A Modular Tide Level Prediction Method Based on a NARX Neural Network

WENHAO WU1, LIANBO LI1, JIANCHUAN YIN2, WENYU LYU3, WENJUN ZHANG1

1Navigation College, Dalian Maritime University, Dalian 116026, China
2Maritime College, Guangdong Ocean University, Zhanjiang 524088, China
3School of Foreign Languages, Dalian Maritime University, Dalian 116026, China

Corresponding author: L. LI (e-mail: lianboli@dlmu.edu.cn), W. ZHANG (e-mail: wenjunzhang@dlmu.edu.cn).

This work was supported in part by the National Key R&D Program of China (Grant No. 2019YFB1600602), the National Natural Science Foundation of China (Grant No. 51879024), the Liaoning Natural Science Planning Project Foundation (Grant Nos. 20180530283, 2020-HYLH-27, 201801704), the Dalian Maritime University Teaching Reform Project (Grant No. 2020Y04), the Transportation Education Scientific Research Project of China Institute of Communications Education (Grant Nos. JTYB20-03, JTYB20-09), and the Fundamental Research Funds for the Central Universities (Grant No. 3132021133).

ABSTRACT

Tide variations are affected not only by periodic movement of celestial bodies but also by time-varying interference from the external environment. To improve the accuracy of tide prediction, a modular tide level prediction model (HA-NARX) is proposed. This model divides tide data into two parts: astronomical tide data affected by celestial tide-generating forces and nonastronomical tide data affected by various environmental factors. Final tide prediction results are obtained using a nonlinear autoregressive exogenous model (NARX) neural network combined with harmonic analysis (HA) data. To verify the feasibility of the model, tide data under different climatic and geographical conditions are used to simulate the prediction of tide levels, and the results are compared with those of traditional HA, the genetic algorithm-back propagation (GA-BP) neural network and the wavelet neural network (WNN). The results show that the greater the influence of meteorological factors on tides, the more obvious is the improvement in accuracy and stability of HA-NARX prediction results compared to traditional models, with the highest prediction accuracy improvement of 234%. The proposed model not only has a simple structure but can also effectively improve the stability and accuracy of tide prediction.

INDEX TERMS GA-BP neural network, Harmonic analysis, WNN, Modular prediction, NARX neural network, Tide prediction

I. INTRODUCTION

Tides are periodic fluctuations of seawater generated by the combined gravitational forces of the Moon and Sun and by the inertial centrifugal force required for the relative motion of the Earth. With the development of science and technology, the influence of tides on navigation is gradually increasing [1-2]. As a traditional technique for tide prediction [3-5], the harmonic analysis method decomposes complex tides into several parts with periodic changes. By analyzing observed tide level data, the constants in the tide harmonic model can be obtained. Then, according to these harmonic constants, the tide components can be calculated and used to predict tides. The main drawback of this method is that a large amount of long-term observation data is needed to obtain a relatively accurate harmonic analysis model [6]. In addition, tides are affected not only by gravity but also by weather, and the harmonic analysis method is unable to consider the influences of complex weather factors. Therefore, it is difficult for the traditional harmonic analysis model with a static structure to provide high-precision tide predictions.

In recent years, artificial intelligence technology has developed rapidly. Intelligent computation techniques have been widely employed in the areas of ocean engineering and marine science [7-10]. For example, wavelet neural networks (WNNs) [11], support vector machines (SVMs) [12], backpropagation (BP) neural networks [4, 13], and long short-term memory (LSTM) neural networks [5, 14] have been extensively applied in coastal and marine engineering due to their strong searching, reasoning, planning and self-learning abilities [15-16]. Qiu et al. [17] proposed an operational evaluation method for tide forecasting based on dynamic weight allocation that realized synchronous forecasting using
multiple forecasters. Although this method reduces unstable factors in the prediction results and improves the accuracy and rationality of tide prediction, the approach is more labor-intensive. Yin et al. [18] proposed a variable-structure radial basis function (RBF) network based on a sliding data window to predict tide levels in real time. The prediction accuracy of this method is improved compared with that of traditional harmonic analysis, but considerable room for improvement remains. Nitsure et al. [19] proposed a method for indirectly predicting sea levels using a generic programming artificial neural network (GPANN) and wind field information, but its prediction accuracy is easily affected by changes in the surrounding environment. Yin [20] proposed an online sequential extreme learning machine (OS-ELM) by introducing the hidden element pruning strategy for online tide prediction. The OS-ELM has improved prediction accuracy and calculation speed, but the lack of neurons in the hidden layer makes the stability of the network vulnerable.

In the last decade, an increasing number of scholars have begun to use combined methods to predict tides. El-Diasty and Al-Harbia [11] proposed a high-precision sea level prediction model by combining harmonic analysis with WNNs. The results showed that the prediction accuracy was improved over that of traditional harmonic analysis; however, the dispersion of the data was relatively large, and the data were analyzed using only RMS, which was not sufficient to quantify the accuracy. Zhang et al. [21] used harmonic analysis and an adaptive network-based fuzzy inference system (ANFIS) to establish a comprehensive and reliable tide level prediction network with improved prediction accuracy. The main drawback of this model is its overly complex structure, which requires tedious computational steps to derive prediction results. Liu et al. [22] proposed a combined tide forecasting model based on harmonic analysis and autoregressive integrated moving-average-support vector regression (ARIMA-SVR), which improves the accuracy of single-prediction models. This combined model shows effectively improved prediction accuracy, but the improvement is limited.

