Applied Soft Computing 62 (2018) 478–490

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

GMDH-based hybrid model for container throughput forecasting: Selective combination forecasting in nonlinear subseries

Lili Mo [a], Ling Xie [a], Xiaoyi Jiang [b], Geer Teng [c], Lixiang Xu [d], Jin Xiao [a,*]

[a] Business School, Sichuan University, Chengdu 610064, China
[b] Department of Mathematics and Computer Science, University Münster, Münster 48149, Germany
[c] Institute of Social Development and Western China Development Studies, Sichuan University, Chengdu 610064, China
[d] Department of Mathematics and Physics, Hefei University, Hefei 230601, China

ARTICLE INFO

Article history:
Received 8 March 2017
Received in revised form 6 October 2017
Accepted 20 October 2017
Available online 5 November 2017

Keywords:
Container throughput forecasting
Hybrid model
GMDH neural network
Selective combination forecasting

ABSTRACT

The accurate forecasting of future container throughput is important for the construction, upgrade, and operation management of a port. This study introduces group method of data handling (GMDH) neural network and proposes a hybrid forecasting model based on GMDH (HFMG) to forecast container throughput. This model decomposes the original container throughput series into two parts: linear trend and nonlinear variation, and uses the seasonal autoregressive integrated moving average (SARIMA) approach to predict the linear trend. Considering the complexity of forecasting nonlinear subseries, the proposed model adopts three nonlinear single models, namely, support vector regression (SVR), back-propagation (BP) neural network, and genetic programming (GP), to predict the nonlinear subseries. Then, the model establishes selective combination forecasting by the GMDH neural network on the nonlinear subseries and obtains its combination forecasting results. Finally, the predictions of two parts are integrated to obtain the forecasting results of the original container throughput time series. The container throughput data of Xiamen and Shanghai Ports in China are used for empirical analysis, and the results show that the forecasting performance of the HFMG model is better than that of SARIMA model, as well as some hybrid forecasting models, such as SARIMA-SVR, SARIMA-GP, and SARIMA-BP. Finally, the monthly out-of-sample forecasts of container throughput for the two ports throughout 2016 are given.

1. Introduction

With the deepening of economic globalization and the frequent exchanges of international trade, container transportation has played an important role in reducing transportation time and trade cost. Therefore, major ports in the world have striven to develop containerization. The World Bank statistics has shown that the global container throughput in 2014 reached 679.2647 million twenty-foot equivalent units (TEU), which has increased by 80% since 2005. With the upsurge of container transport development, the booming construction of ports resulted in a number of problems, such as overcapacity and declined throughput capacity utilization [29]. Container port construction requires a long period and large investment funds. Once over-construction occurs, some issues, such as the excess of capacity and failure to gain the expected profit, will lead to a huge waste of time and capital [51]. Therefore, it is very important to predict the future container throughput for adjusting the port development direction, making port operation schedules, planning the port scale, and reducing resource waste [9].

1.1. Literature review

The research on container throughput forecasting began in the 1980s and has considerably developed. Existing forecasting methods mainly include single models and hybrid approaches. The single models use only one model for prediction and can be divided into three types [50].

(1) Time series models include exponential smoothing (ES), autoregressive integrated moving average (ARIMA) [1,38], seasonal autoregressive integrated moving average (SARIMA) [41,31,14], vector autoregressive (VAR) [45], decomposition approach [35], and grey forecasting (GM) [7,16,28]. For example, Schulz and Prinz [41] forecasted the quarterly container throughput series in Germany using the SARIMA and Holt–Winters ES models, in which the former performed slightly better than the latter. Syafi'i et al. [45] proposed a vector error correction model to forecast container

* Corresponding author.
E-mail address: xjxiaojin@126.com (J. Xiao).

https://doi.org/10.1016/j.asoc.2017.10.033
1568-4946/

throughput based on the VAR model and verified good forecasting performance. Guo, et al. [16] proposed an improved gray Verhulst model and overcame the deficiency of the increasing forecasting error of the GM (1,1) model as the container throughput grows in the S-curve.

(2) Causal analysis models [42,10,34] include regression and elasticity coefficient analysis. For instance, Chou, et al. [10] proposed an improved regression model to predict the container throughput of Taiwan Port; the model performed better than the traditional regression model. Patil and Sahu [34] used a regression model for predicting the freight flow at Mumbai Port and implemented an elasticity analysis to identify the impact factors of freight flow.

(3) Nonlinear dynamic forecasting models include artificial neural network [15,26,43,37,25], genetic programming (GP) [8], and support vector regression (SVR) [30]. For example, Gosasang, et al. [15] achieved better results in predicting the container throughput of Bangkok Port by using a multilayer perceptron neural network than a linear regression model. Chen and Chen [8] adopted the GP, X-11, and SARIMA models to predict the container throughput of major ports in Taiwan, and the comparison results showed that the GP model performed best. Mak and Yang [30] proposed an approximate least squares support vector machine (ALSSVM) model to predict the container throughput of the Hong Kong Port and compared with support vector machine (SVM), least squares SVM, and radial basis function neural network, which showed that the ALSSVM model was the best overall.

The container throughput time series is usually complex; thus, a single model based on linear assumptions or a nonlinear dynamic model often cannot obtain satisfactory forecasting performance [39,21,2]. An increasing number of researchers have constructed hybrid forecasting models [17,13,51,52,9,53] to solve this problem. For example, Xie et al. [52] proposed three hybrid models based on the least squares support vector regression (LSSVR). The empirical analysis based on the container throughput time series of Shanghai and Shenzhen Ports showed that the proposed hybrid models performed better than the single models.

1.2. Problem statement

The above studies have paid significant attention to hybrid models because their prediction performance is usually better than that of the single models. An efficient hybrid forecasting model may be constructed in three ways: (1) The embedded method embeds one model into another (i.e., optimizes the parameters of one model with another, such as the References [17,13]); (2) The divide and rule method, which first decomposes the original time series into several subseries, and then models and predicts each subseries by an appropriate model, finally integrates the prediction results according to certain rules, such as the References [51,52]. This method is popular. However, given that some of the subseries decomposed from the original time series are often highly nonlinear, accurately forecasting for those subseries using a single nonlinear model is difficult, and a large prediction error on some subseries leads to a poor prediction performance in the entire time series. (3) Combination forecasting method adopts several models to predict the original time series and assigns a weight to the forecasting results of each model in order to obtain the final combined results, such as Refs. [9,53]. This method is convenient and simple; however, it usually allows the combination of all trained models, and multi-collinearity among the models may exist, which will degrade the forecasting accuracy of the model. Forecasting performance can be improved by selecting and combining the forecasting results of a subset of the models for a final decision, but the manner in which an optimal subset is derived still remains a challenge.

This study combines the latter two hybrid forecasting methods above and proposes a hybrid forecasting model based on group method of data handling (GMDH) [20], called HFMG. This model first decomposes the original time series into two parts: linear trend and nonlinear variation. The linear trend is the increase of the container throughput caused by long-term economic growth and predicted by the SARIMA model. The nonlinear variation mainly comes from the irregular changes and shocks of various factors in the economic system. Considering the complexity of nonlinear variation, the model trains three nonlinear single models, namely, SVR, BP neural network, and GP, and implements combination forecasting to recover the disadvantages of the second method. The reason why we choose these models is that they have been widely used to forecast container throughput and proven effective [41,31,14,37,25,8,30]. Further in combination forecasting, it introduces GMDH neural network proposed by Ivakhnenko [20] to compensate for the deficiency of the third method and establish the selective combination forecasting. The factor screening function of GMDH neural network can objectively and automatically choose factors that critically influence the research object [49]. Thus, GMDH can reduce the effect of multi-collinearity on the performance of the model to some extent. Finally, the predictions of two parts are integrated to obtain the forecasting results of the original container throughput time series. The empirical analysis verifies the effectiveness of the HFMG model.

This study is organized as follows. Section 2 briefly reviews several commonly used time series forecasting models and the basic theory of the GMDH neural network. Section 3 describes the modeling idea and detailed modeling steps of the proposed HFMG model. Section 4 compares the performance of the HFMG model and other hybrid models in two actual container throughput series and gives further out-of-sample forecasting. Section 5 concludes this study.

2. Related theories

2.1. Commonly used forecasting models

2.1.1. SARIMA

The ARIMA model proposed by Box and Pierce [5] in the 1970s is a commonly used time series forecasting model. Considering the seasonal patterns in the time series, the ARIMA model is extended to the SARIMA model [4]. A SARIMA process can be denoted as $SARIMA(p, d, q)(P, D, Q)_s$, which is expressed as follows.

$$\Phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D Y_t = \Theta(L)\Theta(L^s)\varepsilon_t, \quad (1)$$

where $\{Y_t\}$ denotes the time series, L is the lag operator, s is the period of series, p and q are the autoregressive and moving average process orders, respectively, d is the difference times, P and Q are the seasonal autoregressive and moving average process orders, respectively, D is the seasonal difference times, and ε_t is a white noise with zero-mean and constant variance, $\Phi(L) = 1 - \sum_{i=1}^{p}\varphi_i L^i$, $\Theta(L) = 1 - \sum_{i=1}^{q}\theta_i L^i$, $\Phi(L^s) = 1 - \sum_{i=1}^{P}\varphi_{is} L^{is}$, and $\Theta(L^s) = 1 - \sum_{i=1}^{Q}\theta_{is} L^{is}$, φ_i, θ_i, φ_{is}, θ_{is} are the coefficients.

