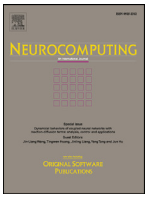
Neurocomputing 356 (2019) 244–256

Contents lists available at ScienceDirect

Neurocomputing

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



Effective passenger flow forecasting using STL and ESN based on two improvement strategies

Lan Qin, Weide Li*, Shijia Li

School of Mathematics and Statistics, Center for Data Science, Laboratory of Applied Mathematics and Complex Systems, Lanzhou University, Lanzhou 730000, China

ARTICLE INFO

Article history:
Received 17 December 2018
Revised 17 March 2019
Accepted 19 April 2019
Available online 11 May 2019

Communicated by Dr. P. Zhang

Keywords:
Passenger flow prediction
Seasonal-trend decomposition procedures based on loess(STL)
Echo state network(ESN)
Grasshopper optimization algorithm(GOA)
Adaptive boosting(Adaboost)

ABSTRACT

Accurate passenger flow prediction is fairly challenging because of chaotic nature of transportation system and influence mechanism originated from multiple factors. It has been found that passenger flow has a nonlinear characteristic and a remarkable seasonal tendency. In this study, two novel hybrid approaches combining seasonal-trend decomposition procedures based on loess(STL) with echo state network(ESN) improved by grasshopper optimization algorithm(GOA) and adaptive boosting(Adaboost) framework respectively are proposed to forecast monthly passenger flow in China. According to the proposed methods(STL-GESN, STL-AESN), the original passenger flow data are firstly decomposed into seasonal, trend and remainder components via STL. Then the improved ESN is adopted to forecast the trend and the remainder components, and the seasonal-naive method is utilized to forecast the seasonal component. Finally, the forecasting results of the three components are summed to obtain the final forecasting of monthly passenger flow. Two passenger flow forecasting applications based on air data and railway data respectively are conducted to verify the effectiveness and scalability of the proposed approaches. The experimental results show that STL-GESN and STL-AESN obtain higher prediction accuracy compared with other forecasting approaches. Application studies also demonstrate that the proposed approaches are practical choice for passenger flow forecasting.

1. Introduction

Rapid development of social economy has promoted the prosperity of transportation in recent years. However, an inevitable problem is that stable operation of transportation system needs much material and financial resources, which forces traffic departments or relevant companies to make comprehensive strategies and arrange funds as reasonably as possible. Passenger flow prediction is a significant and necessary task for upper management of transportation system, because it helps to ease the traffic pressure and avoid meaningless expenses to some extent. The prediction is a challenging topic because traffic flow is influenced by personal factors including population density and consumption level, and impersonal factors, such as weather condition and emergency.

The passenger flow prediction has drawn the attention of researchers in the past few decades, various approaches have been proposed to forecast passenger flow. The typical methods mainly are statistical models, such as autoregressive integrated moving average(ARIMA) [1–3], seasonal autoregressive integrated moving average (SARIMA) [4,5], grey model (GM) [6], Kalman Filtering [7,8], Spectral Analysis [9,10]. While above approachs excel in exploring linear relationships, they have limited ability of capturing nonlinear characteristics. In recent years, many scholars have also turned to some nonlinear methods, such as support vector regression(SVR) [11–13] and artifical neural network(ANN). The common ANN models used in this field include back-propagation neural network(BPNN) [14], extreme learning machine(ELM) [15] and deep belief network(DBN) [16].

Recently, the recurrent neural network(RNN) becomes more popular in many fileds because of its advantage of short-term memory [17–19]. As a variant of RNN, echo state network(ESN) has been widely applied for time-series forecasting tasks [20–23]. The main reasons are that it has strong nonlinear fitting ability and low computational consumption. In order to obtain more satisfactory forecasting results from ESN, two improvement strategies including optimization algorithm and boosting algorithm, are proposed to optimize ESN.

The randomly generated connection weight matrix and reservoir have a great impact on the performance of ESN, so it is necessary to search optimal values of the vital parameters of ESN

* Corresponding author.
E-mail addresses: 1287511124@qq.com (L. Qin), weideli@lzu.edu.cn (W. Li).

https://doi.org/10.1016/j.neucom.2019.04.061
0925-2312/© 2019 Elsevier B.V. All rights reserved.

for higher prediction accuracy. This paper uses a new intelligent algorithm based on the food searching mechanism of the grasshopper in larva and adult phases, named as grasshopper optimization alogorithm(GOA), which was proposed by Saremi et al. [24]. The competitive advantage of GOA lies in the better quality of exploration and exploitation compared to other optimization methods. The GOA has successfully applied for feature selection problem [25], trajectory optimization [26], financial stress prediction [27], signal analysis [28] and base station setting [29]. At present, there are no studies using ESN imporved by GOA for passenger flow forecasting. As one of boosting algorithm, the adaboost framework could improve the prediction accuracy by adjusting adaptively the weights of weak predictors generated by the process of each training [30–32]. This paper tries to employ the adaboost framework to integrate ESN for improving the generalization performance.

Nonlinear trend and remarkable seasonality are significant characteristics of passenger flow, which influence the forecasting accuracy of predictor. According to the previous study conducted by Zhang et al. [33], models that ignore seasonal and trend patterns will obtain poor forecasting performance. In other words, the data preprocessing(detrending and deseasonalizing) before prediction could help improving modeling accuracy. Common methods for dealing with seasonal variations include SARIMA [5], X-11-ARIMA [34], X-12-ARIMA [35] and seasonal exponent adjustment [36]. However, the above-mentioned ways may not be appropriate. The main reasons are as follows: (1) The first three models based on the assumption of linear relationship are not preponderant to handle complex nonlinear problems [33]. In addition, the differencing processing existing in these methods is not always satisfactory to handle nonlinearity and seasonality simultaneously [37]. (2) The decomposition methods including X-11-ARIMA and X-12-ARIMA are widely applied in different fields, but they are complex to achieve, because these approaches need to consider outliers and holiday effects etc. (3) As a multiplicative model, the seasonal exponent adjustment method is easily understood and used, avoiding the limitations of the first three methods, but results in minor improvement in modeling accuracy. It is verified in our two applications. Therefore, a seasonal-trend decomposition technique named STL is introduced to extract the seasonal factor from the raw data in this article. Compared to other decomposition methods such as X-11-ARIMA and X-12-ARIMA, some particular advantages of STL lie in: (1) its strong adaptation to abnormal values guarantees the robustness of the component series, which could enhance in prediction accuracy of modeling. (2) STL could handle time series with any seasonal frequency greater than one [38], which means that it has broad adaptability towards data. (3) STL is based on numerical method and does not need to determine parameters, which means that this method is easy to achieve. As a signal decomposition method, the empirical mode decomposition(EMD) [39] is always applied to decompose data into component series, and indeed facilitates better forecasting performance of modeling. However, this article chooses STL rather than EMD, the main reason is that the interpretability of components called intrinsic mode function(IMF) obtained from EMD is not well.

In this paper, we develop two novel ESN-based passenger flow forecasting methods. The main contributions of this paper are as follows:

(1) The seasonal adjustment method STL is used to deal with the seasonal variations of monthly passenger flow data.
(2) The new evolution algorithm GOA and the integration framework adaboost for regression respectively are introduced to improve the forecasting performance of ESN.
(3) Two real-world passenger flow forecasting applications are conducted to validate the effectiveness and universality of the proposed methods.

The remainder of this paper is organized as follows: Section 2 details the methodology, including STL, the basic ESN, GOA and the proposed STL-GESN, STL-AESN. Section 3 presents two numerical forecasting examples, including air and rail passenger flow in China. Section 4 draws conclusion and puts forward some directions for future research.

2. Methodology

In this section, the existed methods including STL, ESN, GOA and adaboost framework are presented. Then, the proposed methods STL-GESN and STL-AESN are formulated, and the corresponding processes are detailed.

2.1. Seasonal-trend decomposition procedures based on loess

The time-series decomposition method STL is a filtering procedure, which decomposes time-series data into additive variation components, including seasonality, trend and remainder. Compared to other decomposition techniques, such as X-11-ARIMA and X-12-ARIMA, the obvious advantages of STL lie in its robust adaptation to outliers in the data and application for a large number of time-series data without any mathematical modeling [38].

Given monthly passenger flow series X_t, STL could decompose X_t into three additive components of seasonality S_t, trend T_t and remainder R_t: $X_t = S_t + T_t + R_t$. The iteration mechanism of STL composes two recursive procedures, the inner and outer loops. During each iteration in the inner loop, seasonal smoothing and trend smoothing are used to update the seasonal and trend components, respectlvely. After the each inner loop ends, the remainder component is calculated based on the estimation of the seasonal and trend components. Then robust weights that help to reduce the distraction of outliers in the subsequent inner loop, are also computed based on the estimated remainder component. The specific steps of inner loops are shown as follows:

Suppose $S_t^{(k)}$ and $T_t^{(k)}$ are the corresponding seasonal and trend components at the kth iteration of the inner loop, and initialize $T_t^{(k)} = 0$.

Step 1 Detrending. At $(k+1)-th$ internation of the inner loop, the trend component $T_t^{(k)}$ is removed from X_t, i.e $X_t \leftarrow X_t - T_t^{(k)}$.
Step 2 Cycle-subseries smoothing. X_t is smoothed via loess to obtain the series $C_t^{(k+1)}$.
Step 3 Low-Pass Filtering of cycle-subseries. $C_t^{(k+1)}$ is handled using low-pass filtering and loess regression to obtain the series $L_t^{(k+1)}$.
Step 4 Detrending of smoothed cycle-subseries. The seasonal component $S_t^{(k+1)}$ is calculated as: $S_t^{(k+1)} = C_t^{(k+1)} - L_t^{(k+1)}$.
Step 5 Deseasonalizing. The seasonal component $S_t^{(k+1)}$ is removed from X_t, i.e $X_t \leftarrow X_t - S_t^{(k)}$.
Step 6 Trend smoothing. X_t is smoothed via loess to obtain the trend component $T_t^{(k+1)}$.

Upon the finish of the inner loop, the remainder component $R_t^{(k+1)}$ is computed as: $R_t^{(k+1)} = X_t - S_t^{(k+1)} - T_t^{(k+1)}$. Any large values in R_t are regarded as outliers and their weights are calculated. In the following inner loop, the weights are used to down-weight the effect of outliers, which are identified in the previous iteration of the outer loop.

