Estimation of Sea Level Variability in the China Sea and Its Vicinity Using the SARIMA and LSTM Models

Qinting Sun, Jianhua Wan, and Shanwei Liu

***Abstract*—With a gradually rising global average sea level, it is of great significance to predict changes in the sea level. However, sea level variations often exhibit both linear and nonlinear characteristics, complicating the prediction of sea level changes with a single model. The seasonal autoregressive integrated moving average (SARIMA) model can fully consider the linear characteristics of time series, but its nonlinear prediction ability is poor; the long short-term memory (LSTM) model can compensate for this shortcoming. To predict complex sea level changes, we propose a strategy to combine the SARIMA and LSTM models to increase the sea level prediction accuracy. In our method, sea level anomaly (SLA) time series are decomposed into the trend and seasonal term and random term; then, the SARIMA model is used to predict the trend and seasonal term of sea level variations, whereas the random term is predicted by LSTM. Sea surface height data from 1993 to 2018 are used in an experiment. Compared with other models, the performance of the SARIMA+LSTM model is superior in predicting sea level changes with a minimum root mean square error of 1.155 cm and a maximum determination coefficient (R^2) of 0.89 during the testing period. Furthermore, the predicted results are in close agreement with the SLA data, which indicates that the SARIMA+LSTM model could be successfully used for the estimation of sea level variability.**

***Index Terms*—China Sea and its vicinity, long short-term memory (LSTM), sea level prediction, seasonal autoregressive integrated moving average (SARIMA).**

I. Introduction

SEA level rise is slowly threatening human survival and impairing economic development, as this phenomenon has significant impacts on the social economy, natural environment, and ecosystem of coastal areas. First, rising sea levels may submerge some low-lying coastal areas, strengthening the ability of marine dynamic factors to cause beaches to retreat and erode coast [1]. Second, the frequency of storm tides will increase as sea level rise, which will not only endanger the lives and property of the people in coastal areas, but also salinize the land [2]. Sea level rise occurs continuously, and seawater immersion results in the destruction of agricultural production and the ecological environment. The Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR5) stated that the global mean sea level has risen by 0.19 m, estimated from a linear trend based on tide gauge records over the period 1901–2010 as well as satellite data since 1993. The report also showed that the global mean sea level is expected to rise by 0.63–0.98 m by the end of the 21st century [3]. It is highly probable that the mean rate of sea level rise was 1.7 mm/year between 1901 and 2010; between 1993 and 2010, the rate was very likely higher at 3.2 mm/year, and similarly high rates likely occurred between 1930 and 1950 [4]. As a country with a coastline stretching 18 000 km, China's developed areas are located mainly throughout coastal regions. In recent years, an increasing number of studies have shown that the sea level along the coast of China has been rising under the background of increasing global sea level [5]. Specifically, the sea level throughout the China Sea and its vicinity rose rate of 3.3 mm/year, much higher than the global average of 1.7 mm/year from 1980 to 2017 [6]. Therefore, it is necessary to conduct further research on sea level change, especially the periods and prediction of sea level changes over the China Sea and its vicinity.

Traditional measurements of sea level changes rely on tide gauge and ship measurements. These traditional measurements have high accuracy, but they also suffer from some disadvantages, such as a high cost, an uneven distribution, and a long-acquisition period [7]. Different from coastal tide gauges and ship measurements, satellite altimetry provides measurements of sea level change on a near-global scale [8]. Since the first radar altimeter satellite Skylab was launched in 1973, a large amount of satellite altimeter data with high accuracy and high spatiotemporal resolution have been extensively employed to research global sea level change and in other applications related to the marine sciences [9]. Many experts and scholars have calculated the global sea level rise rate and predicted the sea surface height by using satellite altimetry data [10].

With the development of satellite technology, an increasing number of in-depth studies have been conducted on sea level change. In the previous studies, some researchers applied a single prediction method to estimate sea level variability; for

Manuscript received February 29, 2020; revised April 28, 2020; accepted May 15, 2020. Date of publication June 1, 2020; date of current version June 25, 2020. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 17CX02071, in part by the National Science Foundation of China under Grant 61571009, and in part by the Key R&D Program of Shandong Province under Grant 2018GHY115046. *(Corresponding author: Shanwei Liu.)*

Qinting Sun is with the School of Geosciences, China University of Petroleum (East China), Qingdao 266580, China (e-mail: s16010068@s.upc.edu.cn).

Jianhua Wan and Shanwei Liu are with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China (e-mail: wjh66310@163.com; shanweiliu@163.com).

Digital Object Identifier 10.1109/JSTARS.2020.2997817

example, support vector machines [11], semiempirical approaches [12], and artificial neural networks [13] have been utilized to predict the changes in sea level rise. However, sea level changes often exhibit both linear and nonlinear characteristics, so estimating sea level variability with a single method is difficult. In recent years, sea level data have been divided into linear and nonlinear parts, which have been used in different models for sea level predictions. For example, Sun and others predicted sea level anomalies (SLAs) based on a combined least squares and ARMA model [14]. Imani and others predicted the daily sea level by using an extreme learning machine and a relevance vector machine [15]. Zhao and others predicted the sea level with a combined least squares neural network approach [16].

With the development of neural networks and other technologies, the autoregressive integrated moving average (ARIMA) technique developed by Box and Jenkins, which is a mature approach in the field of economics, has been used to predict time series with obvious periodic changes [17]. Furthermore, the seasonal autoregressive integrated moving average (SARIMA) model is an extension of the ARIMA model that unlike the ARIMA technique it considers the seasonal factors of time series, and thus can better predict trends and seasonal terms in time series [18]. SARIMA model is mature in the fields of economy and health care [19], [20]. Long short-term memory (LSTM) can be applied to discover small patterns by using nonlinear factors of time series and thereby obtain predicted values closer to the actual values [21]. For a number of years, LSTM networks have been successfully employed to evaluate horizontal irradiance [22], sea surface temperature [21], and significant wave height [23].

However, the SARIMA and LSTM models have not been widely used to estimate sea level variations based on the satellite altimetry data. Therefore, we propose a strategy to combine the SARIMA and LSTM models to obtain a higher sea surface height prediction accuracy. The current study aims to demonstrate the applicability and capability of the SARIMA+LSTM model to predict the sea level variations in China Sea and its vicinity. In this article, SLA time series are decomposed into a trend and seasonal term (linear parts) and a random term (nonlinear parts); then the SARIMA model is used to predict the trend and seasonal term of the sea level variations, whereas the random term is predicted by LSTM. Furthermore, the SARIMA+LSTM model is compared with the SARIMA model and LSTM model and altimetry-based sea level anomaly (SLA) data from 1993 to 2018 are used to proposed the model.

The following sections describe the data and techniques used in this study, the results obtained, and some concluding remarks.

II. Data and Methods

A. Data Selection

The satellite altimetry data in this article are geophysical data records from the TOPEX/Poseidon (T/P), Jason-1, Jason-2, and Jason-3 satellites from January 1993 to December 2018, which were obtained from the Centre National d'Etudes Spatiales (CNES). China Sea and its vicinity are studied as a whole area in this article. The study area with longitudes ranging from 105°E to 135°E and latitudes ranging from 0°N to 45°N, as shown in Fig. 1. The MSS_CNES_CLS15 model utilized in this study is provided by CNES and is recognized as a highly precise global mean sea surface model.

