



ARTICLE

A Hybrid BPNN-GARF-SVR Prediction Model Based on EEMD for Ship Motion

Hao Han and Wei Wang*

School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, 100191, China

*Corresponding Author: Wei Wang. Email: wongwei@buaa.edu.cn

Received: 18 January 2022 Accepted: 16 March 2022

ABSTRACT

Accurate prediction of ship motion is very important for ensuring marine safety, weapon control, and aircraft carrier landing, etc. Ship motion is a complex time-varying nonlinear process which is affected by many factors. Time series analysis method and many machine learning methods such as neural networks, support vector machines regression (SVR) have been widely used in ship motion predictions. However, these single models have certain limitations, so this paper adopts a multi-model prediction method. First, ensemble empirical mode decomposition (EEMD) is used to remove noise in ship motion data. Then the random forest (RF) prediction model optimized by genetic algorithm (GA), back propagation neural network (BPNN) prediction model and SVR prediction model are respectively established, and the final prediction results are obtained by results of three models. And the weights coefficients are determined by the correlation coefficients, reducing the risk of prediction and improving the reliability. The experimental results show that the proposed combined model EEMD-GARF-BPNN-SVR is superior to the single predictive model and more reliable. The mean absolute percentage error (MAPE) of the proposed model is 0.84%, but the results of the single models are greater than 1%.

KEYWORDS

Back propagation neural network; ensemble empirical mode decomposition; genetic algorithm; random forest; SVR; ship motion prediction

1 Introduction

Under the influence of wind, waves and other factors, the movement of the ship sailing on the sea in six degrees of freedom is randomly complicated and it seriously affects the sport safety of the ship. Since it is difficult to establish an accurate dynamical model for ship motion, most of the existing models for predicting ship motion are based on the measurements data of the ship at historical moments to predict the movement in the future.

There are a lot of studies that have been adopted in ship motion prediction. These methods are mainly categorized into traditional time series analysis methods and intelligent prediction methods. Traditional methods mainly include auto-regressive model (AR), moving-average model (MA) and auto regressive moving average mode (ARMA). Yumori et al. established ship motion prediction model based on AR and ARMA model with measured information, including wave motion and ship



motions [1]. And similar models can be found in Zhang et al. [2,3]. However, these methods generally used the least squares to solve model coefficients but it was inappropriate to use this constant model coefficients to forecast ship motion which is a random process. Therefore, Peng et al. proposed a method for on-line estimation of parameters using recursive least squares and developed an on-line fast ARMA fixed-order algorithm [4]. The methods mentioned above play an important role in predicting the ship motions, but AR, MA and ARMA model are linear models, which are not suitable to forecast the real ship motion which has complex, strong nonlinear and chaos characteristic.

Huang et al. creatively proposed a new processing method called empirical mode decomposition (EMD), which can effectively and adaptively deal with the nonlinearity and non-stationarity existing in actual engineering signals [5]. EMD is to equalize the original time series with the sum of several stationary intrinsic mode function (IMF) components and a trend function component, which is the process of smoothing the original time series. Therefore, it has been popularly employed in developing hybrid models for various time series prediction. Huang et al. proposed AR-EMD technique decomposing the complex ship motion data into a couple of simple intrinsic mode functions and residual, which produces better prediction compared to AR models [6]. Duan et al. developed a hybrid AR-EMD-SVR model which overcomes the non-stationary difficulty and achieves better predictions than other models [7]. However, the EMD method has mode mixing and endpoint effects, which destroys the physical meaning of the IMF. Aiming at the deficiency of EMD, EEMD is adopted in this paper to decompose complex signals of ship motion into multiple simple sub-signals [8,9].

With the development of artificial intelligence technology, machine learning models, which could solve the nonlinear or nonstationary problem, have drawn attention and are used to predict the time series. These models are all data-driven models, such as Artificial Neural Networks (ANNS), Support Vector Regression (SVR), etc. Zhang et al. discussed diagonal recurrent neural network and acquired better results [10]. Peng et al. proposed a novel method of single-output three-layer Back Propagation (BP) neural network to identify Volterra series kernels, which has higher precision, longer prediction time, effectiveness and adaptability [11]. Huang et al. applied radial basis function (RBF) neural network to develop one-step forward prediction of ship pitching motion [12]. Peng et al. proposed Echo State Network (ESN) based on Kalman filter algorithm for modeling in ship motion. However, there are issues that must be considered regarding the actual number of hidden nodes included in an ANN [13]. Liu et al. introduced a prediction method based on extreme learning machine, support vector machine and particle swarm optimization (PSO-KELM), which has a simple structure, fast training speed and simulation results showed high accuracy prediction [14]. Luo et al. adopted support vector machine for the parametric identification of ship coupled heave and pitch motions with real oceanic conditions [15]. Furthermore, Li et al. designed a dynamic seasonal robust v-support vector regression forecasting model (DSRvSVR) and analysis results showed that the hybrid model received better forecasting performance compared with other models [16]. Ensemble learning tree model is another important machine learning model that has achieved great success in predicting weather, stocks, traffic flow [17,18]. For example, Chen designed a random-forest based forecast model that had consistently shown better predictive skills than the ARIMA model for both long and short drought forecasting [19]. Li et al. adopted the ARIMA-XGBoost hybrid model to forecast China's energy supply security level and got accurate result [20]. But tree model is rarely used to predict ship motion.

Due to the uncertainty of ship motion, the methods mentioned above usually use a single complex model for prediction, which becomes unreliable when overfitting or failure occurs. Therefore, an EEMD-GARF-BPNN-SVR combined prediction method is proposed in the paper, which can effectively utilize the information of different single models and has higher precision and credibility.

The final prediction result is obtained by adding the results of various models according to the weight coefficient. And the weights coefficients are determined by the correlation coefficients.

The rest of this paper is structured as follows: First, the EEMD algorithm and various prediction models are introduced. The methods used to predict ship motion are introduced in detail, including time series method and various machine learning algorithms such as BPNN, SVR, RF and XGBoost. Then the GA algorithm for optimizing RF model parameters and the combined prediction model based on correlation coefficient are introduced. Next, all the models are constructed with measured data and the performance of each model is analyzed. The conclusion is drawn finally.

2 Method

In general, the goal of the prediction model is to map historical data $(x_t, x_{t-1}, \dots, x_{t-p})$ to future data x_{t+q} with mapping function f :

$$x_{t+q} = f(x_t, x_{t-1}, \dots, x_{t-p+1}), q > 0, p > 0$$

where p is the number of the inputs, q is the prediction step size. Due to the non-stationary and nonlinear behavior of ship motion, f is a time-varying nonlinear function. Therefore, ship motion prediction is a nonlinear multi-step prediction problem which is more difficult than one-step ahead forecasting, since it has to deal with various additional complications, like reduced accuracy and increased uncertainty [21,22].

In this paper, the EEMD-BPNN-GARF-SVR algorithm is used as the prediction model. Fig. 1 shows the main framework of the algorithm proposed in this paper.