2.1.2. BP neural network

BP neural network is a multi-layer feedforward neural network trained by the error inverse propagation algorithm by Rumelhart et al. [40] in 1986, and it has been widely used in many fields. The model can be obtained by adjusting the connection weights among the different neurons in the network to minimize the measurement error between the actual and expected output. Fig. 1 shows the structure of the BP neural network with three layers [6].

$$c_j = f_1\left(w_{oj} + \sum_{i=1}^{I} w_{ij} x_i\right), j = 1, 2, \ldots, J, \quad (2)$$

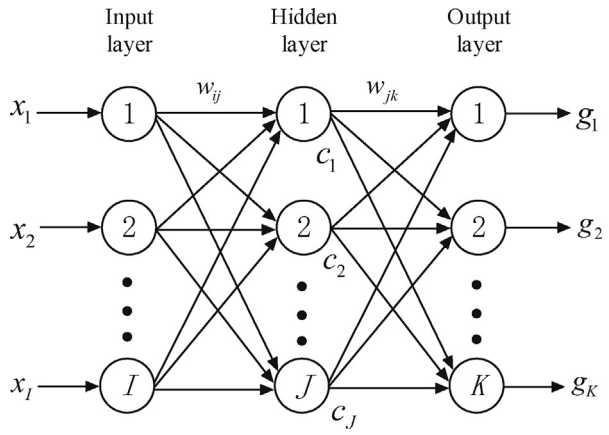


Fig. 1. The structure of three layers BP neural network.

$$g_k = f_2(w_{ok} + \sum_{j=1}^{J} w_{jk} c_j),\ k = 1, 2, \ldots, K, \tag{3}$$

where I, J, and K are the numbers of neuron nodes in each layer, $\{x_1, x_2, \ldots, x_I\}^T$ is the input vector, $\{c_1, c_2, \ldots, c_J\}^T$ is the neuron vector in the hidden layer, $\{g_1, g_2, \ldots, g_K\}^T$ is the final output vector, w_{oj} and w_{ok} are the threshold values in the input and hidden layers, respectively, w_{ij} is the connection weight between the i-th neuron node in the input layer and the j-th neuron node in the hidden layer, w_{jk} is the connection weight between the j-th neuron node in the hidden layer and the k-th neuron node in the output layer, and f_1 and f_2 are the activation functions.

To forecast the time series $\{Y_t\}$, we usually let $K = 1$, $I = r$ (r is a positive integer). Y_t is regarded as the output vector and the lag terms of Y_t from 1 to r order form the input vector $\{x_1, x_2, \ldots, x_I\}^T$. Thus, we can employ the BP neural network to predict the time series $\{Y_t\}$.

2.1.3. Support vector regression

Corinna and Vapnik [11] introduced SVM in 1995. SVM adopts the structural risk minimization principle, maps linear inseparable input vectors to high-dimensional feature space, and constructs linear decision surface to realize nonlinear decision-making function in the original space, and it has strong generalization ability. SVM can be applied to the regression problem (i.e., SVR). By constructing a real-valued function, the sample point is expected as far as possible from the real-valued function, whereas the deviation between the actual value of the sample and the output value of the function is as small as possible. Then, the regression problem is transformed into an optimization problem [44]. Taking the linear regression as an example, for a given set of samples S, $(x_i, y_i) \in S$, $i = 1, 2, \ldots, n$, the optimization problem can be defined as

$$\min\ \frac{1}{2}\|\lambda\|^2 \tag{4}$$

$$s.t.\ |<\lambda, x_i> + b - y_i| \le \xi,\ \ i = 1, 2, \ldots, n, \tag{5}$$

where $f(x) = <\lambda, x_i> + b$ is a linear regression function, $\xi > 0$ is a number given discretionarily, and b is a constant.

2.1.4. Genetic programming

GP was first proposed by Koza [24] based on the Darwinian biological evolution of natural selection theory to simulate biological evolution. The GP model first defines an operation symbol set and a variable set, and it randomly generates a certain number of binary trees as the initial population, where each tree represents a function. Afterwards, the model selects, copies, crossovers, and varieties, and it then calculates the fitness of each individual and eliminates those with poor fitness. The above process is repeated by every generation until the individual whose fitness is larger than the threshold value of the fitness appears. GP has good nonlinear processing ability and can generate the specific function expression; thus, it has been applied in many fields.

Suppose the operation symbol set is $S_F = \{+, -, \times, /\}$, the variable set is $S_V = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, and Fig. 2 shows the crossover process of the GP model. Before the crossover, the two trees are represented by $TR_1 = (x_2 - x_6) \times x_1 + x_4$ and $TR_2 = x_1/x_3 \times (x_2 + x_4)$. During crossing, the subtrees in the dashed box change their positions, and two new binary trees are obtained, which are represented by $TR_1 = (x_2 - x_6) \times x_1 + x_1/x_3$ and $TR_2 = x_4 \times (x_2 + x_4)$.

2.2. GMDH neural network

The GMDH neural network is the core technique of self-organizing data mining [32,48], and it can decide the variables that enter the model, structure, and model parameters in a self-organizing manner. Recently, the GMDH neural network has been successfully applied in engineering [33], medical treatment [23,22], and enterprise management [49,48,47,46].

By the GMDH neural network, a model can be represented as a set of neurons in which different pairs in each layer are connected through a polynomial, and each layer produces new neurons for next layer. For a problem, the training set W is divided into two parts: model learning set A and model selection set B in the GMDH neural network. Let $X = (x_1, x_2, \ldots, x_n)$ and y be the input vector and actual output, respectively, and n is the input vector dimension. Given M observations of multi-input, single-output data pairs $\{(y_i, X_i) | X_i = (x_{i1}, x_{i2}, \ldots, x_{in}),\ i = 1, 2, \ldots, M\}$ in set A, where i represents the i-th data pairs, it is expected to train a GMDH neural network for predicting the output value $\hat{y}_i$:

$$\hat{y}_i = \hat{f}(X_i),\ i = 1, 2, \ldots, M \tag{6}$$

The problem is determining a GMDH neural network to realize

$$\sum_{i=1}^{M} (y_i - \hat{y}_i)^2 \to \min \tag{7}$$

The general connection between the input and output variables can be expressed by a complicated discrete form of the Volterra functional series in the form of

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots, \tag{8}$$

which is also known as the Kolmogorov-Gabor (K-G) polynomial. This study uses the linear K-G polynomial in the form of

$$\hat{y} = a_0 + \sum_{i=1}^{n} a_i x_i \tag{9}$$

Let $V = \{v_1 = a_0, v_2 = a_1 x_1, \ldots, v_{n+1} = a_n x_n\}$ be the initial model set of the GMDH neural network. The following linear function is selected as the transfer function.

$$f(\omega_1, \omega_2) = b_0 + b_1 \omega_1 + b_2 \omega_2 \tag{10}$$

where b_0, b_1, and b_2 are the coefficients, ω_1 and ω_2 represent any two of the input models.

The GMDH neural network modeling process is as follows. First, in the input layer, the initial models in V are fed in pairs as the transfer function input in Eq. (10). They generate $C_{n+1}^2 = \frac{(n+1)n}{2}$

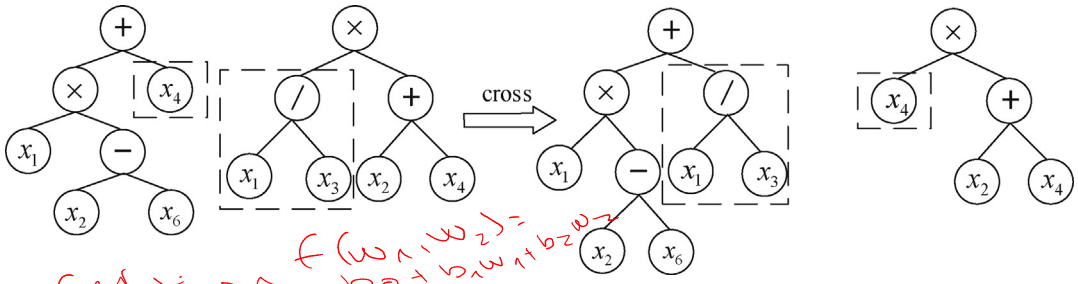


Fig. 2. Crossover of binary trees in GP model.

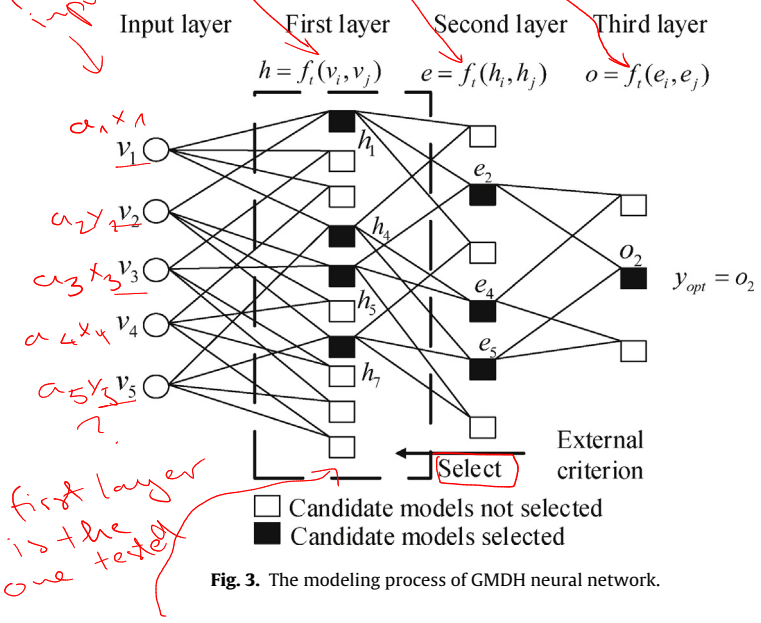


Fig. 3. The modeling process of GMDH neural network.

middle candidate models in the first layer, then, the model parameters are estimated on the model learning set A using the least squares method, and the external criterion values of candidate models are calculated on the model selection set B. Some best middle candidate models with the smallest external criterion values are selected. The selected middle candidate models are fed in pairs as the transfer function input. They generate a series of middle candidate models in the second layer and some superior candidate models are reserved. The process continues and stops when the optimal complexity model is identified using the termination principle, which is presented by optimal complexity theory. With the increased complexity of the middle candidate models, the external criterion values initially decrease and then increase, and the global minimum of the external criterion corresponds to the optimal complexity model y_{opt} (see Fig. 3). Finally, to seek the initial model contained in the optimal complexity model, y_{opt}, the recurrence of the GMDH network structure must be conducted from the last layer until the initial input layer is reached. Fig. 3 shows that the initial input models, v_1, v_2, v_3, and v_5, are chosen, that is, a_0, x_1, x_2, and x_4 are selected. However, v_4 is eliminated during the self-adaption process. Therefore, x_3 is eliminated.

3. Hybrid forecasting model based on GMDH

3.1. Basic idea

The economic system time series, which has linear trend and nonlinear variation, is usually complex. Therefore, the linear trend can be predicted by a linear model, and the nonlinear characteristic hidden in the time series can be captured by some nonlinear forecasting models. Existing studies usually adopt a single nonlinear forecasting model, such as the SVR, GP, and BP neural network. However, these nonlinear models have their own advantages and disadvantages. Combining the forecasting results of several nonlinear models (i.e., combination forecasting) and making them learn from each other can improve forecasting performance. The prediction results of different nonlinear forecasting models often have high multi-collinearity. Significant forecasting bias may be produced if all predicting results are combined. Therefore, selecting a subset of the nonlinear forecasting models for combination may avoid this problem and improve the forecasting performance of the model. Thus, choosing the basic nonlinear forecasting models and determining the optimal combination weights are crucial. The GMDH neural network can find the optimal complexity combined model according to optimal complexity theory and select the basic nonlinear models to automatically enter the optimal complexity model.