Following the above process, the original time series is decomposed into seasonal, trend and remainder components. As for the forecasting of seasonal component, the seasonal-naive method is used to deal with this problem. It is noticable that S_t cycles for 12 months, i.e the forecasting value at time step $(t+12)$ is same

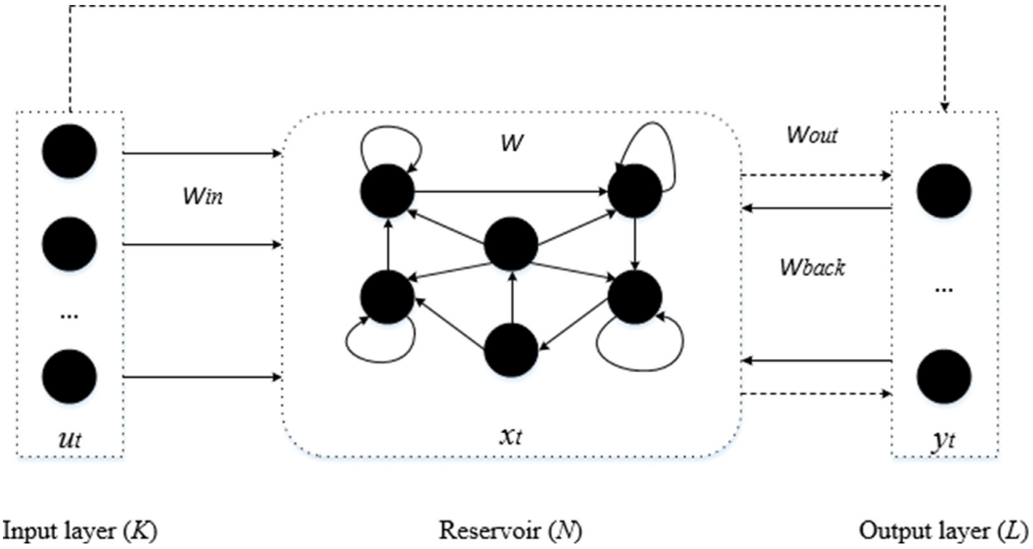


Fig. 1. The structure of ESN.

to historical value at time step t. Therefore, this method can be described with formula as follows:

$$\hat{S}_{t+12} = S_t, t = 1, 2, \dots, N \tag{1}$$

where $\hat{S}_{t+12}$ and S_t are the forecasting values at time step $(t+12)$ and the actual value at time step t, respectively.

2.2. Echo state network

2.2.1. The basic principle of ESN

As one novel variant of RNN, ESN replaces the hidden layer of RNN with a special net structure called reservoir, which is a core area owning the ability of extracting data features and fitting nonlinear patterns. Compared with traditional ANNs, the training of ESN is executed based on least squares method rather than gradient descent, which avoid problems of gradient vanish and gradient explosion existing in RNN [40]. Thus, ESN has much faster training speed and evades the dilemma of trapping in local optimum. Besides, the powerful fitting capacity of ESN is approved in plenty of studies. The basic structure of ESN is shown in Fig. 1.

The structure of ESN has three parts, including input layer with K input units, reservoir with N internal units and output layer with L output units.

At time step t, input sequence u_t, reservoir state sequence x_t and output sequence y_t are formalized as:

$$\begin{cases} u_t = [u_1(t), u_2(t), \dots, u_K(t)]^T \\ x_t = [x_1(t), x_2(t), \dots, x_N(t)]^T \\ y_t = [y_1(t), y_2(t), \dots, y_L(t)]^T \end{cases} \tag{2}$$

At time step $t+1$, reservoir state sequence is updated as follows:

$$x_{t+1} = f(W_{in} \cdot u_{t+1} + W \cdot x_t + W_{back} \cdot y_t) \tag{3}$$

where $W_{in} \in R^{N \times K}$ is the connection weight matrix from input layer to reservoir, $W \in R^{N \times N}$ is the connection weight matrix between internal units in reservoir, $W_{back} \in R^{N \times L}$ is the feedback connection weight matrix from output layer to reservoir. W_{in}, W, W_{back} are randomly initialized ranging from [0,1] and remain unchanged during the training of ESN. f is the activation function of reservoir, such as hyperbolic tangent function or sigmoid function, here is set to be the former.

The output sequence of ESN at time step $t+1$, can be obtained by:

$$y_{t+1} = g(W_{out} \cdot [x_{t+1}; u_{t+1}]) \tag{4}$$

where $W_{out} \in R^{L \times (N+K)}$ is the connection weight matrix from input and internal units to output units. In fact, the training goal of ESN is to determine the matrix W_{out}. g is the activation function of output layer, such as hyperbolic tangent function or linear function, here is set to be the latter. Then, Eq. (4) can be transformed into:

$$y_{t+1} = W_{out} \cdot [x_{t+1}; u_{t+1}] \tag{5}$$

Given the training dataset has T time steps, i.e $t = 1, 2, \dots, T$, Eq. (5) could be transformed into the matrix version, as follows:

$$Y = M \cdot W_{out}^T \tag{6}$$

where $Y = [y_1, y_2, \dots, y_T]^T \in R^{T \times L}$, $M = [m_1, m_2, \dots, m_T]^T \in R^{T \times (N+K)}$ with $m_t = [x_t; u_t], i = 1, 2, \dots, T$.

Thus, the solution of W_{out} is calculated as follows:

$$W_{out} = (M^{-1} \cdot Y)^T \tag{7}$$

where M^{-1} is the Moore-Penrose pseudo inverse of M.

On the basis of previous research, it is a fact that if the ESN wants to own echo state property i.e short-term memory, it starts to train the network after ESN does nothing during certain iterations. The size of the iteration time called washout time step T_0. If the time step $t < T_0$, reservoir state sequence u_t is influenced by initial weights; if $t \geq T_0$, u_t only depends on training samples. Therefore, the time steps before T_0 will be removed, and the remainder time steps will be saved. The training process of ESN is detailed as follows:

Step 1 Initialize the ESN. Set the number of internal units in reservoir N and the washout time step T_0, then randomly generate weight matrix W_{in}, W, W_{back} in the range of [0,1]. Suppose the initial reservoir sequence $x_0 = [0, 0, \dots, 0]^T \in R^{N \times 1}$ and output sequence $y_0 = [0, 0, \dots, 0]^T \in R^{L \times 1}$.

Step 2 Update the state of reservoir. Calculate the new state sequence of reservoir based on Eq. (3).

Step 3 Calculate the connection weight matrix W_{out}. Collect the input sequences $u_t, t = T_0, \dots, T$ and the reservoir state sequences $x_t, t = T_0, \dots, T$ to construct the matrix $M \in R^{(T-T_0+1) \times (N+K)}$, and collect the output sequences $y_t, t = T_0, \dots, T$ to construct the matrix $Y \in R^{(T-T_0+1) \times L}$, then calculate W_{out} based on Eq. (7).

2.2.2. The optimizable parameters of ESN

It can be learned from Section 2.2.1 that the connection weight matrix including W_{in}, W and W_{back} are randomly initialized before training, which influence forecasting output sequences. Thus some researchers tried to optimize these parameters via optimization algorithms and got some improvement in prediction accuracy. In addition, optimizing the vital parameters of the reservoir could result in advancement on the fitting capacity and forecasting performance [41].

The main parameters of the reservoir include: the scale N, the connection rate α and the spectral radius β. The parameter N is relevant to the training samples and the complexity of the research. It has been demonstrated that the value of N is bigger, the fitting ability of ESN is better, but too big value will cause the overfitting problem. Many studies determine the reservoir size via the trial and error method or the empirical formula generated by Kolmogorov Theorem, as shown in Eq. (8).

$$N = min(2K+1, \sqrt{K+L+1}+c), c \in [1, 10] \tag{8}$$

where K is the size of input units, L is the size of output units and c is a random integer in the range of [1,10].

The parameter α ranging from 0 to 1 determines the connection condition of internal units in the reservoir, because not all internal units are interconnected. In general, if the value of α is too small, ESN may lost the function of short-term memory; while the value is too big, it is hard for the reservoir to decode the state. The connection rate α is usually set between 0.01 and 0.05 based on the experience. The parameter β determines whether ESN has echo state property, i.e short-term memory. The value of β is universally set to either the biggest absolute eigenvalue of weight matrix W or a certain value smaller than 1 according to previous studies.

Besides, the washout time step T_0 is influential to the performance of ESN. The common practice is that the value of T_0 is artificially set to a constant in consideration of computing cost.

2.3. Two improvement strategies for better performance

In order to enhance the forecasting performance of ESN, two optimization stratgies including the new evolution algorithm GOA and the effective integration framework adaboost are proposed. GOA is used to search optimal values of the four vital parameters of ESN, and adaboost is introduced to improve the generalization ablity. Although their research priorities are not same, the goal of them is to obtain more accurate forecasting results. The flowchart of two algorithms, including GOA for parameter selection of ESN(GESN) and adaboost for improving fitting capacity of ESN(AESN), is shown as Fig. 2.

2.3.1. Grasshopper optimization algorithm

A novel optimization algorithm named GOA, was proposed by Saremi et al. [24]. The GOA is a heuristic algorithm based on swarm intelligence. It shows competitive advantages on the efficiency of exploration and exploitation compared to other optimization methods. The mechanism of seeking food among swarming grasshoppers is that the swarm moves small steps in the larva phase and moves big steps in the adult phase. It has been known that exploration and exploitation are the two vital phases in the nature-inspired algorithms, the former helps to improve convergence speed while the latter to avoid local optima when searching for a target. In the process of exploration, individuals need to move abruptly, which represents the global search ability; In the process of exploitation, they are expected to move locally, which represents the local search ability. The mathematical model for simulating the swarm behavior is as follows:

$$X_i = r_1 \sum_{j=1, j\neq i}^{N} s(d_{i,j}) \frac{x_j - x_i}{d_{i,j}} - r_2 g e_g + r_3 u e_w \tag{9}$$

$$s(r) = f e^{\frac{-r}{l}} - e^{-r} \tag{10}$$

where X_i is the position of $i-th$ grasshopper, r_1, r_2 and r_3 are random numbers ranging from [0,1], $d_{i,j}$ is the Euclidean distance between the $i-th$ and $j-th$ grasshopper; the function s is to assess the social interaction, f is the intensity of attraction and l is the attractive length scale. According to the original GOA paper, the best values of f and l are 1.5, 0.5 respectively; g is the constant and e_g is a vector towards the center of earth, u is a constant and e_w is a vector towards wind direction.