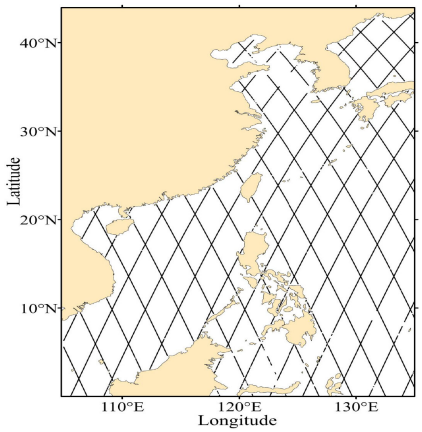


Fig. 1. Research area and tracks of satellites.

B. Normal Morlet Wavelet Transform

Wavelet transform analysis, a powerful statistical tool, was first used in the field of signal processing and analysis. Through noise reduction and the reconstruction and extraction of sound, image, earthquake, and other signals, the vibration periods of different signals can be determined in both the time and the frequency domains. In the geoscience, all kinds of meteorological factors and hydrological processes can be regarded as signals that change periodically with time. Therefore, wavelet analysis is also applicable to the geosciences to analyze the complex time patterns of various processes.

In particular, the continuous wavelet transform (CWT) has found many applications in the field of signal analysis.

For $f(t) \in L(R)$, the normal Morlet wavelet transform is defined as [24]

$$W_g f(a,b) = |a|^{-1} \int_{-\infty}^{+\infty} f(t)\, \overline{\mathrm{g}\left(\frac{t-b}{a}\right)} dt,\ a,b \in R, a \neq 0 \tag{1}$$

where $L(R)$ is a functional space defined by

$$L(R) = \left\{ f(t) \left\| \int_{-l}^{+l} f(t)\, dt \right| < +\infty\ ,\ \forall l \in R^{+} \right\} \tag{2}$$

where the variables a and b are the timescale and time translation indices, respectively. The fundamental kernel, $g(\tau)$, is a normalized simplified Morlet wavelet as

$$\begin{aligned} \mathrm{g}(\tau) &= G(\tau)/\hat{G}(2\pi) \\ &= \left(\sqrt{2\pi}\delta\right)^{-1} \exp -\frac{\tau^2}{2\delta^2} + i2\pi\tau, \qquad (\forall \delta \in R) \geq 1. \end{aligned} \tag{3}$$

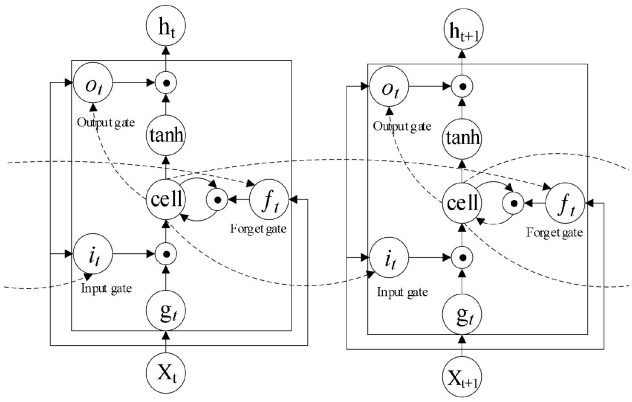


Fig. 2. Unit structure of LSTM.

C. Seasonal Autoregressive Integrated Moving Average

The SARIMA model can fully consider the seasonal characteristics of time series [25]. This model can not only resolve the instability of altimetry data instability but also remove the periodicity of data as a purely stochastic process; thus, the SARIMA model can predict sea level changes well. The SARIMA (p, d, q)(P, D, Q) [s] function is defined as [26]

$$\Phi_p(L)\,A_P(L^{\mathrm{s}})\left(\Delta^d\Delta_s^D SLA\right)=\Theta_q(L)\,B_Q(L^{\mathrm{s}})\,\mu_t \tag{4}$$

$$\begin{cases}\Phi_p(L)=\left(1-\varphi_1\mathrm{L}-\varphi_2\mathrm{L}^2-\cdots-\varphi_{\mathrm{p}}\mathrm{L}^{\mathrm{p}}\right)\\ A_P(L^{\mathrm{s}})=\left(1-\alpha_1\mathrm{L}^{\mathrm{s}}-\alpha_2\mathrm{L}^{2\mathrm{s}}-\cdots-\alpha_{\mathrm{P}}\mathrm{L}^{\mathrm{Ps}}\right)\\ \Theta_q(L)=\left(1+\theta_1\mathrm{L}+\theta_2\mathrm{L}^2+\cdots+\theta_{\mathrm{q}}\mathrm{L}^{\mathrm{q}}\right)\\ B_Q(L^{\mathrm{s}})=\left(1+\beta_1\mathrm{L}^{\mathrm{s}}+\beta_2\mathrm{L}^{2\mathrm{s}}+\cdots+\beta_{\mathrm{Q}}\mathrm{L}^{\mathrm{Qs}}\right)\end{cases} \tag{5}$$

where s is the periodic component, Δ is the nonseasonal difference, Δ_s is periodic seasonal difference, d is the number of nonseasonal differences, D is the number of seasonal differences, SLA is the predicted sea level anomaly, L is a lag operator, μ_t represents white noise, $\Phi_p(L)$ and $A_P(L^s)$ are the characteristic nonseasonal and seasonal autoregression polynomials, respectively, $\Theta_q(L)$ and $B_Q(L^s)$ are the nonseasonal and seasonal moving average operators, respectively, and p, P, q and Q are the maximum lag orders of the nonseasonal, seasonal, autoregressive, and moving average operators, respectively.

D. Long Short-Term Memory

An LSTM model is composed of four main parts, namely, an output gate, an input gate, a forget gate, and a memory cell. The input gate will control whether the data enter the training network at a particular time. The memory cell stores the calculated value for use in the next stage. The output gate controls whether to output the calculated value. The forget gate determines what information is discarded from our cellular state. The LSTM unit structure is shown in Fig. 2 [27], where the symbol $\odot$ represents point-wise scalar multiplication.

Fig. 2 illustrates that the working mechanisms of the gates and the information flow can be expressed using the following equations:

$$\begin{aligned}\mathrm{f}_t&=sigm(W_f x_t+R_f h_{t-1}+b_f)\\ i_t&=sigm(W_i x_t+R_i h_{t-1}+b_i)\\ o_t&=sigm(W_o x_t+R_o h_{t-1}+b_o)\\ g_t&=\tanh(W_g x_t+R_g h_{t-1}+b_g)\\ c_t&=\mathrm{f}_t\cdot c_{t-1}+i_t\cdot g_t\\ h_t&=o_t\cdot\tanh(c_t)\end{aligned} \tag{6}$$

where f_t, i_t and o_t are the outputs of the forget gate, input gate and output gate, respectively, W_{n} and R_{n} are the weights applied to the input x_t and output h_{t-1}, respectively, from the previous cell, and b_n represents the corresponding biases with $n\in(i,g,f,o)$.