It is worth noting that the traditional GMDH neural network does not calculate the external criterion values of the initial models in the input layer (see Fig. 3) and the optimal complexity model may be sourced from all the middle candidate models, which do not include the initial models. For container throughput forecasting, the combination forecasting model can achieve better forecasting performance than the single model in most cases; however, the possibility that some single models outperform the combination forecasting model is not excluded. If a model is constructed based on the GMDH neural network, the nonlinear single models in the initial model set should not be excluded from the candidate set of the optimal forecasting model. Thus, this study improves the traditional GMDH neural network by calculating the external criterion values of the initial models and adding them to the candidate model set.

The proposed forecasting model, HFMG, in this study has three steps: (1) establishing the SARIMA model in the container throughput time series and obtaining the linear trend forecast; (2) generating the nonlinear residual subseries and adopting the BP, SVR, and GP models as the base single models to forecast the nonlinear residual subseries; and (3) constructing the combination forecasting model using the modified GMDH neural network and obtaining the final forecasting result of the container throughput time series.

3.2. External criteria construction

In the actual system modeling, we propose different requirements, which are the purposes of modeling or a prior knowledge of the system. In the GMDH neural network, the external criterion is a mathematical description of these specific requirements, which can select the optimal complexity model from the candidate model set. The numerous external criteria can be divided into two types. (1) Accuracy criteria are concerned with the error of the model in different parts of the sample data, mainly represented by the regularity criterion. (2) Consistency criteria examine the consistency of the models built for the same system in different parts of the

EXTERNAL CRITERION OPTIONS

- Accuracy criterion: model error on sample data represented by regularity criterion
- Regularity criterion: (Mathematics)
  - example from study of cohort patients in the dispensary world:
    - lost to follow-up = patients who abruptly stop going to pharmacies after given period without a known reason. ⇒ criterion of regularity is for what is considered regular time between visits (60, 90, 120, days) ⇒ if patient hasn't visited in 30 days that is regular
- consistency criterion: checking how consistent the model is over different parts of the data sample ⇒ usually represented by minimum bias criterion
- Minimum bias: criterion for minimizing bias.

STUDY CRITERION:

- ARC - average regularity criterion: $\Delta^2 = \frac{1}{N} \sum_{i \in W} (z_i - \hat{z}_i)$ (observation, prediction trained on W; size of W; ? sample in W; training set W)
- SSC
- Symetric stability criterion:
  $\sigma_A^2 = \sum_{i \in W} (z_i - \hat{z}_{i,A})$ trained on: dataset A
  $\sigma_B^2 = \sum_{i \in W} (z_i - \hat{z}_{i,B})$ — " — dataset B
  $s_W^2 = \sigma_A^2 + \sigma_B^2$
  i is the input to $\hat{z}_i$
- SMBC
- Symetric minimum bias criterion:
  $\eta_W^2 = \| \hat{z}_{i,A} - \hat{z}_{i,B} \|^2_{i \in W}$ ← 2-norm

sample data, which is mainly represented by the minimum bias criterion.

This study selected four different external criteria. The training set W is equally divided into model learning set A and model selection set B, and the four external criteria are defined as follows.

(1) Average regularity criterion (ARC)

$$\Delta^2(W) = \frac{1}{N}\sum_{i\in W}(z_i - \hat{z}_i(W))^2 \tag{11}$$

where N is the sample size of training set W, z_i is the observation of i-th sample in W, and $\hat{z}_i(W)$ is the forecasting of i-th sample in dataset W by the model learned in W. This criterion requires learning and selecting the model performed on the entire training set W.

(2) Symmetric stability criterion (SSC)

$$o^2(A) = \sum_{i\in W}(z_i - \hat{z}_i(A))^2 \tag{12}$$

$$o^2(B) = \sum_{i\in W}(z_i - \hat{z}_i(B))^2 \tag{13}$$

$$S^2(W) = o^2(A) + o^2(B) \tag{14}$$

where z_i is the observation of i-th sample in W, $\hat{z}_i(A)$ represents the forecasting of i-th sample in W by the model learned in dataset A, and $\hat{z}_i(B)$ represents the forecasting of i-th sample in W by the model learned in dataset B. Therefore, $o^2(A)$ denotes the forecasting error of the model learned in dataset A over the entire training set W, and $o^2(B)$ refers to the forecasting error of the model learned in dataset B over the entire training set W.

(3) Symmetric minimum bias criterion (SMBC)

$$\eta_{bs}^2(W) = \|\hat{z}_i(A) - \hat{z}_i(B)\|_{i\in W}^2 \tag{15}$$

where $\hat{z}_i(A)$ represents the forecasting of i-th sample in dataset W by the model learned in dataset A, and $\hat{z}_i(B)$ represents the forecasting of i-th sample in dataset W by the model learned in dataset B. This criterion requires learning the model in datasets A and B respectively, predicting the entire training set W, and computing the 2-norm of the difference value of the forecasting by the two models. The criterion indicates the ability of the established model to correctly respond to the samples, which are not used in modeling but have the same regularity and directly reflect the generalization ability of the model.

(4) Absolute noise immune criterion (ANIC)

$$\rho^2(A) = \sum_{i\in A}(\hat{z}_i(A) - \hat{z}_i(W))^2 \tag{16}$$

$$\rho^2(B) = \sum_{i\in B}(\hat{z}_i(B) - \hat{z}_i(W))^2 \tag{17}$$

$$\rho^2(W) = \rho^2(A) + \rho^2(B) \tag{18}$$

where $\hat{z}_i(A)$ represents the forecasting of i-th sample in dataset A by the model learned in the same dataset, $\hat{z}_i(W)$ in Eq. (16) represents the forecasting of i-th sample in dataset A by the model learned in W, $\hat{z}_i(B)$ represents the forecasting of i-th sample in dataset B by the model learned in the same dataset, and $\hat{z}_i(W)$ in Eq. (17) represents the forecasting of i-th sample in dataset B by the model learned in W. Thus, $\rho^2(A)$ refers to the deviation between the forecasts made by the models learned in datasets A and W respectively over dataset A, $\rho^2(B)$ denotes the deviation between the forecasts made by the models learned in datasets B and W respectively over dataset B.

Among the four external criteria, ARC and SSC are the accuracy criteria, whereas SMBC and ANIC are the consistency criteria. We constructed different versions of the combination forecasting models, GMDH-ARC, GMDH-SSC, GMDH-SMBC, and GMDH-ANIC, based on the above external criteria.

3.3. Modeling process

The HFMG modeling process proposed in this study is shown in Fig. 4. The original container throughput time series is $\{Y_t\}(t = 1, 2, \ldots, T_0)$, where T_0 is the sample size of the series. The detailed modeling steps are as follows.

Step 1: The SARIMA model is established in the container throughput time series, and the linear trend forecasting is obtained.

a.The original time series is processed through difference and seasonal difference until it passes the unit root test (i.e., time series is stationary).

b.The appropriate values of the parameters, p, d, q, P, D, and Q, are selected to construct the SARIMA models by analyzing the autocorrelation and partial autocorrelation graphs of the stationary series, and the optimal SARIMA model is selected based on the Akaike Information Criterion (AIC) and the Schwarz Criterion (SC).

c.The selected SARIMA model is adopted to predict the linear trend, and the result is denoted as $\{\hat{l}_t\}$.

Step 2: The nonlinear residual subseries is generated, and the BP, SVR, and GP models are adopted as the base single models to forecast the nonlinear residual subseries.

a. The nonlinear residual subseries is denoted as $\{u_t\} : u_t = Y_t - \hat{l}_t$.

b. The BP, SVR, and GP models are utilized to predict the residual subseries, and the forecasts are denoted as $\{\hat{u}_t^i\}(i = 1, 2, 3)$.

Step 3: The combination forecasting model is constructed by the GMDH neural network, and the final forecasting result of the container throughput time series is obtained.

a. Let u_t be the dependent variable and $X = (\hat{u}_t^1, \hat{u}_t^2, \hat{u}_t^3)$ be the independent variables, the new training set is $W = (u_t, X)$, and W is equally divided into the model learning set A and the model selection set B.

b. The general relationship between the dependent and independent variables is generated as follows. $u_t = a_0 + a_1\hat{u}_t^1 + a_2\hat{u}_t^2 + a_3\hat{u}_t^3$. The subsections are regarded as the initial models of the GMDH neural network: $v_1 = a_0, v_2 = a_1\hat{u}_t^1, v_3 = a_2\hat{u}_t^2$, and $v_4 = a_3\hat{u}_t^3$.

c. Regard the initial models as candidate models, compute their external criterion values in the model selection set B.

d. Combine pairs of the initial models to generate C_4^2 (=6) candidate models in the first layer, estimate their parameters by least squares method in the model learning set A, compute the external criterion values in set B. Further, three candidate models with the lowest external criterion values are selected and combined in pairs to generate C_3^2 (=3) candidate models in the second layer.

e. The process continues and stops when the optimal complexity combination model y_{opt} is identified by optimal complexity theory. $\{\hat{u}_t^i\}(i = 1, 2, 3)$ is then substituted into y_{opt} to obtain the nonlinear combination forecasting result, denoted as $\hat{u}_t$.

f. The final prediction result of the original container throughput time series is obtained: $\hat{Y}_t = \hat{l}_t + \hat{u}_t$.

3.4. Example

To demonstrate the modeling process of the HFMG model, we provide an example on series N0671 from M3-competition [19]. It is a quarterly time series with 36 data points. For convenience and easy to show, here we choose the latest 12 data points of the series to model, and the original data is marked as Y_t. Then the process of HFMG model can be explained as follows:

Step 1: The SARIMA$(0, 0, 1)(0, 1, 1)_4$ model is selected based on AIC and SC criteria after stabilizing process, autocorrelation and partial autocorrelation analysis. Establish the model on Y_t, obtain the linear trend forecasting and denote it as $\hat{l}_t$.

