Engineering Applications of Artificial Intelligence 99 (2021) 104132

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Effective electricity load forecasting using enhanced double-reservoir echo state network

Lu Peng [a], Sheng-Xiang Lv [b], Lin Wang [a,*], Zi-Yun Wang [a]

[a] School of Management, Huazhong University of Science and Technology, Wuhan 430074, China
[b] School of Business Administration, Guangdong University of Finance & Economics, Guangzhou 510320, China

ARTICLE INFO

Keywords:
Electricity load forecasting
Echo state network
Mutual information
Backtracking search optimization algorithm

ABSTRACT

Accurate electricity load prediction is essential to ensure the efficient, reliable, and secure operation of the power system. In this study, a hybrid forecasting method called improved backtracking search optimization algorithm (IBSA)–double-reservoir echo state network (DRESN) (IBSA–DRESN) is proposed on the basis of IBSA and DRESN. Mutual information is utilized to eliminate low-significance input features and retain key input features. The DRESN structure aims to increase the diversity of the network. Roulette strategy, adaptive mutation operator, and niche operator is introduced to improve the standard BSA algorithm. The IBSA is applied to optimize several critical parameters in the DRESN neural network. The proposed IBSA–DRESN method is evaluated using two electricity load datasets, namely, North-America and PJM. Compared with eight popular benchmark models, prediction results show that IBSA–DRESN is more accurate for one-step ahead electricity load forecasting. In one-day ahead forecasting, IBSA–DRESN obtains better prediction performance in most cases.

1. Introduction

Electricity demand is gradually increasing as the main energy source of modern society. The shortage of electricity load will seriously affect the overall efficiency of the power supply chain and bring difficulties to the decision making of the power industry. Thus, accurate electricity load forecasting which is useful for optimal decision-making has an important role in ensuring the security, reliability, and economic operation of the power system. According to Maldonado et al. (2019) and Moon et al. (2020), the load forecasting can be classified into short-term (less than few days), medium-term (few weeks or months), and long-term (few years) load forecasting depend on the forecasting horizon. Medium-term and long-term load forecasting is usually used for integrated resource planning and the construction of electric equipment. According to Maldonado et al. (2019), in the next 24 to 36 h of short-term operational decision-making, a 1% reduction in forecast errors may result in a reduction in operating costs of $10 million. Thus, effective short-term electricity load projection has important application value for improving operational efficiency. Besides, electricity load forecasting becomes a challenging work due to various factors, such as weather conditions, time information, and big events that affect electricity load (Khwaja et al., 2020; Vaghefi et al., 2014). Thus, an advanced forecasting model with high precision is essential to propose in the application of short-term load prediction.

Echo state network (ESN) is a unique recurrent neural network (RNN) that utilizes reservoir structure to learn data characteristics and match complicated patterns. Unlike other artificial neural networks (ANNs), the least square method is employed for ESN training. Thus, ESN trains faster than traditional ANNs using gradient descent. ESN is widely applied in time series data forecasting because of its good learning capacity (Xu and Han, 2016; Hu et al., 2020a,b). The energy-related forecasting studies of ESN are few. Wang et al. (2018b) designed an ensemble sparse AdaBoost with ESN to forecast industrial electricity consumption. Chitsazan et al. (2019) proposed a nonlinear ESN to forecast wind speed and direction. Some techniques were designed to enhance the forecasting performance of ESN.

The nonlinear solving capability of ESN mainly comes from the reservoir. ESN is composed of a group of sparsely connected synapses that are randomly generated during the initialization process and will not be modified in the next training. The reservoir is randomly generated and affects the performance of ESN. The different reservoirs can have various sparse matrices and internal neurons. Thus, the double-reservoir ESN (DRESN) structure is developed in our study. The DRESN structure can increase the diversity of the network and alleviate the drawbacks of the randomness of the internal weight matrix. The reservoir parameters are vital for generating a suitable reservoir. The reservoir scale (N) is larger. Then, its network is more complex. Sparse

* Corresponding author.
E-mail addresses: Pengluhust@163.com (L. Peng), shengxianglv@foxmail.com (S.-X. Lv), wanglin@hust.edu.cn (L. Wang), wangziy94@126.com (Z.-Y. Wang).

