonal term, p: order of autoregressive term; q: order of moving average term, d: difference order; (P, S, Q): corresponding orders for seasonal term;
B = backshift operator;
b = dead lag time;
C = constant;
D = difference operator;
E = expectation operator;
H_t = water level at time t;
S_t = salinity data at time t;
z_t = artificial data series at time t;
δ = Dirac delta function;
$\theta_q(B)$ and $\theta_Q(B^S)$ = series of nonseasonal and seasonal moving average components of orders q and Q;
$\rho_T(\tau)$ and $\rho_p(\tau)$ = total and partial autocorrelation series;
τ = time lag;
$\varphi_p(B)$ and $\varphi_p(B^S)$ = series of nonseasonal and seasonal autoregressive components of order p;
ω and w = frequencies in Fourier fucntion; and
∇ = difference operator.