III. Analysis and Prediction Changes in Sea Level

A. Data Processing

According to the editorial guidelines provided by CNES, the land components, sea-ice targets, and contaminated data were first removed from the time series. Moreover, to obtain high quality altimetry data, we correct the errors caused by sea state deviations, atmospheric reverse pressures, and atmospheric refraction and so on; to rectify these errors, inverse barometer correction, and troposphere and ionosphere corrections and other geophysical corrections described in the literature [28]–[30] are employed.

T/P series satellites adopt repeated periodic orbits; that is, they orbit the Earth from a particular position and then return to their original position after one cycle, thereby moving back and forth. However, the orbits with the same number display some deviations due to the orbit offset and other factors, and thus, the orbits of satellite motion do not strictly coincide. Therefore, a collinear adjustment is necessary. First, we determine a reference orbit with the best quality, extract the largest amount of data in the same number of orbits, and take the observation points on the reference orbit as normal points. Then, we interpolate the sea surface heights on the other orbits with the same number by applying distance weighting to the normal points; in this way, the sea surface heights at the normal points are obtained. Hence, multiple orbits with the same number are fitted into one orbit.

Ascending and descending orbits intersect due to the rotation of the Earth and the perturbation of the orbits. Nevertheless, the sea surface heights at these intersection points are not necessarily equal; that is, there are some deviations because of the influences of the radial orbit error and system error. The first step in correcting this deviation is to determine the location of the intersection. In general, almost no observations are available at these intersections. Therefore, the sea surface heights at intersections need to be interpolated from the adjacent points; then, the inconsistent values at the intersections can be obtained [31]. The radial orbit error of a satellite can be eliminated by crossover adjustment. The mean and root mean square error (RMSE) of all sea surface height data before the

TABLE I
STATISTICAL RESULTS BETWEEN BEFORE AND AFTER THE COLLINEAR AND CROSSOVER ADJUSTMENTS (UNIT: CM)

Satellite	Before collinear adjustment		After collinear adjustment		After crossover adjustment	
	Mean	RMSE	Mean	RMSE	Mean	RMSE
TP	0.7	16.8	3.4	13.4	-0.1	10.7
Jason-1	-4.4	15.9	2.9	12.5	0.0	7.8
Jason-2	-0.1	14.3	4.2	11.6	-0.1	7.1
Jason-3	-2.4	17.4	1.5	10.2	-0.1	6.8

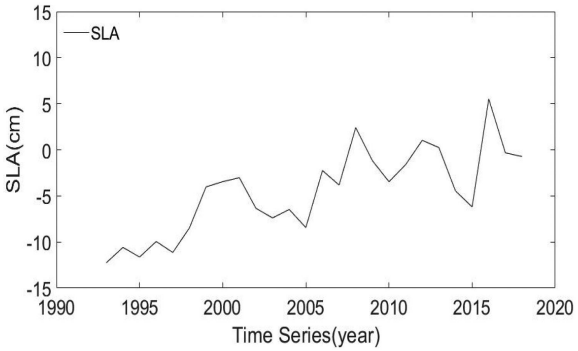


Fig. 3. Mean annual SLA time series in the study area after bias corrections.

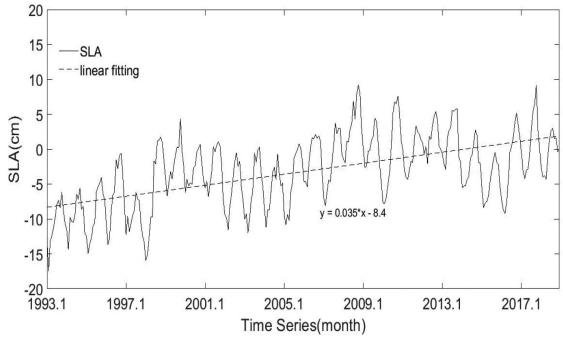


Fig. 4. Mean monthly SLA time series in the study area after bias corrections.

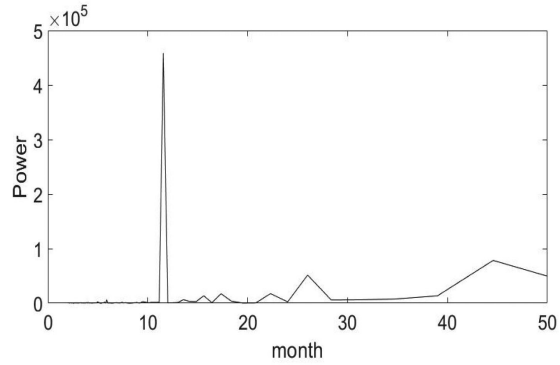


Fig. 5. Power spectra of the SLA time series.

collinear adjustment are obtained, and results are listed in the first three columns of Table I. The intersection discrepancies after the collinear adjustment and before and after the crossover adjustment are counted, and the results are described in Table I. Evidently, the RMSE decreased significantly after the collinear adjustment and crossover adjustment, indicating that a collinear adjustment and a crossover adjustment are very necessary.

B. Periodic Change in Sea Level

A signal that changes with time may have different periods on different time scales, and different periods may exist in the same time domain. In this article, annual and monthly sea level change time series are processed and analyzed, and the periodic changes in the time series are analyzed by the wavelet transform.

The sea surface height of a normal point is obtained by the inverse distance weighting method. All SLAs of normal points can be computed by using sea surface height data along the reference orbit of normal points minus the values calculated by interpolating the mean normal points within the MSS_CNES_CLS15 sea level model. However, due to the uneven distribution of survey points, the SLA values need to be weighted by latitude. The SLA for the overall study area is corrected by correcting the biases and implementing linear fitting [32]. The SLA time series are shown in Fig. 3 (annual) and Fig. 4 (monthly).

Figs. 3 and 4 illustrate that the sea level change shows an increasing trend. On the one hand, the annual SLA data can show only the overall trend; that is, the periodic trend is not shown. On the other hand, the period and seasonal trend in the monthly data are evident. Thus, the monthly data were used to analyze the periodic sea level variation in this article.

As shown in Fig. 4 the SLA in the study area exhibits an increasing trend, and there is an obvious periodic pattern: the sea level decreases in the winter and spring after rising in the summer and autumn. The rate of sea level rise within the study area as determined by linear fitting is 4.16 mm/a, and the rate of sea level rise along the coast of the China Sea (4.64 mm/a) was similar to that obtained by Guo *et al.* [33]. Moreover, sea level fluctuations are also greatly affected by large climate changes, such as the sea level changed significantly when a large El Niño phenomenon occurred from 1997 to 1998. The power spectrum of the monthly variation time series obtained by applying a Fourier transform is plotted in Fig. 5.