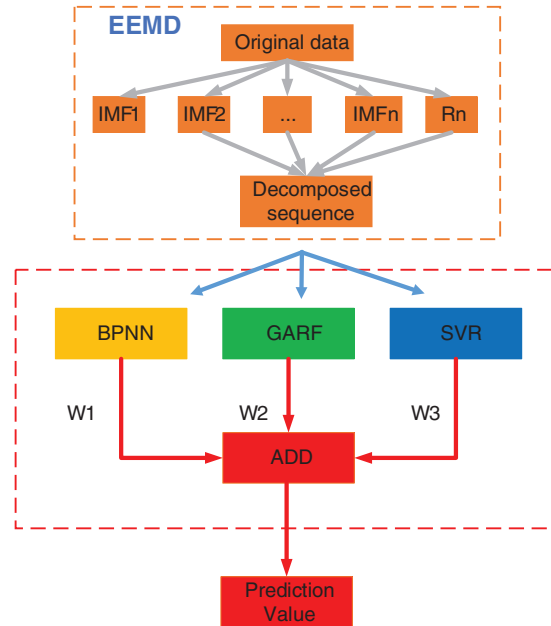


Figure 1: The framework of the EEMD-BPNN-GARF-SVR model

First, original data is processed by EEMD and is divided into $n+1$ subsets including $IMF_i (i = 1, 2, \dots, n)$ and a residual set. IMF_i corresponds to the highest frequency component and IMF_n corresponds to the lowest frequency component. IMF_1 is considered as a noise sequence, thus a

reconstructed sequence can be obtained by removing IMF_i . Then the reconstructed data is divided into training set and test set. training set is used to build BPNN, GARF and SVR models, where the parameters of GARF is optimized by genetic algorithm (GA), and the test data is used to test the prediction performance of the models. The final prediction results of the hybrid model are obtained by combining the results of single model, while the optimal weighting coefficients are determined according to the correlation coefficients between the model prediction values and ship motion test values.

2.1 EEMD

EMD is a processing method for non-stationary nonlinear signals proposed by Huang in 1998 [5] which has been successfully applied in many fields [5]. The purpose of the EMD algorithm is to decompose a poorly performing signal into a set of better performing Intrinsic Mode Function (IMF) functions that satisfy two properties: (1) the number of extreme points (maximum or minimum) of the signal is equal to or at most one different from the number of zero crossings; (2) The average of the upper envelope composed of local maxima and the lower envelope composed of local minima is zero.

The procedure of the EMD algorithm is shown in [Table 1](#).

Table 1: EMD algorithm

-
1. Find all the maxima and minima points of the original data sequence $x(t)$ and fit them to the upper and lower envelopes of the original sequence respectively using the cubic spline function;
 2. Calculate the mean of the upper and lower envelopes m_1 ;
 3. Calculate a new sequence $h_1 = x(t) - m_1$;
 4. If h_1 satisfies the two criteria of the IMF, h_1 is defined as an IMF, $r(t) = x(t) - h_1$; otherwise, let $x(t) = h_1$, and return to the step 1;
 5. Return to the first step unless the last data sequence $r(t)$ cannot be decomposed.
-

The EMD method has been applied quickly and effectively in different engineering fields, such as in oceans, atmosphere, celestial observations and geophysical record analysis. But an inevitable shortcoming of EMD, mode mixing which is caused by the data intermittency, such as intermittent signal, impulse interference and noise, limits its application. Mode mixing refers to an IMF containing extremely different characteristic scales or similar characteristic scales distributed in different IMFs, causing adjacent IMFs to mix together.

Aiming at this phenomenon, Huang proposed the EEMD which use Gaussian white noise in original sequence. When the signal is added to a uniformly distributed white noise background, signal regions of different scales are automatically mapped to the appropriate scale associated with the background white noise. Due to the characteristics of zero-mean noise, the noise will be canceled after multiple averaging calculations.

The steps of EEMD algorithm steps are shown in [Table 2](#).

Table 2: EEMD algorithm

-
1. Add a random white noise with the mean of 0 and the variance of σ to the original signal to obtain a noisy signal $x_w(t)$.
 2. Decompose $x_w(t)$ into k IMFs components via the EMD.
 3. Repeat steps 1 and 2 N times, and each time adds different white noise sequences.
 4. Since the random white noise mean value is 0, the noise component of each IMF component can be eliminated by averaging the corresponding IMF component, and the IMF component decomposed by EEMD is obtained.
-

2.2 Autoregressive Integrated Moving Average (ARMA)

The ARMA model is mostly used in time series forecasting. For a given data set of ship motion $\{x(t), t = 1, 2, \dots, n\}$, ARMA (p, q) can be defined as follows:

$$x(t) = \alpha_1 x(t-1) + \alpha_2 x(t-2) + \dots + \alpha_p x(t-p) + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} \quad (1)$$

where p and q are the model order of the AR and MA, respectively. ε_t is a white noise with finite variance, α_p and β_q are the regression coefficients of the ARMA model. The most important step in model construction is to determine the model orders p and q , which are generally identified by autocorrelation functions and partial autocorrelation functions. In order to fit the data as much as possible and avoid overfitting, Akaike Information Criteria (AIC) was used for model identification in this paper.

$$AIC = 2k - 2\ln(L) \quad (2)$$

where k is the number of parameters, L is the likelihood:

$$L = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\sum_{i=1}^n \frac{(x_i - \hat{x}_i)^2}{2\sigma^2} \right] \quad (3)$$

where x_i is the observed, and \hat{x}_i is the prediction, $(x_i - \hat{x}_i)$ is the prediction residual. In the model identification, the AIC value is minimized, which determines the smallest appropriate order to represent the time series.

2.3 Back Propagation Neural Network (BPNN)

Artificial neural networks have been widely used in recent years to construct time series prediction models because they can fit functions of arbitrary complexity with arbitrary precision. As a classic algorithm, BP network is favored by many researchers. A typical predictive network structure with one input layer, one hidden layer and one output layer is as follows:

In general, the output of the neural network is

$$x(t+1) = f_o \left(\sum_{h=1}^n \left[w_{ho} \left[f_h \left(\sum_{i=1}^p w_{ih} x(t-i) + b_h \right) \right] \right] + b_o \right) \quad (4)$$

where p and n are the number of input layer and hidden layer neurons, respectively. w_{ih} , w_{ho} , b_h , b_o are the coefficients of the input layer and the hidden layer, coefficients of the hidden layer and output layer, offsets of the hidden layer and output layer, respectively. f_h and f_o are the hidden layer activation

function and output layer activation function, respectively. A network structure with one input layer, two hidden layers and one output layer is adopted in this paper.

