

Multiscale attention-based LSTM for ship motion prediction

Tao Zhang ^{a,b,*}, Xiao-Qing Zheng ^{a,b}, Ming-Xin Liu ^c

^a School of Information Science and Engineering, Yan Shan University, Qinhuangdao, 066004, China

^b Hebei Key Laboratory of Information Transmission and Signal Processing, Qinhuangdao, 066004, China

^c College of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang, 524088, China

ARTICLE INFO

Keywords:

Ship motion prediction
Long short-term memory
Multiscale
Attention mechanism
Two-stage training mechanism

ABSTRACT

Ship motion prediction is applied to the shipboard stabilized platform to keep the equipment on the platform stable all the time, which is of great practical significance to the safety and efficiency of shipboard equipment operation. Long Short-term Memory (LSTM) Network is a classic time series prediction method that has made remarkable achievements in this field. However, the dynamic frequency range of single LSTM in ship motion prediction is insufficient to meet the stabilized platform with higher precision requirements. To improve the performance of LSTM in ship motion prediction, this paper presents a novel method named as multiscale attention-based LSTM. At first, wavelet transform is employed to decompose ship motion signals into several frequency scales, which makes LSTM to capture the inherent law of ship motion from each frequency scale. And then the weights of different scales are obtained by attention mechanism, which promote the sensitivity of the whole system by paying more attention to significant information and suppress the interference of noise signals. Both of the steps form a multiscale attention mechanism, which promote the adaptability and improve the performance of the LSTM. In addition, to avoid being trapped in local optimization, the two-stage training mechanism is designed for model training based on the model structure. Ship motion data are used to evaluate the feasibility and effectiveness. The experiments show that the proposed method achieves better performance compared with other popular methods.

1. Introduction

The motions of six degrees of freedom (DOF), such as roll, pitch, yaw, sway, surge and heave, are generated when the ship is sailing on the sea, which continuously disturbs the shipboard equipment and reduces the security and work efficiency of the operation on the sea (Cheng et al., 2019). Ship motion prediction is one of the measurements for reducing the risk by predicting the future motion trend of the ship according to the historical motion data. It can provide the motion trend for shipboard stabilized platform in advance to keep the equipment on the platform stable all the time by adjusting the attitude of the platform (Takami et al., 2021). Therefore, the research on ship motion prediction is of great practical significance.

In the traditional ship motion prediction method, the linear analysis methods such as autoregression (AR) model and autoregressive moving average (ARMA) model based on time series analysis are fast in modeling and prediction (Huang et al., 2014; Jiang et al., 2020), but the prediction accuracy of linear model cannot meet the requirements. With the rapid development of computer technology, the machine learning

method has been employed to predict the time series of ship motion (Hou and Zou, 2015). Li et al. proposed a model based on support vector regression (SVR) to improve the prediction accuracy of ship motion (Li et al., 2016). Moreover, neural network has been also applied to ship motion prediction because of its ability to approximate various nonlinear mappings (Li et al., 2017; Yin et al., 2013, 2014). Huang et al. proposed a wavelet neural network for online ship roll motion prediction (Huang et al., 2018). Yao et al. used long short-term memory (LSTM) network to predict the ship motion with high accuracy (Yao et al., 2018). Yin et al. constructed a variable radial basis function (RBF) network based on sliding data window to predict ship roll dynamics online (Yin et al., 2018b). Among them, the recurrent neural network represented by LSTM has made remarkable achievements in time series prediction fields due to its temporal structure with memory capacity (Karevan and Suykens, 2020; Li et al., 2019; Yang et al., 2019).

However, ship motion is affected by many factors, such as navigation state, sea conditions, which makes its dynamics more complex (Bulian and Francescutto, 2011; Fossen, 2011; Vidicperunovic and Jensen, 2009). The entanglement of various factors with different characteristics

* Corresponding author. College of Information Science and Engineering, Yan Shan University, Qinhuangdao, 066004, China.

E-mail address: zhtao@ysu.edu.cn (T. Zhang).

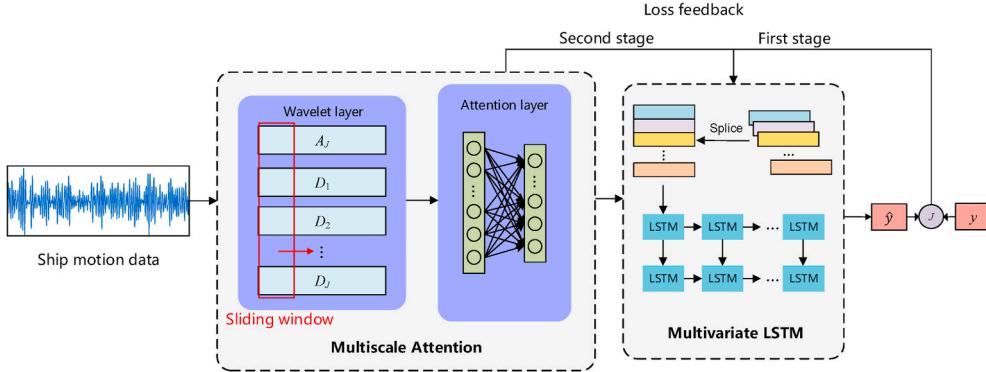


Fig. 1. MSA-LSTM model architecture.

increases the difficulty of ship motion prediction (Li et al., 2016). Even though LSTM has good performance for time series problems, it still has the problem of frequency dynamic range for complex ship motion, which makes it unable to meet the accuracy requirements of stabilized platform (Wang, 2020). Single neural network model is not enough to capture all the important features and it is very difficult to accurately predict the ship motion (Duan et al., 2016). The hybrid prediction methods combining pre-processing techniques and neural networks have become more effective (Cao et al., 2019; Hong et al., 2019).

Wavelet analysis is a multiscale analysis method in the time domain and frequency domain (Mallat, 2009; Yu et al., 2006). It can be applied to decompose the original signals into sub-signals at different frequency scales, and the sub-signals can be analyzed to obtain the detailed information of the original signals (Chang et al., 2019; Khare and Bajaj, 2020). Yin et al. put forward a prediction method combining wavelet transform and RBF network for ship roll prediction (Yin et al., 2018a). This method predicts each decomposed component and summarizes the prediction results to get the final prediction, which improves the performance. Among other decomposition prediction methods, each component after decomposition is usually modeled and predicted separately, and the final prediction results are obtained through reconstruction (Chang et al., 2019; Liu et al., 2014, 2017, 2020a). However, it is difficult to accurately predict each component, especially the high-frequency component after the wavelet decomposition, which leads to the accumulation of prediction error of each component and the increase of prediction output error. Moreover, we think that the multiple frequency information of ship motion data contains some information that may interfere with the prediction, which affects the prediction accuracy.

In addition, neural networks based on attention mechanism have been well applied in natural language processing in recent years (Choi et al., 2018; Usama et al., 2020; Vaswani et al., 2017; Yin et al., 2016). By assigning different probability weights to the hidden layer units of the neural network, the hidden layer can focus on the more critical information. With the wide application of attention mechanism, it achieved good performance in the field of time series prediction too (Du et al., 2020; Hu and Zheng, 2020; Liu et al., 2020b). The combination of wavelet analysis and attention mechanism in ship motion prediction is a method to solve the frequency disturbance.

In this paper, a multiscale attention-based LSTM (MSA-LSTM) model is proposed for the shipboard stabilized platform and the two-stage training mechanism is designed according to the proposed model structure to train the model, so as to avoid the model falling into local optimality and improve the convergence rate. In this proposed model, wavelet decomposition is adopted to decompose ship motion into different frequency scales to reduce the complexity, which is helpful for LSTM to learn the inherent laws of ship motion. On the basis of wavelet multiscale analysis, attention is introduced to form the multiscale attention mechanism, which enables the LSTM network to pay more

attention to significant information for prediction, and suppress the noise interference of useless information. These modules are correlated and combined to improve the performance of ship motion prediction and make the proposed MSA-LSTM model as a novel method for high precision shipboard stabilized platform. In order to verify the feasibility and effectiveness of the proposed model, the MSA-LSTM model is used to the ship motion prediction experiment by using the ship motion data collected under different sea conditions. The results show that the MSA-LSTM model can adapt to the prediction of ship motion under different sea conditions, and the prediction accuracy is better than other models.

The rest of this paper is arranged as follows. Section 2 introduces the architecture and theoretical formula of the MSA-LSTM model. The optimization strategy and the two-stage training mechanism is presented in section 3. In Section 4, the results of ship motion prediction experiment and the analysis of prediction results are presented. Section 5 summarizes the results of this research.

2. Methodology

In this section, each module of the proposed MSA-LSTM model is described in detail and then the optimization strategy and training mechanism are introduced. The model architecture is shown in Fig. 1. It includes two main modules: Multiscale attention and multivariate LSTM. Firstly, wavelet transform decomposes the ship motion data to different scales, which reduces the complexity of input data. Secondly, The Multiscale attention is designed to adaptively adjust weights of different scale components. Finally, the multivariate LSTM prediction network structure is constructed to predict the final predicted results. In addition, the two-stage training mechanism designed according to the model structure is applied to the model training.