Kumar et al. [23] proposed a model based on the strong coupling between fully nonlinear potential flow theory (FNPT) in the far field and the Navier-Stokes (NS) equations in the nearshore region to estimate the run-up of tsunami-like waves. The focus of this study was not tides, but it provided ideas for predicting tides using hybrid models. A tide level prediction model with a simple structure, high-accuracy prediction results, a short running time and the ability to overcome the influence of atmospheric effects is proposed to address the main drawbacks of previously proposed models. Tide level data can be regarded as time series in forecasting. The structural characteristics of a nonlinear autoregressive exogenous (NARX) neural network model provide better learning efficiency and higher prediction accuracy for time series [24]. In addition, NARX neural networks have been implemented in modeling and prediction in several research areas, such as by Lou et al. [25], Buevich et al. [26] and Shahbaz et al. [27]. It is worth noting that there is no existing research on tide level prediction with NARX, which involves a dynamic regression network consisting of static neurons and network output feedback that outperforms full regression neural networks. The details are shown in Section 2.

To further improve the prediction accuracy of tide data, HA-NARX is proposed. The results are compared with those of the traditional harmonic analysis method, a WNN, and the genetic algorithm (GA)-BP neural network.

The rest of the paper is organized as follows. Section 2 presents the methods, including the comparison methods and proposed method. Section 3 describes the selection of tide data and quantifiers for accuracy. Section 4 contains the prediction results obtained using the different methods. Section 5 contains the analysis and discussion of the prediction results. Section 6 contains the conclusion and recommendation.

II. METHODS
This section focuses on the NARX neural network, reconciliation analysis method and HA-NARX, including the algorithm structure, operation steps and parameter settings. The remaining two comparison algorithms, the GA-BP neural network and WNN, are also briefly introduced.

A. NARX NEURAL NETWORK
A NARX neural network is a nonlinear autoregressive model for describing nonlinear discrete systems [35]. It is the most widely used type of neural network in nonlinear dynamic systems and is suitable for time series prediction. Consequently, NARX neural networks have been applied to solve nonlinear sequence prediction problems in many fields. The memory effect of a NARX neural network on historical data enhances its processing ability for dynamic data and improves its prediction performance for complex series. Furthermore, NARX neural networks have a stronger mirroring capability for nonlinear fitting than other neural networks and are more suitable for the analysis and prediction of time series data such as tide level data [35-37].
In this model, the gradient descent method is used to update the weight vector until the model converges to the target error. The weighted sum of all inputs determines the activation state of the neuron and is expressed as follows:

$$net = a \sum \omega_i x_i + \sum \omega_j y_j + b$$ \hfill (1)

The activation state can be described by the activation function, which can be expressed as:

$$\theta = \psi (net)$$ \hfill (2)

The activation function $\psi$ can amplify the output of neurons or limit the output to an appropriate range. In this paper, the S-type function is selected as the activation function, and the output layer is a linear function of the activation function. A typical NARX neural network consists of an output layer, an input layer, a hidden layer and output and input delays. However, the parameters of each part of the corresponding neural network should be determined before applying the network model. The basic structure is shown in Figure 7.

In this model, the two $y(t)$ terms in the structure mean the output of the neural network at the next moment (the right-hand $y(t)$) and the output of the neural network at the previous $(t-n)$ time (the left-hand $y(t)$); $W$ means the connection weight; $b$ means the threshold; and $1:2$ means the delay order, where the analog number of the next output layer refers to the number of the first two input layers, with the mathematical expression $y(t) = f(x(t-1), x(t-2))$. The NARX neural network model can be expressed as the following equation [24, 38]:

$$y(t) = f(y(t-1), y(t-2), \ldots, y(t-n_y), u(t-1), u(t-2), \ldots, u(t-n_u))$$ \hfill (3)

where $u$ is an externally determined variable. According to the formula, the value of $y(t)$ at the next moment depends on the input value $x(t)$ and the previous output $y(t)$.

Similarly, to predict the tide level, it is necessary to set the initial input parameters of the NARX neural network, including the numbers of nodes in the input layer, hidden layer and output layer and the delay order of the input and output. In this paper, considering many nonlinear factors that affect tide level data, five input parameters, namely, the wind speed, wind direction, gust speed, air temperature and air pressure, are selected to predict the tide level. Therefore, the number of input nodes is 5; the number of output nodes is 1; the number of neurons in the hidden layer is determined to be 10 according to empirical Equation (3); and the default delay order of the input and output is 1:2 (meaning that the simulation data of the next output layer refers to the data of the first two input layers).