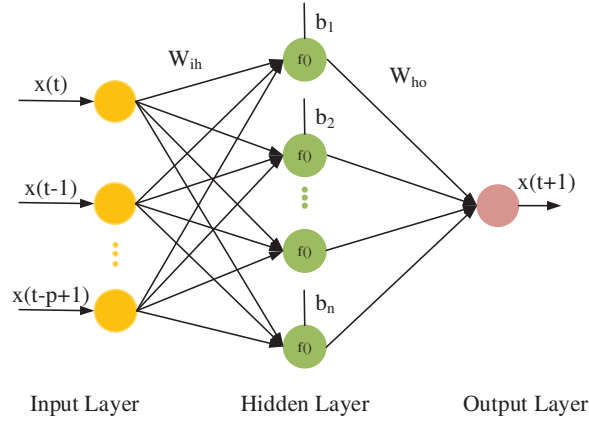


Figure 2: The structure of BPNN

2.4 Support Vector Regression

For a given training set $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$

where $x_i \in R^n$ is n-dimensional input of the ship motion, $y_i \in R$ is the prediction value, the aim of SVR is to find a hyperplane that fits y without errors under the given accuracy ε , that is, the distance from all sample points to the optimal hyperplane is not greater than ε . SVR tries to map initial data from the sample space into a higher dimensional feature space R^m through a nonlinear mapping function and thus converts a nonlinear problem into linear problem to get the optimal solution. The SVR model is constructed as follows:

$$y = f(x) = \omega \cdot \varphi(x) + b \quad (5)$$

where $\varphi(x)$ is the nonlinear mapping function, ω is the weight vector, $b \in R$ is the threshold, “ \cdot ” is the dot-product in the feature space. Taking into account the allowable error, we introduce slack variables ξ, ξ^* and penalty parameter C , therefore transform the above problem to a convex optimization problem:

$$\begin{aligned} & \min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ & s.t. \begin{cases} y_i - \omega \cdot \varphi(x_i) - b \leq \varepsilon + \xi \\ \omega \cdot \varphi(x_i) + b - y_i \leq \varepsilon + \xi^* \\ \xi, \xi^* \geq 0 \end{cases} \end{aligned} \quad (6)$$

After converting the optimization problem into a dual problem, the optimal regression function can be solved as

$$y = f(x) = \omega \cdot \varphi(x) + b = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (7)$$

where α_i and α_i^* are the Lagrange multipliers, $K(x_i, x) = \varphi(x_i) \cdot \varphi(x)$ is the kernel function, and gaussian function is choose as the kernel function in the paper.

2.5 Ensemble Trees

Decision tree is a well-known predictive model with many advantages such as intelligibility and small amount of computation, therefore, it is widely used in data science and competition. But a simple regression tree has the disadvantage of easily fitting the training data and unstable, therefore, little change in training data may result in very different trees and predictions. Ensemble trees are one solution to improve robustness and reliability of regression trees by combining multiple regression trees produced with different generation methods. RF and XGBoost are two widely used ensemble models but with different features.

The RF first randomly samples the original sample set by the Bootstrap sampling method to obtain t samples containing m samples, so that 63.2% of the samples in the initial training set appear in all sampling sets. Based on each sample set, a regression tree can be built. The process of RF regression tree is different from the traditional decision tree. It randomly extracts k features from all features as the split feature set of the current node. Each regression tree grows from top to bottom recursively, and all decision trees are combined. Random forests. The final predicted result is the average of the t regression trees predictions.

XGBoost is a scalable end-to-end tree boosting system, which is used widely by data scientists to achieve state-of-the-art results on many machines learning challenges. The goal of this model is to predict the output y with input x ,

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in \Gamma, \tag{8}$$

where K is the number of the functions, Γ represents a set of regression functions.

The objective function of XGBoost is as follows:

$$Obj = \sum_I l(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \text{ where } \Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \tag{9}$$

where l is the loss function indicating the difference between the prediction \hat{y}_i and the target y_i . Ω is the complexity of a single tree including the number of leaves T and the regular term of the residual ω to prevent overfitting.

During the training process, the training objective function of each tree is shown in

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_i(x_i)) + \Omega(f_i) \tag{10}$$

The second-order Taylor expansion of the objective function is

$$Obj^{(t)} = \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i)] + \Omega(f_i) \tag{11}$$

where $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$

f is defined as

$$f_i(x) = \omega_{q(x)}, \omega \in R^T, q: R^d \rightarrow \{1, 2, \dots, T\} \tag{12}$$

where q is a structure function that maps input x to the index of the lead node. w gives the leaf score corresponding to each index number.

I is defined as a sample set above each leaf:

$$I_j = \{i/q(x_i) = j\} \tag{13}$$

therefore,

$$\begin{aligned}
 Obj^{(n)} &= \sum_{i=1}^n \left[g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \Omega(f_i) \\
 &= \sum_{i=1}^n \left[g_i \omega_q(x_i) + \frac{1}{2} h_i \omega_q^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \\
 &= \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \\
 &= \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T
 \end{aligned} \tag{14}$$

where $G_j = \sum_{i \in I_j} g_i$, $H_i = \sum_{i \in I_j} h_i$

Then the optimal ω_j^* can be obtained as follows:

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \tag{15}$$

$$Obj = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \tag{16}$$

The good structure can be obtained by the objective function. In order to balance the new node complexity and objective function, exact greedy algorithm was adopted by the authors to calculate the gain in the loss reduction from introducing the split:

$$Gain = -\frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \tag{17}$$

L, R represents the left and right nodes of the leaf.

2.6 Genetic Algorithm

Genetic algorithm is a randomized search algorithm generated by evolutionary theory and genetic mechanism, which adaptively adjust the search direction and can be used to optimize parameters without falling into the local optimal solution. Genetic algorithms mainly include chromosome coding, evaluation of fitness, selection, and hybrid mutation operations. In the paper, the GA algorithm is used to optimize the two parameters of the RF prediction model, which are the maximum depth of the tree and the number of base classifiers. First, set the range of values for the two parameters, it determines the required number of binary coded, which map the parameter space into to the chromosomal space. An initial binary coded chromosome of a certain size is then randomly generated, and fitness of each individual in the initial chromosome population is calculated. Roulette wheel selection, deterministic or tournament selection are adopted to select individuals with higher fitness values from the group they have higher opportunity to transmit the good genes to the next generation as parents. Then the core operational variations and intersections in the genetic algorithm are executed, improving the search ability of the genetic algorithm and promoting the group to continuously develop toward the optimal solution. Fig. 3 shows the algorithm structure of EEMD-GARF.