2.1. Problem statement

In this paper, we take each dimension of the six degrees of freedom ship motion as the input of the MSA-LSTM model. Separate models are built for each dimension for training and prediction. Given one of the dimensions of the ship motion signal $X = (x_1, x_2, \dots, x_T)$, we aim at predicting a series of future ship motion data in the method of sliding window. Therefore, we use the historical ship motion data $X_t = (x_{t-d+1}, x_{t-d+2}, \dots, x_t) \subseteq X$ at current time stamp t to predict x_{t+q} , where d is the tunable window size and q is the tunable interval for prediction. Moreover, we denote the corresponding predicted value as \hat{y}_{t+q} and the target value as y_{t+q} . By constantly moving the sliding window to predict the value after interval q , we can obtain a series of future ship motion data. In this paper, according to the response time of the stabilized platform and the time of model prediction, we set the prediction interval q as 1s and set the step size of the sliding window to 1. We learn a nonlinear mapping function by using the historical ship motion data X_t and its corresponding target value y_{t+q} to obtain the predicted value \hat{y}_{t+q} .

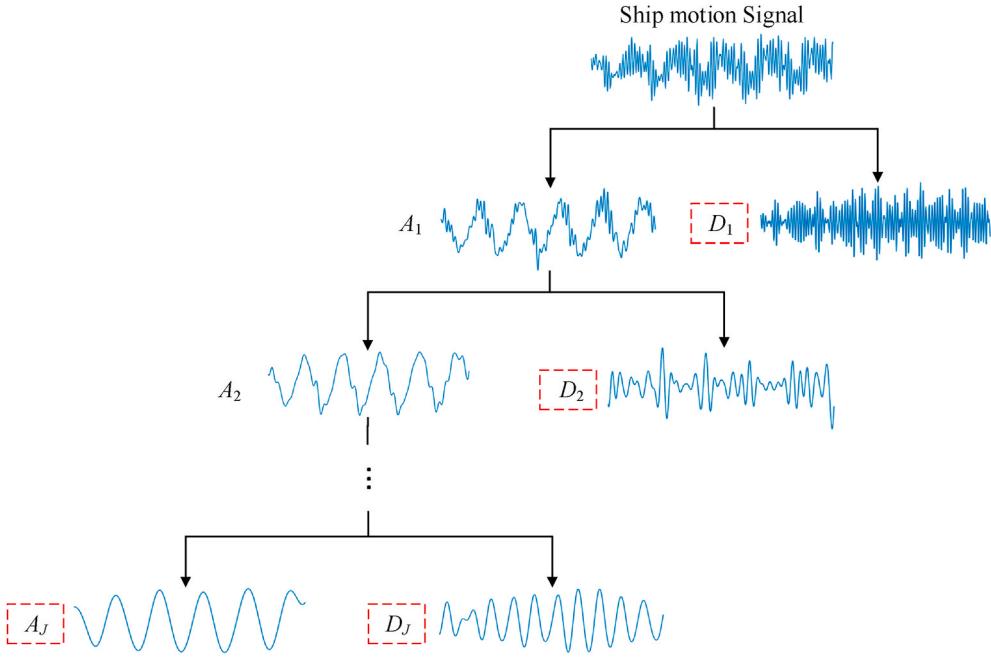


Fig. 2. Wavelet decomposition schematic diagram.

with the following formulation:

$$\hat{y}_{t+q} = f(X_t, y_{t+q}; w, b) \quad (1)$$

where w and b represent all the trainable weights and biases in our proposed model, respectively. $f(\cdot)$ is the nonlinear mapping function to fit all the ship motion data by learning the appropriate w and b .

2.2. Model architecture

2.2.1. Wavelet transform

The first part of MSA-LSTM model is multiscale analysis of ship motion signals with wavelet transform. The input is one of the six dimensions of ship motion, which aims to decompose the ship motion signal into different frequency scales. By learning the signals of different frequency scales, it is easier for our proposed model to grasp the inherent laws. Wavelet analysis decomposes the input ship motion signal X into approximate component and detail component with the time domain information preserved. The wavelet decomposition can be calculated as:

$$\begin{cases} A_n^1 = \sum_k H_{n-2k} X_k; D_n^1 = \sum_k G_{n-2k} X_k, j = 1 \\ A_n^j = \sum_k H_{n-2k} A_k^{j-1}; D_n^j = \sum_k G_{n-2k} A_k^{j-1}, j = (2, 3, \dots, J) \end{cases} \quad (2)$$

where J presents the number of layers of decomposition; H is the low-pass wavelet filter coefficient and G is the high-pass wavelet filter coefficient; $A^j \in R^{1 \times T}$ is the wavelet coefficient of the approximate component of the signal at layer j and $D^j \in R^{1 \times T}$ is the wavelet coefficient of the detail component. Fig. 2 shows the effect of wavelet decomposition of partial ship motion data. The final output after wavelet decomposition is expressed as:

$$\tilde{X} = \begin{bmatrix} A_1^J & A_2^J & \dots & A_T^J \\ D_1^1 & D_2^1 & \dots & D_T^1 \\ \vdots & \ddots & \ddots & \vdots \\ D_1^J & D_2^J & \dots & D_T^J \end{bmatrix} \quad (3)$$

2.2.2. Multiscale attention

Ship motion signal contains a variety of frequency information. Under different sea conditions, different frequency scales are of different importance to the prediction, some of which may cause interference. To solve this issue, we introduce an attention mechanism on the basis of multiscale analysis to make the proposed model adaptive to focus on more critical frequency information. Given the multiscale sequence \tilde{X}_t and the original sequence X_t in the sliding window at time stamp t , the following attention mechanism is used:

$$s(\tilde{X}_t^i, X_t) = v_s^\top \tanh(W_{sc}\tilde{X}_t^i + W_{so}X_t) \quad (4)$$

$$\alpha_t^i = \frac{\exp[s(\tilde{X}_t^i, X_t)]}{\sum_{j=1}^{J+1} \exp[s(\tilde{X}_t^j, X_t)]} \quad (5)$$

where $s(\cdot)$ represents scoring function that evaluates the relevance between the multiscale components and the original sequence. $\tilde{X}_t^i \in R^d$ is the i^{th} row of $\tilde{X}_t \in R^{(J+1) \times d}$. $v_s \in R^d$, W_{sc} and $W_{so} \in R^{h \times d}$ are trainable parameters, h represents the size of the hidden layer and $\tanh(\cdot)$ is the activation function. The attention weights are determined by the multiscale sequence and the original sequence, which indicate the importance of each frequency scale for ship motion prediction. Since any scale in the sliding window under the current time stamp has its corresponding weight, the output of the multiscale attention layer is defined as follows:

$$X'_t = \begin{bmatrix} \alpha_t^1 A_{t-d+1}^J & \alpha_t^1 A_{t-d+2}^J & \dots & \alpha_t^1 A_t^J \\ \alpha_t^2 D_{t-d+1}^1 & \alpha_t^2 D_{t-d+2}^1 & \dots & \alpha_t^2 D_t^1 \\ \vdots & \ddots & \ddots & \vdots \\ \alpha_t^{J+1} D_{t-d+1}^J & \alpha_t^{J+1} D_{t-d+2}^J & \dots & \alpha_t^{J+1} D_t^J \end{bmatrix} = (x'_{t-d+1}, x'_{t-d+2}, \dots, x'_t) \quad (6)$$

2.2.3. Prediction with multivariate LSTM

In the traditional decomposition prediction method, each decomposed component is predicted by a separate model, the prediction results

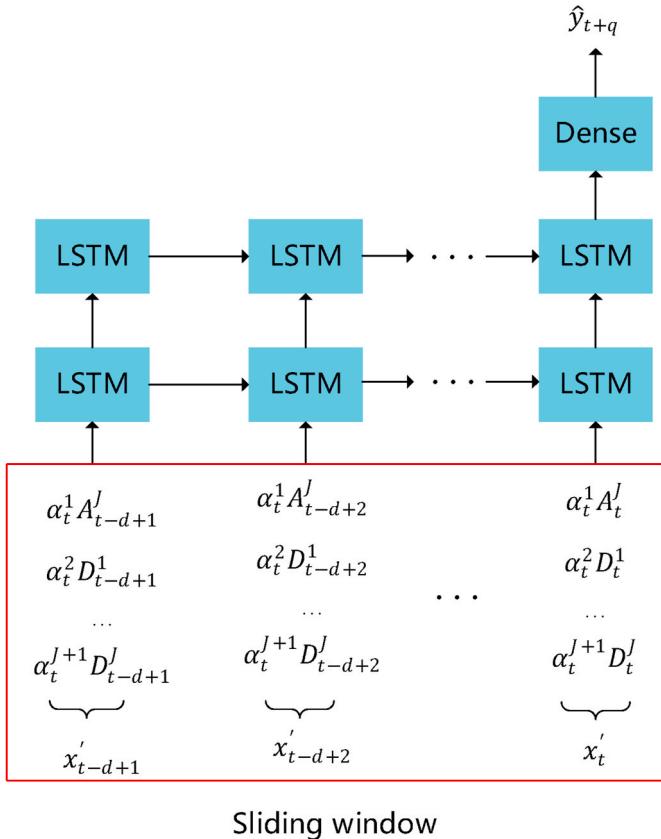


Fig. 3. Multivariate LSTM layer structure.

of each model will produce errors, and then the prediction results of each component are reconstructed to obtain the final prediction results, which will result in the accumulation of errors. In this module, we splicing all the components after multiscale attention into a matrix and input it into LSTM layer for training and prediction. In this way, the cumulative prediction errors are eliminated, and the cost of model parameter adjustment for each decomposed component are avoided.