Step 2: Generate the nonlinear residual subseries $u_t = Y_t - \hat{l}_t$ first, next utilize the BP, SVR, and GP models to predict the residual

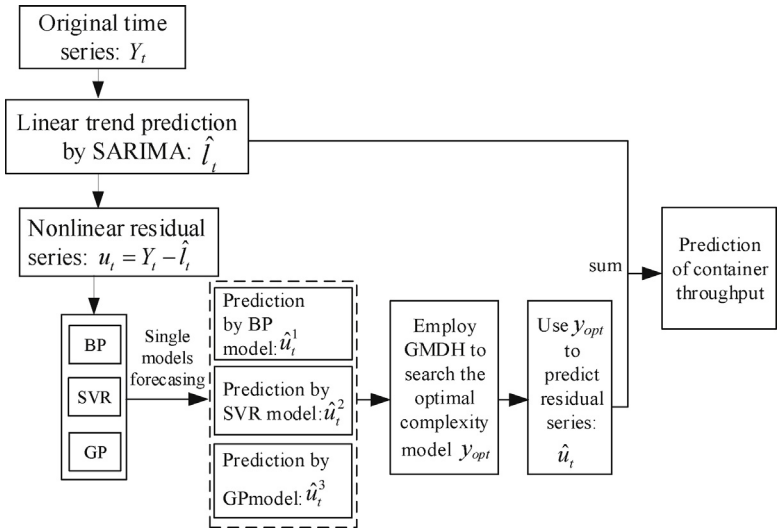


Fig. 4. Flowchart of the HFMG model.

Table 1
The original data and prediction of the time series N0671.

Original data: Y_t	2116	2321	2424	3259	2228	2419	2684	3577	2490	2667	2983	3894
Linear trend prediction: $\hat{l}_t$					2310.5	2463.3	2674.9	3668.9	2434.9	2750.8	2959.8	3912.2
Nonlinear subseries: u_t					−82.5	−44.3	9.1	−91.8	55.1	−83.8	23.2	−18.2
Prediction of u_t by BP: $\hat{u}_t^1$						1.4	−12.1	−55.3	48.2	−83.4	5.0	−55.7
Prediction of u_t by SVR: $\hat{u}_t^2$						−41.7	6.5	−89.1	57.7	−81.1	20.4	−20.8
Prediction of u_t by GP: $\hat{u}_t^3$							38.2	−41.1	72.3	−86.1	48.7	−47.7
Final prediction of HFMG: $\hat{Y}_t$						2421.7	2681.4	3579.7	2492.5	2669.7	2980.2	3891.4

subseries respectively, and the forecasting results are denoted as $\{\hat{u}_t^i\}$, $i = 1, 2, 3$. The Y_t, $\hat{l}_t$, u_t, and $\{\hat{u}_t^i\}$ are all shown in Table 1. It is worth noting that the SARIMA, BP, SVR and GP models need the lag terms of the dependent variable to model, thus, some data points do not have forecasting results for lack of lag terms values. Specifically, the SARIMA model contains a four-order lag term, hence, the former four data of $\hat{l}_t$ do not have values; both the BP and SVR models contain a one-order lag term, so the first value of $\hat{u}_t^1$ and $\hat{u}_t^2$ do not have predictions; the GP model has a two-order lag term, thus it loses two prediction values in $\hat{u}_t^3$.

Step 3: Construct the combination forecasting model by GMDH neural network. Here the ANIC is selected as the external criterion.

a. Let u_t be the dependent variable and $X = (\hat{u}_t^1, \hat{u}_t^2, \hat{u}_t^3)$ be the independent variable; the new training set is $W = (u_t, X)$. Divide W equally into model learning set A and model selection set B.

b. Select K-G polynomial to describe the relationship between u_t and X: $u_t = a_0 + a_1\hat{u}_t^1 + a_2\hat{u}_t^2 + a_3\hat{u}_t^3$, and get four initial models: $v_1 = a_0$, $v_2 = a_1\hat{u}_t^1$, $v_3 = a_2\hat{u}_t^2$, and $v_4 = a_3\hat{u}_t^3$.

c. Regard the initial models (except v_1) as the candidate models of the input layer, let a_1, a_2, a_3 be 1, compute their external criterion values on the model selection set B according to Eq. (18), and finally denote the smallest value (here is the candidate model v_3, i.e., the SVR model) as the external criterion value of the input layer, S_0 (see Fig. 5a).

d. Combine pairs of initial models to generate C_4^2 (=6) candidate models in the first layer, estimate their parameters on the model learning set A, compute the external criterion values on set B and denote the smallest value as the external criterion value of the first layer, S_1 (see Fig. 5b).

e. Select 3 candidate models with the lowest external criterion values to enter the second layer, combine them to generate C_3^2 (=3) candidate models, estimate the parameters, compute the external criterion values and regard the smallest value as the external criterion value of the second layer, S_2; similarly, we can obtain the external criterion value of the third layer, S_3 (see Fig. 5c).

f. Notice that $S_0 < S_3 < S_2 < S_1$, thus, the optimal complexity model y_{opt} is just the SVR model in the input layer. The nonlinear combination forecasting results are equal to the prediction of the SVR model, i.e., $\hat{u}_t = \hat{u}_t^2$.

Step 4: Obtain the final prediction result of the time series: $\hat{Y}_t = \hat{l}_t + \hat{u}_t$, and the result is shown in the last row of Table 1.

4. Empirical analysis

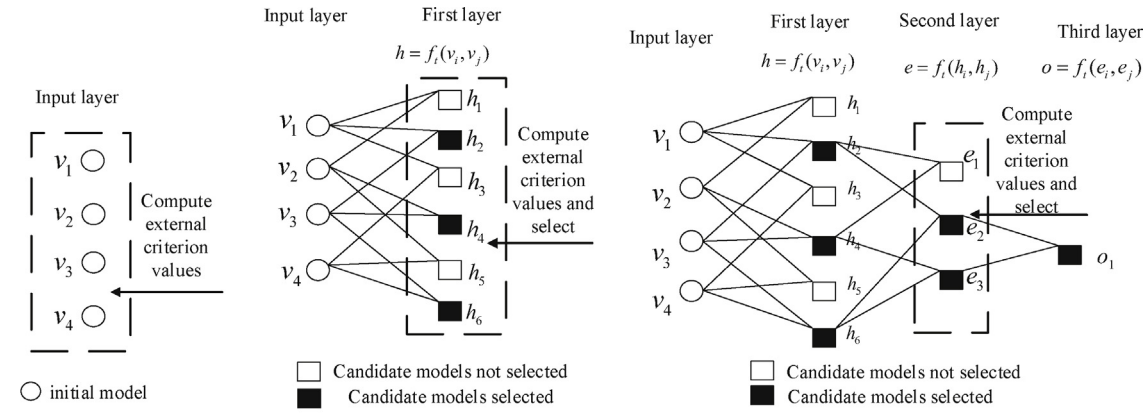
4.1. Data and evaluation criteria

This study utilizes the monthly container throughput data of the Xiamen and Shanghai Ports in China from January 2001 to December 2015 for the experiment, and takes the data from January 2001 to December 2014 as the training set and the data in 2015 as the test set. Data are collected from the official website of the Ministry of Transport of the People's Republic of China. Fig. 6 shows the original container throughput time series of the two ports. The two series have strong cyclicality and volatilities, especially in 2008, they significantly declined.

To validate the existence of seasonality in data scientifically, the HEGY test [12] is employed. The original HEGY test is used for quarterly data; Beaulieu and Miron [3] later derived the HEGY procedure for the monthly data that can be used in this study. The assisted regression of the HEGY test is constructed as

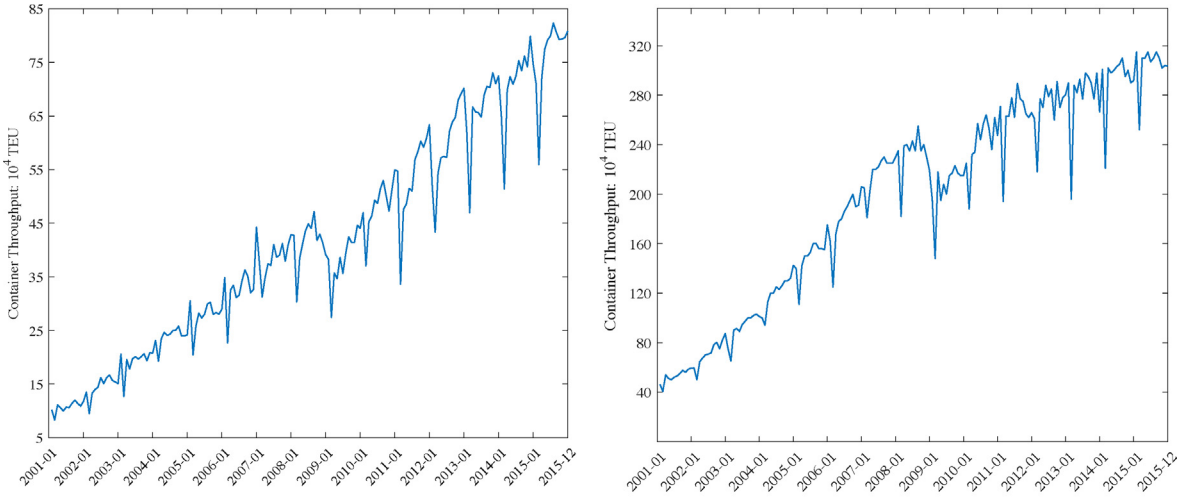
$$(1 - L^{12})Y_t = \sum_{k=1}^{12}(\alpha_k LY_{k,t}) + \sum_{k=2}^{12}(\beta_k S_k) + \beta_0 t + c + \varepsilon_t \quad (19)$$

where L is the lag operator; t is the period; Y_t is the data of period t; S_k is the seasonal dummy variable with a value of 0 or 1; c is a constant; ε_t is the white noise with zero-mean and constant variance;



(a) Initial models (b) The first layer of the neural network (c) The complete GMDH neural network

Fig. 5. The process of the GMDH neural network of the example.



(a) Xiamen Port (b) Shanghai Port

Fig. 6. Container throughput time series of two ports from 2001 to 2015.