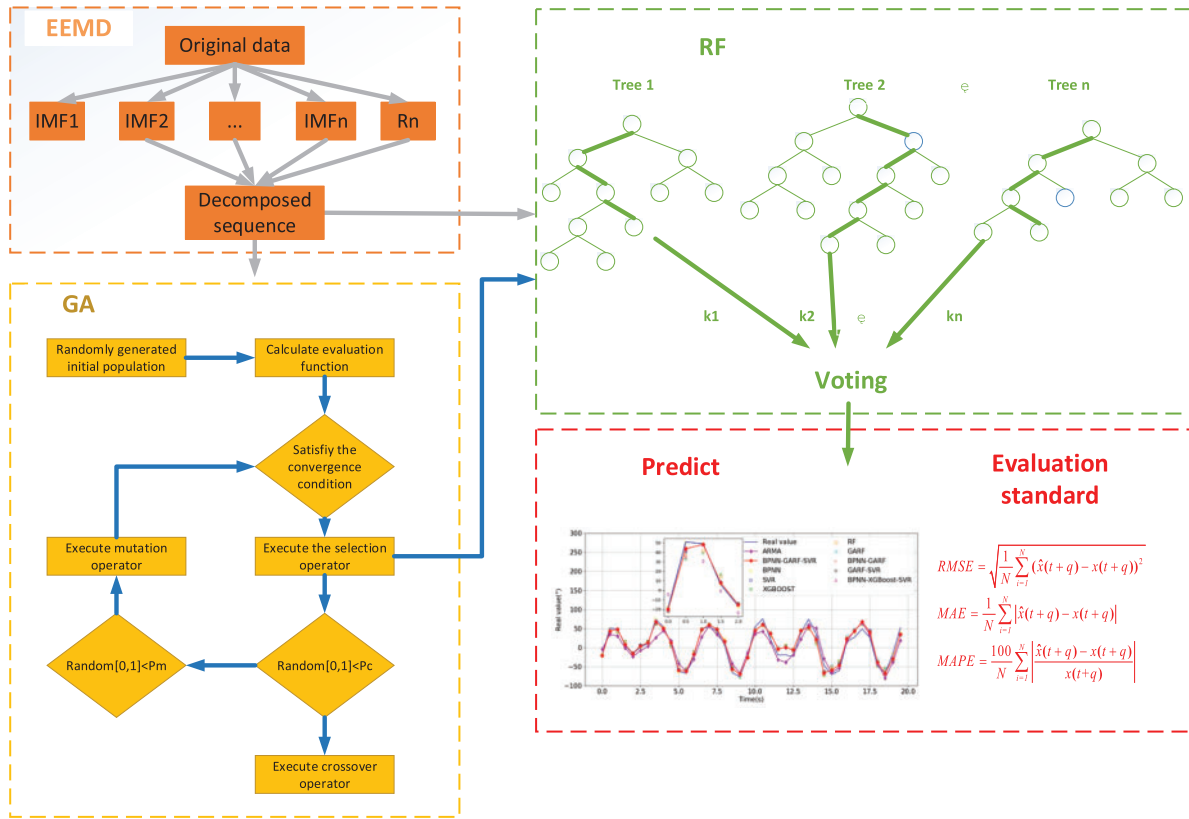


Figure 3: The framework of the EEMD-GARF model

2.7 Combined Prediction Based on Correlation Coefficient

In order to reduce the risk of prediction, a combined forecasting method came into being, the most concerned issue of which is how to find the weighted average coefficient. In most cases, the time-varying weight combination method produces better results than the constant weight combination method. A combined prediction model based on correlation coefficients is adopted in this paper, and the optimal solution of weighting coefficients is determined by the error square sum criterion.

We assume the observations are $\{x_t, t = 1, 2, \dots, N\}$, the predicted values of the prediction models are

$$\{x_{it}, t = 1, 2, \dots, N, i = 1, 2, \dots, m\}$$

where m is the number of the model. Combined forecast value is as follows:

$$\hat{x}_i = \sum_{i=1}^m l_i x_{it}, i = 1, 2, \dots, m, \sum_{i=1}^m l_i = 1 \tag{18}$$

l_i is the weight coefficient of the i -th prediction method.

The correlation coefficient between the predicted sequence of the i -th single prediction method and the actual observed sequence is

$$R_i = \frac{\sum_{t=1}^N (x_t - \bar{x})(x_{it} - \bar{x}_i)}{\sqrt{\sum_{t=1}^N (x_t - \bar{x})^2} \sqrt{\sum_{t=1}^N (x_{it} - \bar{x}_i)^2}}, i = 1, 2, \dots, m \quad (19)$$

The correlation coefficient between the combined predicted value and the actual observed value is

$$R = \frac{\sum_{t=1}^N (x_t - \bar{x})(\hat{x}_t - \bar{\hat{x}})}{\sqrt{\sum_{t=1}^N (x_t - \bar{x})^2} \sqrt{\sum_{t=1}^N (\hat{x}_t - \bar{\hat{x}})^2}} = \frac{\sum_{t=1}^N e_t \sum_{i=1}^m l_i e_{it}}{\sqrt{\sum_{t=1}^N e_t^2} \sqrt{\sum_{t=1}^N (\sum_{i=1}^m l_i e_{it})^2}}, i = 1, 2, \dots, m \quad (20)$$

where $e_t = x_t - \bar{x}$, $x_{it} - \bar{x}_i$

We define

$$\varepsilon_i = (e_{i1}, e_{i2}, \dots, e_{iN})^T \quad E_{ij} = \varepsilon_i^T \varepsilon_j = \sum_{t=1}^N e_{it} e_{jt}, i, j = 1, 2, \dots, m \quad E = (E_{ij})_{m \times m} \quad L = (l_1, l_2, \dots, l_m)^T$$

Then the (21) can be obtained

$$R = \frac{\sum_{i=1}^m l_i \sum_{t=1}^N e_t e_{it}}{\sqrt{\sum_{t=1}^N e_t^2} \sqrt{L^T E L}}, i = 1, 2, \dots, m \quad (21)$$

E_{ij} represents the covariance of the predicted values of the i -th and j -th prediction methods, and E_{ij} represents the variance of the predicted value sequence of the i -th method. E is the covariance information matrix of the combined prediction.

It is obvious that the larger R is, the more effective the prediction method is, so the solution of the weighting coefficient can be defined as an optimization problem as follows:

$$\text{Max } R(l_1, l_2, L, l_m) = \frac{\sum_{i=1}^m l_i \sum_{t=1}^N e_t e_{it}}{\sqrt{\sum_{t=1}^N e_t^2} \sqrt{L^T E L}} \quad (22)$$

$$s.t. \begin{cases} \sum_{i=1}^m l_i = 1, \\ l_i \geq 0, i = 1, 2, L, m \end{cases}$$

When R is the largest, the l is the weight coefficient of each model. When the optimal solution has more than two components that are not zero, that is, two or more single prediction methods provide effective information, the combined prediction method corresponding to the optimal solution is better than the best one of the single prediction methods.

3 Result and Experiment

In the paper, ARMA, BPNN, SVR, RF, GARF, XGBOOST are selected as the single prediction model to forecast the ship motion and BPNN-GARF, GARF-SVR, BPNN-XGBoost-SVR and BPNN-GARF-SVR are designed as the hybrid models. The data used in this paper is the measured data of the ship sailing in the sea, whose sampling frequency is 10 Hz. Three common criteria, root mean square error (RMSE), the mean absolute deviation (MAE) and the mean absolute percentage

error (MAPE) were adopted to measure the prediction performances of the models. The definition of the RMSE, MAE, MAPE are as follows:

$$\begin{aligned}
 RMSE &= \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}(t+q) - x(t+q))^2} \\
 MAE &= \frac{1}{N} \sum_{i=1}^N |\hat{x}(t+q) - x(t+q)| \\
 MAPE &= \frac{100}{N} \sum_{i=1}^N \left| \frac{\hat{x}(t+q) - x(t+q)}{x(t+q)} \right|
 \end{aligned} \tag{23}$$

where $\hat{x}(t+q)$ and $x(t+q)$ represent predicted and actual values at time $t+q$, respectively. N is the number of forecasting samples.

For the model mentioned in the article, parameter adjustment is very important, which helps to improve the prediction accuracy of the model. The parameters of used models were set and determined as the following. For ARMA model, orders p and q were determined by Akaike Information Criterion (AIC) with the coefficients α and ε were calculated by recursive least square method. The decomposition parameters of EEMD are set according to common settings where $\sigma = 0.1$, $N = 100$. In the BPNN model, the parameters with better performances are adopted as the number of neurons. Therefore, a BPNN structure containing an input layer, two hidden layers, and an output layer is designed in the text, where the number of neurons is 100, 10, 10, 100, respectively. The weight coefficients and offset coefficients between the layers are first randomly initialized and then optimized by the gradient descent method. The Gaussian function is used as the kernel function in the SVR model. The XGBoost use a grid search method to determine multiple parameters. In the GARF model, the range of values of the two parameters are taken as [10,160] and [1,64] and intervals are 10 and 1, respectively, therefore, the number of coded bits is 10 bits. The number of populations and the number of iterations is both set to 50, the mean square error of predicted and true values is adopted as the fitness function of the model.