https://doi.org/10.1016/j.engappai.2020.104132
Received 10 March 2020; Received in revised form 5 December 2020; Accepted 8 December 2020
Available online 16 December 2020

degree (SD) refers to the number of non-zero connection that reflects the connection between neurons in the reservoir. Input scaling (IS) influenced the reconstruction of the input vector. Before the input vector is input into the reservoir, it is scaled by the scale factor, that is input scaling. IS can adjust the nonlinear capacity of the reservoir. Spectral radius (SR) refers to the maximum eigenvalue of the interconnect weight matrix in the reservoir. The SR can be randomly set, and the interconnect weight matrix can be generated. All reservoir parameters affect the performance of DRESN. Thus, the DRESN parameters are essential to be improved by other procedures to increase forecasting performance (Hu et al., 2020a,b).

The task of this study aims to optimize important hyperparameter values of DRESN and test the performance of enhanced DRESN in solving electricity load prediction problems. According to Zhong et al. (2017) and Wang et al. (2018a), sparse degree, input scaling, and spectral radius are usually among the range of [0,1]. The reservoir scale is an integer in the range of [1, 100] in our study. According to Srivastava and Fahim (2007), the optimization of DRESN parameters can be classified as a mixed discrete-continuous problem. Evolutionary methods based on optimization algorithms are capable of solving such computational challenges of neural networks efficiently. Chouikhi et al. (2017) pre-trained some fixed weight values of ESN by particle swarm optimization algorithm. Zeng et al. (2017) proposed an improved differential evolution (DE) algorithm to optimize a back-propagation neural network (BPNN). Peng et al. (2018) verified that DE helped select the hyper-parameters of long short-term memory (LSTM) to predict electricity prices. The backtracking search optimization algorithm (BSA) is a novel evolutionary algorithm (Civicioglu, 2013). Compared with other algorithms, BSA is easy to implement, less computationally costly, more efficient, and requires fewer parameters. BSA has been applied in considerable fields, such as power flow (Chaib et al., 2016), automatic generation control (Madasu et al., 2017), and time series forecasting (Wang et al., 2019). Therefore, BSA represents an efficient and effective model to optimize DRESN parameters.

Many factors, such as day types, weather conditions, and historical load, affect the forecasting accuracy of the electricity load. Hence, the selection of the input variables is a critical process. Mutual information (MI) is an advantage in evaluating the dependence between variables with non-linear relationships. Therefore, MI is used to select suitable influencing factors for the neural network. In this study, DRESN is designed as the prediction model, and the important DRESN parameters are optimized by improved BSA (IBSA). As some operations are simply randomized in the standard BSA, the IBSA is proposed by introducing the adaptive mutation factor, adding the selection of round roulette, and designing the adaptive niching process to balance exploration and exploitation. Thus, IBSA can improve local development capabilities and maximize the population diversity of the algorithm. Effectiveness of IBSA–DRESN in electricity load forecasting problems is tested subsequently (Fig. 1).

The contributions are presented as follows:

(a) A hybrid forecasting model based on double-reservoir ESN structure and improved BSA is presented. The hybrid IBSA–DRESN structure can increase the diversity of the network and alleviate the drawbacks of the randomness of the internal weight matrix to obtain a better prediction performance.

(b) Roulette strategy, adaptive mutation operator, and niche operator are introduced to enhance the capability of standard BSA. The IBSA is applied to optimize several vital parameters in the DRESN structure.

(c) The prediction results of two electricity load datasets, namely, North-America and PJM, show that the proposed IBSA–DRESN outperforms other popular forecasting approaches.

The remainder of this work is structured as follows: The literature review is present in Section 2. Section 3 introduces the adopted methods. Section 4 presents the experiments. Finally, Section 5 presents the discussions and Section 6 concludes this work.