α_k, β_0, and β_k are the coefficients of each term; and $Y_{k,t}$ represents a series of polynomial of Y_t and its lagged terms, which can be referred to Beaulieu and Miron's paper. The hypotheses corresponding to the unit roots at frequencies 0, π, $\frac{\pi}{2}$, $\frac{2\pi}{3}$, $\frac{\pi}{3}$, $\frac{5\pi}{6}$, and $\frac{\pi}{6}$ are respectively expressed as follows:

$$
\begin{aligned}
&H_0 : \alpha_1 = 0, H_1 : \alpha_1 < 0;\\
&H_0 : \alpha_2 = 0, H_1 : \alpha_2 < 0;\\
&H_0 : \alpha_3 = \alpha_4 = 0, H_1 : \alpha_3 \neq 0 \quad or \quad \alpha_4 \neq 0;\\
&H_0 : \alpha_5 = \alpha_6 = 0, H_1 : \alpha_5 \neq 0 \quad or \quad \alpha_6 \neq 0;\\
&H_0 : \alpha_7 = \alpha_8 = 0, H_1 : \alpha_7 \neq 0 \quad or \quad \alpha_8 \neq 0;\\
&H_0 : \alpha_9 = \alpha_{10} = 0, H_1 : \alpha_9 \neq 0 \quad or \quad \alpha_{10} \neq 0;\\
&H_0 : \alpha_{11} = \alpha_{12} = 0, H_1 : \alpha_{11} \neq 0 \quad or \quad \alpha_{12} \neq 0;
\end{aligned}
\tag{20}
$$

The unit root at frequency 0 is nonseasonal, whereas the others are seasonal. A one-sided *t*-test is used to examine the former two hypotheses. We accept the null hypothesis of the existence of unit root if the *t*-statistic is larger than the corresponding critical value. A two-sided *F*-test is employed to examine the other hypotheses. We accept the null hypothesis if the *F*-statistic is smaller than the corresponding critical value. We let the significance level $\alpha = 0.05$. Table 2 shows the test values, from which we can conclude the following. (1) in Xiamen Port, the *t*-statistic for α_1 is larger than the critical value −3.28 [3], which indicates a nonseasonal unit root at 95% confidence level at frequency 0. The one for α_2 is smaller than the critical value −2.75. Thus, we reject the null hypothesis at frequency π. The F-statistics at frequencies $\frac{\pi}{2}$ and $\frac{\pi}{3}$ are smaller than the critical value 6.23, which means that we should accept the null hypothesis at 95% confidence level and that seasonal unit roots exist. (2) in Shanghai Port, we accept the null hypotheses at frequencies 0, π, and $\frac{5\pi}{6}$ by comparing the statistics with the critical values. Thus, seasonality exists in the two container throughput time series, and seasonal difference is necessary.

To evaluate the forecasting performance of the HFMG model proposed in this study, we introduce four evaluation criteria [52,19]: root mean square error (*RMSE*), absolute mean percentage error (*MAPE*), mean absolute scaled error (*MASE*), and predictive directionality index D_{stat}. The *RMSE*, *MAPE*, and *MASE* are error indexes, so the smaller they are, the better the performance of the

Table 2
Results of HEGY Test for Seasonal Unit Roots.

Frequency	0	π	$\frac{\pi}{2}$		$\frac{2\pi}{3}$		$\frac{\pi}{3}$		$\frac{5\pi}{6}$		$\frac{\pi}{6}$	
t-statistic values	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	α_{11}	α_{12}
Xiamen Port	−1.11	−4.19	−3.07	−1.11	−4.75	0.18	−2.53	−1.47	−5.66	−1.00	−1.72	−3.33
Shanghai Port	−1.20	−1.87	−4.47	0.32	−4.66	0.97	−4.42	−2.32	−3.45	−0.90	−2.60	−5.42

F-statistic values	$F_{3,4}$	$F_{5,6}$	$F_{7,8}$	$F_{9,10}$	$F_{11,12}$
Xiamen Port	5.39	11.28	4.38	16.80	7.36
Shanghai Port	10.05	11.44	13.34	6.22	19.33

Table 3
Parameters in BP model.

Port	Lag orders	Neurons number in hidden layer	Neurons number in output layer	Transfer function in input layer	Transfer function in hidden layer
Xiamen Port	3	7	1	tan-sigmoid	linear
Shanghai Port	9	4	1	tan-sigmoid	linear

Table 4
Parameters in GP model.

Port	Initial population	Crossover probability	Mutation probability	Max iteration times	Fitness
Xiamen Port	40	0.95	0.01	120	0.4591
Shanghai Port	40	0.95	0.05	120	0.5962

model is; while the D_{stat} is just the opposite. The four criteria are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{m}(\hat{Y}_t - Y_t)^2}{m}} \tag{21}$$

$$MAPE = \frac{\sum_{t=1}^{m}\left|\frac{\hat{Y}_t - Y_t}{Y_t}\right|}{m} \times 100\% \tag{22}$$

$$MASE = \frac{1}{m}\frac{\sum_{t=1}^{m}|\hat{Y}_t - Y_t|}{\frac{1}{m-1}\sum_{t=2}^{m}|Y_t - Y_{t-1}|} \tag{23}$$

$$D_{stat} = \frac{1}{m}\sum_{t=2}^{m} a_t,\ a_t = \begin{cases} 1 & (Y_t - Y_{t-1})(\hat{Y}_t - \hat{Y}_{t-1}) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

where Y_t is the actual value of the t-th sample, $\hat{Y}_t$ is the predicted value of the t-th sample, m is the sample size of the test set.

4.2. Experimental setup

The optimal values of the SARIMA model parameters can be determined based on the AIC and SC criteria. It was found that SARIMA$(1,1,1)(0,1,2)_{12}$ and SARIMA$(1,0,1)(2,1,1)_{12}$ can achieve the best forecasting performance in the Shanghai and Xiamen Ports, respectively.

For the nonlinear residual subseries, we utilized three single models (i.e., BP, GP, and SVR) for the initial prediction. The BP neural network setup for the two ports is shown in Table 3, where the lag orders and the neurons number in the hidden layer are the optimal values from the repeated experiments. The optimal parameters setup of the GP model is displayed in Table 4. The fitness values are relatively small because of the limited useful information in the nonlinear residual subseries. An oversize threshold value of the fitness in experiments may be over-fitting. The experiments also demonstrated that it cannot perform well in this case.

In addition, we used the Libsvm-3.1 toolbox to implement the SVR model. We chose the most commonly used RBF as the kernel function because of its nonlinear mapping ability. The two important parameters are penalty parameter C and kernel width γ. We introduced the grid computing method in the toolbox to search for the best parameter values. The SVR model setup is shown in Table 5.

Table 5
Parameters in SVR model.

Port	Lag orders	Penalty parameter C	Kernel width γ
Xiamen Port	1	0.1	20.86
Shanghai Port	1	8.52	65.70

Finally, the experiments of each model are repeated 10 times, and the final forecasting result is the mean value.

4.3. Experimental results and analysis

4.3.1. Linear trend forecasting

Fig. 7 shows the results of the SARIMA model fitting to the original container throughput time series of the two ports. The nonlinear residual subseries are displayed in Fig. 8. The two figures show that the SARIMA models fit the original time series well, and the residual series appear to have a certain gap with white noise. We introduce Ljung-Box Q test [27] to check whether the residual subseries are white noise or not. The p values of the residual white noise test for the Xiamen and Shanghai Ports are 0.0400 and 0.0281, respectively, which are both less than 0.05. Thus, the residual subseries in the two ports are not white noise at 95% confidence level, which means that useful information is left in them.

4.3.2. Nonlinear residual subseries forecasting

We first employed three nonlinear single models to predict the nonlinear subseries, and the results are shown in Table 6. In Appendix A, we display that the *MAPE* value on the nonlinear residual subseries may be larger than 100%, thus, it is not suitable to use *MAPE* criterion to evaluate the forecasting performance of different models on the nonlinear residual subseries and we only use the other three criteria, i.e., *RMSE*, *MASE*, and D_{stat} in this subsection. However, the *MAPE* criterion will be used in the next subsection to evaluate the forecasting performance of different models on the original time series because it is a commonly used and recommended criterion [19]. Table 6 shows that the SVR model in Xiamen Port is optimal for the nonlinear subseries forecasting because it

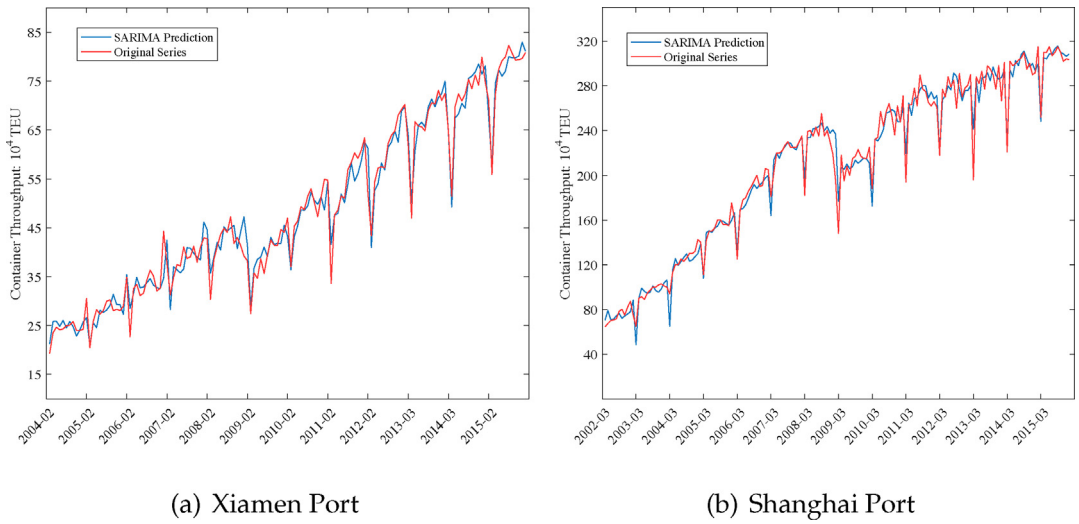


Fig. 7. Linear trend prediction by the SARIMA model.