A set of 100 s with a sampling frequency of 10 Hz is created for EEMD decomposition and the Fig. 4 is the result of EEMD. The ship motion series is divided into 8 IMFs and one residual. The original sequence is divided into 8 IMFs and one residual, and all IMFs are arranged in the order in which they are extracted, with frequencies from high to low. It can be seen from the decomposition results that the original sequence is mainly controlled by low frequency sequences such as IMF2, IMF3, and IMF4. In all IMFs, the IMF1 with the highest frequency is considered as noise, and the remaining IMFs and residuals are re-added into a new approximate sequence, the complexity of which is reduced.

The Fig. 5 shows the instantaneous frequency of each IMF sequence. The fold lines are IMF1~IMF5 from top to bottom, and the frequency is also from high to low, which is consistent with the decomposition result of the Fig. 5. It can be seen that the main IMFs are clearly separated into different frequency components and there is no mode mixing.

This paper used 0.1 s as the initial prediction step, and each prediction is increased by 0.1 s until 9.9 s. The Fig. 5 describes the true value and the predicted value curve for the 50-step prediction of different models. The Table 3 shows the prediction accuracy of the 10 prediction models from the three evaluation criteria RMSE, MAE, MAPE when predicting 50 steps.

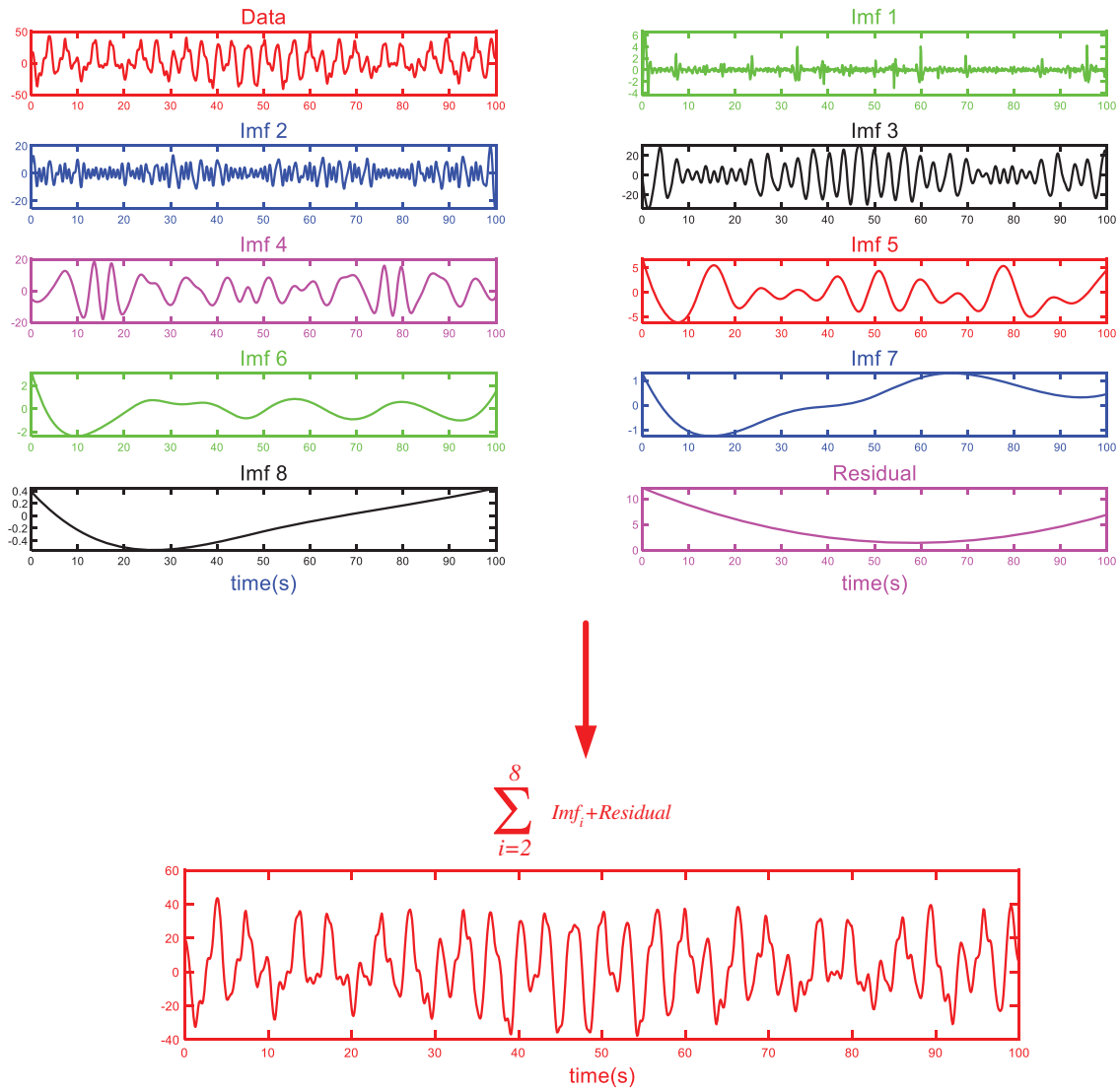


Figure 4: EEMD decomposition results and reconstruction sequences

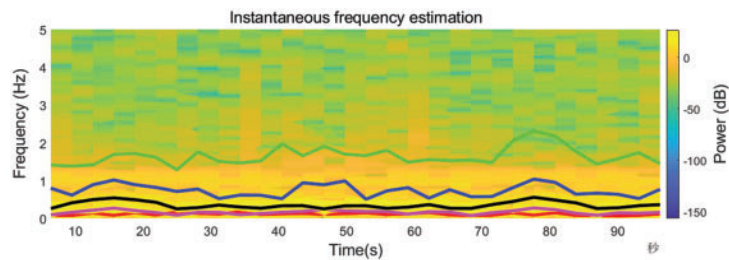


Figure 5: Instantaneous frequency estimation of the IMFs

As can be seen from the Fig. 6 and Table 3, as a linear model which is easy to construct, the prediction accuracy of ARMA model is inferior to all other models when it is used to predict the

complex nonlinear ship motion. As a model with strong nonlinear fitting ability, BPNN and XGBoost models behave much better than ARMA model, but they are inferior to SVR and RF models in a single model. The RF model optimized by the genetic algorithm has better performance, but the improvement effect brought by the genetic algorithm cannot make the RF model exceed the SVR model, whose performance is the best in the single model when predicting 50 steps, and the RMSE, MAE, MAPE of which is 10.26, 7.47, 0.91.