In general, the output of a neural network is fed back as input, this method has two advantages: first, the NARX neural network is known, that is, measured tide level data are selected to predict the tide level. Therefore, the number of steps is 1; the number of output nodes is 1; the number of neurons in the next output layer refers to the data of the first two input layers.

In this model, the gradient descent method is used to update the weight vector until the model converges to the target error. The weighted sum of all inputs determines the activation state of the neuron and is expressed as follows:

$$net = a \sum \omega_i x_i + \sum \omega_j y_j + b$$ \hfill (1)

The activation state can be described by the activation function, which can be expressed as:

$$\theta = \psi (net)$$ \hfill (2)

The activation function $\psi$ can amplify the output of neurons or limit the output to an appropriate range. In this paper, the S-type function is selected as the activation function, and the output layer is a linear function of the activation function. A typical NARX neural network consists of an output layer, an input layer, a hidden layer and output and input delays. However, the parameters of each part of the corresponding neural network should be determined before applying the network model. The basic structure is shown in Figure 7.

In this model, the two $y(t)$ terms in the structure mean the output of the neural network at the next moment (the right-hand $y(t)$) and the output of the neural network at the previous $(t-n)$ time (the left-hand $y(t)$); $W$ means the connection weight; $b$ means the threshold; and $1:2$ means the delay order, where the analog number of the next output layer refers to the number of the first two input layers, with the mathematical expression $y(t) = f(x(t-1), x(t-2))$. The NARX neural network model can be expressed as the following equation [24, 38]:

$$y(t) = f(y(t-1), y(t-2), \ldots, y(t-n_y), u(t-1), u(t-2), \ldots, u(t-n_u))$$ \hfill (3)

where $u$ is an externally determined variable. According to the formula, the value of $y(t)$ at the next moment depends on the input value $x(t)$ and the previous output $y(t)$.

Similarly, to predict the tide level, it is necessary to set the initial input parameters of the NARX neural network, including the numbers of nodes in the input layer, hidden layer and output layer and the delay order of the input and output. In this paper, considering many nonlinear factors that affect tide level data, five input parameters, namely, the wind speed, wind direction, gust speed, air temperature and air pressure, are selected to predict the tide level. Therefore, the number of input nodes is 5; the number of output nodes is 1; the number of neurons in the hidden layer is determined to be 10 according to empirical Equation (3); and the default delay order of the input and output is 1:2 (meaning that the simulation data of the next output layer refers to the data of the first two input layers).
The larger the delay order is, the more data are referenced in the prediction process, and the better the prediction effect is. In this paper, the delay order is set as 1:20. The structure of the NARX neural network after setting these parameters is shown in Figure 4.

![Figure 4](image)

**FIGURE 4.** The structure of the NARX neural network.

### B. HARMONIC ANALYSIS

According to long-term observations, tides are composed of a series of harmonic vibrations (tide components). The period of tide components corresponds to the period of each component field of the tide force.

Each tide component can be expressed by the following formula:

\[ h = R \cos(q_0 + V_0 + u - K) \]  \hspace{1cm} (4)

In Equation (4), \( h \) means the height of the component, \( R \) means the amplitude of the component, \( q_0 \) means the angular rate of the component (which can be determined according to the component), \( V_0 + u \) means the phase angle of the imaginary celestial body at zero universal time at the beginning of the observation period, and \( K \) means the phase angle at which the high tide lags behind the moon culmination due to seabed friction and inertia.

The tide component expression can be further written as follows:

\[ h = f H \cos(q_0 + V_0 + u - K) \]  \hspace{1cm} (5)

In Equation (5), \( f \), \( q_0 \) and \( V_0 + u \) are all known. If the values of \( H \) and \( K \) for each component are obtained, then the tide component can be obtained. \( H \) and \( K \) are called the tide harmonic constants of tide components. If the harmonic constants are known, the tide height of each component can be acquired, and the future tide can be calculated by adding all of the components.

In this paper, the T-TIDE tool package [28] is selected to perform tide harmonic analysis. The tide level data to be analyzed are input, the data interval is set as 1 hour, and the latitudes of the tide stations are set (Table II). The starting time of the data is GMT 0000 on January 1, 2019.

The output parameters include the names of the tide components obtained by harmonic analysis, which depend on the length of the data; that is, the longer the data, the more tide components there are. The tide components selected in this paper include M2, S2, K1, O1, N2, Q1, P1, K2, Ss, and Sa. In addition, the angular rate, amplitude and amplitude error, and delay angle and delay angle error of the tide component are also included. The signal-to-noise ratio (SNR) is used in T-TIDE to determine whether the tide component is significant, as calculated by the square of the ratio of the tide component amplitude to the amplitude error. Generally, a tide component with an SNR > 2 is considered significant.

The T-TIDE tool package is run in MATLAB to obtain \( H \) and \( K \), and the final tide level predicted by harmonic analysis is obtained by inserting the obtained values into Equation (4) or Equation (5).