According to the relationship between the time and frequency of the peak in Fig. 5, the most significant period of sea level change in the study area is 1a, but periods of 2.2 and 3.7 a are also observed. However, these simple time series and spectrum plots can provide only rough information about sea level change, as it is difficult to accurately reflect the periodic characteristics accurately. Fortunately, the wavelet transform can analyze time series in the time domain simultaneously and has adaptive time-frequency window transform, local analysis, and refinement functions. Therefore, Daubechies4 (DB4) wavelet is selected in this article to perform a one-dimensional CWT on multiple scales (32, 64, 128 months) for the monthly SLA time

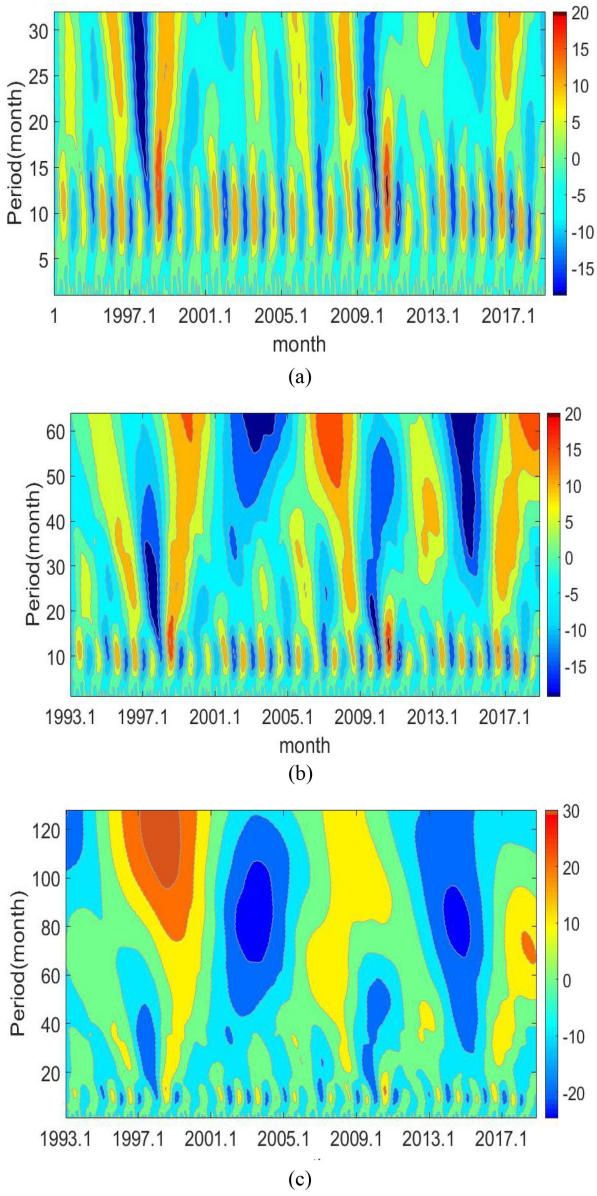


Fig. 6. Wavelet spectra for the mean monthly SLA in the study area (scales = 32 (a), 64 (b), 128, and (c) months).

series. A contour map of the wavelet coefficient is shown in Fig. 6.

In Fig. 6, the red corresponds to a high value, whereas blue represents a low value; the darker the color is, the greater the amplitude of the fluctuation. Taking Fig. 6(a) (scale = 32 months) as an example, there are dozens of wave energy concentration centers of SLAs (i.e., where the isolines are dense and closed). Each of these concentration centers near the vertical coordinate of ten months is analyzed. Among them, at the center (7, 9), on the time-domain scale (abscissa), the influence ranges from April 1993 to October 1993; on the frequency-domain scale (ordinate), the center spans approximately nine months. The variation in wave energy around the concentration center tends to be gentle, and the wavelet coefficient is positive, indicating that the sea level is rising. Next, at the center of the contour at (13,

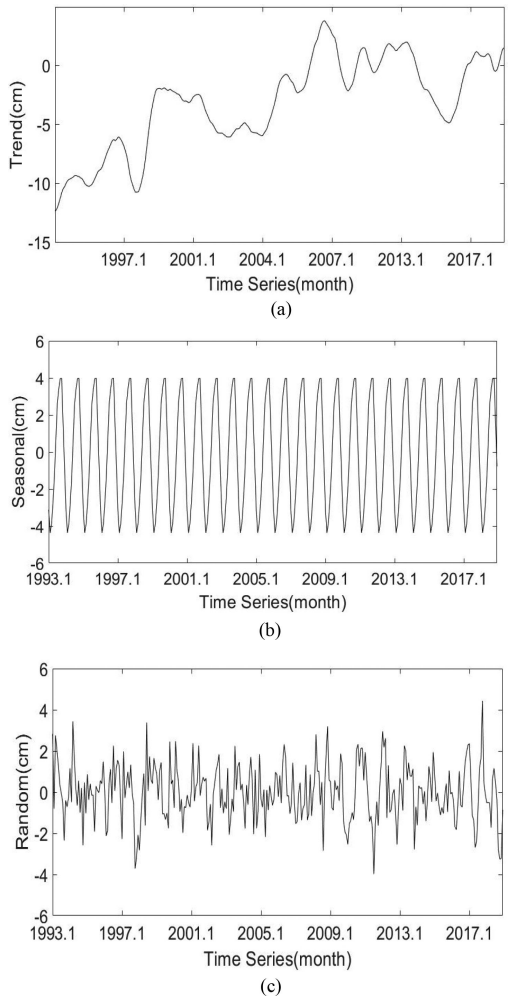


Fig. 7. Signal decomposition of the SLA time series.

10), the influence ranges from October 1993 to April 1994 on the time-domain scale, and on the frequency-domain scale, the center spans approximately ten months. The wavelet coefficient is negative, which indicates that the sea level is dropping.

Fig. 6(a) demonstrates that the average coordinates of all the central points are approximately 12 months; that is, the sea level change in the study area has a period of 1a and is characterized by alternating rising and falling trends. The sea level is low in winter and spring and high in summer and autumn, which is confirmed by the above analysis of the time series. We also find that periods of 1.5a, 2.2a [see Fig. 6(b)], 4.6a, and 7a [see Fig. 6(c)] in the sea level change within the study area.

C. Prediction of Sea Level Change Using the SARIMA+LSTM Model

The periodic and trend terms of sea level variations are fitted and extrapolated using the SARIMA model, whereas the random term is predicted by LSTM. First, the SLA time series data are decomposed, and the characteristics are extracted. Fig. 7 shows the decomposition results of the monthly SLA time series

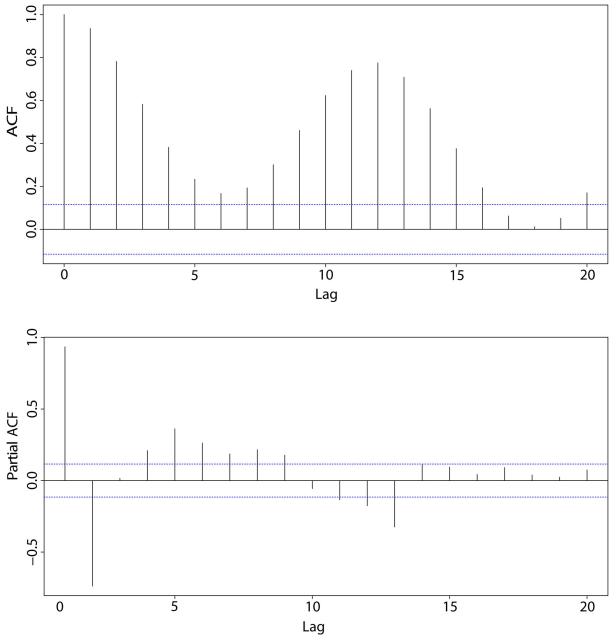


Fig. 8. Autocorrelation function and partial autocorrelation function of the trend and seasonal term.