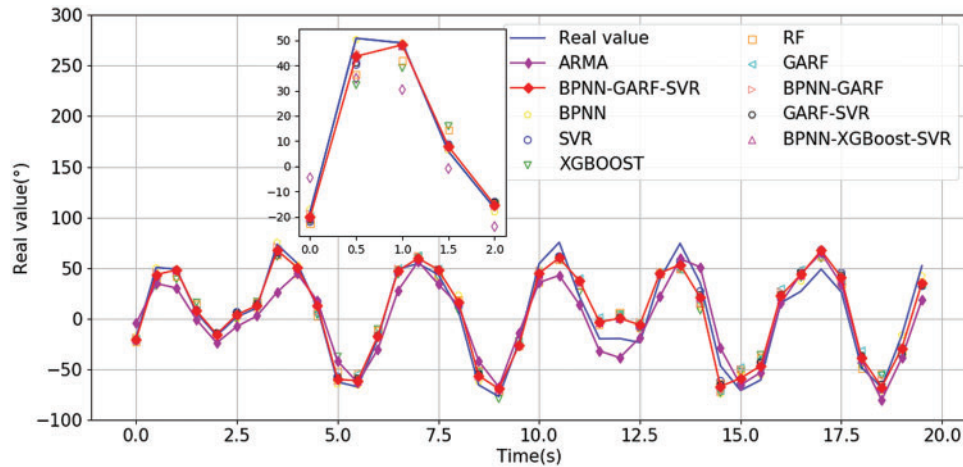


Figure 6: Ship motion prediction curve for all models

Table 3: Prediction accuracy of each model when the forecast step is 50

Forecasting models	RMSE	MAE	MAPE (%)
ARMA	25.17	19.41	3.92
BPNN	12.15	8.16	0.98
SVR	10.26	7.47	0.91
XGBoost	12.40	9.27	1.44
RF	11.38	8.58	1.18
GARF	10.86	7.91	1.15
BPNN-GARF	10.22	7.24	1.01
GARF-SVR	10.08	7.34	0.95
BPNN-XGBoost-SVR	9.71	6.84	0.84
BPNN-GARF-SVR	9.69	6.84	0.85

Among the four combined models used in the article, BPNN-GARF has similar performance with SVR while the other three models GARF-SVR, BPNN-GARF-SVR, BPNN-XGBoost-SVR have higher precision than the SVR model, which indicates the prediction method designed in the paper is at least non-inferior combination. GARF-SVR can be regarded as a redundant combination forecast model because compared to the SVR model (RMSE = 10.26, MAE = 7.46, MAPE = 0.91) and the GARF model (RMSE = 10.87, MAE = 7.91, MAPE = 1.15), three evaluation criteria of GARF-SVR model (RMSE = 10.08, MAE = 7.34, MAPE = 0.95) only increased by 0.18, 0.12, -0.04 (1.7%,

1.7%, -5%) and 0.79, 0.57, 0.20 (7.3%, 7.2%, 1.7%), respectively. The RMSE and MAE criteria only increased by less than 2% compared with the SVR, and the MAPE even decreased by 5%.

From the Fig. 5, it can be seen that the combined prediction results in the GARF-SVR mainly depend on the SVR, and the weight coefficient of GARF, which cannot provide much useful information for the final forecast result, is only 0.25 while the coefficient of SVR is 0.75. BPNN-GARF is a typical superior combination model. Compared with BPNN and GARF, the performance of BPNN-GARF (RMSE = 10.22, MAE = 7.24, MAPE = 1.01) is increased by 1.94, 0.09, -0.03 (16.0%, 11.3%, -3.0%) and 0.65, 0.68, 0.14 (6.0%, 8.6%, 1.3%), respectively, at least to the extent of SVR. BPNN-GARF combination model can be depicted as an excellent combination that achieve better results than a single model by combining a better model with a poorer model. The other two combined models BPNN-GARF-SVR and BPNN-XGBoost-SVR are basically similar in performance, because the most useful information of these two models is contributed from SVR, and GARF and XGBoost only provide a little effective information, which can be seen from the weighting coefficients. In the BPNN-GARF-SVR combination model, the weight coefficients of BPNN, GARF, and SVR are 0.30, 0.18, and 0.52, which are the same as the BPNN, XGBoost, and SVR weight coefficient bases in the BPNN-GARF-SVR model, 0.32, 0.09, and 0.59. Since XGBoost is inferior to GARF, the latter has a higher weighting coefficient than the former. Compared with the SVR model, the BPNN-GARF-SVR model has improved by 0.57, 0.63, 0.53, (5.5%, 8.4%, 5.9%). For the BPNN and GARF models, the improvement of the combined model is more obvious that the performance is increased by -2.46 , -1.32 , -0.12% (20.2%, 16.1%, 12.2%) and -1.17 , -1.078 , -0.29 (10.8%, 13.6%, 25.6%), respectively. From the above conclusions, by combining a plurality of individual prediction models, an integrated model with better performance can be obtained.

Figs. 7a–7c show the relationship between prediction accuracy and prediction step size. It can be observed that when the prediction step size is less than 10, the BPNN effect is the best as as the prediction step size increases, the SVR and GARF models gradually achieve better results. GARF model is significantly better than RF, but the optimization effect gradually decreases with time. This is because the genetic algorithm optimization result is adopted as a fixed parameter for RF model which does not change with the prediction step size increases. The prediction result curve from GARF-SVR also concludes that this is a redundant prediction model because the performance curve of GARF-SVR almost coincides with the curve of SVR.

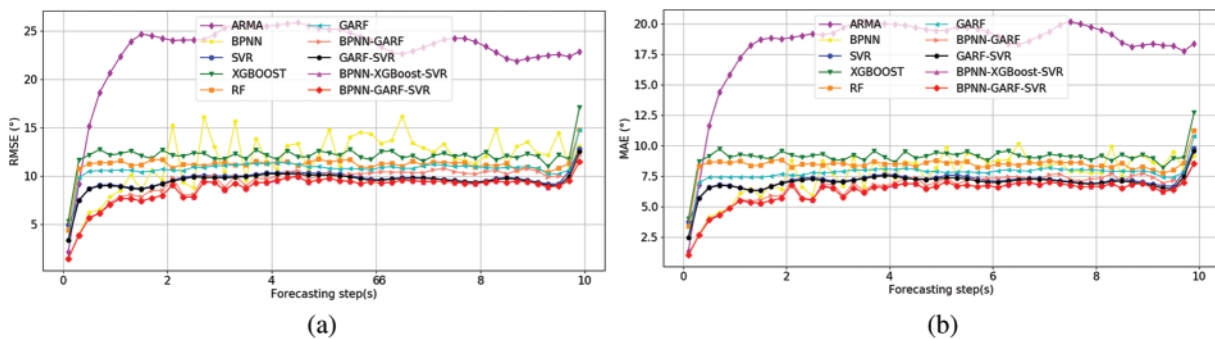


Figure 7: (Continued)