Eq. (9) cannot be utilized directly to solve optimization problems, thus a modified version, i.e., position updating formula is proposed as follows:

$$X_i^d = c\left(\sum_{j=1, j\neq i}^{N} c\frac{ub_d - lb_d}{2} s\left(\left|x_j^d - x_i^d\right|\right) \frac{x_j - x_i}{d_{i,j}}\right) + T_d \tag{11}$$

$$c = c_{max} - l \cdot \frac{c_{max} - c_{min}}{L} \tag{12}$$

where X_i^d is the new position of $i-th$ grasshopper based on its present position x_i, the target position T_d(current best solution) and the positions of other grasshoppers $x_j, j = 1, 2, \ldots, N, j \neq i$. N is the number of grasshoppers; ub_d is the upper bound in the $d-th$ dimension and lb_d is the lower bound in the $d-th$ dimension; The function s is similar to s the in Eq. (10); c is the decreasing coefficient for shrinking the areas including comfort zone, attraction region and repulsion region, c_{min} is the minimum of c, c_{max} is the maximum, l is the current iteration and L is the maximum number of iterations. Based on settings of the original GOA paper, c_{min} and c_{max} are set to 0.0001 and 1.0 respectively. The pseudocode of GOA is shown as follows:

Algorithm 1 Pseudocode of the GOA algorithm.

```
Input: The population of solutions $X_i (i = 1, 2, ..., N)$;The boundaries of solutions, $lb_d$ and $ub_d$;The max iteration $Maxgen$;The minimum and maximum of decreasing coefficient $c$, $c_{min}$ and $c_{max}$
Output: The best solution $T$
while $l \leq Maxgen$ do
    Update $c$ based on Eq. (12)
    for each solution do
        Normalize the distances between individuals in [1,4]
        Update the solution of the present individual based on Eq. (11)
        Bring the present solution back if it goes outside $[lb_d, ub_d]$
        Update $T$ if there is a better solution
    end for
    $l = l + 1$
end while
Return $T$
```

As stated in Section 2.2.2, the predictive performance of ESN is mainly influenced by six parameters including three connection weight matrix(W_{in}, W, W_{back}), the number of internal units N, the connection rate α and the spectral radius β. The value of N could determine the size of weight matrix and the values of α and β also influence the weight matrix W, so it is not easy to simultaneously optimize six parameters. In this paper, N is set to a constant and W is initialized randomly, then other four parameters including W_{in}, W_{back}, α and β are optimized by GOA.

2.3.2. Adaptive boosting framework

Adaboost is an iterative algorithm, which helps to improve the fitting performance of training set. The main idea of it is that a powerful predictor is generated by repeatedly adding basic predictors [42]. In this paper, the detailed procedure of adaboost used for regression is shown as follows:

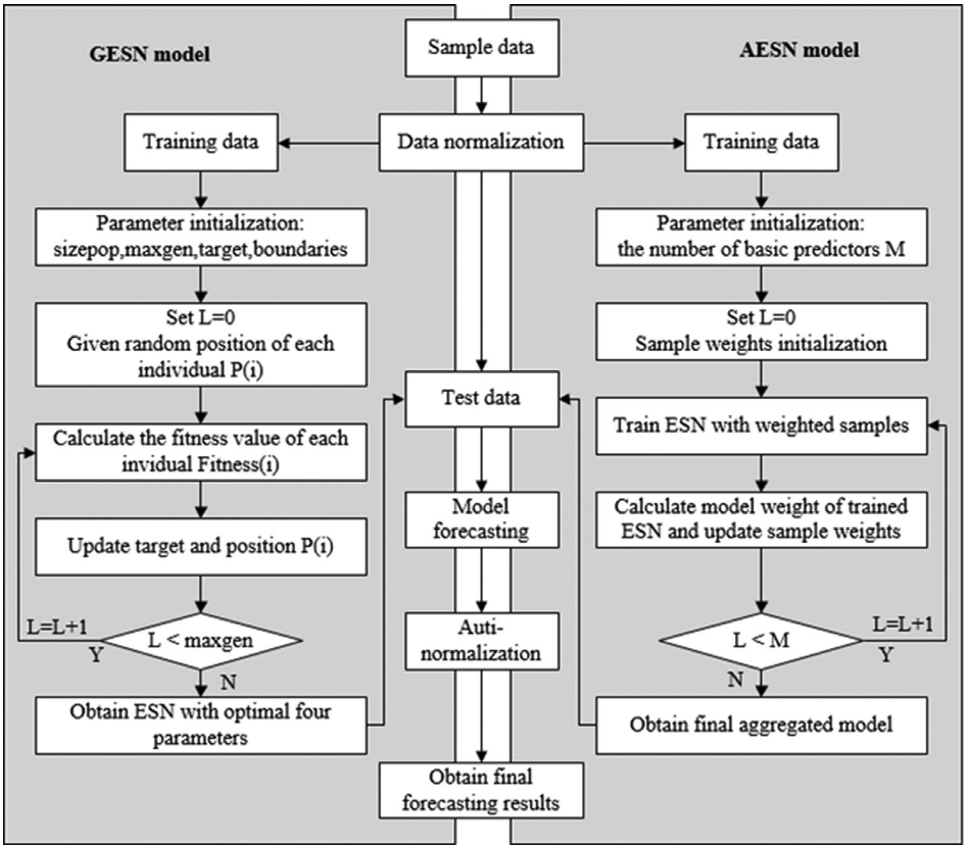

GESN model
Sample data
AESN model
Training data
Data normalization
Training data
Parameter initialization:
sizepop,maxgen,target,boundaries
Parameter initialization:
the number of basic predictors M
Set L=0
Given random position of each individual P(i)
Set L=0
Sample weights initialization
Test data
Calculate the fitness value of each invidual Fitness(i)
Train ESN with weighted samples
Model forecasting
Update target and position P(i)
Calculate model weight of trained ESN and update sample weights
L=L+1
Y
L < maxgen
N
Auti-normalization
L < M
L=L+1
Y
N
Obtain ESN with optimal four parameters
Obtain final aggregated model
Obtain final forecasting results

Fig. 2. The flowchart of the GESN and AESN models.

Step 1 Suppose the size of training set is N and the number of basic predictors is M. Initialize the weight of samples in training set $w_i = \frac{1}{N}, i = 1, 2, \ldots, N, w = [w_1, w_2, \ldots, w_N]$.

Step 2 For $m = 1, 2, \ldots, M$

(1) Obtain the basic predictor G_m after training.

(2) Calculate relative error of each training sample: $RE_i = \left|\frac{G_m(x_i) - y_i}{y_i}\right|, i = 1, 2, \ldots, N.$

(3) Calculate average forecasting loss of G_m: $acc_m = \sum_{i=1}^{N} w_i RE_i$.

(4) Calculate the weight of G_m: $\alpha_m = log(\frac{1}{acc_m})$.

(5) Update the training sample weights:

$$w_i = w_i \exp\left(\alpha_m RE_i\right), i = 1, 2, \ldots, N$$
$$w_i = \frac{w_i}{\|w\|}$$

Step 3 The final predictor is generated as:

$$f(x) = \frac{1}{\alpha_1 + \alpha_2 + \cdots + \alpha_m} \sum_{m=1}^{M} \alpha_m G_m(x)$$

The vital parameter of adaboost is the number of basic predictors, because it could influence the performance of models and computation consumption. While adaboost could enhance the forecasting performance of training set, overfitting problem will happen and computational cost will rise with the increasing number of basic predictors. Thus, a suitable number of basic predictors in adaboost needs to guarantee the balance between the good predictive effect and the less computation time.

2.4. Proposed STL-GESN and STL-AESN

The procedure of the proposed methods STL-GESN and STL-AESN for passenger flow prediction is shown in Fig. 3. The steps are listed as follows:

Step 1 *STL for decomposition*. The Chinese monthly passenger flow data X_t is decomposed into three components, including seasonality S_t, trend T_t and R_t.

Step 2 *ESN for modeling and forecasting*. ESN improved by GOA or adaboost is used to model and forecast the trend component T_t and the remainder component R_t. In addition, the seasonal-naive method is used to forecast the seasonal component S_t.

Step 3 *Integration for final forecasting results*. The forecasting results of three components are summed as the final forecasting of monthly passenger flow.

3. Empirical analysis

In this section, two real-world examples are conducted to validate the effectiveness and scalability of the proposed methods STL-GESN and STL-AESN, which are the monthly air passenger flow and rail passenger flow of China from January 2005 to August 2018.

3.1. Comparative example: monthly air passenger flow forecasting in China

3.1.1. Data description and indices of performance assessment

The monthly air passenger flow data from January 2005 to August 2018 is acquired from Chinese National Bureau of Statistics (http://www.stats.gov.cn/) with a total of 164 time-series records, as shown in Fig. 4. It can be seen from Fig. 4 that the data have obvious growth trend and seasonal characteristic. In this paper, data from January 2005 to February 2018 is utilized as training set, and March 2018 to August 2018 as test set. In view of the influence of great numeric range, the original data need to be normalized. In this paper, a linear transformation is used to put

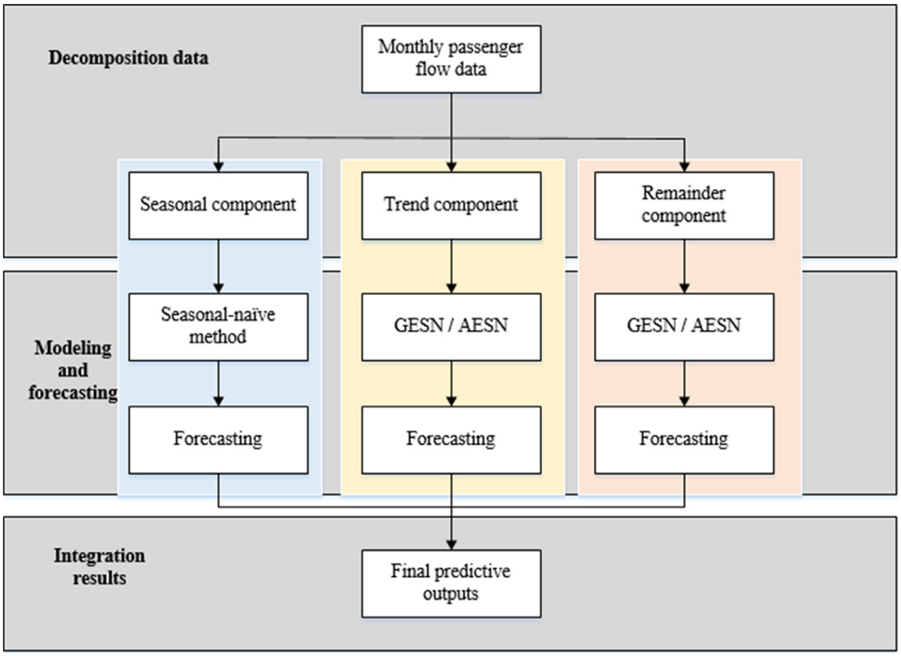


Fig. 3. The flowchart of proposed two hybrid methods for forecasting of monthly passenger flow in China.