We use a two-layer LSTM network for prediction, so as to achieve a high prediction accuracy within a reasonable training time. Given a multiscale weighted ship motion signal at time stamp t , $X'_t = (x'_{t-d+1}, x'_{t-d+2}, \dots, x'_t)$ where $x'_t \in R^{J+1}$ represents the input for any time step in the sliding window. The formulas describing an LSTM cell at time step τ are presented as:

$$i_\tau = \sigma(W_{xi} [\alpha_i^1 A_\tau^J; \alpha_i^2 D_\tau^1; \dots; \alpha_i^{J+1} D_\tau^J] + W_{hi} h_{\tau-1} + b_i) \quad (7)$$

$$f_\tau = \sigma(W_{xf} [\alpha_i^1 A_\tau^J; \alpha_i^2 D_\tau^1; \dots; \alpha_i^{J+1} D_\tau^J] + W_{hf} h_{\tau-1} + b_f) \quad (8)$$

$$c_\tau = f_\tau \odot c_{\tau-1} + i_\tau \odot \tanh(W_{xc} [\alpha_i^1 A_\tau^J; \alpha_i^2 D_\tau^1; \dots; \alpha_i^{J+1} D_\tau^J] + W_{hc} h_{\tau-1} + b_c) \quad (9)$$

$$o_\tau = \sigma(W_{xo} [\alpha_i^1 A_\tau^J; \alpha_i^2 D_\tau^1; \dots; \alpha_i^{J+1} D_\tau^J] + W_{ho} h_{\tau-1} + b_o) \quad (10)$$

$$h_\tau = o_\tau \odot \tanh(c_\tau) \quad (11)$$

where i_τ represents input gate state, f_τ is forgetting gate state, c_τ is cell state, o_τ is output gate state, $h_{\tau-1}$ and h_τ are the hidden layer of the previous and current time step. In addition, W and b with subscript are trainable parameters, \odot is the element-wise product. The formulas describing two-layer LSTM at time step t are presented as:

$$H_t = LSTM^{(2)} [(h_{t-d}, x'_{t-d+1}), (h_{t-d+1}, x'_{t-d+2}), \dots, (h_{t-1}, x'_t)] \quad (12)$$

where $H_t = (h_{t-d+1}, h_{t-d+2}, \dots, h_t) \in R^{r \times d}$ is the output hidden layer of the two-layer LSTM, r represents the size of LSTM hidden layer. We take the

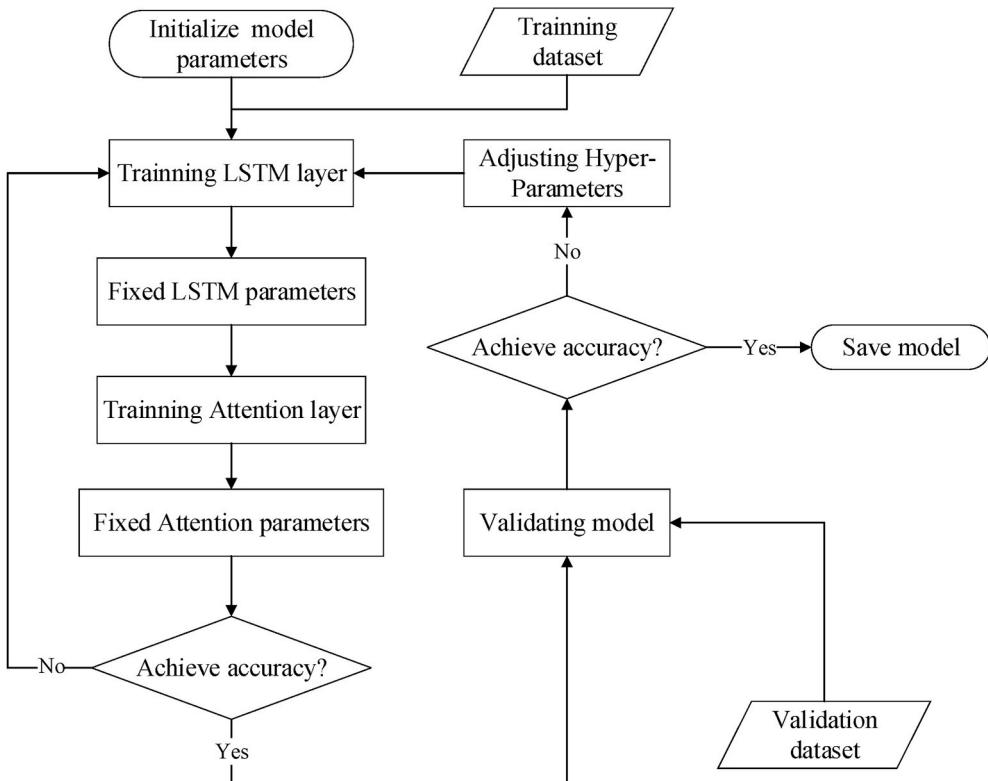


Fig. 4. Flow chart of training MSA-LSTM model.

Table 1

Specific ship parameters information.

Parameters	Overall length (m)	Length between perpendiculars (m)	Molded breadth (m)	Molded depth(m)	Designed draft(m)	Displacement (t)
Values	216.0	202.0	28.6	9.8	6.4	18174.9

Table 2

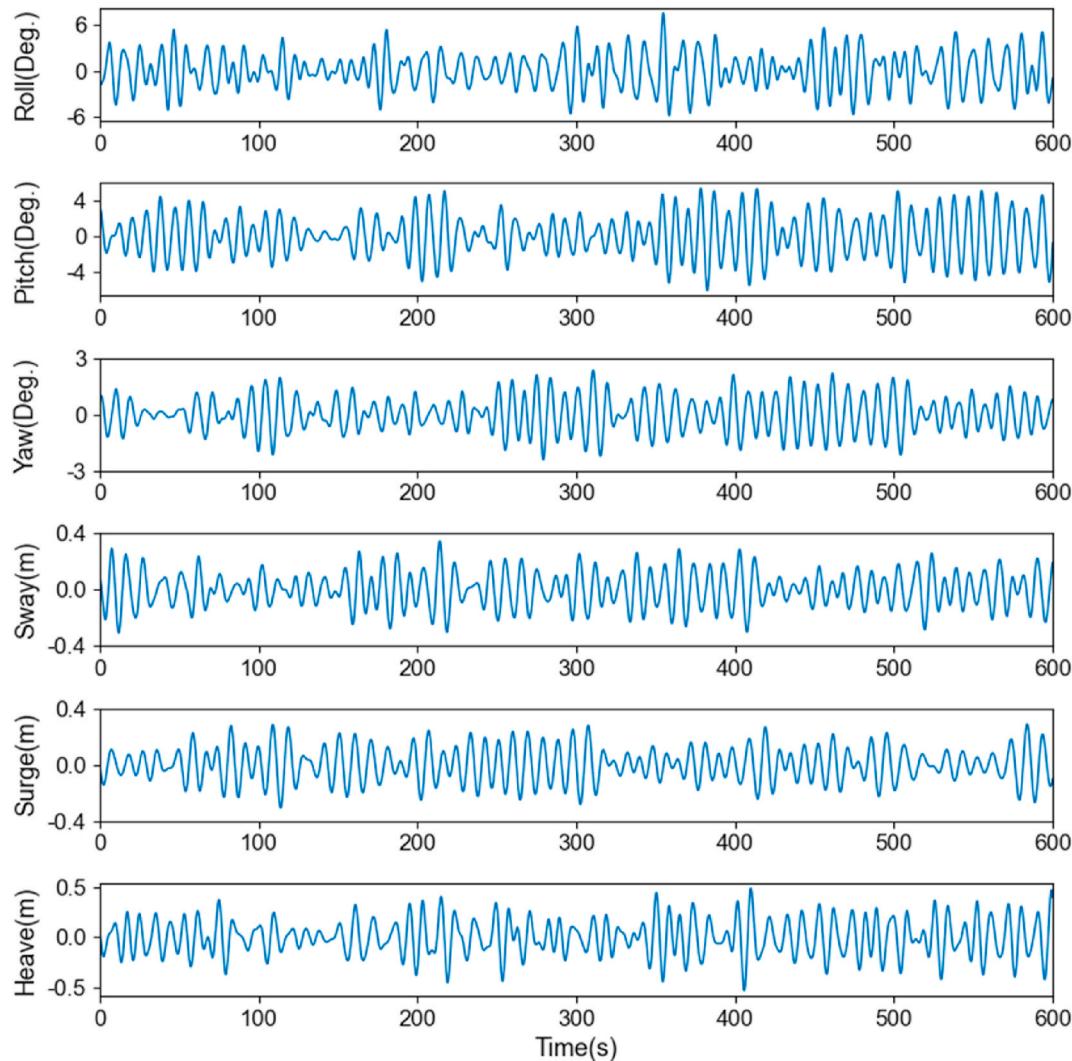
Conditions and environment of data collection.

Datasets	Speed (kn)	Wave direction (deg.)	wave height (m)	Sampling frequency (Hz)	Sampling points
Dataset I	15	135	1.5	10	36024
Dataset II	10	150	2	10	24152

hidden layer of the last time step in the sliding window as the output of the LSTM layer at time stamp t . Then the final output of the MSA-LSTM model can be computed as:

$$\hat{y}_{t+q} = \tanh(W_p h_t + b_p) \quad (13)$$

where $\hat{y}_{t+q} \in R$ is the final predicted value, $W_p \in R^{1 \times r}$ and $b \in R$ are trainable parameters, q is the adjustable interval for prediction. All the predicted value can be contacted a vector $\hat{Y} \in R^M$, M represents the number of testing set samples. The structure of multivariate LSTM is shown in Fig. 3.