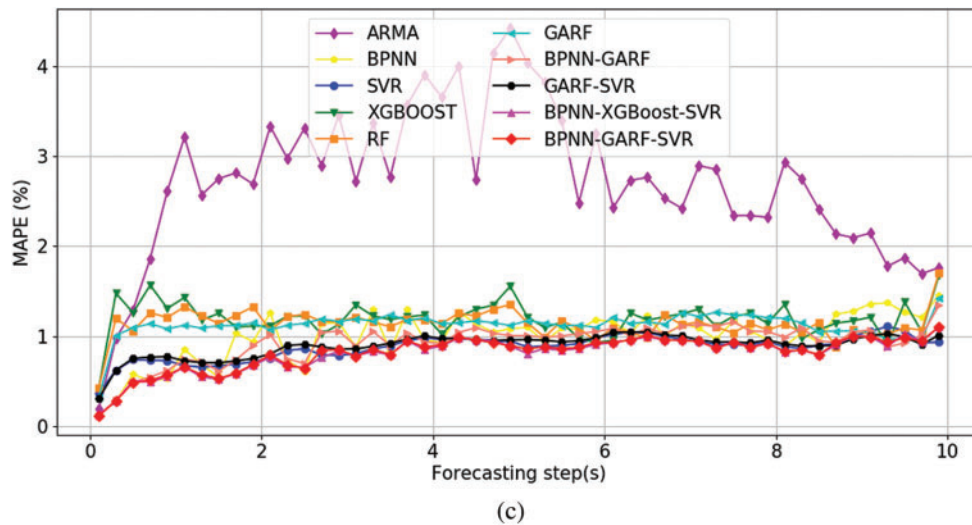


Figure 7: The curve of prediction accuracy and forecast step size

In the BPNN-GARF/XGBoost-SVR model, the performance is always superior to the current single prediction model. When the prediction step is small, BPNN is better than SVR and GARF and the combined model is better than BPNN. With the increase of the prediction step size, SVR is better than BPNN and GARF, and the combined prediction model is superior to SVR. Compared to BPNN and SVR, GARF and XGBoost have small weight coefficients in combined forecasting, but they also provide some useful information to the final result.

The Fig. 8 shows the variation of the weight coefficient with the predicted step size. At the beginning, BPNN’s weight coefficient is the highest, then gradually decreases and tends to stabilize, while the SVR coefficient is gradually increasing. The coefficient of GARF is relatively low, and sometimes close to zero. Therefore, SVR provides the main valid information, BPNN provides secondary useful information, and GARF also provides useful information, but sometimes it may be redundant.

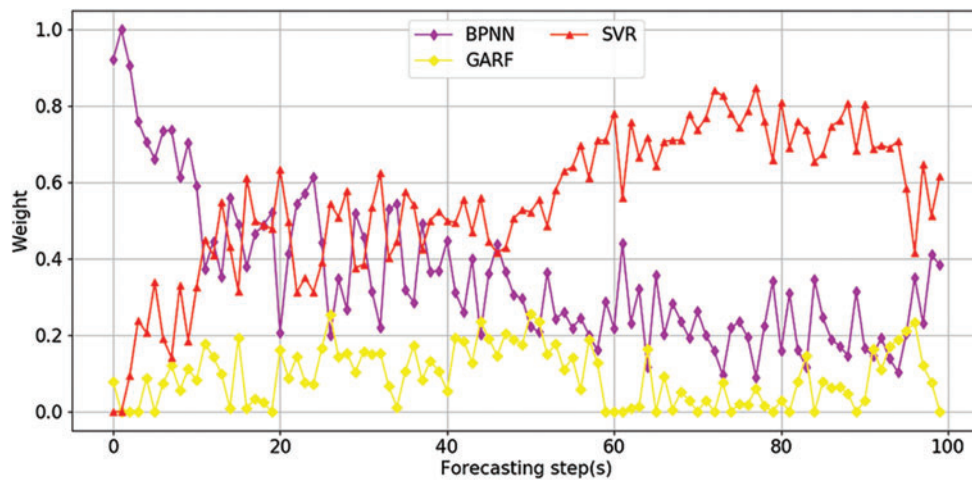


Figure 8: The curve of BPNN-GARF-SVR weight coefficients and forecast step size

The Fig. 9 shows the correlation coefficient between the predicted value and the real value, which provides another perspective for evaluating the performance of the model, and the weight coefficient in the combined model is also determined according to the correlation coefficient of each single model. The correlation coefficient of the combined model is obviously higher than the single model, which is consistent with the results of three evaluation criteria.

The Figs. 10a–10d show the correlation coefficients of BP, GARF, SVR and BPGARF when predicting 50 steps. The correlation coefficient of the combined model is higher than the correlation coefficient of the single model, which is consistent with the previous analysis.

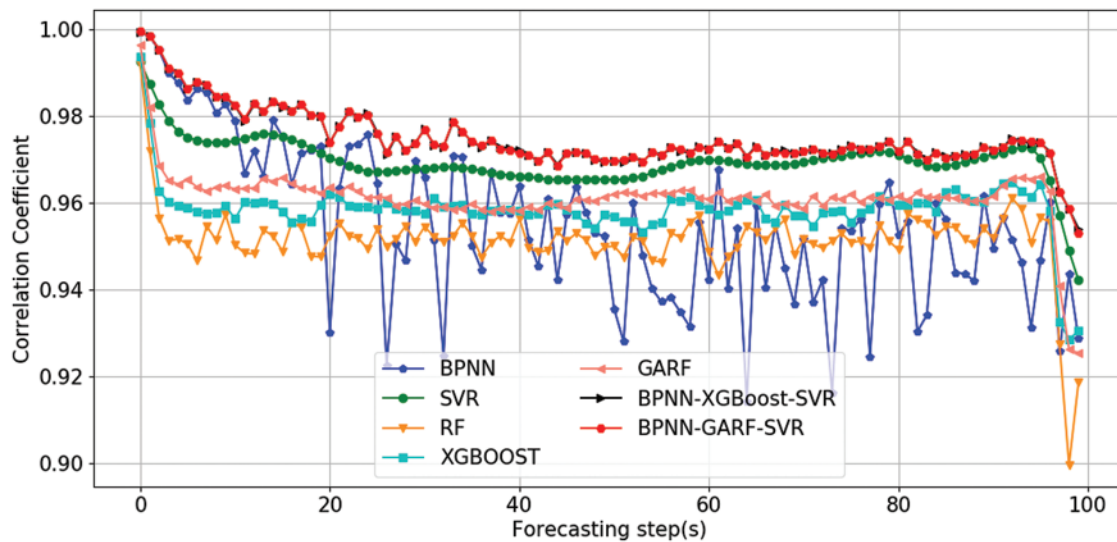
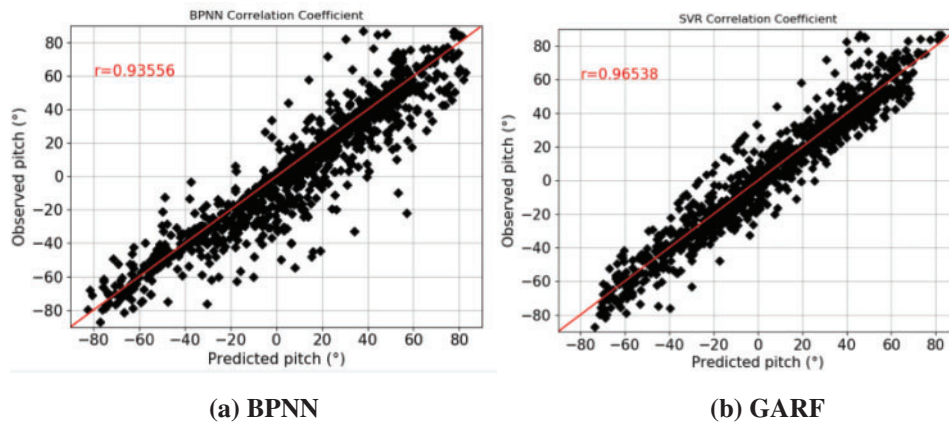