### C. MODULAR TIDE PREDICTION METHOD

Theoretically, tides are periodic fluctuations of sea water caused by the gravitational influences of the Moon and Sun, but other factors also affect tides, such as the air temperature, air pressure, and wind. Therefore, tide data can be divided into astronomical tides and nonastronomical tides in tide forecasts. Astronomical tides are caused mainly by the tidal forces of celestial bodies and exhibit obvious variation trends. In contrast, nonastronomical tides are affected by environmental factors and do not display regular changes, instead showing strong randomness. Therefore, there is a large difference between these two tide types, and using only a single method may not reflect the complete law of tides [39], resulting in relatively large errors.

Based on the above information, the modular/ensemble tide prediction method is added to the NARX model [40-41]. This model divides tide data into two parts: astronomical tide data affected by celestial tide-generating forces and nonastronomical tide data affected by various environmental factors. Using the characteristics of harmonic analysis, the astronomical tidal component of the tide is obtained by reconciling the results of the analysis, and then the NARX neural network is used to predict the nonastronomical tide component and modify the prediction results to improve the accuracy of tide prediction. The specific steps are depicted in Figure 5.

![Figure 5](image)

**FIGURE 5.** A block diagram of the HA-NARX prediction model.

The measured tide level data \( y(t) \) are set as the input of the modular prediction model. Because the harmonic analysis prediction model can be used for long-term tide prediction, \( y_0(t) \) is the tide prediction value obtained from the harmonic analysis module, and \( y_1(t) \) is the difference between \( y \) and \( y_0 \). Since the harmonic analysis method considers the influences of celestial bodies on tides, the difference \( y_1(t) \) between the
verified data and the data predicted by the harmonic analysis method can be regarded as the nonastronomical component of the tide level data affected by various uncertainties and nonlinear factors such as hydrometeorology. In the NARX model, according to the model structure (Figure 4), if the model outputs the prediction data \( y_2(t+1) \) at the next moment, the input terminal actually measures five kinds of meteorological data \( y_1(t), y_1(t-1) \ldots y_3 (t-N+1) \), and the prediction \( y_2(t) \) is output in the last step. The final prediction result \( Y(t+N) \) is obtained by adding the output data from the NARX model to that of the harmonic analysis module.

This table shows the improvement obtained using HA-NARX compared to simple NARX neural networks with the same testing data. The tide data of Miami Biscayne Bay (25° 43.9' N, 80° 9.7' W), USA, from June 1, 2020 GMT 0000 to July 30, 2020 GMT 2300 are selected as the test data, and the observations are sampled every 1 h, for a total of 1440 sets. Consistent with the formal prediction experiment, the first 1200 groups of data are used as training data, and the remaining 240 groups of data are used as prediction data. The accuracy indicators in the table are described in detail in Section 3.

| Prediction Models | RMSE  | CC    | NSE   | MAPE (%) |
|-------------------|-------|-------|-------|----------|
| HA-NARX           | 0.0196| 0.9951| 0.9935| 0.0074   |
| Simple NARX       | 0.0255| 0.9934| 0.8793| 0.0171   |

It can be seen from the indicators that HA-NARX has a better prediction effect than simple NARX, the accuracy and robustness of the model are improved, and the comprehensive optimization degree is calculated as approximately 23.3%. The working principle of the proposed prediction method is presented above, and the remainder of this section will briefly introduce the tide level prediction methods used for comparison.

D. GA-BP NEURAL NETWORK

The BP neural network is a multilayer feedforward neural network trained according to the error reverse propagation algorithm and is the most widely used neural network at present [29-30]. The BP neural network mainly includes two aspects: signal forward transmission and error BP. In forward transmission, the input signal is processed layer by layer from the input layer to the output layer by the hidden layer. If the output layer cannot obtain the actual output, BP is initiated to adjust the weight and threshold of the whole network according to the prediction error so that the predicted output of the BP neural network gradually approaches the actual output.

To fundamentally improve the prediction accuracy of the BP neural network, this paper adopts the genetic algorithm (GA) to optimize the neural network. The GA, which was originally proposed by Professor Holland of Michigan University [31], is a method that simulates the biological evolution mechanism in nature; that is, useful features are retained, while useless features are removed in the optimization process. When solving complex combinatorial optimization problems, compared with some conventional optimization algorithms, the GA can usually obtain better optimization results quickly.

![A structural diagram of the GA-BP model.](image)

To complete the tide level prediction, it is necessary to set the initial input parameters of the GA-BP neural network. The main contents include the number of layers of the BP neural network, the numbers of nodes in the input layer, hidden layer and output layer, and the initial parameters of the genetic optimization algorithm. First, the topological structure of the neural network model should be determined. The number of nodes in the input layer is determined by the number of input parameters. In this paper, a BP neural network is trained with the data obtained by the harmonic analysis method to predict tides; thus, the number of nodes in the input layer is taken as 1. The number of nodes in the hidden layer is the main factor affecting the performance of the BP neural network. If the selected number of hidden layer nodes is not appropriate, it is difficult for the trained network to output accurate prediction data. To solve this problem, this paper adopts Equation (6).

\[
M = \sqrt{m + n + a}
\]  

(6)

In Equation (6), \( M \) means the number of hidden layer nodes, \( m \) means the number of input layer nodes, \( n \) means the number of output layer nodes, and \( a \) is a random natural number between 0 and 10.