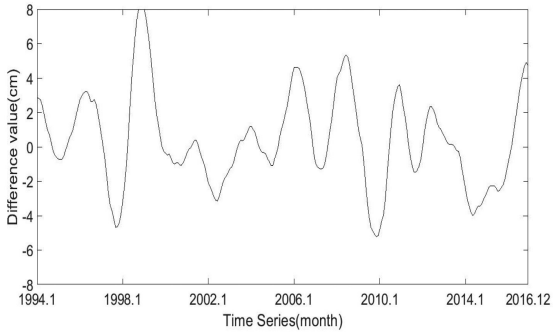


Fig. 9. Difference values of the first-order seasonal difference.

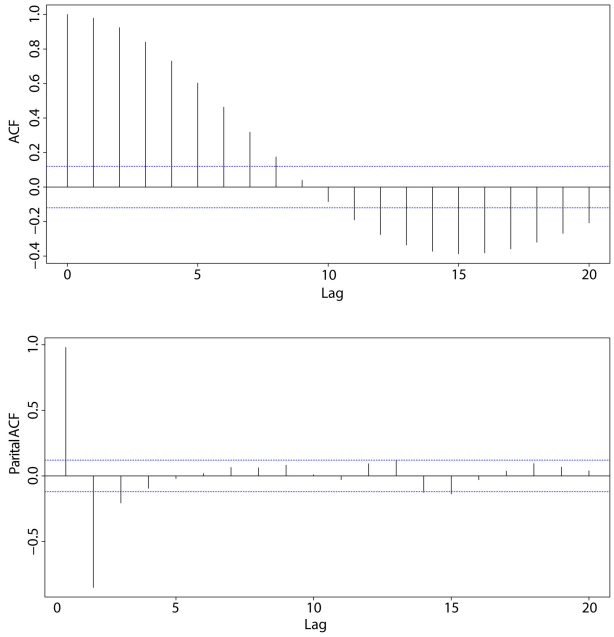


Fig. 10. Autocorrelation function and partial autocorrelation function after removing the first-order seasonal difference.

from 1993 to 2018. The decomposition results include three components: "trend", "season," and "random."

Fig. 7 (a) shows that the sea level anomaly shows a slowly rising trend, which is in accordance with the situation of global sea level rise. Fig. 7(b) demonstrates that the seasonal fluctuation of the time series is relatively with a periodicity of one year, which is consistent with the wavelet analysis results. The random term plotted in Fig. 7(c) is the random signal and noise of the SLA value generated under uncertain or sudden conditions. In this article, the SARIMA model is used to predict the trend and seasonal term; then the LSTM model is used to predict the random term. We use the data from 1993 to 2016 as training samples and the data from 2017 and 2018 as testing samples.

1) Predicting the Trend and Seasonal Term: First, we need to determine the parameters of the SARIMA (P, D, q) (P, D, q) [S] model. The autocorrelation coefficients graphs and partial autocorrelation coefficients of the trend and seasonal term are shown in Fig. 8, which suggests that the partial autocorrelation function has a third-order truncation, and the number of nonseasonal differences d is 0.

Because the data to be predicted are not a stationary sequence, the periodicity of the data can be eliminated by using a first-order seasonal difference. The resulting time series obtained by using this difference is shown in Fig. 9. The time series in Fig. 9 is obviously a stationary sequence, and the p-value after removing the seasonal difference is less than 0.01, which confirms the validity of this result. At this time, the number of seasonal differences D $= 1$ can be determined. Therefore, another difference does not need to be applied to the data.

The autocorrelation coefficients and partial autocorrelation after removing the first-order seasonal difference are shown in Fig. 10, which illustrates that the autocorrelation coefficients decay periodically with time. The function is tailed, and the maximum lag order of the seasonal term P is 0. Additionally, the partial autocorrelation coefficients are truncated and tend to 0 after the third order. Hence, the partial autocorrelation function has a third-order truncation, and the maximum lag order of the moving average operator Q is 3.

From the above, we can select D $= 1$, P $= 0$, q $= 3$, and d $=$ 0. p and q can each be taken as 0, 1, and 2, and can be tested one by one from low to high. The RMSE is the square root of the ratio between the square sum of the deviation between the observed value and the true value and the number of observations m. This metric is used to measure the deviation between the observed value and the true value. The R^2 reflects the degree of data dispersion; the larger the R^2 is, the better the data are grouped. The Akaike information criterion (AIC) is a standard for measuring the goodness of statistical model fit. The smaller the AIC, the better the fit. When we choosing the best model from a group of models, we consider the RMSE, R^2 and AIC.

TABLE II
FITTING RESULTS OF SELECTED MODELS FOR THE TREND AND SEASONAL TERM OF SEA LEVEL ANOMALY (UNIT: CM)

Type of model	RMSE	R^2	AIC
SARIMA (0,0,0) (0,1,3) [12]	2.751	0.151	1326.28
SARIMA (0,0,1) (0,1,3) [12]	2.614	0.233	966.91
SARIMA (0,0,2) (0,1,3) [12]	2.412	0.347	619.21
SARIMA (1,0,0) (0,1,3) [12]	0.602	0.959	145.66
SARIMA (1,0,1) (0,1,3) [12]	1.092	0.866	-200.86
SARIMA (1,0,2) (0,1,3) [12]	0.999	0.888	-505.45
SARIMA (2,0,0) (0,1,3) [12]	1.071	0.871	-487.55
SARIMA (2,0,1) (0,1,3) [12]	0.753	0.936	-567.75
SARIMA (2,0,2) (0,1,3) [12]	0.834	0.922	-580.23

Nine SARIMA models are selected, and the statistical results of each model are presented in Table II.

The RMSE is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{7}$$

wherey_i is actual the value, $\hat{y}_i$ is the predicted value, and m is the number of predictions.

R^2 is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{8}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}_i$ is the mean of actual value, and *n* is the number of predictions.

The AIC is defined as

$$\mathrm{AIC} = 2k - 2ln(L) \tag{9}$$

where K is the number of parameters and *L* is the likelihood function.

Table II demonstrates that the RMSE of the SARIMA (1, 0, 0)(0, 1, 3) model is the lowest, but it is not the best model because the AIC is large. The discreteness of the SARIMA (2, 0, 1) (0, 1, 3) [12] model is the lowest, but the degree of fitting is not very good. Furthermore, the SARIMA (2, 0, 2) (0, 1, 3) [12] model has better aggregation and the best fitting degree(its AIC is the lowest), so we choose the SARIMA (2, 0, 2) (0, 1, 3) [12] model as the prediction model.

For the obtained SARIMA (2, 0, 2) (0, 1, 3) [12] prediction model, the mean value of the residual is 0.0002, the standard deviation is 0.069, the mean value falls within ± 0.25 cm, and the original assumption that the mean value of the residuals is 0 is accepted. The Ljung-box method and autocorrelation function diagram are used to test the time independence of the residuals. The results are shown in Fig. 11.