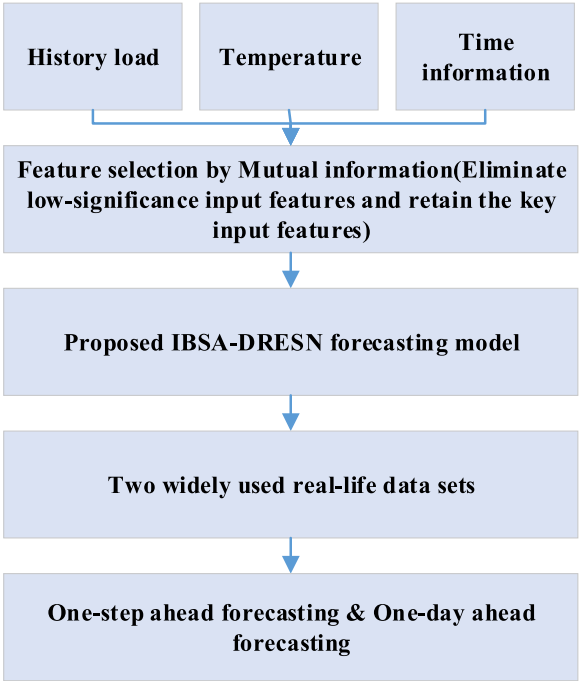


Fig. 1. Flowchart of this study.

2. Literature review on electricity load forecasting

Earlier researches showed that electricity-related prediction is challenging as the demand for electricity is greatly dependent on various factors, such as weather, time information, and daily activities. Moreover, forecasting the electricity load accurately is quite difficult because the electricity load dataset displays clearly nonlinear characteristics. Table 1 shows that the typical methods of electricity load forecasting are categorized into statistical, artificial intelligence-based, and hybrid models.

Multiple linear regression (MLR), autoregressive model (AR), and other statistical models have been used in the field of electricity load forecasting. Tascikaraoglu and Sanandaji (2016) compared the forecasting accuracy of AR and its varieties using a historical load. Lusis et al. (2017) compared MLR, regression trees, support vector regression (SVR), and neural networks using the data of weather, historical load, and calendar effects. However, the prediction accuracy is generally low and the parameter estimation is commonly difficult.

Widely used methods for artificial intelligence-based models include SVR and ANN. For instance, Shi et al. (2018) designed a pooling-based deep RNN for load forecasting based on historical load. Ribeiro et al. (2019) used a wavelet neural network (WNN) for load forecasting based on historical load. Kim et al. (2019) combined RNN and convolution neural network(CNN) for load forecasting based on historical load. Maldonado et al. (2019) proposed the kernel-based SVR model for load forecasting considering temperature, historical load, and calendar data. Elhendawi and Wang (2020) designed a full WNN model for load prediction considering temperature, humidity, historical load, and calendar data. Moon et al. (2020) used deep neural network (DNN) for short-term load forecasting based on weather factor, calendar information, and historical electric load. However, the prediction accuracy of SVR is sensitive to the selection of parameters and is difficult to implement for large-scale training samples. In addition, for ANN models, determining the network structure is usually difficult and it is usually time-consuming for the training.

Many researchers also attempted to use hybrid models to increase prediction accuracy. For example, Jiang et al. (2018) combined a grid

Table 1
Typical methods used for electricity load forecasting.

Category	Paper	Method	Inputs	Best accuracy
Statistical models	Tascikaraoglu and Sanandaji (2016)	AR	Historical load	Mean absolute error (MAE) = 1.2308, Root mean square error (RMSE) = 1.6243, Normalized root mean square error (NRMSE) = 0.1550.
	Lusis et al. (2017)	MLR	Weather, historical load, calendar effects	RMSE = 0.18, NRMSE = 0.607
Artificial intelligence-based models	Shi et al. (2018)	RNN	Historical load	MAE = 0.2913, RMSE = 0.5280, NRMSE = 0.1076.
	Kim et al. (2019)	RNN, CNN	Historical load	RNN:RMSE = 58.039 MAPE = 4.751%; CNN:RMSE = 78.016, MAPE = 5.071%.
	Ribeiro et al. (2019)	WNN	Historical load	The coefficient of determination R^2. The specific value is not provided.
	Maldonado et al. (2019)	SVR	Temperature, historical load, calendar data.	MAPE = 2.12%
	Elhendawi and Wang (2020)	WNN	Temperature, humidity, historical load, calendar data.	MAE = 232.8, MAPE = 1.52%
	Moon et al. (2020)	DNN	Weather, calendar, historical load.	MAPE = 5.64%, MAE = 2.35
Hybrid models	Jiang et al. (2018)	GTA, SVR	Historical load	MAPE = 2.53%
	Singh and Dwivedi (2018)	FLT, ANN	Temperature, historical load, calendar data.	MSE=6. 3603 $\times 10^5$, NMSE=0.2564, RMSE = 797.51, MAE = 618.61, MAPE = 7.35%, Maximum error percentage=20.207%, Daily peak MAPE = 6.05%.
	Zhang et al. (2018)	IEMD, ARIMA, WNN, FOA	Historical load	MAPE = 0.82%, MAE = 69.03, RMSE = 2.15.
	Singh et al. (2019)	ANN, Evolutionary algorithm	Temperature, historical load, calendar data.	MAE = 251.236, MAPE = 3.180%, Daily Peak MAPE = 4.121%