By solving of the empirical formula combined with the multiple test prediction method, the number of hidden layer nodes is finally determined to be 10. For the GA-BP neural network model used in this paper, the output is the predicted tide level at a certain time; thus, the number of nodes in the output layer is set as 1. For the introduced GA, four parameters need to be set in advance.
The algorithm is already times have the highest correlation. Therefore, the final set of parameters is as follows: the population size is set as 50, the number of iterations is set as 100, the crossover probability is set as 0.5, and the mutation probability is set as 0.005.

### E. WAVELET NEURAL NETWORK

The WNN was first proposed by Zhang et al. [33] in 1992 as a neural network model. The algorithm combines the advantages of the wavelet function and BP neural network. WNNs are based on the topological structure of the BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP, a BP neural network and employ a wavelet neural function to replace the sigmoid function of the hidden layer nodes for mapping. On the basis of retaining the characteristics of BP neural networks, such as signal forward propagation and error BP,
network is combined with wavelet analysis theory, giving the model numerous advantages over traditional BP neural networks, namely, fault tolerance, convergence and approximation ability [34]. The training steps and prediction algorithm flow of the WNN are as follows:

1. Network initialization. The scaling factor \( a_j \), translation factor \( b_j \) and network connection weights \( W_{ij} \) and \( W_{jk} \) of the wavelet function are initialized randomly.

2. Sample classification. The samples are divided into training samples and test samples. The training samples are used to train the network, and the test samples are used to test the prediction accuracy of the network.

3. Output value prediction. The training samples are input into the network to calculate the network prediction output and the error between the network output and the expected output.

4. Weight correction. According to the error, the network weights and wavelet basis function parameters are modified so that the network predictions are close to the expected values.

5. Whether the algorithm ends is determined. If not, the algorithm returns to step 3.

III. MATERIALS

This section presents information on the data selected for the simulation, including the time, location, sampling interval and additional meteorological data of the selected tide data. The selected measures of prediction effectiveness, including \( RMSE \), \( CC \), \( NSE \), and \( MAPE \), are also presented.

Astronomical tides can be divided into atmospheric tides and ocean tides; that is, tides are affected not only by gravity but also by the atmosphere, which is related to thermal excitation by the sun, and this atmospheric effect is more obvious at middle and low latitudes than at high latitudes. Due to differences in atmospheric conditions under the influence of different climate types, this paper selects tide stations in the eastern and western United States to obtain observation data and selects 90 days of data from these two tide stations, namely, Yorktown in the southeastern United States and San Francisco in the western United States, from GMT0000 on June 1, 2020, to GMT2300 on July 30, 2020. The East Coast of the United States is characterized by a humid subtropical monsoon climate with an average temperature above zero degrees in the coldest month in winter. The climate is hot and rainy in summer, warm and dry in winter, and has four distinct seasons. In contrast, the West Coast of the United States has a Mediterranean climate that is hot and dry in summer and mild and rainy in winter.

In addition, considering the impact of extreme weather, 90 days of data from Matagorda Bay in the southern United States affected by the Hurricane Delta from GMT0000 on August 1, 2020, to GMT2300 on September 30, 2020, are added. The tide and meteorological data used in this paper are from the website https://tidesandcurrents.noaa.gov/. The tide level is expressed in units of meters, the starting surface of the tide level is the average low tide level, and the observation interval is 1 hour. Three sets of data sets are listed, and each set contains 1440 tide levels. By listing these data, the tide level time series of this period can be obtained.

On the other hand, this paper also collects meteorological data corresponding to each tide level, including temperature, pressure, wind direction, wind speed and gust speed data.

In the neural network model, one-step prediction is carried out first. This paper takes the first 1200 groups of data as training data and the remaining 240 groups of data as prediction data. Due to the large amount of data required by the harmonic analysis method to predict tide levels, this paper also collects historical data from three tide stations from January 2019 to June 2020, which satisfies the required amount of historical data (more than 18 months) required for the harmonic components selected in this paper. The tide gauge information and tide level data selected in this paper are described below.

| Region | Station | Latitude | Longitude | Station ID |
|--------|---------|----------|-----------|-----------|
|        |         |          |           |           |

Before using the WNN to forecast tides, it is necessary to set various parameters. The input data structure used in this paper is 2-10-1: the input layer has two nodes, which represent the tide data of the first two time points of the prediction time node; the hidden layer has 10 nodes, which are determined by an empirical formula; and the output layer has one node, which is the tide level predicted by the WNN. The network weight and wavelet basis function are randomly obtained during the parameter initialization step. The WNN is trained 100 times. Then, the trained WNN is used to predict the tide data at the next moment.
The relationship between these two variables is
\begin{equation}
\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{n} (y_k - \tilde{y}_k)^2}{n}}
\end{equation}

where \( n \) is the number of samples or interpreted as a time index in time series analysis; \( y_k \) and \( \tilde{y}_k \) represent the observed values and the average of the observed values, respectively; and \( \hat{y}_k \) and \( \tilde{y}_k \) represent the predicted values and the average of the predicted values, respectively.