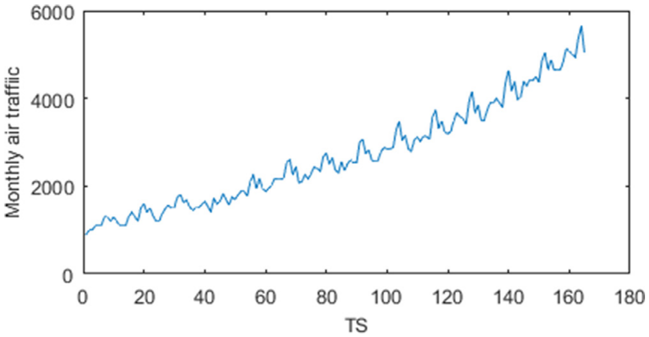


Fig. 4. Monthly air passenger flow in China.

the original data scaled into the range [0,1], as shown in Eq. (13).

$$\hat{x}_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{13}$$

where $\hat{x}_i$ is the normalized value, x_i is the unnormalized value, x_{min} and x_{max} are the minimum and maximum of the time-series data respectively.

The proposed hybrid approaches STL-GESN and STL-AESN are used for one-step-ahead prediction. It is known that the smaller forecasting error indicates the better performance of the corresponding models. For the purpose of evaluating and comparing the forecasting performance of the proposed models and other alternatives, two common forecasting errors are selected as indices of performance assessment, i.e root mean square error(RMSE) and mean absolute percentage error(MAPE). The RMSE and MAPE can be calculated by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2} \tag{14}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t)}{y_t}\right| \times 100\% \tag{15}$$

where y_t is the actual value, $\hat{y}_t$ is the forecasting value and n is the size of the actual values.

3.1.2. Parameter setting

In this paper, STL is implemented with the software R3.3.3, and ESNs based on two improvement strategies are programmed with the software Matlab 2016a.

The optimal time lag length K, i.e the number of input units in ESN, is determined by trial and error method. Given K is a integer that belongs to [1,12], the goal is to determine the optimal value of K, which renders the forecasting performance of training set best. When the values of K are 12, 12 and 7, ESN shows the well-pleasing performance on the original data, trend component and remainder component respectively. In addition, the number of internal units in reservoir N is set to 10 based on Eq. (8). Therefore, the structure of ESN consists of 12 or 7 input units, 10 internal units and 1 iutput units. Other parameter settings of ESN are as follows: the washout time step T_0 is 25, the activation functions of reservoir and output layer are tangent and linear function, respectively.

As stated in Section 2.3.1, the weight matrix W is initialized randomly, the GOA is used to search optimal values of four vital parameters of ESN, including W_{in}, W_{back}, α and β. The specific parameter settings of GOA are as follows: the population size *sizepop* is 20, the dimension of individuals in the swarm D is 4, the values in weight matrix W_{in} and W_{back} belong to [-1,1], the connection rate α belongs to [0.01,0.1], the spectral radius β belongs to [0.1,0.9], the max iterations is 100. After four optimal parameters are obtained, W is updated as: $W \leftarrow W^{*}\alpha/\beta$. In this way, the optimal parameters of ESN are determined.

The only parameter of adaboost is the number of basic predictors M, which could influence the forecasting performance of STL-AESN. If the value of it is too small, the enhancing effect of performance on ESN may not be evident, while too big value could increase computational cost. Based on several trials, the value of M is set to 5.

3.1.3. Method comparison

To validate the optimization ablity of GOA, other three evolution algorithms including particle swarm optimization(PSO) [43], grey wolf optimization(GWO) [44] and whale optimization algorithm(WOA) [45] are introduced as the benchmarks for comparison. In addition, the Friedman test is used to detect significant

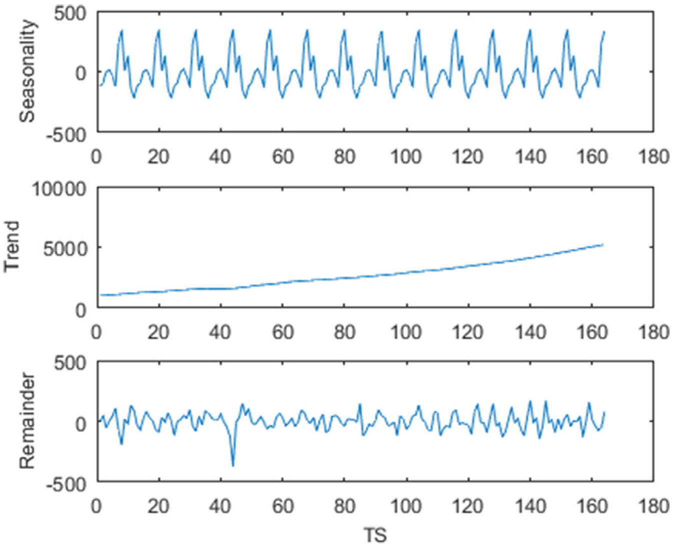


Fig. 5. The decomposition results of monthly air passenger flow using STL.

Table 1
The forecasting performance of ESN, PSO-ESN, GWO-ESN, WOA-ESN and GOA-ESN models for monthly air passenger flow forecasting.

Model	Training set		Test set	
	RMSE	MAPE(%)	RMSE	MAPE(%)
ESN	102.3439	3.4017	143.4939	2.2637
PSO-ESN	94.1676	3.4472	135.3856	2.3785
GWO-ESN	83.0460	2.8393	143.3343	2.2646
WOA-ESN	92.0846	3.1104	108.8308	1.7304
GOA-ESN	92.6213	3.1697	116.584	1.8496

$F = 5.7 > F_{0.05}(5, 8) = 3.687$, reject H_0

Table 2
The forecasting performance of ESN, EMD-ESN, SEAM-ESN and STL-ESN models for monthly air passenger flow forecasting.

Model	Training set		Test set	
	RMSE	MAPE(%)	RMSE	MAPE(%)
ESN	99.8265	3.4623	159.2997	2.6360
EMD-ESN	63.7810	2.2477	148.3139	2.2852
SEAM-ESN	99.8265	3.4623	66.8646	4.7069
STL-ESN	56.5053	1.9058	83.1109	1.4827

Table 3
The seasonal indexes for each month in air passenger flow forecasting.

Month	Seasonal index	Month	Seasonal index
Jan.	1.0487	Jul.	0.9866
Feb.	1.0455	Aug.	0.9956
Mar.	0.9589	Sep.	0.9876
Apr.	1.0544	Oct.	1.0011
May.	1.0280	Nov.	0.9997
Jun.	0.9863	Dec.	1.0073

differences among the results of different algorithms [46]. In this paper, the null hypothesis of Friedman test is defined as no difference in the predictive performance of models, and the alternative hypothesis means that the negation of the null hypothesis. The statistic F of Friedman test is shown as Eq. (16):

$$F = \frac{12}{nk(k+1)} \sum_{j=1}^{k} rank_j^2 - 3n(k+1) \tag{16}$$

where k is the number of compared models; n is the total number of forecasting results; $rank_j$ is the total rank of $j-th$ model concerning forecasting error.

To validate the contribution of STL, the common seasonal exponential adjustment method(SEAM) for dealing with seasonal variations of data and the frequently applied signal-decomposition method named empirical mode decomposition(EMD) are introduced as the benchmarks for comparison.

To validate the effectiveness of proposed methods STL-GESN and STL-AESN, other common time-series forecasting models including seasonal autoregressive integrated moving average(SARIMA), X-12-ARIMA, least square support vector regression(LSSVR), back propagation neural network(BPNN) and extreme learning machine(ELM) are utilized as the benchmarks for comparison.

The experimental environment includes Matlab 2016a, R3.3.3, self-written Matlab programs and a computer with an Inter(R) Core(TM)-i7 and Windows 10 operating system.

3.1.4. Forecasting results

The decomposition results of monthly air passenger flow data via STL are shown in Fig. 5. It can be learned that the original data is decomposed into seasonal, trend and remainder components. The seasonal component shows a 12-month cycle, and the trend component shows a growth trend.

Table 1 shows the two performance assessment indices RMSE and MAPE of ESN, PSO-ESN, GWO-ESN, WOA-ESN and GOA-ESN on the original data of air passenger flow, and the optimization processes of PSO, GWO, WOA and GOA for searching optimal parameters of ESN are shown in Fig. 6. It can be learned from Table 1 that: (i)On both training set and test set, the two values(RMSE and MAPE) obtained from GOA-ESN are smaller than those calculated by ESN, which proves the effectiveness of GOA for improving the performance of ESN. (ii) The two values of GWO-ESN are the smallest on training set but relatively large in test set among models, which means GWO-ESN may cause over-fitting. While the performance of GOA-ESN is little inferior to WOA-ESN on test set, the former performs much better than PSO-ESN and GWO-ESN, which indicates the cetain superiority of GOA-ESN. The result of Friedman test ($F = 5.7$) at the 0.05 significance level shows that GOA-ESN is significant superior to other models. In a word, GOA has the advantage in optimizing the four parameters of ESN for obtaining higher prediction accuracy.