traverse algorithm (GTA) with SVR to forecast the short-term load based on historical load. Singh and Dwivedi (2018) combined the follow the leader algorithm (FLT) with ANN for load forecasting based on temperature, historical load, and calendar data. Zhang et al. (2018) proposed a hybrid model based on improved empirical mode decomposition (IEMD), ARIMA, and WNN optimized by fruit fly optimization algorithm (FOA) for load forecasting based on historical load. Singh et al. (2019) integrate ANN with an enhanced evolutionary algorithm for short-term load prediction based on temperature, historical load, and calendar data. However, the hybrid models are more complex and it is difficult to determine a suitable hybrid model for specific prediction problems.

In summary, for one-step ahead forecasting, the mean absolute percentage error (MAPE) of forecasting models is usually below 5%. However, it is usually more difficult for one-day ahead forecasting and the MAPE of the forecasting models is usually up to 10%. Besides, electricity load projection is challenging because electricity load is extensively affected by numerous factors, including temperature, time information, and historical loads. Scholars tend to use more data and more complex methods to improve prediction accuracy. All methods have inferiority and advantages and cannot always obtain the best result in every case. The proposed IBSA–DRESN model shows its remarkable learning and accurate forecasting capabilities in the applications of electricity load forecasting.

3. Proposed methods

In this section, the details of IBSA–DRESN are provided. Briefly, the IBSA is applied to optimize several vital parameters in the DRESN structure. The details of improved BSA are also introduced.

3.1. Mutual information

MI shows an advantage in evaluating the dependence between variables with non-linear relationships and has been widely applied for feature selection. Fundamentally, MI calculates the quantity of information contained in a set of variables (Qian and Shu, 2015; Bouzgou and Gueymard, 2019).

X and Y are set as two random variables. The joint probability density function of X and Y is set as $\mu_{X,Y}$. Their marginal density functions of $\mu_X(x)$ and $\mu_Y(y)$ are represented as:

$$\mu_X(x) = \int \mu_{X,Y}(x,y)dy; \mu_Y(y) = \int \mu_{X,Y}(x,y)dx. \quad (1)$$

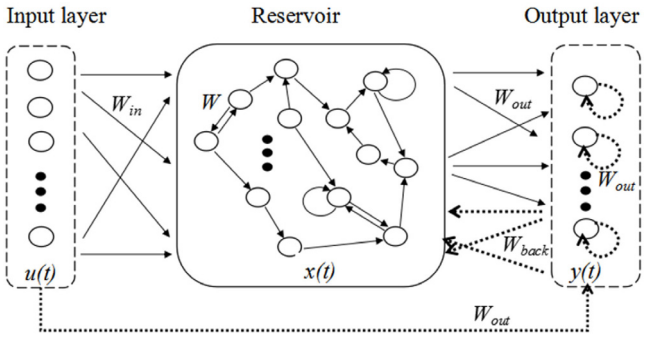


Fig. 2. Typical ESN structure.