The \( \text{RMSE} \) can effectively reflect the measurement precision. In contrast, the \( CC \) is a statistical index reflecting the closeness of the degree of correlation between variables and is calculated according to the product difference method. Based on the deviations between two variables and their respective averages, the degree of correlation between these two variables is reflected by multiplying the two deviations. Generally, a \( CC \) above 0.7 indicates that the relationship is very close; 0.4~0.7 indicates a close relationship; and 0.2~0.4 shows that the relationship is general. The smaller the \( \text{RMSE} \) is, the larger the \( CC \), indicating a better prediction effect.

The \( \text{NSE} \) is generally used to verify the goodness of hydrological model prediction results. The \( \text{NSE} \) takes a value from negative infinity to 1. If the \( \text{NSE} \) is close to 1, the model is of good quality and credible; if the \( \text{NSE} \) is close to 0, the simulation results are close to the mean level of the observed values (i.e., the overall results are credible, but the process simulation error is large); and if the \( \text{NSE} \) is much less than 0, the model is not credible.

The \( \text{MAPE} \) can be used to measure the goodness of a model's prediction results, considering not only the error between the predicted value and the true value but also the ratio between the error and the true value in the range of \([0, +\infty)\). A \( \text{MAPE} \) of 0% indicates a perfect model, and a \( \text{MAPE} \) greater than 100% indicates a poor model. However, there is a drawback: the \( \text{MAPE} \) is asymmetric and imposes a greater penalty for negative errors (when the predicted value is higher than the actual value) than for positive errors. Therefore, the \( \text{MAPE} \) will favor models that underpredict rather than overpredict.

### IV. SIMULATIONS

In this section, Figures 11, 12, 13 and 14 show the prediction results of each of the four prediction models under different atmospheric effects, and Table IV compares the error extremes of each prediction.

#### A. HARMONIC ANALYSIS

|                | San Francisco | Yorktown | Virginia | Texas |
|----------------|---------------|----------|----------|-------|
| California     | 37°           | 122°     | 48.4N    | 13.6N |
| Virginia       | 37°           | 76°28.7W | 13.6°    | 42.6N |
| Texas City     | 28°           | 95°54.8W | 28°      | 37°   |

\[ CC = \frac{\sum_{k=1}^{n} (y_k - \tilde{y}_k) (\hat{y}_k - \tilde{y}_k)}{\sqrt{\sum_{k=1}^{n} (y_k - \tilde{y}_k)^2 \sum_{k=1}^{n} (\hat{y}_k - \tilde{y}_k)^2}} \]

\[ NSE = 1 - \frac{\sum_{k=1}^{n} (y_k - \hat{y}_k)^2}{\sum_{k=1}^{n} (y_k - \tilde{y}_k)^2} \]

\[ MAPE = \frac{100\%}{n} \sum_{k=1}^{n} \left| \frac{\hat{y}_k - y_k}{y_k} \right| \]
The tide levels predicted by the harmonic analysis method and the observed tide levels are compared in Figure 11.

![Figure 11](image1.png)

The tide levels predicted by the GA-BP neural network and the observed tide levels are compared in Figure 12.

![Figure 12](image2.png)

C. WAVELET NEURAL NETWORK
The tide levels predicted by the WNN and the observed tide levels are compared in Figure 13.

![Figure 13](image3.png)
D. HA-NARX NEURAL NETWORK
The predicted tide levels and the observed tide levels after training the HA-NARX neural network are compared in Figure 14.

V. ANALYSIS
In this section, Figures 15, 16, and 17 show the prediction results and comparisons of the error curves for the four tide level prediction methods at the three tide stations. Table V shows a comparison of the calculated CC, NSE, RMSE, and MAPE values of the four models for convenience.
At the San Francisco tide station, which is characterized by a Mediterranean climate, the data were acquired in summer (June and July), just after the rainy season had ended. During this period, under the control of the subtropical high, air masses sink, the climate is hot and dry with little rain, there are few clouds and sufficient sunshine, and the climate is relatively stable. Hence, the influences of nonlinear climate factors are weak, and the most important factor affecting tides is the gravitational force between celestial bodies. Moreover, the observation data (Figure 8) demonstrate that the measured curve exhibits periodic oscillation; therefore, among the four models, the HA-NARX neural network provides the most accurate prediction results.

In summary, for tide data that are minimally affected by atmospheric conditions, the advantage of the HA-NARX neural network is not obvious.

The data from the Yorktown tide station influenced by a humid subtropical monsoon climate were similarly acquired during summer (June and July). During this period, which coincides with the annual rainy season, the temperature and precipitation change violently. Excessive precipitation leads to a high actual local tide level, causing the traditional adjustment and analysis method to provide inaccurate predictions. This phenomenon continues until the end of the rainy season, and thus, the error in the former part of the forecast data is obviously larger than that in the latter part.