Fig. 11 demonstrates that the autocorrelation coefficient of the residual is 1 at lag $= 0$, whereas the remaining coefficients all fall into the confidence interval, indicating that the residuals

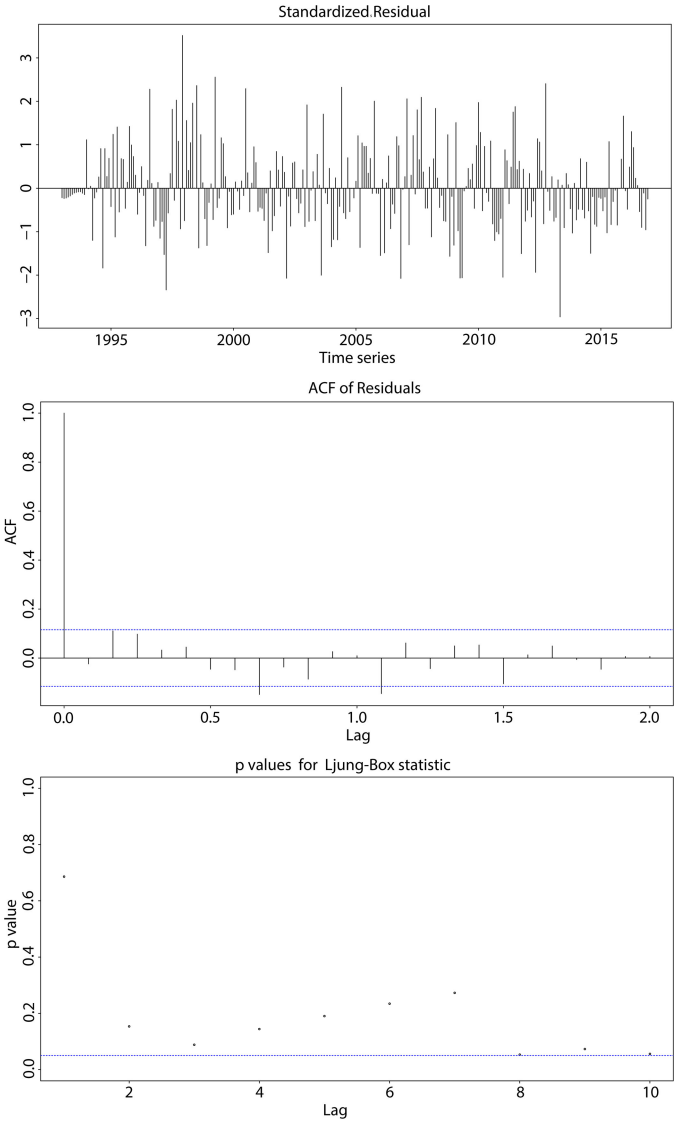


Fig. 11. Time independence test for the residuals.

before and after lag $= 0$ have no correlation. The p-value of the Ljung-box test is 0.23, and all of the p-value fall above the 0.05 level, further proving that the residual sequence is independent of time.

This time independence test shows that the SARIMA (2, 0, 2) (0, 1, 3) [12] prediction model is effective for fitting SLA changes and can be used for prediction.

The trends and seasonal terms of the SLAs in 2017 and 2018 are predicted by using the SARIMA model with the data from 1993 to 2016, and the results are shown in Fig. 12, in which the black line represents the reconstructed result of the decomposed seasonal and trend signal, and the red line represents the predicted value by using the SARIMA model.

Fig. 12 shows that the prediction effect of some values cannot reach the real value, and the values predicted with the model as a whole have a high degree of fit with the real value. The maximum absolute error is -1.146 cm, the average absolute error is -0.560 cm, and the overall prediction effect is good. However,

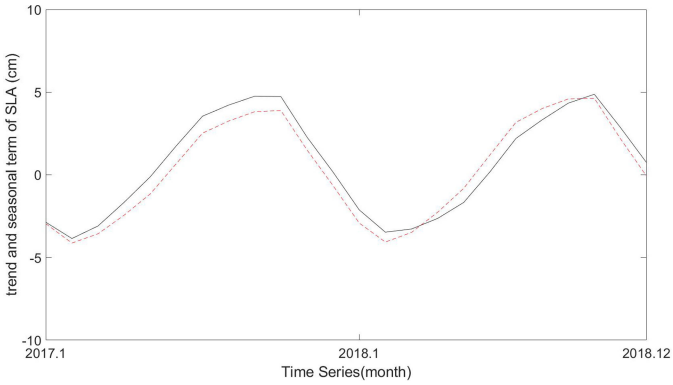


Fig. 12. Actual and predicted value of SLA by using the SARIMA model (red line represents predicted values, and black line represents actual values).

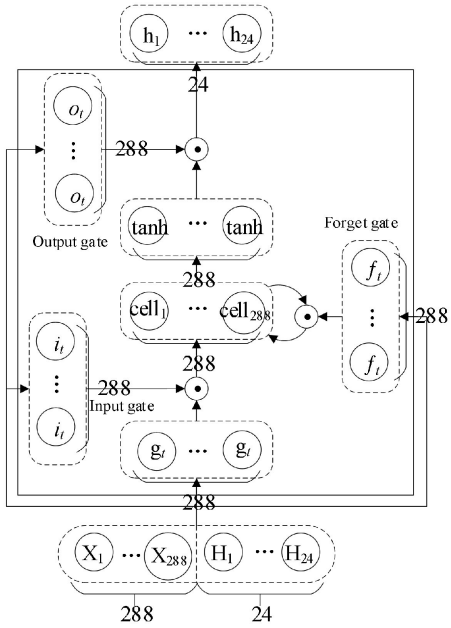


Fig. 13. LSTM network training process plot (X denotes the training data, H is the actual data, and h is the predicted data).

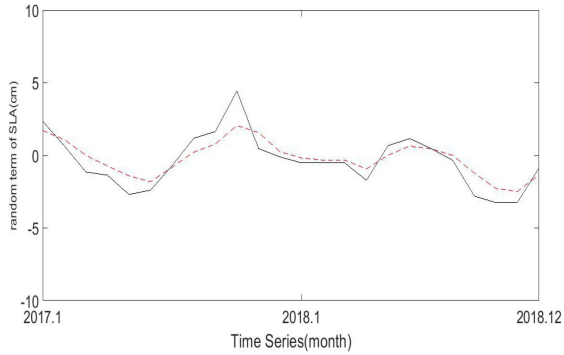


Fig. 14. Actual and predicted values of the SLA by using the LSTM model (red line represents predicted values, and black line represents actual values).

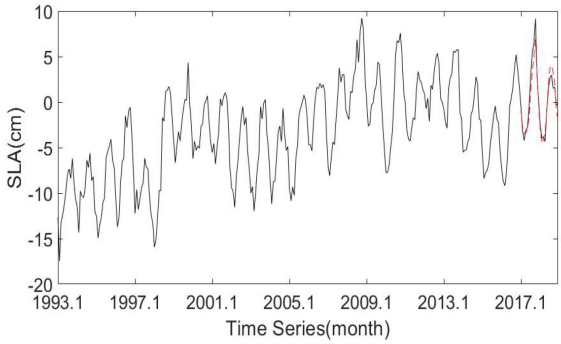


Fig. 15. Actual and predicted values of the SLA by using the SARIMA+LSTM model (red line represents predicted values, and black line represents the SLA).