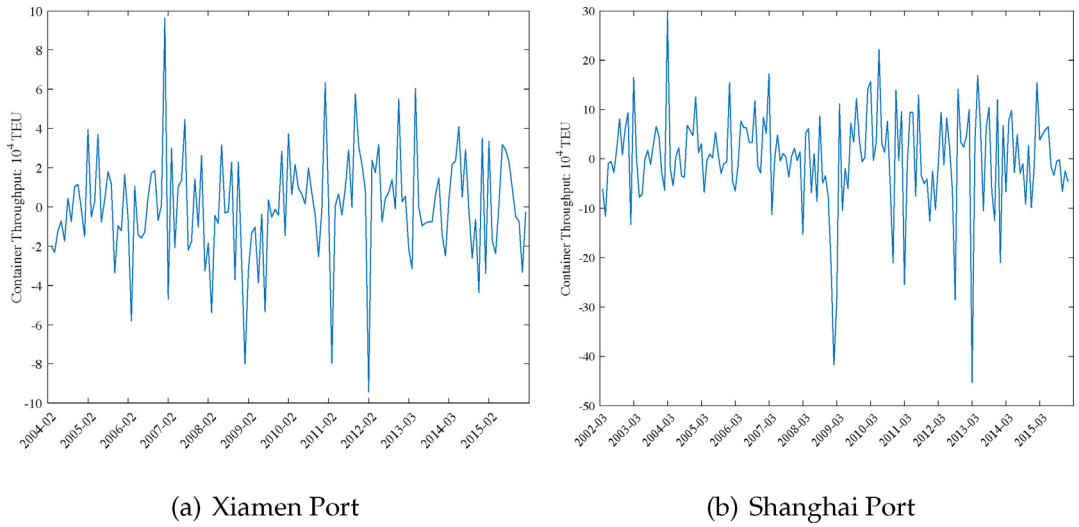


Fig. 8. The residual subseries of the two ports.

Table 6
Performance of three models on the nonlinear subseries.

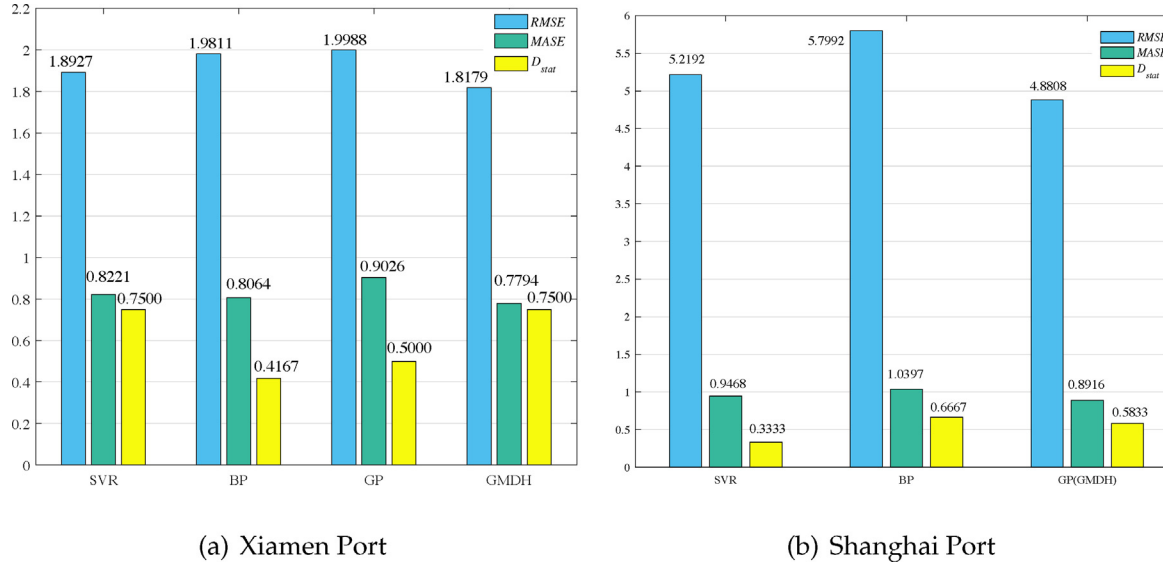
Port	Model	*RMSE*	Rank	*MASE*	Rank	D_{stat}	Rank
Xiamen Port	SVR	**1.8927**	1	0.8221	2	**0.7500**	1
	BP	1.9811	2	**0.8064**	1	0.4167	3
	GP	1.9988	3	0.9026	3	0.5000	2
Shanghai Port	SVR	5.2192	2	**0.9468**	1	0.3333	3
	BP	5.7992	3	1.0397	3	**0.6667**	1
	GP	**4.8808**	1	0.8916	2	0.5833	2

performs best, except in the *MASE* criterion, where it ranks second. It means that the SVR model has the smallest average predicting error and the best predictive directionality according to *RMSE* and D_{stat} respectively. The BP model is optimal only in the *MASE* criterion which indicates it has the smallest error compared to the one-step naïve forecast among the three single models, while the GP model performs worst in Xiamen Port according to the three criteria. The *MASE* values are all less than 1 indicates that the three models are all better than one-step naïve forecast. The optimal nonlinear single model in Shanghai Port is GP. It has the smallest *RMSE* value, and is suboptimal in other two evaluation criteria. The SVR model performs best in the *MASE* criterion but worst in the D_{stat} criterion. The BP model is the worst in Shanghai Port because its prediction error is the largest according to the *RMSE* and *MASE* values though it has the best predictive directionality.

Moreover, we adopted the modified GMDH neural network for the combination forecasting. Table 7 reports the performance of four optimal complexity models, y^1_{opt}, y^2_{opt}, y^3_{opt}, and y^4_{opt}, learned by different versions of the GMDH combination forecasting model. Table 7 shows that model y^4_{opt} selected by the GMDH-ANIC combination model has better combination forecasting performance than the ones selected by other three combination models in all the three evaluation criteria on the residual subseries of the Xiamen Port. In addition, model y^4_{opt} only selects the SVR and BP models for

Table 7
Performance of different versions of GMDH combination forecasting models on the nonlinear residual subseries.

Port	Index	GMDH-ARC	GMDH-SMBC	GMDH-SSC	GMDH-ANIC
Xiamen Port	Optimal complexity model	y^1_{opt}	y^2_{opt}	y^3_{opt}	y^4_{opt}
	Selected single models	SVR, BP, GP	BP, GP	SVR, BP, GP	SVR, BP
	RMSE	1.8266	2.0714	1.8266	**1.8179**
	MASE	0.7882	0.8214	0.7882	**0.7794**
	D_{stat}	0.6667	0.4167	0.6667	**0.75**
Shanghai Port	Optimal complexity model	y^1_{opt}	y^2_{opt}	y^3_{opt}	y^4_{opt}
	Selected single models	SVR, BP, GP	SVR, BP	SVR, BP	GP
	RMSE	5.6403	5.6732	5.6732	**4.8808**
	MASE	0.9127	0.9245	0.9245	**0.8916**
	D_{stat}	**0.6667**	**0.6667**	**0.6667**	0.5833



(a) Xiamen Port (b) Shanghai Port

Fig. 9. Performance of different models on the nonlinear residual series.

combination, which reflects the selective combination forecasting characteristic of the GMDH neural network. The two combination models, GMDH-SSC and GMDH-ARC, both combine all single models. Thus, the forecasting results of models y^1_{opt} and y^3_{opt} are the same. However, their combination forecasting performance is poorer than that of model y^4_{opt}, which shows the necessity of the selective combination forecasting we proposed. For the residual subseries of the Shanghai Port, model y^4_{opt} selected by the GMDH-ANIC combination model still has the best evaluation criteria values except in D_{stat}. Therefore, model y^4_{opt} has the optimal combination forecasting performance overall. However, model y^4_{opt} only selects one single model, GP, which verifies our modification to the traditional GMDH neural network (i.e., computing the external criteria values of the initial models and placing them into the candidate model set) is necessary and reasonable. Therefore, the following experiments regard ANIC as the external criterion of the GMDH combination model for the nonlinear residual subseries forecasting.

Fig. 9 compares the forecasting performance between the nonlinear single models and the GMDH combination forecasting model proposed in this study. It is worth noting that in Shanghai Port, the GMDH combination model only selects the GP model. Thus, the combination model is the same as the GP model and they appear in the same histograms in Fig. 9b. Fig. 9 shows that the GMDH combination forecasting model performs best in Xiamen Port because of the smallest *RMSE* and *MASE* criteria values, and the largest D_{stat} value. Further, in Shanghai Port it ranks first in the *RMSE* and *MASE* criteria and ranks second in the D_{stat} criterion, which shows that it still performs best overall. Therefore, the GMDH combination forecasting model has better overall forecasting performance than the nonlinear single models in the residual subseries of the two ports.

4.3.3. Performance comparison between the HFMG and other models

This section compares the hybrid forecasting model HFMG with the SARIMA approach and the three hybrid models, SARIMA-BP, SARIMA-SVR, and SARIMA-GP. The three hybrid models first adopt the SARIMA approach to predict the linear trend of the container throughput time series, and then employ the BP, SVR, and GP models to forecast the nonlinear residual subseries respectively, finally integrate the results of two parts. Table 8 displays the values and ranks of the four evaluation criteria for different models in Xiamen and Shanghai Ports, respectively, where the parallel ranks are provided if the values are equal.

Table 8 shows that, in Xiamen Port, the HFMG model has the smallest *RMSE*, *MAPE*, *MASE* values and the largest D_{stat} value, indicating the smallest prediction error and the best predictive directionality respectively. Thus, the HFMG model has the best forecasting performance. The SARIMA model performs poorest, whose *RMSE*, *MAPE*, and *MASE* criteria values are the largest. In Shanghai Port, the combination forecasting for the nonlinear subseries in the HFMG model only selects the GP model, and the forecasting result of the HFMG model is the same as that of the SARIMA-GP model. From Table 8, it can be seen that the HFMG and SARIMA-GP models in Shanghai Port have the smallest *RMSE* and *MASE* criteria values, and their *MAPE* and D_{stat} criteria values are only poorer than that of the SARIMA-SVR and SARIMA-BP models, respectively. Thus, they perform best. The SARIMA model is still the worst for the

Table 8
Performance comparison of different models on the original container throughput series.

Port	Model	*RMSE*	Rank	*MAPE*(%)	Rank	*MASE*	Rank	D_{stat}	Rank
Xiamen Port	SARIMA	2.1698	5	2.4083	5	0.4313	5	**0.7500**	2.5
	SARIMA-BP	1.9811	3	1.9845	2	0.3642	2	**0.7500**	2.5
	SARIMA-SVR	1.8927	2	2.0769	3	0.3713	3	**0.7500**	2.5
	SARIMA-GP	1.9988	4	2.2138	4	0.4076	4	0.5833	5
	HFMG	**1.8179**	1	**1.9039**	1	**0.3520**	1	**0.7500**	2.5
Shanghai Port	SARIMA	6.0405	5	1.5224	5	0.3250	5	0.5833	4.5
	SARIMA-BP	5.7992	4	1.4589	4	0.3136	4	**0.8333**	1
	SARIMA-SVR	5.2192	3	**1.2289**	1	0.2848	3	0.5833	4.5
	SARIMA-GP	**4.8808**	1.5	1.2944	2.5	**0.2787**	1.5	0.6667	2.5
	HFMG	**4.8808**	1.5	1.2944	2.5	**0.2787**	1.5	0.6667	2.5

Table 9
The annual growth rate of container throughput of the two ports from 2010 to 2016.