Table 2 shows the RMSE and MAPE of ESN, STL-ESN, EMD-ESN and SEAM-ESN on the original data. In addition, the decomposition results of monthly air passenger flow data via EMD are shown in Fig. 7, and the calculated seasonal indexes based on the forecasting values and actual values of training set are shown in Table 3. It can be learned from Table 2 that: (i) On both training set and test set, the two values(RMSE and MAPE) obtained from STL-ESN are smaller than those calculated by ESN, which proves the effectiveness of STL for improving the forecasting performance. (ii) The forecasting performance of EMD-ESN indeed enhances compared to ESN, and the value of RMSE from SEAM-ESN is smaller than that from ESN, however the value of MAPE becomes bigger unexpectedly. The two values of STL-ESN are the smallest among chosen models, which proves the applicability of STL-ESN in this paper.

Fig. 7 displays that the monthly air passenger flow data is decomposed into three intrinsic mode functions(IMFs) and a residual component via EMD. It is clear that the residual component shows a upward trend but IMFs do not have regular features. Table 3 lists 12 seasonal indexs for adjusting the seasonal variations in the forecasting values. The seasonal indexes larger than 1 mean the forecasting values of monthly air passenger flow from ESN are underestimated while less than 1 mean the values are overestimated.

The forecasting results and two performance assessment indices of the proposed methods(STL-GESN and STL-AESN), and other models including SARIMA, X-12-ARIMA, LSSVR, BPNN, ELM,

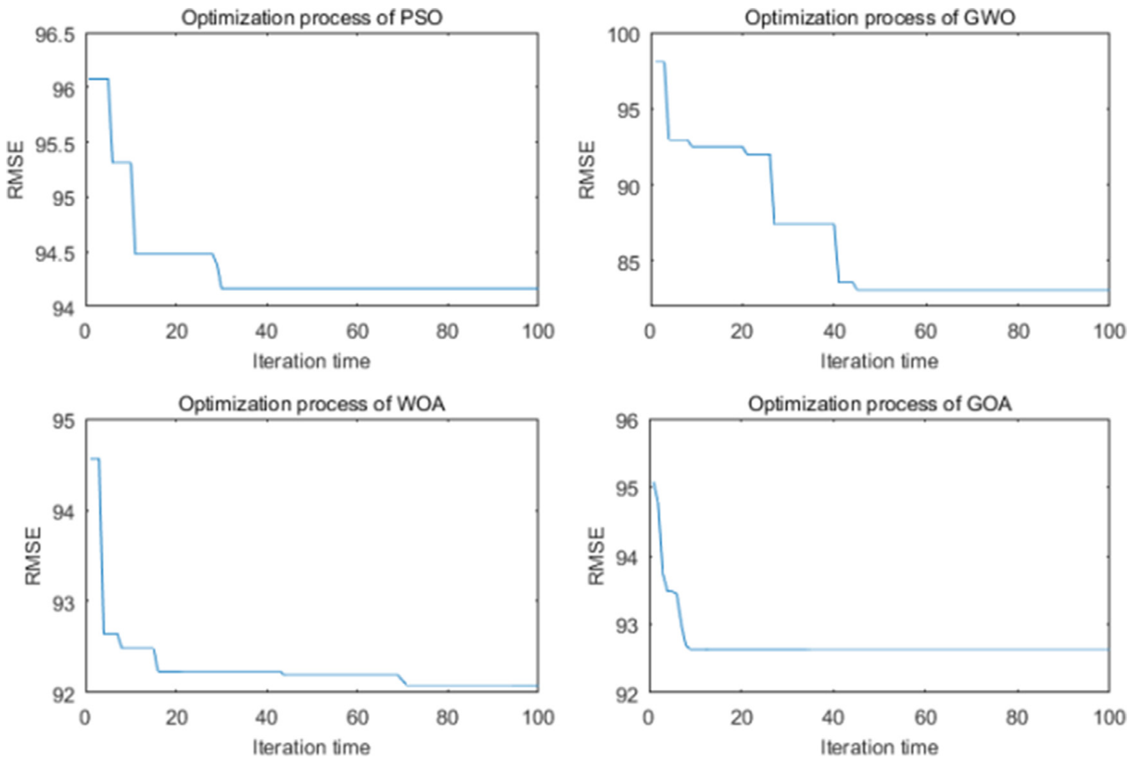


Fig. 6. The optimization processes of PSO, GWO, WOA and GOA for monthly air passenger flow forecasting.

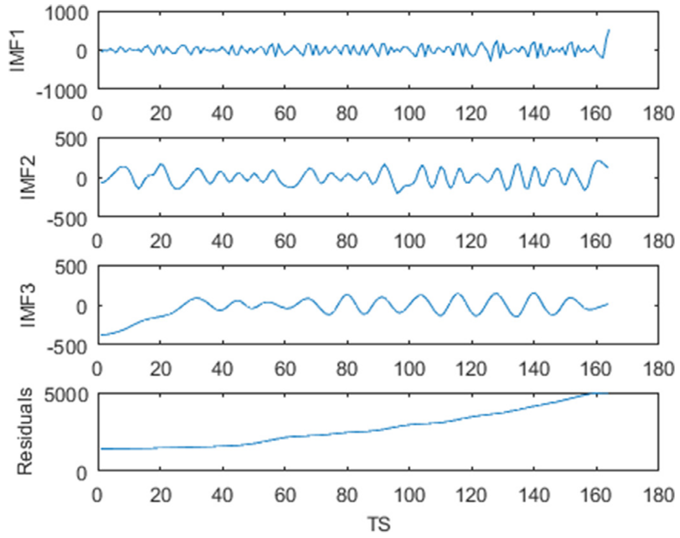


Fig. 7. The decomposition results of monthly air passenger flow using EMD.

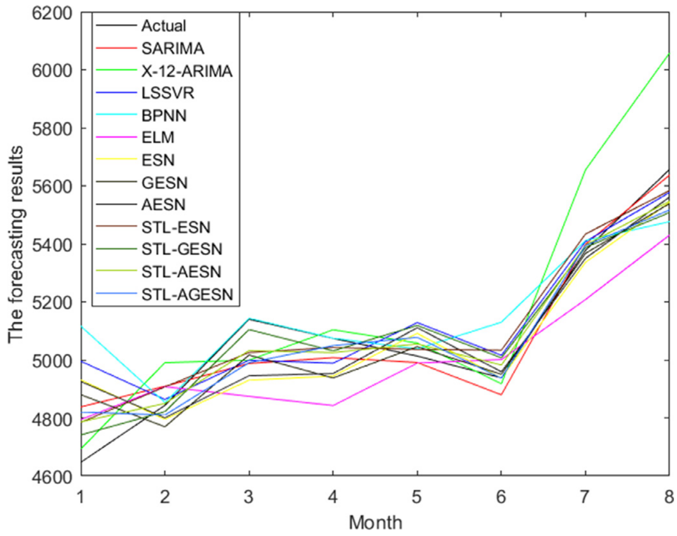


Fig. 8. The forecasting results of monthly air passenger flow in China.

ESN, GESN, AESN, STL-ESN and STL-AGESN, are shown in Table 4. It can be learned from Table 4 that: (i) The values of two indices given SARIMA are 95.6635 and 1.5142%, smaller than other single models, which shows the good performance of this model in time series forecasting. (ii) The values of MAPE given by X-12ARIMA and proposed method STL-AESN are 2.6137% and 1.2746% respectively, which manifests that the proposed methods have more obvious advantage to forecast air passenger flow compared to widely used X-12-ARIMA. (iii) The values of indices given by STL-GESN are 79.5883 and 1.2921%, and by STL-AESN are 78.7483 and 1.2746% respectively, all smaller than those of other models. The results show the contribution made by two improvement strategies, and the superiority of the proposed methods. In addition, the model STL-AGESN(98.8721 and 1.4893%) has worse forecasting performance than STL-GESN and STL-AESN, which indicates that it is inadvisable to combine GOA and adaboost for enhancing the forecasting performance of STL-ESN.

Fig. 8 displays the forecasting and actual values of monthly air passenger flow in China from January 2018 to August 2018 obtained by 6 single models and 6 hybrid models. It is obvious that the forecasting values of STL-GESN and STL-AESN are much closer to the actual values, which shows better forecasting performance of proposed methods.

In conclusion, the proposed STL-ESNs based on two improvement strategies apparently obtain the approving forecasting

Table 4
The forecasting results of SARIMA, X-12-ARIMA, LSSVR, BPNN, ELM, ESN, GESN, AESN, STL-ESN, STL-GESN, STL-AESN and STL-AGESN models for air passenger flow(unit: people).

Month	Actual	SARIMA	X-12-ARIMA	LSSVR	BPNN	ELM	ESN
2018.1	4647	4837.8550	4692.9888	4995.3648	5116.3898	4794.0046	4931.0080
2018.2	4843	4908.8930	4990.3041	4863.6305	4853.2781	4907.6417	4794.9979
2018.3	5140	4987.4410	4998.9105	4998.7772	5143.2268	4874.1936	4929.9073
2018.4	5074	5007.5550	5103.6406	4988.6649	5074.5964	4842.5023	4944.1261
2018.5	5013	4990.8940	5058.0773	5128.2514	5037.8731	4988.8928	5091.2331
2018.6	4938	4879.4210	4917.7465	5015.7339	5129.9799	5001.1560	4937.0461
2018.7	5378	5399.0640	5655.1016	5407.9574	5411.2258	5208.0056	5336.5259
2018.8	5657	5637.2150	6057.9341	5577.3416	5476.5724	5430.4529	5543.3709
RMSE	–	95.6635	188.6047	148.1460	190.9110	171.3274	143.4939
MAPE(%)	–	1.5142	2.6137	2.2738	2.3224	2.8947	2.2637
Month	Actual	GESN	AESN	STL-ESN	STL-GESN	STL-AESN	STL-AGESN
2018.1	4647	4880.0159	4925.4029	4784.5851	4741.4926	4786.5177	4820.2710
2018.2	4843	4769.5149	4800.2843	4906.8805	4821.3986	4849.7070	4809.7262
2018.3	5140	5016.5013	4945.3326	5025.3789	5104.3229	5031.9379	4990.3214
2018.4	5074	4937.6345	4952.9444	5042.4544	5031.1218	5025.2296	5048.9642
2018.5	5013	5045.5215	5110.8741	5036.3950	5118.5427	5054.7790	5077.5622
2018.6	4938	4950.1336	4958.8273	5033.3196	5006.9622	4982.1205	4936.6292
2018.7	5378	5367.3091	5350.1067	5433.9103	5387.8683	5395.5174	5394.2589
2018.8	5657	5539.2222	5561.6136	5583.8099	5508.0871	5547.2490	5516.9623
RMSE	–	116.5840	137.7372	83.1109	79.5883	78.7483	98.8721
MAPE(%)	–	1.8496	2.2031	1.4827	1.2921	1.2746	1.4893