TABLE III
FITTING RESULTS OF SELECTED MODELS UNIT: CM

Type of model	MAE	RMSE	R^2
SARIMA	1.236	1.667	0.773
LSTM	1.251	1.718	0.759
SARIMA+LSTM	0.837	1.155	0.893
LS+RBF	1.933	2.221	0.597

the prediction of individual values is not ideal, especially the values in May 2017, December 2017, June 2018, and December 2018.

2) Predicting the Random Term: The LSTM model can be used to predict the random term of SLA. The advantage of LSTM is that it can learn and remember long time series and does not rely on prespecified window lag observations as input. The LSTM network training process as shown in Fig. 13. We use the SLA of 288 months as input data, and then the SLA of 24 months are predicted by using the LSTM model.

The test set data are substituted into the model for prediction, and the data prediction results from 2017 to 2018 are shown in Fig. 14, which demonstrates that the prediction effect of the random term is good, and the predicted RMSE is 0.88 cm. Although the predicted monthly sea level maximum is not exactly the same as the actual value, the predicted values fall into the confidence interval range of the actual values, denoting a high accuracy for the short-term prediction of sea level.

3) Prediction of Sea Level Change Using the SARIMA+LSTM Model: Finally, the results based on the SARIMA+LSTM model are shown in Fig. 15. Evidently, the proposed method can predict the trend of the sea level height well. The RMSE after the prediction is 1.155 cm.

D. Analysis of Periodic Models

Next, the results of this study are compared with those from SARIMA and LSTM models. These two models have been shown to be some of the most effective models for time series prediction [17], [18]. The SLA time series are predicted by using the SARIMA model and LSTM model separately, and the prediction results are described in Table III. In recent studies,

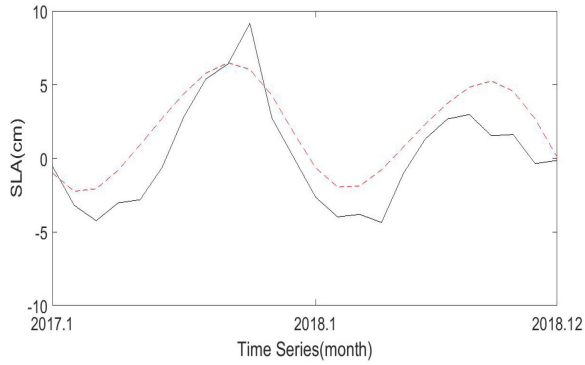


Fig. 16. Actual and predicted values of the SLA by using the LS+RBF model (red line represents predicted values, and black line represents actual values).

Zhao *et al.* [16] predict sea level change by using combined method of least squares and a neural network, among which the least squares (LS) and radial basis function (RBF) method present best fitting results. Therefore, LS+RBF model and SARIMA+LSTM model were used to predict the sea level height. The results are listed in Table III and Fig. 16. The mean absolute error (MAE) is the average of all the absolute values of differences between the individual observations and true values. The MAE can avoid the problem of offsetting each deviation.

The MAE is defined as

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} (|y_i - \hat{y}_i|)}{N} \tag{10}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of predictions.

Table III and Fig. 16 show that the MAE of the predicted sea level height obtained by the combined SARIMA+LSTM model is the smallest, and its degree of dispersion is small. Its RMSE is also the smallest, as is the deviation between the predicted and the real values. The SARIMA+LSTM model has better fitting results among the four methods.

IV. Conclusion

The study aims to evaluate the capability of the SARIMA+LSTM model in the short-term prediction of sea level variability in China Sea and its vicinity. The method proposed in this study decomposes SLA time series into the trend and seasonal term and random term; then the SARIMA model is used to predict the trend and seasonal term of sea level variations, whereas the random term is predicted by LSTM. A test is carried out in the China Sea and its vicinity by using 25 years of sea surface height data from the T/P, Jason-1, Jason-2 and Jason-3 satellites. The data for the first 23 years (January 1993 to December 2016) are used for training, whereas those from January 2017 to December 2018 are used for testing.

The results of the study are compared with those obtained by using the SARIMA and LSTM models. The findings indicate that the SARIMA+LSTM model is a useful tool for predicting short-term sea level variations with centimeter-level precision. Although SLA time series are complex, sea level variations can be predicted quickly and successfully while considering only time variables by using the SARIMA+LSTM model. Future studies may consider topographical details or other meteorological data, which were not assessed in the current work.

Acknowledgment

The authors would like to thank the editors and would also like to thank AVISO website of the French Space Center (CNES) for providing the satellite data.

References

[1] N. Mimura, "Sea-level rise caused by climate change and its implications for society," *Proc. Jpn. Acad. Ser. B-Phys. Biol. Sci.*, vol. 89, no. 7, pp. 281–301, Jul. 2013.

[2] J. Yin, Z. Yin, J. Wang, and S. Y. Xu, "National assessment of coastal vulnerability to sea-level rise for the Chinese coast," *J. Coastal Conservation*, vol. 16, no. 1, pp. 123–133, Mar. 2012.

[3] C. P. Field *et al.*, "Climate change 2014: Impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects. Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change," IPCC, Cambridge, NY, USA, IPCC Assessment Rep., 2014, pp. 1–32.

[4] J. A. Church *et al.*, "Sea level change climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change," Cambridge Univ. Press, Cambridge, U.K., Assessment Rep., Sep. 2013, pp. 12.

[5] J. Zhang and M. Q. Fang, "Sea level trends of China seas from 1993 to 2012," *Periodical Ocean Univ. China*, vol. 45, no. 1, pp. 121–126, Jan. 2015.

[6] State Oceanic Adminitration, People's Republic of China, Apr. 23, 2018. "China sea level bulletin 2017," [Online]. Available: http://gc.mnr.gov.cn/201806/t20180619_1798298.html

[7] X. F. Duan *et al.*, "Methodology and case study of sea level prediction based on secular tide gauge data," *Acta Sci. Nat. Univ. Pekinensis*, vol. 50, no. 6, pp. 1065–1070, Jun. 2014.

[8] O. P. Francis, G. G. Panteleev, and D. E. Atkinson, "Ocean wave conditions in the Chukchi Sea from satellite and in situ observations," *Geophys. Res. Lett.*, vol. 38, no. 24, pp. 1–5, Dec. 2011.

[9] A. Cazenave, H. Palanisamy, and M. Ablain, "Contemporary sea level changes from satellite altimetry: What have we learned? What are the new challenges?" *Adv. Space Res.*, vol. 62, no. 7, pp. 1639–1653, Jul. 2018.

[10] C. K. Shum and C. Y. Kuo, "Observation and geophysical causes of present-day sealevel rise," in *Climate Change and Food Security in South Asia*. Dordrecht, the Netherlands: Springer, 2011, ch. 7, pp. 85–104.

[11] T. Asefa *et al.*, "Multi-time scale stream flow predictions: The support vector machines approach," *J. Hydrol.*, vol. 318, no. 1–4, pp. 7–16, Mar. 2006.