Year	2010	2011	2012	2013	2014	2015	2016
Xiamen Port	25.405%	11.065%	11.414%	10.881%	7.183%	7.311%	6.087%
Shanghai Port	3.995%	28.358%	12.342%	5.852%	7.961%	9.048%	6.396%

largest prediction error and the worst predictive directionality in Shanghai Port. In a word, the HFMG model proposed in this study has better overall forecasting performance than the SARIMA model and other three hybrid models in the container throughput time series of the two ports. Meanwhile, the forecasting performance of hybrid models are all better than SARIMA model in both two ports, which demonstrates the necessity of predicting the nonlinear residual series.

4.4. Out-of-sample forecasting

We utilize the HFMG model to predict the container throughput of the Shanghai and Xiamen Ports in 2016, and the results are shown in Fig. 10.

Fig. 10 shows that the container throughput reaches the lowest level in February in the two ports and shows a sharp increase in March; then the throughput growth fluctuates. The peak occurs in November for the Xiamen Port and August for the Shanghai Port. Table 9 shows the annual growth rate of the container throughput of the two ports from 2010 to 2016, where the annual container throughput in 2016 is derived from the sum of the monthly forecasts by the HFMG model. The forecasts in this study indicate that the annual growth rates are 6.087% and 6.396% in the Xiamen and Shanghai Ports, respectively. Compared with the growth rates 7.311% and 9.048% in 2015, they decrease by 1.224% and 2.652%, respectively. The results indicate that the container throughput of the two ports have increased steadily but still suffer from some resistance. The Shanghai and Xiamen Ports are important hubs in China's transport system and the main trunk ports of container transportation, also the world's top container transportation ports. They take on a large number of cargo transportation tasks in China and play a significant role in the global cargo trade. Especially for Shanghai Port, the annual total container throughput has won the first prize in six years worldwide. It has international routes globally and it is an important port in the international economic cycle. The decrease of container throughput growth rates of the two ports may be attributed to the slowdown of the international trade growth. The latest World Economic Outlook from the International Monetary Fund states that, the growth of world trade in goods and services has been around 3% since 2012 and is less than the half of that in the past 30 years as influenced by the global financial crisis. Moreover, China's GDP growth decreased in 2016. Thus, the pressure of economic downturn from the macroeconomic environment is also an important reason.

Table 10
Comparison of computation time of different models for the nonlinear subseries (unit: seconds).

Port	BP	SVR	GP	GMDH combined forecasting
Shanghai	0.92	6.15	28.76	0.0648
Xiamen	1	5.34	35.95	0.0648

4.5. Discussion

We can conclude the following according to the forecasting performance comparison of the different models in Section 4.3: (1) For container throughput forecasting, the nonlinear residual subseries obtained by some models based on linear assumption, such as the ARIMA and SARIMA, is usually not a white noise series, and useful information for prediction modeling still exists. (2) The performance of the selective combination forecasting by GMDH on nonlinear residual subseries is better compared with the commonly used single nonlinear forecasting models, SVR, BP, and GP. (3) The whole forecasting performance of the proposed hybrid model HFMG on the container throughput time series of the Xiamen and Shanghai Ports is better than the SARIMA model, as well as the three hybrid models, SARIMA-BP, SARIMA-SVR, and SARIMA-GP.

The simplest way to compare the time complexity of different forecasting models is to compare computation time in the same conditions. The SARIMA model is used for predicting the linear trend first in the proposed hybrid model HFMG. The computation time of the SARIMA model is not discussed because it is simple and implemented by Eviews software. The computation time of three nonlinear single models and selective combination forecasting by GMDH neural network on the nonlinear subseries are discussed in this study. Table 10 shows the average operation time at 10 times for each model. It shows that the time complexity of the BP model is the lowest among the three nonlinear forecasting models, whereas that of the GP model is the highest. The last column is the computation time of the GMDH combined forecasting. The time complexity of the selective combination forecasting by GMDH neural network is much lower compared with the three single models. Therefore, the time complexity of the hybrid model HFMG lies mainly in the training of the used single nonlinear prediction models.

The proposed model uses the GMDH neural network for combination forecasting on the nonlinear subseries, and the experimental results show that the optimal combination model does not select all trained nonlinear single models in both two ports. Thus, this model can successfully achieve the purpose of selective combination forecasting, and reduce the influence of multi-collinearity on

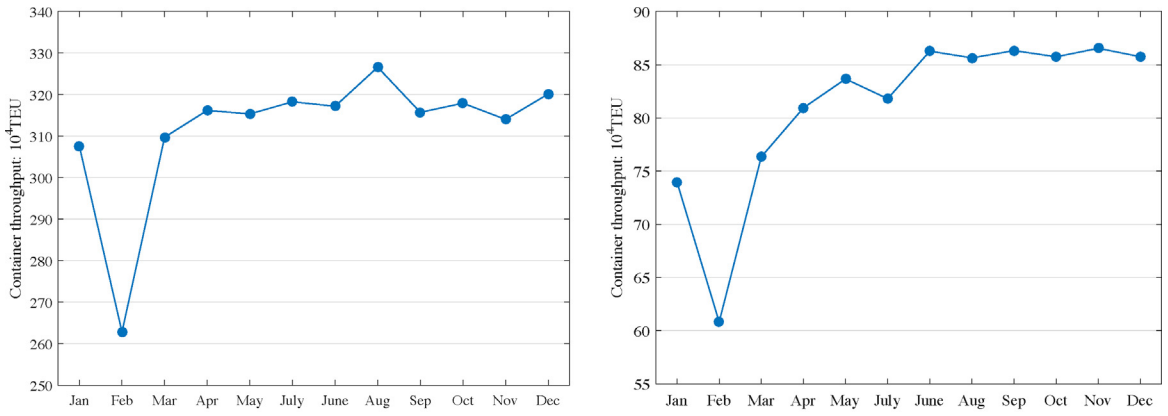


Fig. 10. Out of sample predictions of the container throughput in 2016.

the prediction performance. Meanwhile, the optimal combination model only selects the GP model in Shanghai Port, which shows that the improvement of GMDH neural network is very necessary, and can recover the disadvantage of the traditional GMDH neural network effectively.

In a word, the hybrid forecasting model based on GMDH has some advantages. First, the model structure is simple and understandable. Second, it can be easily extended to any number of input models in the GMDH neural network (i.e., increase the number of single nonlinear models), and it automatically selects the most suitable models for combination forecasting and determines the combination weights. However, it also has some disadvantages. First, if the number of the single models that participate in the combination is large, the time complexity may be high. Second, our research is problem-oriented. We have noticed many container ports, such as Xiamen and Shanghai Ports in this paper, in their container throughput time series (see Fig. 6), it exhibits not only a gradual increased long-term linear trend, but also an obvious nonlinear fluctuation, especially the decline after the global financial crisis in 2008. Therefore, the HFMG decomposes the original time series into linear trend and nonlinear variation first, which may be not suitable for those ports whose container throughput increases as approximate linearity and the nonlinear fluctuation is not obvious.

5. Conclusion

The accurate forecasting of container throughput is important for the investment and construction of port infrastructure, as well as in the operation and management of ports. This study proposes a hybrid forecasting model HFMG, which decomposes the container throughput time series into linear and nonlinear parts. The linear part is predicted by the SARIMA model. The model adopts three nonlinear single models, SVR, BP, and GP, to predict the residual subseries considering the complexity of forecasting the nonlinear subseries. Then, the model establishes selective combination forecasting by the improved GMDH neural network on the nonlinear subseries and obtains its combination forecasting result. Finally, the predictions of two parts are integrated to obtain the forecasting result of the original container throughput time series. The container throughput data of the Xiamen and Shanghai Ports are used for empirical analysis, and the results show that the forecasting performance of the HFMG model is better than that of the SARIMA model and other hybrid forecasting models. The proposed model can be applied to improve the accuracy of container throughput forecasting, assist daily port operation and management activities, such as port transportation capacity allocation and pricing strategy, and guide future planning and development. The model can also be applied to other time series forecasting problems, such as energy demand, crude oil price, and electricity load forecasting.

The proposed hybrid model HFMG in this study decomposes the original container throughput time series into only two parts, which may not be the optimal solution in practice. Therefore, in the future, we will attempt to decompose the container throughput time series into multiple subseries using other decomposition methods, such as wavelet decomposition [36] and empirical mode decomposition [18] and compare the forecasting performance of the HFMG model with that of other models.

Acknowledgements

This research is partly sup-ported by the Natural Science Foundation of China [grant numbers 71471124 and 71501136]; Excellent Youth Fund of Sichuan University [grant numbers skqx201607, sksyl201709, and skzx2016-rcrw14]; MOE Youth Project of Humanities and Social Sciences [grant number 15YJC860034]; Natural Science Foundation of Anhui Higher Education Institutions [grant number KJ2016A604]; Youth Backbone Visiting Research Key Project at Abroad [grant number gxfxZD2016219]; and Key Project of Hefei University [grant number 16ZR21ZDA].

Appendix A. An example of calculating the *MAPE*

This section describes the calculation of the *MAPE* in detail. It takes the Xiamen Port as an example (see Table 11), and the former

Table 11
Calculation of the MAPE on the nonlinear residual subseries for Xiamen Port in 2015.