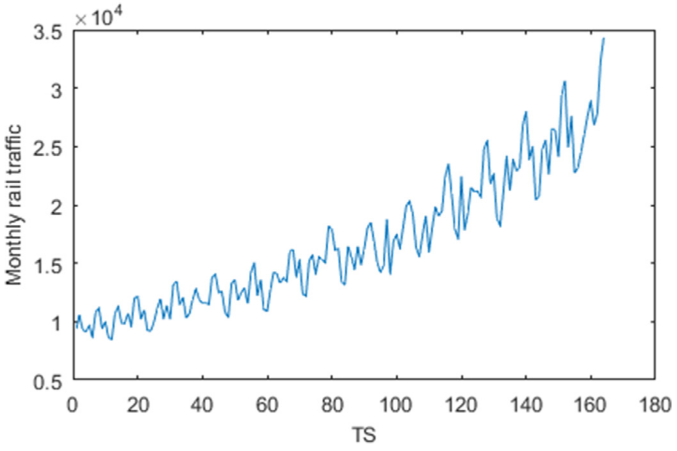


Fig. 9. Monthly rail passenger flow in China.

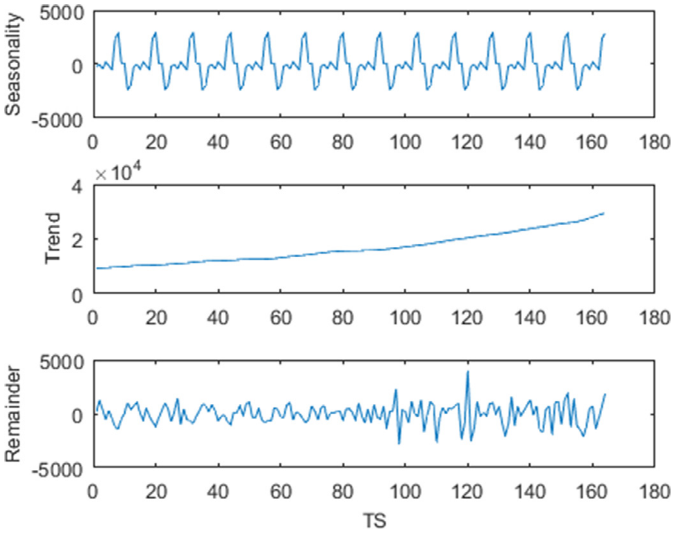


Fig. 10. The decomposition results of monthly rail passenger flow using STL.

results, and outperform other alternatives in Chinese monthly air passenger flow forecasting.

3.2. Extended example: monthly rail passenger flow forecasting in China

The monthly rail passenger flow data from January 2005 to August 2018 are also acquired from Chinese National Bureau of Statistics (http://www.stats.gov.cn/) with a total of 164 time-series records, as shown in Fig. 9. It is applied to compare the proposed STL-GESN and STL-AESN, and other single or hybrid models. Similarly, the data from January 2005 to February 2018 are utilized as training set, and March 2018 to August 2018 as test set. Because the magnitude of rail passenger flow data is large, this paper shrinks the data tenfold and then implements data normalization based on Eq. (13).

The decomposition results of monthly rail passenger flow data via STL are shown in Fig. 10. It can be learned that the original data is decomposed into seasonal, trend and remainder components. The seasonal component shows a 12-month cycle, and the trend component shows a growth trend, which indicates the traffic flow data have some intercommunity.

Table 5
The forecasting performance of ESN, PSO-ESN, GWO-ESN, WOA-ESN and GOA-ESN models for monthly rail passenger flow forecasting.

Model	Training set		Test set	
	RMSE	MAPE(%)	RMSE	MAPE(%)
ESN	96.5759	4.4615	182.7676	5.112
PSO-ESN	90.8546	4.2429	180.6008	5.4676
GWO-ESN	84.0645	4.1525	189.6735	5.7611
WOA-ESN	89.5219	4.3750	167.2039	4.6660
GOA-ESN	91.4610	4.3195	158.3179	4.2422

$F = 4.7 > F_{0.05}(5, 8) = 3.687$, reject H_0

Table 5 shows two performance assessment indices RMSE and MAPE of ESN, PSO-ESN, GWO-ESN, WOA-ESN and GOA-ESN on the original data of rail passenger flow, and the optimization processes of PSO, GWO, WOA and GOA for searching optimal parameters of

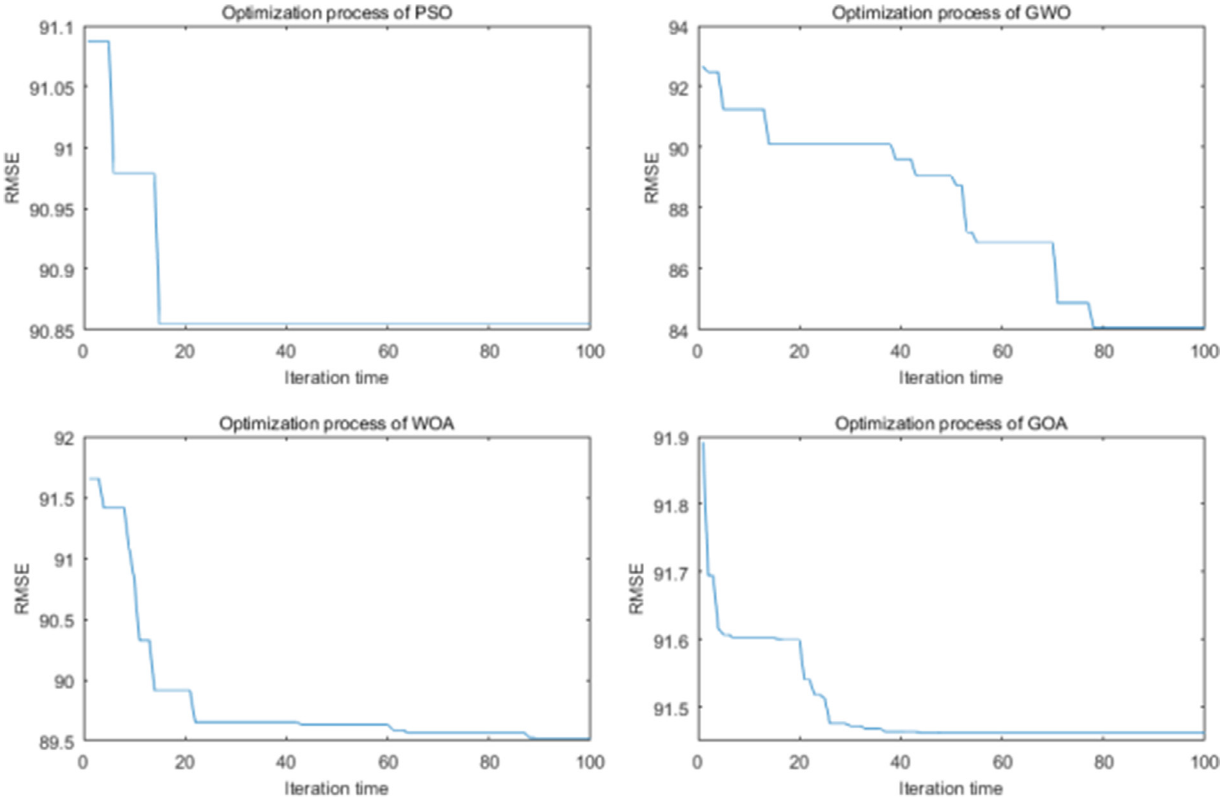


Fig. 11. The optimization processes of PSO, GWO, WOA and GOA for for monthly rail passenger flow forecasting.

Table 6
The forecasting performance of ESN, EMD-ESN, SEAM-ESN and STL-ESN models for monthly rail passenger flow forecasting.

Model	Training set		Test set	
	RMSE	MAPE(%)	RMSE	MAPE(%)
ESN	96.9146	4.3659	167.4080	4.5908
EMD-ESN	84.3447	4.0154	160.5913	4.9292
SEAM-ESN	96.9146	4.3659	73.7718	6.0458
STL-ESN	78.6781	3.8470	149.8055	4.6190

ESN are shown in Fig. 11. It can be learned from Table 5 that: (i) On both training set and test set, the two values(RMSE and MAPE) obtained from GOA-ESN are smaller than those calculated by ESN, which proves the effectiveness of GOA for improving the performance of ESN. (ii) The two values of GWO-ESN are the smallest on training set but the biggest in test set among models, which means GWO-ESN may not effectively improve the performance of ESN in this paper. In terms of RMSE, GOA-ESN performs worst among four hybrid models, but the gap is not overt based on the criterion of MAPE. Besides, GOA-ESN has better performance on test set than PSO-ESN, GWO-ESN and WOA-ESN. The result of Friedman test($F = 4.7$) at the 0.05 significance level shows that GOA-ESN is significant superior to other models. In short, it is feasible to optimize the four parameters of ESN for obtaining higher prediction accuracy using GOA.

Table 6 shows the RMSE and MAPE of ESN, STL-ESN, EMD-ESN and SEAM-ESN on the original data of rail passenger flow. In addition, the decomposition results of monthly rail passenger flow data via EMD are shown in Fig. 12, and the calculated seasonal indexes based on the forecasting values and actual values of training set are shown in Table 7. It can be learned from Table 6 that: (i) In terms of RMSE, on both training set and test set, the value obtained from STL-ESN are smaller than those calculated by ESN, however the forecasting performance of the former on test set is little inferior to the latter in consideration of MAPE. In general, the comparative analysis still sure the effectiveness of STL for improving the forecasting performance. (ii) The forecasting performance of EMD-ESN indeed enhances on training set and test set, the value of RMSE from SEAM-ESN is smaller than that from ESN, however the value of MAPE becomes bigger unexpectedly. The value of MAPE obtained from STL-ESN is the smallest among chosen models, which proves the applicability of STL-ESN in this paper.