Likewise, the prediction results of the GA-BP neural network selected in this paper can also be divided into two parts. In the former part, the influences of nonlinear climate factors are fully considered, thereby increasing the prediction accuracy. However, in the latter part, the influences of climate factors on the tide level basically disappear, but the GA-BP neural network does not make timely adjustments. As a result, the predicted levels are higher than the actual measured levels. This is because the input data of the GA-BP neural network are based on the prediction data of the harmonic analysis method. Although the error of the harmonic analysis method is reduced to a certain extent, due to the limitation of the neural network structure, the GA-BP neural network cannot adjust the prediction output in time to change the data. Hence, the GA-BP neural network improves the overall prediction accuracy only in comparison with the traditional harmonic analysis method and does not completely overcome the influences of climate factors.

However, in the WNN prediction results, the prediction curve basically conforms to the measured values, and the error distribution is uniform. This is because the hidden layer of the WNN uses a wavelet neural function, which has a better mapping performance for tide data, and the prediction accuracy is improved compared with that of the traditional harmonic analysis method and GA-BP neural network. Therefore, the WNN can basically overcome the influences of nonlinear climate factors.

Finally, compared with all previous models, the HA-NARX neural network exhibits the best prediction effect. As a kind of dynamic neural network, it can remember the tide level at a previous time and apply it to future predictions. Consequently, the HA-NARX neural network can handle complex climate
change and make accurate predictions of tide levels. Furthermore, the tide levels predicted by the HA-NARX neural network are consistent with the measured values, and the error distribution is uniform and smaller than that of the WNN. Therefore, the HA-NARX neural network not only overcomes the influences of nonlinear climate factors but also further reduces the error based on the WNN to improve the prediction accuracy.

To highlight the influences of atmospheric factors on tide level data, this paper introduces tide data measured during an extreme weather event in August and September at the Matagorda tide station. Hurricane Delta originated in the North Atlantic Ocean and landed near Creole, Louisiana, on October 9, 2020, approximately 310 km from the tide gauge station. The measured data were greatly affected by the hurricane and exhibited obvious instability that was significantly different from the tide level data affected periodically by climatic factors. The former part of the data is relatively stable and exhibits periodicity, whereas the latter part displays an obvious increase in the tide level, increasing the difficulty of accurately predicting the tide level.

The prediction data selected in this paper can be divided into a water increase component and a periodic change component. Figure 17 demonstrates that the prediction results of the harmonic analysis method are completely inconsistent with the actual situation. The overall predicted tide levels are low, with periodic fluctuations. Therefore, the trend found by the harmonic analysis method from past data is not suitable for predicting tide levels during extreme weather. Furthermore, the GA-BP neural network is trained and predicted on the basis of the harmonic analysis data, which reduces part of the error, but the overall output is still periodic due to the abovementioned limitation of the harmonic analysis method. Moreover, the prediction data from the WNN are close to the real values, and the increasing trend of the tide level is predicted well, but the predicted peak value is not accurate. Ultimately, the HA-NARX neural network prediction results are most in line with the real situation, and accurate results are obtained for both the water increasing component and the periodic change component because the HA-NARX neural network considers the tidal forces of celestial bodies and the influences of atmospheric factors.

![FIGURE 17. Prediction results and error comparison for the Matagorda tide station.](image)

| Station ID | Method         | RMSE  | CC     | NSE   | MAPE (%) |
|-----------|----------------|-------|--------|-------|----------|
| 9414290   | Harmonic Analysis | 0.057 | 0.99893| 0.9899| 0.0235   |
|           | GA-BP           | 0.025 | 0.99899| 0.9980| 0.0064   |
|           | WNN             | 0.055 | 0.99586| 0.9906| 0.0406   |
|           | HA-NARX         | 0.021 | 0.94103| 0.9987| 0.0102   |
| 8637689   | Harmonic Analysis | 0.141 | 0.98068| 0.9381| 0.0647   |
|           | GA-BP           | 0.030 | 0.97386| 0.9723| 0.0078   |
|           | WNN             | 0.027 | 0.99609| 0.9978| 0.0051   |
|           | HA-NARX         | 0.019 | 0.98597| 0.9989| 0.0030   |
| 8773146   | Harmonic Analysis | 0.617 | 0.12671| -0.8277| 0.2538   |
|           | GA-BP           | 0.442 | 0.17876| 0.0627| 0.1679   |
|           | WNN             | 0.204 | 0.64527| 0.8001| 0.0137   |
|           | HA-NARX         | 0.022 | 0.99323| 0.9976| 0.0108   |

By comparing the quantifiers of prediction results at the three tide stations, it can be calculated that the greater the influence of meteorological factors on the tide, the more obvious the improvement in the accuracy and stability of HA-NARX prediction results compared with the traditional model. The comprehensive improvement in prediction accuracy is 23.87% and 40.03% in San Francisco and Yorktown, respectively, where astronomical tides dominate, while in Matagorda City,
which is affected by extreme weather conditions, i.e., nonastronomical tides dominate, the prediction accuracy is improved by 234%.