Time	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Residual subseries: u_t	3.34	−1.73	−2.38	0.13	3.18	2.92	2.3	0.93	−0.49	−0.73	−3.33	−0.27
BP prediction of u_t: $\hat{u}_t$	0.52	−1.1	−1.63	0.51	0.14	0.61	0.71	0.27	−0.83	−0.14	1.07	0.56
Difference: δ_t	−2.82	0.63	0.75	0.38	−3.04	−2.31	−1.59	−0.66	−0.34	0.59	4.4	0.83
$\lvert Quotient\rvert$: τ_t	0.84	0.36	0.32	2.92	0.96	0.79	0.69	0.71	0.69	0.81	1.32	3.07

two rows of the table show the original nonlinear residual subseries u_t and the corresponding prediction result $\hat{u}_t$ by the BP model on the test set respectively; the final two rows show the difference $\delta_t = \hat{u}_t - u_t$, and the absolute value of quotient $\tau_t = |\delta_t/u_t|$, respectively. Thus, according to Eq. (22), the *MAPE* of BP model on the nonlinear residual subseries is calculated as:

$$MAPE = (0.84 + 0.36 + 0.32 + 2.92 + 0.96 + 0.79 + 0.69 + 0.71 + 0.69 + 0.81 + 1.32 + 3.07)/12 * 100\% = 112.47\% \quad (25)$$

As we can see, the *MAPE* value is larger than 100%. There may be two reasons. On the one hand, the limitation of the forecasting ability of the model causes a large deviation between the prediction and actual values in some data points; the largest deviation 4.4 in November corresponding to the large absolute value of quotient 1.32 may be due to this reason. On the other hand, the residual subseries contains some data close to zero. We note that the deviations between predictions and actual values in April and December are 0.38 and 0.83 respectively, which are small, but the absolute values of quotients, 2.92 and 3.07, are large because the divisors (i.e., the residual values 0.13 and -0.27) are close to zero. Compare the data in November with those in April and December, the smaller deviation has a larger quotient resulted from a smaller divisor; thus, the *MAPE* of a forecasting series contains numbers close to zero is more likely to be large for it is the average of these quotients.

References

[1] A. Johnson, Forecasting Australia's International container trade, in: 25th Australian Transport Research Forum, Canberra, 2002.
[2] C. Narendra Babu, B. Eswara Reddy, A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data, Appl. Soft Comput. 23 (2014) 27–38.
[3] J. Joseph Beaulieu, J.A. Miron, Seasonal unit roots in aggregate U.S. data, J. Econ. 55 (1) (1992) 305–328.
[4] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time Series Analysis: Forecasting and Control, John Wiley & Sons, 2015.
[5] G.E.P. Box, D.A. Pierce, Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, J. Am. Stat. Assoc. 65 (332) (1970) 1509–1526.
[6] T.C. Chang, R.J. Chao, Application of back-propagation networks in debris flow prediction, Eng. Geol. 85 (3–4) (2006) 270–280.
[7] C.P. Chen, Q.J. Liu, P. Zheng, Application of grey-Markov model in predicting container throughput of Fujian Province, Adv. Mater. Res. 779–780 (2013) 720–723.
[8] S.H. Chen, J.N. Chen, Forecasting container throughputs at ports using genetic programming, Expert Syst. Appl. 37 (3) (2010) 2054–2058.
[9] Z. Chen, Y. Chen, T. Li, Port cargo throughput forecasting based on combination model, Joint International Information Technology, Mechanical and Electronic Engineering Conference (JIMEC 2016) (2016) 148–154.
[10] C.C. Chou, C.W. Chu, G.S. Liang, A modified regression model for forecasting the volumes of Taiwan's import containers, Math. Comput. Modell. 47 (9) (2008) 797–807.
[11] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[12] S. Hyllebergr, F. Engle, C.W.J. Grangerb, S. Yoo, Seasonal integration and cointegration, J. Econ. 44 (1C2) (1990) 215–238.
[13] J. Geng, M.W. Li, W.C. Hong, T.J. Zheng, Port throughput forecasting by using PPPR with chaotic efficient genetic algorithms and CMA, IEEE International Conference on Systems, Man, and Cybernetics (2015) 1633–1638.
[14] S.W. Gikungu, A.G. Waititu, J.M. Kihoro, Forecasting inflation rate in Kenya using SARIMA model, Am. J. Theor. Appl. Stat. 4 (1) (2015) 15–18.
[15] V. Gosasang, W. Chandraprakaikul, S. Kiattisin, A comparison of traditional and neural networks forecasting techniques for container throughput at Bangkok Port, Asian J. Ship. Logist. 27 (3) (2011) 463–482.
[16] Z. Guo, X. Song, J. Ye, A Verhulst model on time series error corrected for port throughput forecasting, J. East. Asia Soc. Transp. Stud. 6 (2005) 881–891.
[17] A. Huang, K. Lai, Y. Li, S. Wang, Forecasting container throughput of Qingdao Port with a hybrid model, J. Syst. Sci. Complex. 28 (1) (2015) 105–121.
[18] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, in: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 454, The Royal Society, 1998, pp. 903–995.
[19] R.J. Hyndman, A.B. Koehler, Another look at measures of forecast accuracy, Int. J. Forecast. 22 (4) (2006) 679–688.
[20] A.G. Ivakhnenko, The group method of data handling-a rival of the method of stochastic approximation, Sov. Autom. Control 13 (3) (1968) 43–55.
[21] M. Khashei, M. Bijari, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, Appl. Soft Comput. 11 (2) (2011) 2664–2675.
[22] T. Kondo, GMDH-type neural network algorithm self-selecting optimum neural network architecture and its application to medical image recognition, Ieice Tech. Rep. Theor. Found. Comput. 108 (2008) 17–24.
[23] T. Kondo, J. Ueno, S. Takao, Hybrid GMDH-type neural network using artificial intelligence and its application to medical image diagnosis of liver cancer, IEEE/SICE International Symposium on System Integration (2011) 1101–1106.
[24] J.R. Koza, Genetic programming as a means for programming computers by natural selection, Stat. Comput. 4 (2) (1994) 87–112.
[25] C.J. Liu, Q.N. Zhang, Dynamic prediction of container throughput based on the time series BP neural network (BP NN), Port Waterway Eng. 1 (2007) 4–8.
[26] L. Liu, The forecast of Qinhuangdao logistic based on BP neural network, Int. J. Digital Content Technol. Appl. 6 (8) (2012) 8–16.
[27] G.M. Ljung, G.E.P. Box, On a measure of lack of fit in time series models, Biometrika 29 (1978) 7–303.
[28] H.A. Lu, S.Y. Chen, Y.C. Yu, An application of grey theory to global trade predictions based on airport cargo traffic, J. Grey Syst. 15 (4) (2012) 195–204.
[29] W.X. Luan, X.H. Ma, Study on supply and demand balance of container port capacity in China, Econ. Geogr. 31 (11) (2011) 1774–1780.
[30] K.L. Mak, D.H. Yang, Forecasting Hong Kong's container throughput with approximate least squares support vector machines, World Congress on Engineering (2007) 7–12.
[31] K.C. Min, H.K. Ha, Forecasting the Korea's port container volumes with SARIMA model, J. Korean Soc. Transp. 32 (6) (2014) 600–614.
[32] J.A. Mueller, F. Lemke, Self-Organising Data Mining: An Intelligent Approach to Extract Knowledge from Data, Hamburg, Libri, 2000.
[33] M. Najafzadeh, G.A. Barani, M.R.H. Kermani, GMDH based back propagation algorithm to predict abutment scour in cohesive soils, Ocean Eng. 59 (2013) 100–106.
[34] G.R. Patil, P.K. Sahu, Estimation of freight demand at Mumbai Port using regression and time series models, KSCE J. Civil Eng. 20 (5) (2016) 2022–2032.
[35] W.Y. Peng, C.W. Chu, A comparison of univariate methods for forecasting container throughput volumes, Math Comput. Modell. 50 (7C8) (2009) 1045–1057.
[36] D.B. Percival, A.T. Walden, Wavelet Methods for Time Series Analysis, 4, Cambridge University Press, 2006.
[37] F.F. Ping, F.X. Fei, Multivariant forecasting mode of Guangdong Province port throughput with genetic algorithms and back propagation neural network, Procedia Soc. Behav. Sci. 96 (2013) 1165–1174.
[38] Y. Rashed, H. Meersman, E. Van de Voorde, T. Vanelslander, Short-term forecast of container throughout: an ARIMA-intervention model for the port of Antwerp, Marit. Econ. Logist. (2016), http://dx.doi.org/10.1057/mel.2016.8.
[39] J.J. Ruiz-Aguilar, I.J. Turias, M.J. Jiménez-Come, Hybrid approaches based on SARIMA and artificial neural networks for inspection time series forecasting, Transp. Res. E: Logist. Transp. Rev. 67 (2014) 1–13.
[40] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536.
[41] P.M. Schulze, A. Prinz, Forecasting container transshipment in Germany, Appl. Econ. 41 (22) (2009) 2809–2815.
[42] W. Seabrooke, E.C.M. Hui, W.H.K. Lam, G.K.C. Wong, Forecasting cargo growth and regional role of the Port of Hong Kong, Cities 20 (1) (2003) 51–64.
[43] C.H. Shin, S.H. Jeong, A study on application of ARIMA and neural networks for time series forecasting of port traffic, J. Navig. Port Res. 35 (1) (2011) 83–91.
[44] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (3) (2004) 199–222.
[45] K. Kuroda Syafi'i, M. Takebayashi, Forecasting the demand of container throughput in Indonesia, Mem. Constr. Eng. Res. Inst. 47 (2005) 227–231.
[46] G. Teng, C. He, J. Xiao, Y. He, B. Zhu, X. Jiang, Cluster ensemble framework based on the group method of data handling, Appl. Soft Comput. 43 (2016) 35–46.
[47] J. Xiao, C. He, X. Jiang, Structure identification of Bayesian classifiers based on GMDH, Knowl. Based Syst. 22 (6) (2009) 461–470.
[48] J. Xiao, C. He, X. Jiang, D. Liu, A dynamic classifier ensemble selection approach for noise data, Inf. Sci. 180 (18) (2010) 3402–3421.
[49] J. Xiao, X. Jiang, C. He, G. Teng, Churn prediction in customer relationship management via GMDH-based multiple classifiers ensemble, IEEE Intell. Syst. 31 (2) (2016) 37–44.
[50] J. Xiao, Y. Xiao, J. Fu, K.K. Lai, A transfer forecasting model for container throughput guided by discrete PSO, J. Syst. Sci. Complex. 27 (1) (2014) 181–192.
[51] Y. Xiao, S. Wang, M. Xiao, J. Xiao, Y. Hu, The analysis for the cargo volume with hybrid discrete wavelet modeling, Int. J. Inf. Technol. Decis. Mak. 16 (3) (2017) 851–863.
[52] G. Xie, S. Wang, Y. Zhao, K.K. Lai, Hybrid approaches based on LSSVR model for container throughput forecasting: a comparative study, Appl. Soft Comput. 13 (5) (2013) 2232–2241.
[53] Y. Zhou, P. Zhao, An application of genetic programming in nonlinear combining forecasting, Int. J. Hybrid Inf. Technol. 9 (6) (2016) 443–454.