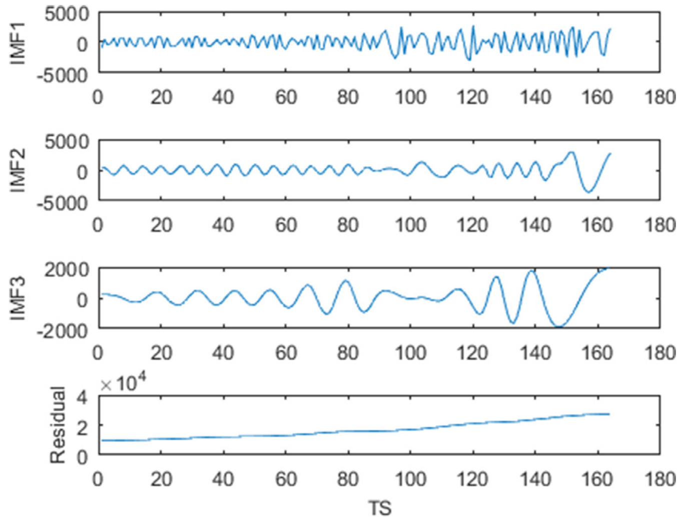


Fig. 12. The decomposition results of monthly rail passenger flow using EMD.

Fig. 12 displays that the monthly rail passenger flow data is decomposed into three intrinsic mode functions(IMFs) and a residual component via EMD. It is clear that the residual component shows

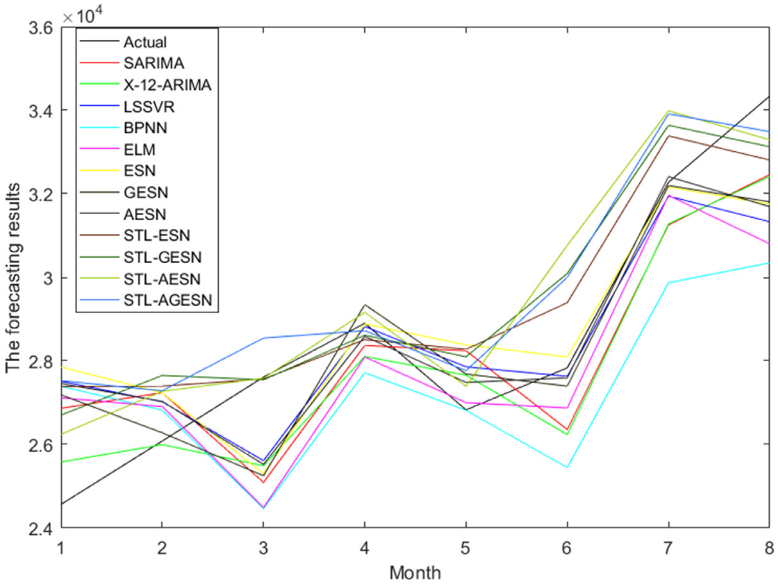


Fig. 13. The forecasting results of monthly rail passenger flow in China.

Table 7
The seasonal indexes for each month in rail passenger flow forecasting.

Month	Seasonal index	Month	Seasonal index
Jan.	1.0287	Jul.	0.9557
Feb.	1.0152	Aug.	1.0048
Mar.	0.9968	Sep.	1.0244
Apr.	0.9756	Oct.	0.9885
May.	1.0238	Nov.	1.0122
Jun.	1.0244	Dec.	0.9946

a upward trend but IMFs do not have regular features. Table 7 lists 12 seasonal indexes for adjusting the seasonal variations in the forecasting values. The seasonal indexes larger than 1 mean the forecasting values of monthly rail passenger flow from ESN are underestimated while less than 1 mean the values are overestimated.

The forecasting results and two performance assessment indices of the proposed methods(STL-GESN and STL-AESN), and other models including SARIMA, X-12-ARIMA, LSSVR, BPNN, ELM, ESN, GESN, AESN, STL-ESN and STL-AGESN, are shown in Table 8. It can be learned from Table 8 that: (i) The values of RMSE and MAPE given by X-12-ARIMA are 132.8794 and 4.0563%, smaller than other single models, which shows the good performance of this model in rail passenger flow forecasting. In addition, the values of MAPE given by X-12-ARIMA is little smaller than the proposed method STL-AESN(4.1724), which shows that the proposed methods have no advantages compared to widely used X-12-ARIMA in this scene. (ii) The values of RMSE and MAPE given by SARIMA are 165.9477 and 5.4990%, significantly greater than those of STL-GESN and STL-AESN, which manifests that SARIMA has poor performance in rail passenger flow forecasting. (iii) The values of indices given by STL-GESN are 146.4895 and 4.5695%,

Table 8
The forecasting results of SARIMA, X-12-ARIMA, LSSVR, BPNN, ELM, ESN, GESN, AESN, STL-ESN, STL-GESN, STL-AESN and STL-AGESN models for rail passenger flow(unit: people).

Month	Actual	SARIMA	X-12-ARIMA	LSSVR	BPNN	ELM	ESN
2018.1	24564	26865.5900	25578.8870	27510.7433	27392.1589	27116.3428	27853.1552
2018.2	26081	27238.6300	25994.8880	27020.7891	26820.4327	26905.3919	27236.9882
2018.3	27612	25091.7400	25489.2520	25611.3675	24467.4762	24486.3089	25298.1423
2018.4	28900	28370.1400	28102.8470	28821.0911	27717.1329	28092.8678	28884.0511
2018.5	26827	28242.0400	27658.7290	27857.9163	26815.0579	27000.4299	28383.0857
2018.6	27834	26356.0800	26233.2180	27629.9021	25446.4244	26872.6309	28093.8836
2018.7	32276	31253.3600	31277.8480	31942.0244	29870.7996	31961.3582	32159.7803
2018.8	34340	32461.5300	32417.4910	31328.7877	30346.9509	30800.2803	31749.3068
RMSE	–	165.9477	132.8794	172.6942	243.0600	197.4791	182.7676
MAPE(%)	–	5.4990	4.0563	4.6872	7.1915	5.3809	5.1120

Month	Actual	GESN	AESN	STL-ESN	STL-GESN	STL-AESN	STL-AGESN
2018.1	24564	27187.1654	27456.8419	27385.6652	26702.2327	26241.1099	27524.2104
2018.2	26081	26278.6592	27029.8828	27395.9381	27650.5848	27271.3457	27284.1645
2018.3	27612	25250.1259	25522.3858	27564.8848	27547.5001	27585.7761	28545.8683
2018.4	28900	29346.9682	28587.3692	28508.9402	28613.9764	29161.1262	28721.9212
2018.5	26827	27685.2385	27483.8309	28276.1359	28097.7591	27393.4850	27766.6741
2018.6	27834	27394.7192	27593.3224	29392.9101	30091.3642	30764.8638	29998.5526
2018.7	32276	32196.5034	32409.6015	33384.5576	33638.9125	33984.8319	33912.2285
2018.8	34340	31806.7238	31690.9199	32807.6025	33124.3984	33291.5138	33486.9928
RMSE	–	158.3179	163.0106	149.8055	146.4895	146.7515	158.4590
MAPE(%)	–	4.2422	4.4382	4.6190	4.5695	4.1724	4.9369

and by STL-AESN are 146.7515 and 4.1724%, respectively, all smaller than those of other models expect X-12-ARIMA. The results show the contribution made by two improvement strategies, and the limited advantages of the proposed methods. In addition, the model STL-AGESN(158.4590 and 4.9369%) has worse forecasting performance than STL-GESN and STL-AESN, which indicates that it is inadvisable to combine GOA and adaboost for enhancing the forecasting performance of STL-ESN. It is noted that the advantage of proposed methods is less overt for forecasting rail passenger flow than air passenger flow.

Fig. 13 displays the forecasting and actual values of monthly rail passenger flow in China from January 2018 to August 2018 obtained by SARIMA, X-12-ARIMA, LSSVR, BPNN, ELM, ESN, GESN, AESN, STL-ESN, STL-GESN and STL-AESN. Without considering X-12-ARIMA, it is obvious that the forecasting values of STL-GESN and STL-AESN have the smallest gaps with the actual values, which shows excellent forecasting performance of proposed models.

In total, the proposed STL-GESN and STL-AESN perform well in Chinese monthly rail passenger flow forecasting.

4. Conclusions and future research

The passenger flow forecasting could provide valid guide for the normal operation of transportation system and the rational investment of resources. That is to say, it is greatly significant to enhance the prediction accuracy of passenger flow. In our study, the data concerning monthly pasenger flow have features of obvious seasonality and nonlinearity, which generates the difficulty of accurate forecasting. Thus, we proposed STL-ESNs based on two improvement strategies(GOA and adaboost ensemble). The basic model ESN is used to capture the complex nonlinear relationship in data, the time-series decomposition method STL is utilized to handle the seasonal variations, the evolution algorithm GOA is applied to search the optimal values of the four parameters in ESN, and the adaboost framework is introduced to improve the generalization ability of ESN. The main conclusions are as follows:

(1) In consideration of the remarkable seasonality of passenger flow, thus it is imperative to handle seasonal variations of data before modeling for better forecasting performance. The experimental results manifest that STL is the most appropriate decomposition method used for deseasonalizing and detrending in this study. Compared to widely used X-12-ARIMA, it is easier to deal with seasonality with STL. Because STL only focus on data without considering other elements, such as holiday effect and outliers. Compared to seasonal index adjustment method, the additive decomposition method STL unioning ESN could achive higher forecasting accuracy.

(2) The hybrid methods using STL-ESN based on two improvement strategies are firstly proposed for monthly passenger flow forecasting. In general, the proposed methods including STL-GESN and STL-AESN are easy to be realized and pretty effective for passenger flow forecasting. The conditions under which they work better are as follows: In terms of data, the passenger flow have remarkable seasonality and upward trend, and the outliers can be present in original time series. Because these conditions could guarantee the satisfactory decomposition effect of seasonal adjustment method(STL). In terms of model, STL-GESN and STL-AESN need to determine the optimal values of vital parameters about ESN, and the suitable number of basic predictors in adaboost framework respectively for ensuring that neural network model(ESN) works best.