**Fig. 5.** 600s of 6-DOF ship motion data for dataset I

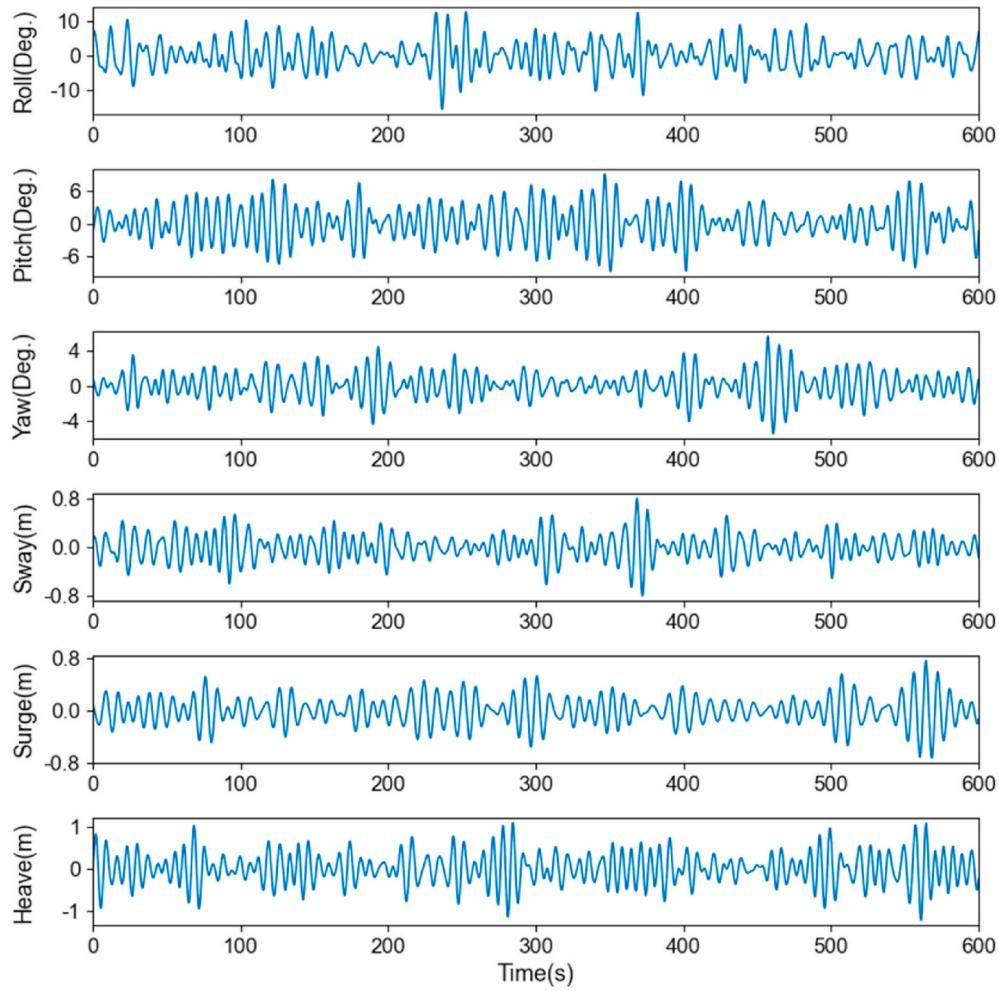


Fig. 6. 600s of 6-DOF ship motion data for dataset II

3. Optimization and training mechanism

3.1. Optimization problem

In this paper, we define a sliding window of length d , and use the historical ship motion data $X_t = (x_{t-d+1}, x_{t-d+2}, \dots, x_t)$ in the window to predict the future value of x_{t+q} . Then through the way of sliding window to achieve the purpose of continuous prediction. Therefore, ship motion prediction becomes a regression task, and we use the mean squared error as the loss function of our proposed model. The objective function is formulated as:

$$(w_{attn}, w_{lstm}, w_{dense}, b_{lstm}, b_{dense}) = \underset{\theta}{\operatorname{argmin}} J \quad (14)$$

$$J = \frac{1}{2N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (15)$$

where θ represents all the trainable weight matrix and bias term in the MSA-LSTM model, N is the number of training samples, Y and \hat{Y} are the sets of true and predicted values, respectively. In the proposed model, we take the method of mini-batch gradient descent with Adam optimizer to iteratively update the parameters (Kingma and Ba, 2015; Ruder, 2016). The dropout strategy is also applied to improve the performance of our proposed model (Srivastava et al., 2014).

3.2. Two-stage training mechanism

Algorithm 1. Two-stage training algorithm

Require: ship motion dataset: $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)$, learning rate: γ , batch size: m ;

Input: original dataset: $\{\mathbf{X}_t\}_{t=1}^N$, multiscale dataset: $\{\tilde{\mathbf{X}}_t\}_{t=1}^N$, target dataset: $\{\mathbf{y}_{t+q}\}_{t=1}^N$;

Output: trained MSA-LSTM model;

- 1 Initialize all the trainable parameters θ in MSA-LSTM model;
- 2 **repeat**
- 3 //training LSTM layer;
- 4 **while** stopping criterion is not satisfied **do**
- 5 Calculate gradient of mini batch of m_l samples: $\mathbf{g}_l \leftarrow \partial J_{m_l}(\theta) / \partial \theta_l$;
- 6 New parameters: $\theta'_l = \theta_l - \gamma \cdot \mathbf{g}_l$;
- 7 Update parameters: $\theta_l \leftarrow \theta'_l$;
- 8 **end while**
- 9 Fixed LSTM layer parameters θ_l ;
- 10 //training Attention layer;
- 11 **while** stopping criterion is not satisfied **do**
- 12 Calculate gradient of mini batch of m_a samples: $\mathbf{g}_a \leftarrow \partial J_{m_a}(\theta) / \partial \theta_a$;
- 13 New parameters: $\theta'_a = \theta_a - \gamma \mathbf{g}_a$;
- 14 Update parameters: $\theta_a \leftarrow \theta'_a$;
- 15 **end while**
- 16 Fixed Attention layer parameters θ_a ;
- 17 **until** $J(\theta)$ starts to converge
- 18 //training all parameters θ until $J(\theta)$ stops decreasing;
- 19 **return** MSA-LSTM model;

The LSTM structure with multiple inputs and single output easily makes the model fall into local optimum. In addition, the introduction of attention mechanism increases the computational complexity, which makes the model difficult to converge. To solve this issue, we propose the two-stage training mechanism to train the proposed model. In the first stage, the sets of feature-value pairs $\{\mathbf{X}_t, \tilde{\mathbf{X}}_t, \mathbf{y}_{t+q}\}$ are input into LSTM layer to train parameters. The LSTM layer parameters are fixed until the loss function is no longer decreasing. In the second stage, the parameters of the attention layer are trained until the loss function no longer decreases, and the attention parameters are fixed. These two stages are repeated until there is a tendency to converge and all parameters are trained. Two-stage training mechanism is outlined in Algorithm 1 and the flow chart is shown in Fig. 4.

4. Experiments

4.1. Datasets

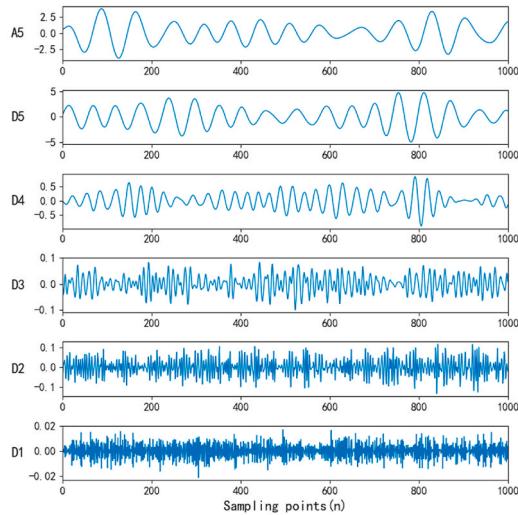
In order to test the feasibility and effectiveness of the MSA-LSTM model, we use two different sea conditions of ship motion data for prediction experiments. The datasets are collected by China Ocean Shipping Company (COSCO) and Yanshan University (YSU) which

include the ship's 6-DOF motion data of roll, pitch, yaw, sway, surge and heave. The specific ship parameters are shown in Table 1. The conditions and environment for data collection are shown in Table 2. We take 0.1s as the sampling interval, dataset I samples continuously for 1 h and dataset II samples continuously for 40 min. During the data collection, the conditions such as course, speed and external environment remain stable. Fig. 5 and Fig. 6 respectively show the six degree of freedom ship motion data of 600s collected in two datasets. The ship motion data are provided to the shipboard stabilized platform, which enabling the platform to perform compensation control in advance. These data have removed the direct current component from the ship motion signals to meet the requirements of the stabilized platform.

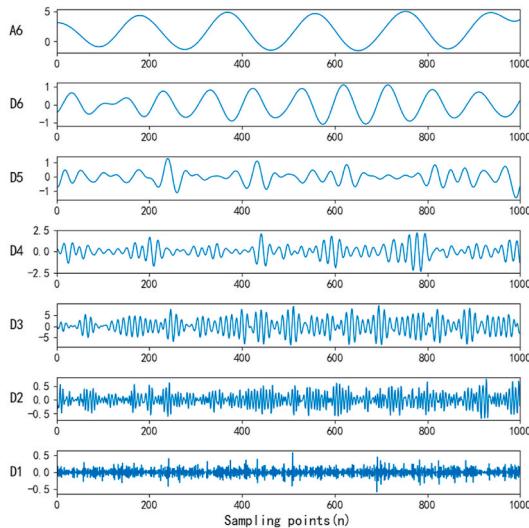
In order to train and test the MSA-LSTM model, six degree of freedom ship motion dataset needs to be divided into training set, verification set and test set for each dimension. In the experiment, we set 70% of the ship motion dataset as the training sets, 10% of the verification sets and 20% of the testing sets. Before the training of proposed model, in order to speed up the convergence of the model, we need to normalize the ship motion data in the interval (0,1). The calculation process is as follows:

$$x'_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}, \quad i = (1, 2, \dots, T) \quad (16)$$

where $\max(x_i)$ and $\min(x_i)$ represent the maximum and minimum values



(a) Decomposed approximation and detail components of ship roll data in dataset I



(b) Decomposed approximation and detail components of ship roll data in dataset II

Fig. 7. The effect of wavelet decomposition on the roll data of 1000 points in two datasets.**Table 3**
Hyper-parameters setting.