It is undeniable that in San Francisco, all three conventional models outperform HA-NARX in the comparison of $CC$ values; however, when evaluating the prediction results in terms of $CC$, 0.7 is used as the cutoff point. The $CC$ values of all four models exceed 0.9, indicating that the prediction results are very closely related to the actual observations; thus, they are calculated at the same level in this index. Similarly, in Yorktown, WNN outperforms HA-NARX in the $CC$ comparison and was also calculated at the same level. However, GA-BP outperforms HA-NARX in the $MAPE$ comparison. In balance, GA-BP outperforms HA-NARX at the San Francisco tide station.

In Matagorda City, for the harmonic analysis method with $CC < 0.2$ and $NSE < 0$, the prediction made can no longer be used as a reference. The positive $NSE$ of GA-BP is close to 0, indicating a good overall prediction trend but a large error. For WNN, the combined improvement of HA-NARX in prediction results is calculated to be 47%. In summary, HA-NARX shows good prediction accuracy and stability and good generalizability under any conditions.

### VI. CONCLUSIONS

A modular tide prediction model based on a NARX network method is proposed and developed in this paper. Considering the influences of atmospheric factors on the atmospheric tidal components of astronomical tides, observation data from three tide stations with different atmospheric conditions are selected and input into different prediction models to predict tide levels. The influences of atmospheric factors on the three tide stations gradually increase. The results show that the proposed HA-NARX model has robustness as well as high accuracy, and the majority of error values and the differences between observed and predicted tide levels fall within the range of $-6.1$ cm and $+6.7$ cm. The prediction accuracy is improved by 20% to 40% compared with the traditional method, the prediction data has high correlation with and low dispersion relative to observation data, the error is stable, and the effect is more prominent under extreme weather conditions. The prediction accuracy is comprehensively improved by 234% compared with the traditional method. In addition, the proposed method has the advantages of a simple structure, short runtime and short calculation time compared with the traditional method. It is worth noting that there is a drawback of the HA-NARX model is that it requires a large amount of real-time weather data, which requires multiple observation devices to work simultaneously, and the cost of forecasting is thereby increased. Additionally, the effectiveness of multistep forecasting by HA-NARX needs to be further investigated.

### REFERENCES

[1]. B. Chen, H. Wang, H and C. Chu, Wavelet and artificial neural network analyses of tide forecasting and supplement of tides around Taiwan and South China sea. OCEAN ENG, vol. 34, pp. 2161-2175, 2007. DOI: 10.1016/j.oceaneng.2007.04.003.

[2]. A. Riazi, Accurate tide level estimation: A deep learning approach. OCEAN ENG, vol. 198, 2020. DOI: 10.1016/j.oceaneng.2020.107013.

[3]. Z. Chen, et al., A. j. v model for the analysis and prediction of tides. ACTA OCEANOLOGICA SIN., vol. 09, no. 04, pp. 475-486, 1990.

[4]. W. Wang, H. Yuan, and X. Wang. A tidal level prediction approach based on BP neural network and cubic B-Spline curve with Knot Insertion Algorithm. MATH PROBL ENG, 2018. DOI: 10.1155/2018/9855079.

[5]. G. Zhu, Research on short-term tide forecast based on Bi-LSTM recurrent neural network. JIOSSERS, 03(04), 19-29, 2020. DOI: 10.6918/JIOSSERS.202004.3(4).0004.

[6]. T. Lee, Back-propagation neural network for long-term tidal predictions. OCEAN ENG, vol. 31, no. 02, pp. 225-238, 2004. DOI: 10.1016/S0029-8018(03)00115-X.
WENHAO WU received a B.Eng. degree in nautical science and maritime science and technology from Dalian Maritime University (DMU), Dalian, China, in 2016. He is currently pursuing a M.Eng. degree in traffic information engineering and control. His research interests include neural network algorithms and tide prediction.
LIANBO LI received a B.Eng. degree in nautical science and a M.Eng. degree in traffic information technology and control from Dalian Maritime University (DMU), Dalian, China, in 2008 and 2010, respectively. He also received a master's degree in maritime safety and environment management from World Maritime University (WMU), Malmö, Sweden, in 2020. He is currently an Associate Professor and Master Supervisor with DMU. His current research interests include unmanned ship autonomous collision avoidance, ship collision risk and navigation risk in complex waters.

JIANCHUAN YIN (M’10) received a Ph.D. in traffic information engineering from Dalian Maritime University (DMU), Dalian, China, in 2007. He is an Associate Professor at the Navigation College, DMU. He is currently a Visiting Scholar at the Department of Naval Architecture and Marine Engineering, University of Michigan, Ann Arbor, MI, USA. His current research interests include computational intelligence and ocean engineering.

WENJUN ZHANG received B.Eng., M.Eng., and Ph.D. degrees from Dalian Maritime University (DMU), Dalian, China, in 2000, 2007, and 2014, respectively. He is currently an Associate Professor and Dean of the Navigation College at DMU. His current research interests include ship motion prediction, marine wireless communication, intelligent transportation systems, and global maritime distress and safety systems.

WENYU LYU received a Ph.D. from Korea Maritime & Ocean University (KMOU), Busan, Korea, in 2017. She also received a master's degree in maritime safety and environment management from World Maritime University (WMU), Malmö, Sweden, in 2019. Her current research interests include traffic planning, ESP and corpus linguistics.