(3) The widely used model SARIMA yields good forecasting performance in the first application but performs poorly in the second application. First, as one of multiplicative model, SARIMA could handle seasonal variations of data during modeling, which is something that other models do not have. Second, other single models like neural networks may perform poorly with parameters out of optimization. Third, the two time series used in this paper are well behaved, which is also a main reason that the limited improvement in forecasting performance between SARIMA and the proposed methods. SARIMA is indeed more time-saving compared to the proposed methods. However, as one of linear time series models, the capacity of it is fairly limited to deal with complicated nonlinear issues. In addition, SARIMA only models single time series and could not introduce exogenous variables. In other words, facing more complex nonlinear problems, it is advisable to apply the proposed methods rather than SARIMA for getting higher modeling accuracy under sacrificing certain computational cost.

As for the future research directions, some problems could be further studied, as follows:

(a) The proposed STL-ESNs based on two improvement strategies need to be applied in more forecasting problems for strengthening their expandability.

(b) Considering some factors indeed impact passenger flow, so exogenous variables can be introduced to model and forecast for obtaining better results.

Conflict of interest

The authors declare no confllict of interest.

Acknowledgements

This study is supported by National Nature Science Foundation of China (Grant No. 41571016) and the National Key Research and Development Program of China (Grant No. 2018YFC0406606). The authors would like to thank the anonymous referees for their valuable comments.

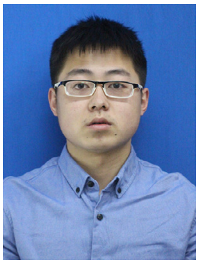
References

[1] N.L. Nihan, K.O. Holmesland, Use of the box and Jenkins time series technique in traffic forecasting, Transportation 9 (2) (1980) 125–143.
[2] B.M. Williams, Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling, J. Transportat. Res. Board 1776 (1) (2001) 194–200.
[3] D. Pavlyuk, Short-term traffic forecasting using multivariate autoregressive models, Proc. Eng. 178 (2017) 57–66.
[4] B.M. Williams, L.A. Hoel, Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results, J. Transp. Eng. 129 (2003) 664–672.
[5] A. Rusyana, N. Marzuki, M. Flancia, Sarima model for forecasting foreign tourists at the Kualanamu International Airport, in: Proceedings of the International Conference on Mathematics, Statistics, and Their Applications, 2017, pp. 153–158.
[6] H. Guo, X.P. Xiao, J. Forrest, et al., Urban road short-term traffic flow forecasting based on the delay and nonlinear grey model, J. Transp. Syst. Eng. Inf. Technol. 13 (6) (2013) 60–66.
[7] J.H. Guo, W. Huang, B.M. Williams, Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification, Transp. Res. Part C 43 (2014) 50–64.
[8] S.V. Kumar, Traffic flow prediction using Kalman filtering technique, Proc. Eng. 187 (2017) 582–587.
[9] T.T. Tchrakian, B. Basu, M. O'Mahony, Real-time traffic flow forecasting using spectral analysis, IEEE Trans. Intell. Transp. Syst. 13 (2) (2012) 519–526.
[10] Y.R. Zhang, Y.L. Zhang, A. Haghani, A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model, Transp. Res. Part C 43 (1) (2014) 65–78.
[11] M. Castro-Neto, Y.S. Jeong, M.K. Jeong, et al., Online-svr for short-term traffic flow prediction under typical and atypical traffic conditions, Expert Syst. Appl. Int. J. 36 (3) (2009) 6164–6173.

[12] W.C. Hong, Y.C. Dong, F.F. Zheng, et al., Forecasting urban traffic flow by SVR with continuous ACO, Appl. Math. Model. 35 (3) (2011) 1282–1291.
[13] Y.L. Cong, J.W. Wang, X.L. Li, Traffic flow forecasting by a least squares support vector machine with a fruit fly optimization algorithm, Proc. Eng. 137 (2016) 59–68.
[14] S.W. Li, T. Chen, L. Wang, et al., Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index, Tour. Manag. 68 (2018) 116–126.
[15] Z.Y. Ma, G.C. Luo, D.J. Huang, Short term traffic flow prediction based on on-line sequential extreme learning machine, in: Proceedings of the Eighth International Conference on Advanced Computational Intelligence, 2016, pp. 143–149.
[16] F. Kong, J. Li, B. Jiang, et al., Short-term traffic flow prediction in smart multimedia system for internet of vehicles based on deep belief network, Fut. Gener. Comput. Syst. 93 (2018) 460–472.
[17] M.F. Zhou, X.B. Qu, L.X. Peng, A recurrent neural network based microscopic car following model to predict traffic oscillation, Transp. Res. Part C Emerg. Technol. 84 (2017) 245–264.
[18] X.L. Ma, Z.M. Tao, Y.H. Wang, et al., Long short-term memory neural network for traffic speed prediction using remote microwave sensor data, Transp. Res. Part C 54 (2015) 187–197.
[19] Y. Tian, K.L. Zhang, J.Y. Li, et al., Lstm-based traffic flow prediction with missing data, Neurocomputing 318 (2018) 297–305.
[20] S.S. Zhong, X.L. Xie, L. Lin, et al., Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction, Neurocomputing 238 (2017) 191–204.
[21] R. Marco, B. Piero, Z. Enrico, et al., Ensemble of optimized echo state networks for remaining useful life prediction, Neurocomputing 281 (2017) 121–138.
[22] A.C. Mohammad, M.S. Fadali, M.T. Andrzej, Wind speed and wind direction forecasting using echo state network with nonlinear functions, Renew. Energy 131 (2018) 979–889.
[23] L. Wang, S.X. Lv, Y.R. Zeng, Effective sparse adaboost method with ESN and FOA for industrial electricity consumption forecasting in china, Energy 155 (2018) 1013–1031.
[24] S. Saremi, S. Mirjalili, A. Lewis, Grasshopper optimisation algorithm: theory and application, Adv. Eng. Softw. 105 (2017) 30–47.
[25] M. Mafarja, I. Aljarah, H. Faris, et al., Binary grasshopper optimisation algorithm approaches for feature selection problems, Expert Syst. Appl. 117 (2019) 267–286.
[26] J.F. Wu, H.L. Wang, N. Li, et al., Distributed trajectory optimization for multiple solar-powered UAVs target tracking in urban environment by adaptive grasshopper optimisation algorithm, Aerosp. Sci. Technol. 70 (2017) 497–510.
[27] J. Luo, C.H. Ling, Q. Zhang, et al., An improved grasshopper optimization algorithm with application to financial stress prediction, Appl. Math. Model. 64 (2018) 654–668.
[28] X. Zhang, Q. Miao, H. Zhang, et al., A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery, Mech. Syst. Signal Process. 108 (2018) 58–72.
[29] U. Sultana, B. Khairuddin, Sultana, Placement and sizing of multiple distributed generation and battery swapping stations using grasshopper optimizer algorithm, Energy 165 (2018) 408–421.
[30] H. Liu, H.Q. Tian, Y.F. Li, et al., Comparison of four adaboost algorithm based artificial neural networks in wind speed predictions, Energy Convers. Manag. 92 (92) (2015) 67–81.
[31] L. Li, C.Y. Wang, W. Li, et al., Hyperspectral image classification by adaboost weighted composite kernel extreme learning machines, Neurocomputing 275 (2017) 1725–1733.
[32] Z. Zakaria, S.A. Suandi, M.S. Junita, Hierarchical skin-adaboost-neural network (h-skann) for multi-face detection, Appl. Soft Comput. 68 (2018) 172–190.
[33] G.P. Zhanga, Neural network forecasting for seasonal and trend time series, Eur. J. Oper. Res. 160 (2) (2005) 501–514.
[34] D. Pfeffermann, M. Morry, P. Wong, Estimation of the variances of x-11 ARIMA seasonally adjusted estimators for a multiplicative decomposition and heteroscedastic variances, Int. J. Forecast. 11 (2) (1995) 271–283.
[35] Z.B. Zhou, X.C. Dong, Analysis about the seasonality of China's crude oil import based on x-12-ARIMA, Energy 42 (1) (2012) 281–288.
[36] G.H. Cao, L.J. Wu, Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting, Energy 115 (2016) 734–745.
[37] C. Plosser, Trends and random walks in macroeconomic time series: some evidence and implications, J. Monet. Econ. 10 (2) (1982) 139–162.
[38] M. Theodosiou, Forecasting monthly and quarterly time series using STL decomposition, Int. J. Forecast. 27 (4) (2011) 1178–1195.
[39] A.O. Boudraa, J.C. Cexus, Emd-based signal filtering, IEEE Trans. Instrum. Meas. 56 (6) (2007) 2196–2202.
[40] S. Hochreiter, The vanishing gradient problem during learning recurrent neural nets and problem solutions, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 6 (02) (1998) 107–116.
[41] L. Wang, H.L. Hu, X.Y. Ai, et al., Effective electricity energy consumption forecasting using echo state network improved by differential evolution algorithm, Energy 153 (153) (2018) 801–815.
[42] T. Zhou, G.Q. Han, X.M. Xu, et al., δ -agree adaboost stacked autoencoder for short-term traffic flow forecasting, Neurocomputing 247 (2017) 31–38.
[43] B. Maryam, Optimizing multi-objective PSO based feature selection method using a feature elitism mechanism, Expert Syst. Appl. 113 (2018) 499–514.
[44] M.E. Ali, M.H. Hassan, Farh, Dynamic global maximum power point tracking of the PV systems under variant partial shading using hybrid GWO-FLC, Solar Energy 177 (2018) 306–316.
[45] M.M. Mafarja, S. Mirjalili, Hybrid whale optimization algorithm with simulated annealing for feature selection, Neurocomputing 260 (2018) 302–312.
[46] G.F. Fan, L.L. Peng, W.C. Hong, Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model, Appl. Energy 224 (2018) 13–33.

Lan Qin received bachelor's degree in bioinformatics from Chongqing University of Posts and Telecommunications in 2016, China. He is pursuing his Master degree in Lanzhou university, China. His research interests are in data mining and machine learning.

Weide Li received his Ph.D. degree in ecology from Lanzhou university in 2004. He is a professor in the school of mathematics and statistics of Lanzhou university, China. His research interests include data analysis and mathematical modeling.



Shijia Li received bachelor's degree in statistics from Harbin University of Science and Technology in 2016, Harbin, China. He is pursuing his Master degree in Lanzhou university, China. His research interests are in machine learning and actuarial science.