Parameters	Dataset I setting	Dataset II setting
Attention Layer	Units = 16	Units = 32
LSTM Layer1	Units = 64, dropout = 0.2	Units = 128, dropout = 0.2
LSTM Layer2	Units = 32, dropout = 0.1	Units = 64, dropout = 0.2
Training Parameters	First-stage batchsize = 64 Second-stage batchsize = 32	First-stage batchsize = 64 Second-stage batchsize = 32

of dataset, T is the number of samples in the dataset. After the prediction, the predicted value is inversely normalized to get the final prediction result. The formula is presented as:

$$\hat{y}_i = [\max(x_i) - \min(x_i)]\hat{y}'_i + \min(x_i) \quad (17)$$

where \hat{y}'_i and \hat{y}_i represent the prediction values before and after the anti-normalization respectively.

4.2. Wavelet decomposition parameters setting

The ship motion data need to be preprocessed for transforming that into the format required by MSA-LSTM model. The wavelet decomposition is applied to the ship motion data. Through the spectrum of two datasets, we get the frequency distribution of the ship motion. From this we can find the optimal number of decomposition layers to ensure that the ship motion data can be decomposed in different frequency scales. The decomposition level of dataset I is 5 and dataset II is 6, and the wavelet basis function is 'db10'. Fig. 7 shows the effect of wavelet decomposition on the roll data of 1000 points in two datasets. In addition, the sliding window method is used to generate training features and labels after wavelet decomposition. Prediction experiments with prediction interval of 1s are performed on two datasets, and the sliding window sizes were set to 10 and 15 respectively.

4.3. Hyper-parameters setting

In this section, in order to achieve the optimal performance of the MSA-LSTM model, the hyper-parameters of each part are optimized. It includes three main parts: LSTM layer, attention layer and training parameters. These include hidden layer units, dropout rate and training batch size, which are hyperparameters related to model tuning. The method to determine the optimal parameters is to adjust only the relevant parameters to obtain the optimal prediction performance under the condition of keeping other parameters unchanged. The optimal hyper-parameters of each part are given in Table 3.

4.4. Training MSA-LSTM model

We use the proposed two-stage training mechanism to train the MSA-LSTM model. In the first stage, LSTM layer is trained. In the second stage, we train the attention layer. Repeat these two stages until the model is about to converge, we train all the parameters until loss stops falling. The experiment shows that convergence can be achieved in two rounds. When the curve of the training set and the curve of the verification set continue to decline until it stops falling, it indicates that the model converges and the training is completed. The training set loss curve and verification set loss curve of the two datasets in the training process are shown in Fig. 8 and Fig. 9.

Since the LSTM layer has been trained in the first stage, the model has a priori knowledge of multiscale data. Therefore, the convergence speed of the second stage training is accelerated. Through the attention layer, the proposed model can adaptively learn the degree of correlation between the sub-signals of different frequency scales and the original ship motion signals, and assign the attention weight according to the importance of different sub-signals to the predicted results, so that the model can focus on the more significant frequency scales. The attention

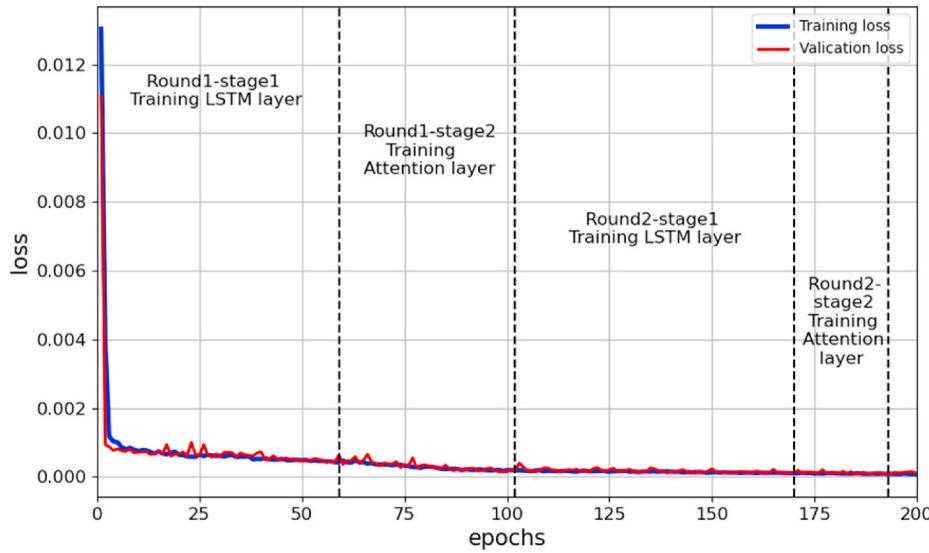


Fig. 8. The training set loss curve of roll data dataset I

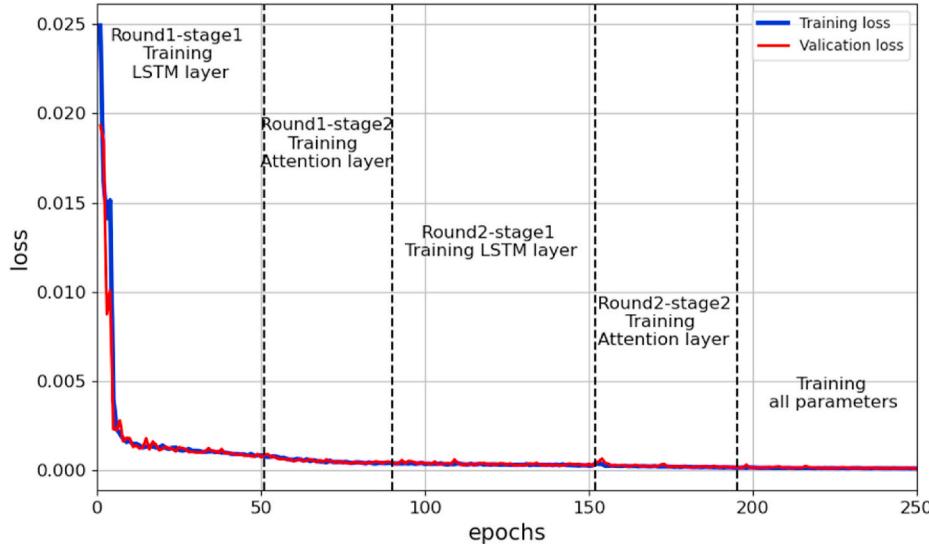


Fig. 9. The training set loss curve of roll data in dataset II

distributions of the two datasets are shown in Fig. 10 and Fig. 11 respectively. It can be found that the proposed model pays different attention to different frequency scales under different sea conditions. From the perspective of the overall distribution of attention, the low-frequency and medium-frequency components are more significant for prediction, while the high-frequency components have relatively less attention weights.

4.5. Evaluation metrics

In order to evaluate the prediction performance of the model and ensure the effectiveness of the error measurement results, Root mean square error (RMSE) and Mean absolute percent error (MAPE) are used as the evaluation metrics of the model prediction results in this experiment. The calculation is as follows:

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (\hat{y}_i - y_i)^2} \quad (18)$$

$$MAPE = \frac{100}{M} \sum_{i=1}^M \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (19)$$

where \hat{y}_i represents the predicted value, y_i is the true value, and M represents the number of testing set samples.

4.6. Results and analysis

In this section, we compared MSA-LSTM model with traditional machine learning methods and classic time series prediction model on two datasets to evaluate the performance of proposed model. In the experiment, SVR model, Prophet model, LSTM model, WT-LSTM model