Figure 9: The curve of correlation coefficient and forecast step size



(a) BPNN

(b) GARF

Figure 10: (Continued)

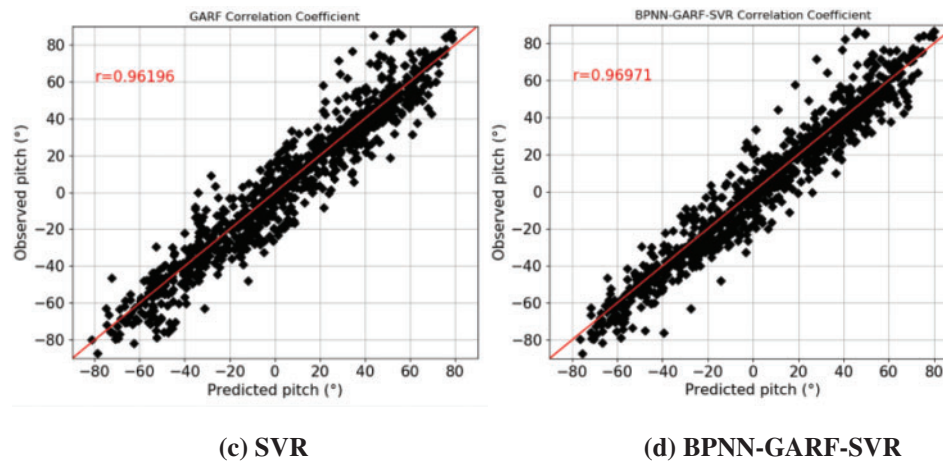


Figure 10: The correlation coefficient of the BPNN, GARF, SVR and BPNN-GARF-SVR when forecast step is 50

4 Conclusion

In this paper, we proposed a hybrid forecasting method EEMD-BPNN-GARF-SVR for ship motion. EEMD is used to remove high frequency noise components from the original sequence and obtain reconstructed sequences. Three independent predictive models were first established based on the reconstructed data, including BPNN, SVR and RF models. The final prediction results are obtained by combining the prediction results of three independent prediction models. Experiments show that for the measured data, when the prediction compensation is 50 steps, the RMSE, MAE and MAPE of the BPNN-GARF-SVR model are 9.69, 6.84 and 0.85, respectively. The RMSE of the single models is greater than 10, the MAE is greater than 8, and the MAPE is greater than 1. It can be seen that through the combination of multiple models, the prediction result is improved compared to the single prediction model.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Yumori, I. S. A. O. (1981). Real time prediction of ship response to ocean waves using time series analysis. *OCEANS 81*, pp. 1082–1089. Piscataway: IEEE.
2. Zhang, X. S., Peng, X. Y., Zhao, X. R. (2002). Diagonal recurrent neural network algorithm for extreme short prediction of ship motion. *Acta Simulata Systematica Sinica*, 2002(5), 641–642+649. DOI 10.16182/j.cnki.joss.2002.05.029.
3. Peng, X. Y., Zhao, X. R., Wei, N. X. (2001). AR algorithm for extremely short-term prediction of large ship's motion. *Ship Engineering*, 2001(5).
4. Peng, X. Y., Zhao, X. R., Gao, Q. F. (2007). Research on real-time prediction algorithm of ship attitude motion. *Journal of System Simulation*, 19(2), 267–271. DOI 10.16182/j.cnki.joss.2007.02.011.
5. Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H. et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal*

- Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903–995. DOI 10.1098/rspa.1998.0193.
6. Huang, L. M., Duan, W. Y., Han, Y., Yu, D. H., Elhandad, A. (2015). Extending the scope of AR model in forecasting non-stationary ship motion by using AR-EMD technique. *Journal of Ship Mechanics*, 19(9), 1033–1049. DOI 10.3969/j.issn.1007-7294.2015.09.002.
 7. Duan, W. Y., Huang, L. M., Han, Y., Zhang, Y. H., Huang, S. (2015). A hybrid AR-EMD-SVR model for the short-term prediction of nonlinear and non-stationary ship motion. *Journal of Zhejiang University-Science A*, 16(7), 562–576. DOI 10.1631/jzus.A1500040.
 8. Wu, Z., Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1), 1–41. DOI 10.1142/S1793536909000047.
 9. Zhou, J., Yu, X., Yuan, X. (2018). Predicting the carbon price sequence in the Shenzhen emissions exchange using a multiscale ensemble forecasting model based on ensemble empirical mode decomposition. *Energies*, 11(7), 1907. DOI 10.3390/en11071907.
 10. Zhang, Y. S., Zhou, L. X., Cai, F., Shi, A. G., Yang, B. Z. (2002). A new technology of forecasting the takeoff/touchdown opportunity of ship-born helicopter in heavy sea. *Navigation of China*, 2002(4), 7–12.
 11. Peng, X., Men, Z., Wang, X., Jia, S. (2014). The ship motion prediction approach based on BP neural network to identify Volterra series kernels. *The 26th Chinese Control and Decision Conference (2014 CCDC)*, pp. 2324–2328. Piscataway, IEEE.
 12. Huang, B. G., Zou, Z. J. (2016). Short-term prediction of ship pitching motion based on artificial neural networks. *International Conference on Offshore Mechanics and Arctic Engineering*, Busan South Korea, V007T06A007. American Society of Mechanical Engineers. DOI 10.1115/OMAE2016-54317.
 13. Peng, X., Dong, H., Zhang, B. (2017). Echo state network ship motion modeling prediction based on Kalman filter. *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 95–100. Piscataway, IEEE.
 14. Liu, X., Huang, Y., Wang, Q., Song, Q., Zhao, L. (2016). A prediction method for deck-motion of air-carrier based on PSO-KELM. *2016 10th International Conference on Sensing Technology (ICST)*, pp. 1–5. Piscataway, IEEE.
 15. Luo, W., Ren, J. (2016). On the identification of coupled pitch and heave motions using support vector machine. *2016 Chinese Control and Decision Conference (CCDC)*, pp. 3316–3321. Piscataway, IEEE.
 16. 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(2), 661–680. DOI 10.1016/j.neucom.2015.09.089.
 17. Chen, J., Li, M., Wang, W. (2012). Statistical uncertainty estimation using random forests and its application to drought forecast. *Mathematical Problems in Engineering*, 2012(1), 1–12. DOI 10.1155/2012/915053.
 18. Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system ACM. *SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. New York, ACM.
 19. Chen, H. Y. (2006). Research on properties of superior combined forecasting models based on correlation coefficients. *Journal of Systems Engineering*, 21(4), 353–360. DOI 10.1007/s11769-006-0026-1.
 20. Li, P., Zhang, J. S. (2018). A new hybrid method for China's energy supply security forecasting based on ARIMA and XGBoost. *Energies*, 11(7), 1687. DOI 10.3390/en11071687.
 21. Tiao, G. C., Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series. *Journal of Forecasting*, 13(2), 109–131. DOI 10.1002/(ISSN)1099-131X.
 22. Taieb, S. B., Bontempi, G., Atiya, A. F., Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications*, 39(8), 7067–7083. DOI 10.1016/j.eswa.2012.01.039.