[12] S. Rahmstorf, "A semi-empirical approach to projecting future sea-level rise," *Science*., vol. 315, no. 5810, pp. 368–370, Oct. 2007.

[13] D. Makarynska and O. Makarynskyy, "Predicting sea-level variations at the Cocos (Keeling) Islands with artificial neural networks," *Comput. Geosci.*, vol. 34, pp. 1910–1917, Dec. 2007.

[14] W. Sun and Q. B. Wang, "Sea level anomaly forecasting based on combined model of least square and ARMA," *J. Geodesy Geodyn.*, vol. 32, no. 5, pp. 91–94, Oct. 2012.

[15] M. Imani *et al.*, "Daily sea level prediction at Chiayi coast, Taiwan using extreme learning machine and relevance vector machine," *Global Planet. Change*, vol. 161, no. 17, pp. 211–221, Dec. 2017.

[16] J. Zhao, Y. G. Fan, and Y. X. Mu, "Sea level prediction in the Yellow Sea from satellite altimetry with a combined least squares neural network approach," *Mar. Geodesy*, vol. 42, no. 4, pp. 344–366, Jun. 2019.

[17] F. M. Tseng, H. C. Yu, and G. H. Tzeng, "Combining neural network model with seasonal time series ARIMA model," *Technol. Forecasting Soc. Change*, vol. 69, no. 1, pp. 71–87, Jan. 2002.

[18] Q. Man *et al.*, "Forecasting the incidence of tuberculosis in China using the seasonal auto-regressive integrated moving average (SARIMA) model," *J. Infection Public Health*, vol. 11, no. 5, pp. 707–712, Oct. 2018.

[19] Q. Y. Li and J. Y. Wang, "Prediction on the number of discharged patients with the application of SARIMA mode," *Chin. Med. Rec.*, vol. 17, no. 7, pp. 32–35, Aug. 2016.
[20] Q. Mao *et al.*, "Forecasting the incidence of tuberculosis in China using the seasonal auto-regressive integrated moving average (SARIMA) model," *J. Infection Public Health*, vol. 11, no. 5, pp. 701–712, Sep. 2018.
[21] C. J. Xiao, N. C. Chen, and C. L. Hu, "Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-Ada Boost combination approach," *Remote Sens. Environ.*, vol. 233, no. 7, pp. 281–301, Nov. 2019.
[22] S. Srivastava and S. Lessmann Xiao, "A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data," *Sol. Energy*, vol. 162, no. 1, pp. 232–247, Mar. 2018.
[23] S. Fan, N. Xiao, and S. Dong, "A novel model to predict significant wave height based on long short-term memory network," *Ocean Eng.*, vol. 205, no. 1, pp. 1–13, Jun. 2020.
[24] L. T. Liu, H. Hsu, H. L. Liu, and E. W. Grafarend, "Normal morlet wavelet transform and its application to the Earth's Polar motion," *J. Geophys. Res.*, vol. 112, no. B8, pp. 1–14, Aug. 2007.
[25] M. Valipour, "Long-term runoff study using SARIMA and ARIMA models in the United States," *Meteorol. Appl.*, vol. 22, no. 3, pp. 592–598, Feb. 2015.
[26] L. T. Liu, H. Hsu, H. L. Liu, and E. W. Grafarend, "Combining neural network model with seasonal time series ARIMA model," *Technol. Forecasting Soc. Change*, vol. 69, no. 1, pp. 71–87, Jan. 2002.
[27] S. Srivastava and S. Lessmann, "A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data," *Sol. Energy*, vol. 162, no. 1, pp. 232–247, Mar. 2018.
[28] E. Bronner *et al.*, "Jason-1 products handbook," CNES, NASA, JPL, CLS, PODAAC, issue: 5.1, Accessed: Apr. 4, 2016. [Online]. Available: https://www.aviso.altimetry.fr/fileadmin/documents/data/tools/hdbk_j1_gdr.pdf
[29] J. P. Dumont *et al.*, "OSTM/Jason-2 products handbook," CNES, NASA, JPL, CLS, NOAA, EUMETSAT, issue: 1.11, Accessed: Jan. 13, 2017. [Online]. Available: https://www.aviso.altimetry.fr/fileadmin/documents/data/tools/hdbk_j2.pdf
[30] J. P. Dumont *et al.*, "Jason-3 products handbook," CNES, NASA, JPL, CLS, EUMETSAT, issue: 1.4, Accessed: Jan. 16, 2017. [Online]. Available: https://www.aviso.altimetry.fr/fileadmin/documents/data/tools/hdbk_j3.pdf
[31] W. P. Jiang, J. C. Li, and Z. T. Wang, "Combining Multi-satellite data to compute the global average sea surface WHU2000," *Chin. Sci. Bull.*, vol. 47, no. 15, pp. 1187–1191, Aug. 2002.
[32] J. H. Wan *et al.*, "Sea-level change over the china sea and its vicinity derived from 25-Year T/P series altimeter data," *J. Indian Soc. Remote Sens.*, vol. 46, no. 12, pp. 1939–1947, Dec. 2018.
[33] J. Y. Guo *et al.*, "Temporal-spatial variations of sea level over China seasderived from altimeter data of TOPEX/Poseidon, Jason-1 and Jsaon-2 from 1993 to 2012. Chinese," *J. Geophys.*, vol. 58, no. 9, pp. 3104–3120, Sep. 2015.

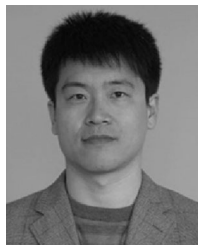
Qinting Sun received the bachelor's degree from the College of Information Science and Engineering, Shandong Agricultural University, Tai'an, China, in 2016. She is currently working toward the Graduate degree with the School of Geosciences, China University of Petroleum (East China), Qingdao, China.

Her research interests include satellite altimetry, deep learning, and sea level change.

Jianhua Wan received the Ph.D. degree from Wuhan University, Wuhan, China, in 2001.

He is currently a Professor with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China. He has authored/coauthored more than 30 high-level papers, and he is the inventor or coinventor of five patents. His research interests include geographic information, ocean remote sensing, and smart city.

Dr. Wan is also the Director of the Chinese Society for Geodesy, Photogrammetry, and Cartography (CSGPC), a member of the theory and method working committee of China Association for Geographic Information System Society, a member of the engineering survey special committee of the Engineering Surveying Special Committee of CSGPC, a member of the oceanology and limnology information technology special committee of Chinese Society for Oceanology and Limnology, and the Vice Chairman of the smart city working committee of CSGPC. He was also the Evaluation Expert of National Science and Technology and the recipient of the National Science and Technology Award, Provincial Science and Technology Award, and National Natural Science Foundation. He was also the Editorial Board Member of the Geomatics World and the Instructor of huitiandi, a well-known wechat platform in the field of Surveying and mapping geographic information.



Shanwei Liu received the Ph.D. degree from the Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Beijing, China, in 2011.

He is currently an Assistant Professor with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China. He has authored/coauthored more than ten high-level papers, and he is the inventor or coinventor of five patents. His research interests include satellite altimetry, ocean remote sensing, and GIS application.