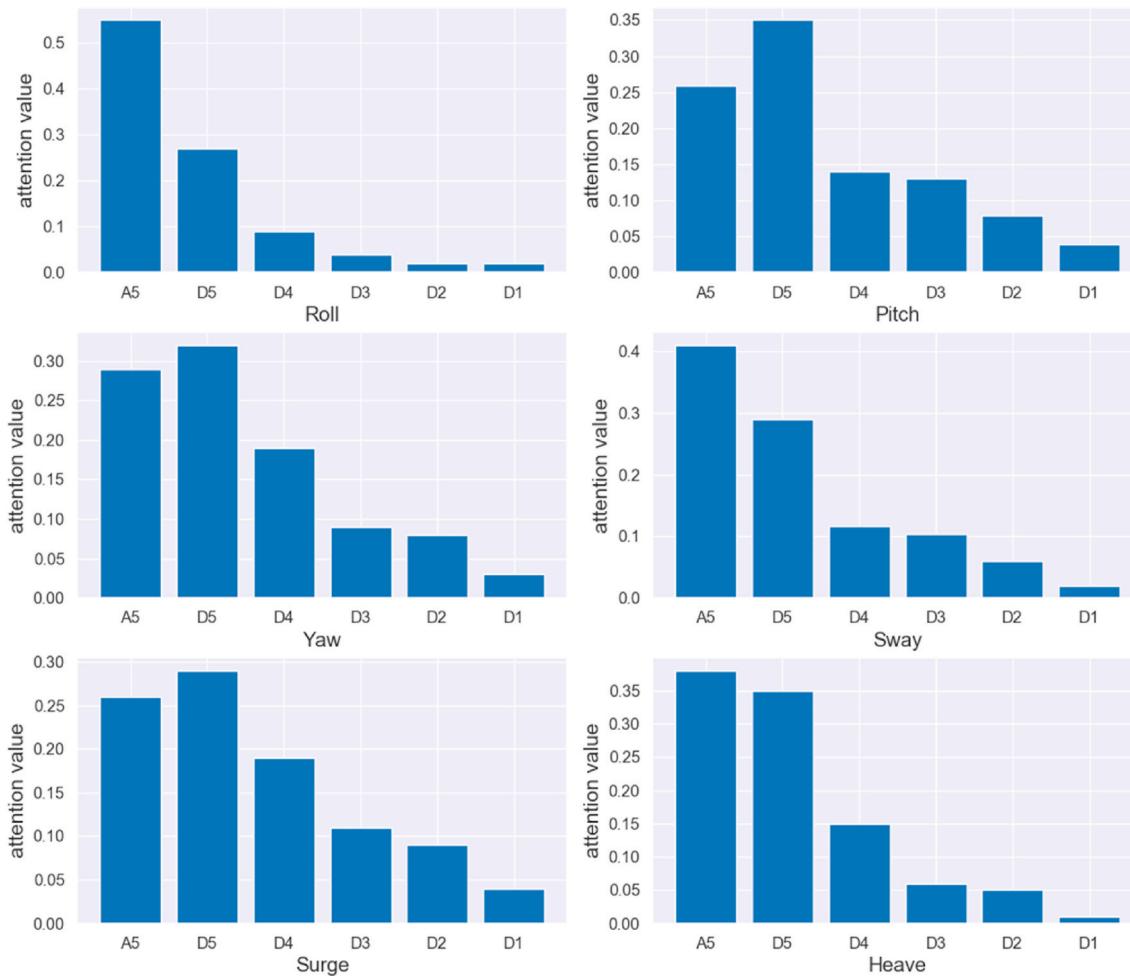


Fig. 10. Attention distribution of dataset I

and WT-RBF model are used as comparison objects and parameters are optimized respectively. Among them, Prophet is a time series prediction framework based on time series decomposition and machine learning proposed by Facebook (Taylor and Letham, 2018), WT-LSTM is a prediction method combining wavelet transform and LSTM, and WT-RBF is a method proposed in reference (Yin et al., 2018a) for ship roll motion prediction.

In order to ensure the fairness of the prediction experiment, all models are trained and tested using the same training set and test set, and the model parameters are optimized according to the verification set. Table 4 and Table 5 show the RMSE and MAPE evaluation index values of the predicted results of the five models on the test set of two datasets. Fig. 12 and Fig. 13 respectively describe the predicted value curve and true value curve of different models for the 6-DOF ship motion data under two datasets of different sea conditions. The predicted value is obtained by combining the predicted single value of each sample. The samples obtained from the sliding window are input into the trained model for prediction, and the predicted single value is obtained each time. The prediction trajectory is formed by iterating the trained model.

The results of the above evaluation indexes show that the RMSE and MAPE values of the proposed MSA-LSTM model are smaller than the other five competing models in predicting the 6-DOF ship motion of

Roll, Pitch, Yaw, Sway, Surge, and Heave on dataset I and II.

The prediction curve comparison results in Figs. 12 and 13 demonstrate that the prediction curve of proposed MSA-LSTM model is closer to the original motion data in 6-DOF motion prediction than the other five competition models. The above results show that the proposed MSA-LSTM model is effective and feasible in the accurate prediction of ship motion.

The detailed prediction results are analyzed as follows. Prophet model is based on historical data to fit, directly predict the future time. It has the characteristics of high modeling speed and short training time. But for the complex nonlinear ship motion, the prediction performance of Prophet model is not as good as other models using sliding window method. Especially in the complex sea conditions, it is difficult to fit the fast time-varying ship motion, resulting in the phenomenon of prediction lag. Compared with the Prophet model, the prediction results of SVR model and LSTM model are closer to the real value. Therefore, SVR model and LSTM model have more advantages in dealing with nonlinear problems and can more accurately fit ship motion data. Moreover, the prediction accuracy of LSTM model is better than that of SVR model, which indicates that LSTM model is more effective for ship motion prediction. Furthermore, the RMSE and MAPE values of the WT-LSTM model and WT-RBF model in six-dimensional motion are both smaller

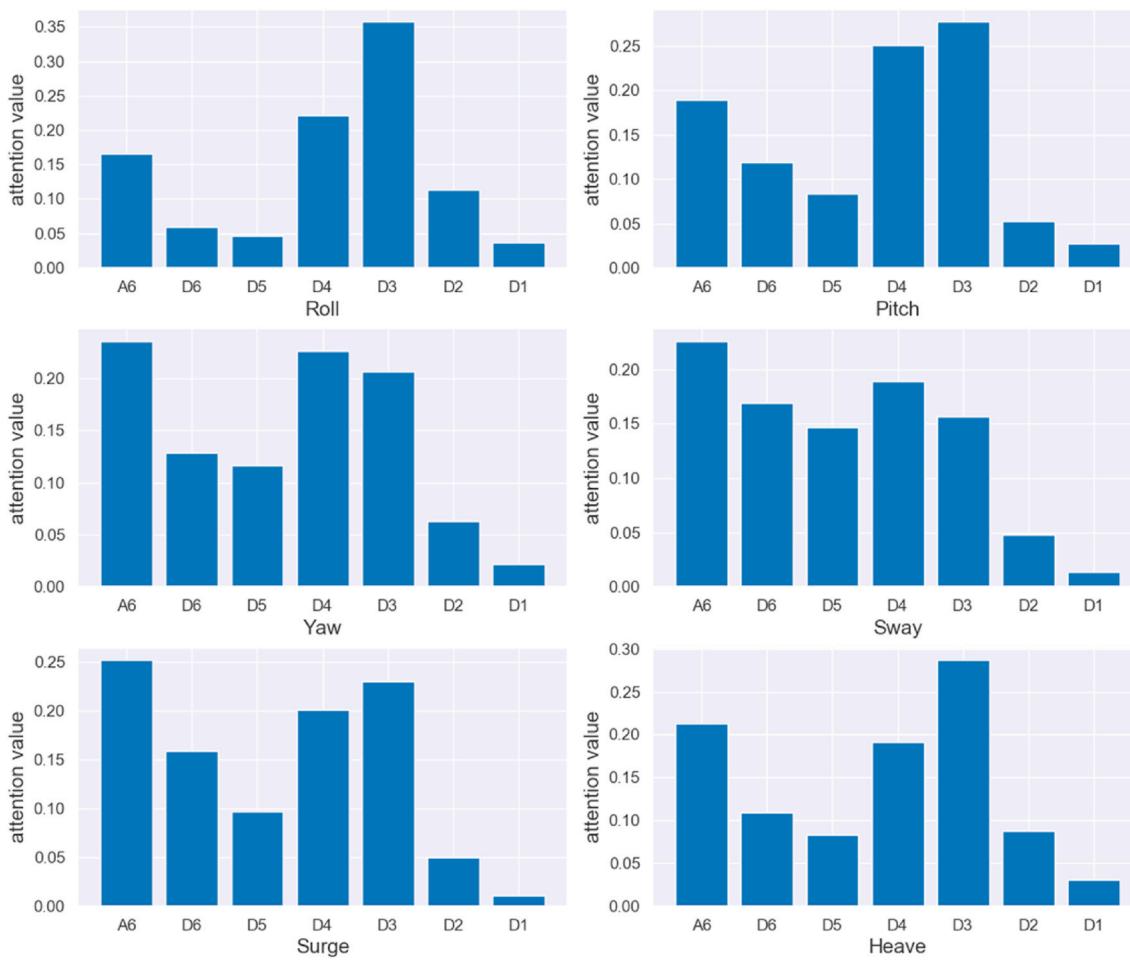


Fig. 11. Attention distribution of dataset II

Table 4

Evaluation metrics comparison of the proposed model and other models using the six degrees of freedom motion data based on the test set of dataset I.

Model	Evaluation metrics	Roll	Pitch	Yaw	Sway	Surge	Heave
SVR	RMSE	0.4146	0.2196	0.3466	0.0176	0.0226	0.0213
	MAPE	58.1117	36.0561	50.1487	39.5318	56.7346	46.1642
Prophet	RMSE	0.5699	0.4868	0.4402	0.0196	0.0178	0.0264
	MAPE	42.4130	34.8920	56.1252	65.7877	38.9227	58.3888
LSTM	RMSE	0.2178	0.2877	0.2357	0.0130	0.0146	0.0138
	MAPE	22.7077	25.3950	46.7423	22.1054	27.5252	39.3704
WT-LSTM	RMSE	0.1162	0.1193	0.1201	0.0058	0.0079	0.0092
	MAPE	15.3218	17.8059	23.8740	18.4125	24.8354	20.4785
WT-RBF	RMSE	0.1679	0.2274	0.1757	0.0096	0.0105	0.0093
	MAPE	20.2794	21.4571	29.2875	19.7673	22.4843	21.4527
MSA-LSTM	RMSE	0.0435	0.0248	0.0264	0.0038	0.0046	0.0059
	MAPE	6.8801	8.2095	8.3743	7.7474	7.3132	9.4548

than those of the LSTM model. Therefore, using wavelet transform to decompose the ship motion to different frequency scales can effectively reduce the complexity of the ship motion, so that the model can capture the ship motion law from simpler sub-signals of different scales, so as to better fit the ship motion. Besides, the prediction accuracy of WT-LSTM model is better than that of WT-RBF model, which indicates that LSTM network is more suitable for short-term ship motion prediction than RBF network. In addition, the MSA-LSTM model proposed in this paper further improves the prediction accuracy, and proves that the prediction method combining multiscale attention mechanism and multivariable LSTM can effectively make the model pay attention to more critical information in different frequency scales, and suppress the interference

of noise signal, which verifies the feasibility and effectiveness of the proposed model.

The sea conditions of dataset II is more complex than that of dataset I, and the ship motion contains more scale frequency information. The prediction accuracy of all models for dataset II is significantly lower than that of dataset I, which indicates that the accuracy of the prediction model will decrease with the increase of the complexity of sea condition. However, MSA-LSTM model can decompose the ship motion into different frequency scales and learn the more important scales that affect the prediction results. Therefore, the prediction accuracy of MSA-LSTM model in complex sea condition is significantly lower than that of the other five models, which shows that the proposed model can adapt to

Table 5

Evaluation metrics comparison of the proposed model and other models using the six degrees of freedom motion data based on the test set of dataset II.

Model	Evaluation metrics	Roll	Pitch	Yaw	Sway	Surge	Heave
SVR	RMSE	0.6042	0.3641	0.3992	0.0956	0.0890	0.1543
	MAPE	39.5719	59.1882	61.3244	74.3501	86.5853	57.6621
Prophet	RMSE	0.7244	0.5249	0.4672	0.0919	0.0890	0.1487
	MAPE	64.8943	58.8570	74.7049	75.6513	86.5607	71.5236
LSTM	RMSE	0.3753	0.3495	0.3616	0.0409	0.0248	0.0662
	MAPE	37.2881	54.1240	72.8033	48.7500	62.6754	57.2565
WT-LSTM	RMSE	0.1554	0.1208	0.2277	0.0332	0.0186	0.0354
	MAPE	27.0347	40.7564	25.0889	24.5466	54.9039	46.9026
WT-RBF	RMSE	0.2254	0.5249	0.2473	0.0362	0.0231	0.0435
	MAPE	34.6024	47.8529	31.6316	25.6314	55.4167	52.4216
MSA-LSTM	RMSE	0.0652	0.0601	0.0437	0.0079	0.0068	0.0118
	MAPE	13.8814	13.9651	16.3743	14.2789	17.1388	18.2886

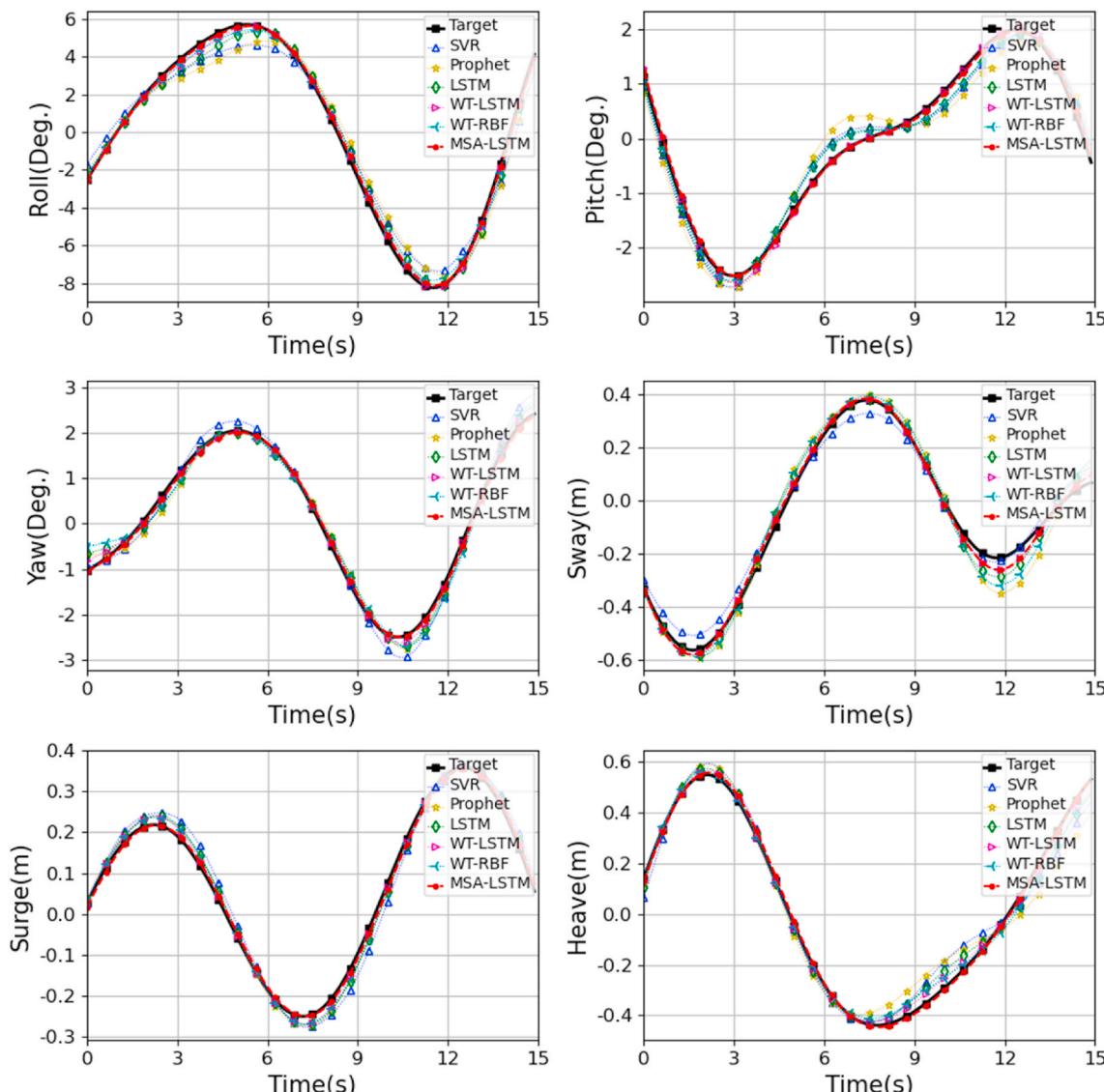


Fig. 12. 15s prediction comparison for the proposed model and other models using the six degrees of freedom motion data based on the test set of dataset I.

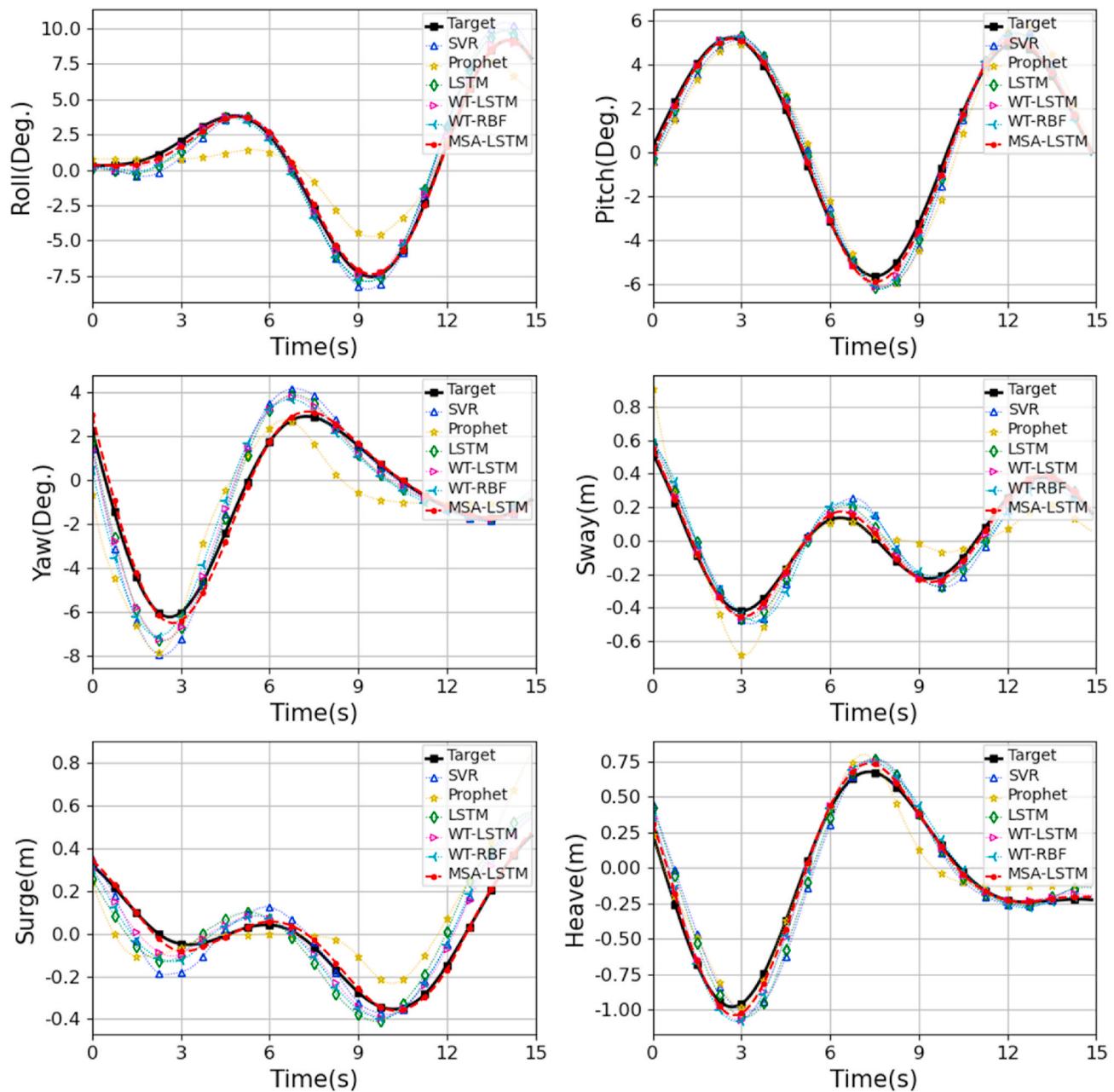


Fig. 13. 15s prediction comparison for the proposed model and other models using the six degrees of freedom motion data based on the test set of dataset II.

different sea conditions and still maintain the prediction performance in complex sea condition.

5. Conclusion

A multiscale attention-based LSTM model for ship motion prediction is proposed, which is applied to the shipboard stabilized platform to keep the equipment on the platform stable. Moreover, the two-stage training mechanism is designed for the model structure. The proposed model captures the motion law of the ship from different frequency scales and can adaptively focus on the more important frequency information for the prediction, so as to improve the prediction accuracy. Under two different sea conditions datasets, the proposed method carries out the prediction experiments of ship roll, pitch, yaw, sway, surge and heave at 1s intervals respectively. Compared with SVR model, Prophet model, LSTM model, WT-LSTM model and WT-RBF model,

MSA-LSTM model has the most accurate prediction results, and in complex sea conditions, the proposed model can still provide the optimal prediction performance. The experimental results prove the feasibility and effectiveness of the proposed method.

In the future research, considering that the number of wavelet decomposition layers needs to be determined by spectral analysis, an adaptive wavelet decomposition algorithm will be designed to determine the number of decomposition layers to improve the efficiency. Furthermore, adaptive selection of sliding window size according to different sea conditions needs further study. In addition, longer period of time prediction for other offshore applications is also the next research direction. When faced with the problem of long period of time prediction, we will adopt the method of optimizing model structure and decreasing sampling of input data to avoid overfitting.

CRediT authorship contribution statement

Tao Zhang: Conceptualization, Methodology, Supervision. **Xiao-Qing Zheng:** Software, Writing – original draft, Validation. **Ming-Xin Liu:** Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is financially supported by the National Natural Science Foundation of China (under Grants 61871465), the Equipment Advance Research Joint Technology Project Foundation of 13th Five-year Plan (under Grants 41412040302) and the Natural Science Foundation of Hebei Province (under Grants F2020203010).

References

- Bulian, G., Francescutto, A., 2011. Effect of roll modelling in beam waves under multi-frequency excitation. *Ocean Eng.* 38 (13), 1448–1463.
- Cao, J., Li, Z., Li, J., 2019. Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A* 519, 127–139.
- Chang, Z., Zhang, Y., Chen, W., 2019. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. *Energy* 187, 115804.
- Cheng, X., Li, G., Skulstad, R., Major, P., Chen, S., Hildre, H.P., Zhang, H., 2019. Data-driven uncertainty and sensitivity analysis for ship motion modeling in offshore operations. *Ocean Eng.* 179, 261–272.
- Choi, H., Cho, K., Bengio, Y., 2018. Fine-grained attention mechanism for neural machine translation. *Neurocomputing* 284, 171–176.
- Du, S., Li, T., Yang, Y., Horng, S.-J., 2020. Multivariate time series forecasting via attention-based encoder-decoder framework. *Neurocomputing* 388, 269–279.
- Duan, W.Y., Han, Y., Huang, L.M., Zhao, B.B., Wang, M.H., 2016. A hybrid EMD-SVR model for the short-term prediction of significant wave height. *Ocean Eng.* 124, 54–73.
- Fossen, T.I., 2011. *Handbook of Marine Craft Hydrodynamics and Motion Control*. John Wiley & Sons.
- Hong, W.-C., Li, M.-W., Geng, J., Zhang, Y., 2019. Novel chaotic bat algorithm for forecasting complex motion of floating platforms. *Appl. Math. Model.* 72, 425–443.
- Hou, X., Zou, Z., 2015. SVR-based identification of nonlinear roll motion equation for FPSOs in regular waves. *Ocean Eng.* 109, 531–538.
- Hu, J., Zheng, W., 2020. Multistage attention network for multivariate time series prediction. *Neurocomputing* 383, 122–137.
- Huang, B.-G., Zou, Z.-J., Ding, W.-W., 2018. Online prediction of ship roll motion based on a coarse and fine tuning fixed grid wavelet network. *Ocean Eng.* 160, 425–437.
- Huang, L.M., Duan, W.Y., Han, Y., Chen, Y.S., 2014. A review of short-term prediction techniques for ship motions in seaway. *J. Ship Mech.* 18 (12), 1534–1542.
- Jiang, H., Duan, S., Huang, L., Han, Y., Yang, H., Ma, Q., 2020. Scale effects in AR model real-time ship motion prediction. *Ocean Eng.* 203, 107202.
- Karevan, Z., Suykens, J.A.K., 2020. Transductive LSTM for time-series prediction: an application to weather forecasting. *Neural Network*. 125, 1–9.
- Khare, S.K., Bajaj, V., 2020. Constrained based tunable Q wavelet transform for efficient decomposition of EEG signals. *Appl. Acoust.* 163, 107234.
- Kingma, D.P., Ba, J., 2015. Adam: A Method for Stochastic Optimization, International Conference on Learning Representations.
- Li, G., Kawan, B., Hao, W., Zhang, H., 2017. Neural-network-based modelling and analysis for time series prediction of ship motion. *Ship Technol. Res.* 64 (1), 30–39.
- Li, M.-W., Geng, J., Han, D.-F., Zheng, T.-J., 2016. Ship motion prediction using dynamic seasonal RvSVR with phase space reconstruction and the chaos adaptive efficient FOA. *Neurocomputing* 174, 661–680.
- Li, Y., Zhu, Z., Kong, D., Han, H., Zhao, Y., 2019. EA-LSTM: evolutionary attention-based LSTM for time series prediction. *Knowl. Base Syst.* 181, 104785.
- Liu, S.B., Shi, P.A., Wu, L., 2014. Short-term prediction of ship motion based on EMD-SVM. *Appl. Mech. Mater.* 571–572, 252–257.
- Liu, T., Wei, H., Zhang, C., Zhang, K., 2017. Time series forecasting based on wavelet decomposition and feature extraction. *Neural Comput. Appl.* 28 (1), 183–195.
- Liu, X., Liu, H., Guo, Q., Zhang, C., 2020a. Adaptive wavelet transform model for time series data prediction. *Soft Comput* 24 (8), 5877–5884.
- Liu, Y., Gong, C., Yang, L., Chen, Y., 2020b. DSTP-RNN: a dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. *Expert Syst. Appl.* 143, 113082.
- Mallat, S., 2009. *A Wavelet Tour of Signal Processing: the Sparse Way*.
- Ruder, S., 2016. An Overview of Gradient Descent Optimization Algorithms arXiv: Learning.
- Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15 (1), 1929–1958.
- Takami, T., Nielsen, U.D., Jensen, J.J., 2021. Real-time deterministic prediction of wave-induced ship responses based on short-time measurements. *Ocean Eng.* 221, 108503.
- Taylor, S.J., Letham, B., 2018. Forecasting at scale. *The American Statian* 72 (1), 37–45.
- Usama, M., Ahmad, B., Song, E., Hossain, M.S., Alrashoud, M., Muhammad, G., 2020. Attention-based sentiment analysis using convolutional and recurrent neural network. *Future Generat. Comput. Syst.* 113, 571–578.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need, *Neural Information Processing Systems*, pp. 5998–6008.
- Vidicperunovic, J., Jensen, J.J., 2009. Parametric roll due to hull instantaneous volumetric changes and speed variations. *Ocean Eng.* 36 (12), 891–899.
- Wang, M., 2020. Study on Extremely Short-Term Prediction of Ship Sway Motion. Yanshan university.
- Yang, B., Sun, S., Li, J., Lin, X., Tian, Y., 2019. Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing* 332, 320–327.
- Yao, Y., Han, L., Wang, J., 2018. LSTM-PSO: long short-term memory ship motion prediction based on particle swarm optimization. In: 2018 IEEE CSAA Guidance, Navigation and Control Conference. GNCC.
- Yin, J.-C., Perakis, A.N., Wang, N., 2018a. A real-time ship roll motion prediction using wavelet transform and variable RBF network. *Ocean Eng.* 160, 10–19.
- Yin, J.-c., Zou, Z.-j., Xu, F., 2013. On-line prediction of ship roll motion during maneuvering using sequential learning RBF neuralnetworks. *Ocean Eng.* 61, 139–147.
- Yin, J.-C., Zou, Z.-J., Xu, F., Wang, N.-N., 2014. Online ship roll motion prediction based on grey sequential extreme learning machine. *Neurocomputing* 129, 168–174.
- Yin, J., Wang, N., Perakis, A.N., 2018b. A real-time sequential ship roll prediction scheme based on adaptive sliding data window. *IEEE Trans. Syst., Man, Cybern., Syst.* 48 (12), 2115–2125.
- Yin, W., Schutze, H., Xiang, B., Zhou, B., 2016. ABCNN: attention-based convolutional neural network for modeling sentence pairs. *TACL* 4 (1), 259–272.
- Yu, Y., Shenoi, R.A., Zhu, H., Xia, L., 2006. Using wavelet transforms to analyze nonlinear ship rolling and heave-roll coupling. *Ocean Eng.* 33 (7), 912–926.