Then, $H(y)$ is the edge of entropy and the uncertainty in Y can be obtained as:

$$H(y) = -\int \mu_Y(y) log \mu_Y(y) dy. \tag{2}$$

If Y is obtained by X indirectly, then $H(Y|X)$ is the conditional entropy and the uncertainty of Y can be represented as:

$$H(Y|X) = -\int \mu_X(x) \int \mu_Y(y|X=x) log \mu_Y(y|X=x) dy dx. \tag{3}$$

If considered jointly, $H(X,Y)$ is the joint entropy and the uncertainty of the pair (X, Y) can be represented as:

$$H(X,Y) = -\int \mu_{X,Y}(x,y) \log \mu_{X,Y}(x,y) dx dy. \tag{4}$$

Then, the MI between X and Y is calculated as:

$$M(X,Y) = H(Y) - H(Y|X). \tag{5}$$

If Y is the output in a prediction problem, the MI can suitably compute the importance of X to model Y, where $Y = X+H$, and H is the prediction horizon. The MI can be further represented as:

$$M(X,Y) = H(X) + H(Y) - H(X,Y). \tag{6}$$

According to Battiti (1994), Eq. (6) can be represented by:

$$M(X,Y) = \int \mu_{X,Y}(x,y) \log \frac{\mu_{X,Y}(x,y)}{\mu_X(x)\mu_Y(y)} dx dy. \tag{7}$$

Similarly, for discrete variables x and y, the MI is represented by:

$$M(x,y) = \sum_{i,j} \mu(x_i, y_j) \log \frac{\mu(x_i, y_j)}{\mu(x_i)\mu(y_j)}. \tag{8}$$

3.2. Echo state network

The standard ESN includes three parts: an input layer (K units), a reservoir (N units), and an output layer (L units) (Fig. 2). The activation of the input, the internal units, and the output at the time step t is represented by Eqs. (9), (10), and (11), respectively.

$$u(t) = \left[u_1(t), u_2(t), \ldots, u_K(t)\right]^T. \tag{9}$$

$$x(t) = \left[x_1(t), x_2(t), \ldots, x_N(t)\right]^T. \tag{10}$$

$$y(t) = \left[y_1(t), y_2(t), \ldots, y_L(t)\right]^T. \tag{11}$$

The network is driven by an input series $u(t)$ through random synaptic input connections composed in the input $N \times K$ weight matrix W_{in}. The $N \times N$ weight matrix W represents the connection weights between reservoir units. The $L \times (K + N + L)$ weight matrix W_{out} represents the connection weights between all units in the network and output units. The connection weights from the output to the reservoir are given in a $L \times N$ weight matrix W_{back}.

The activation of the internal units at time step $t+1$ is computed by Eq. (12):

$$x(t+1) = f(W_{in} \times u(t+1) + W \times x(t) + W_{back} \times y(t)), \tag{12}$$

where f is the *tanh* function.

The network output is computed by Eq. (13).

$$y(t+1) = f_{out}(W_{out} \times (u(t+1), x(t+1), y(t))), \tag{13}$$

where f_{out} is the readout function of the output units and is the hyperbolic tangent function, $(u(t+1), x(t+1), y(t))$ is the vertical connection of $u(t+1)$, $x(t+1)$ and $y(t)$.

According to Wang et al. (2019), for the training of ESN, the inputs are fed to the reservoir and mapped to a high-dimensional state space by an activation function. Suppose there is a series of Q desired inputs and corresponding outputs $\{(u(1), y(1)), (u(2), y(2)), \ldots, (u(Q), y(Q))\}$. The input sequence is assumed to be fed to the reservoir to gather a sequence of internal states $x(1), x(2), \ldots, x(Q)$. Then, the state matrix X and output vector Y are represented by Eqs. (14) and (15), respectively.

$$\mathrm{X} = \begin{bmatrix} u(1) & \cdots & u(Q) \\ x(1) & \cdots & x(Q) \\ (f^{out})^{-1}(y(1)) & \cdots & (f^{out})^{-1}(y(Q)) \end{bmatrix}^T. \tag{14}$$

$$\mathrm{Y} = \left[f_{out}^{-1}(y(1)) \ \ldots \ f_{out}^{-1}(y(Q))\right]^T. \tag{15}$$

Subsequently, W_{out} is calculated by the least square method.

$$W_{out} = \underset{W_{out}}{\operatorname{argmin}} \left\| W_{out} \times X - Y \right\|^2. \tag{16}$$

Furthermore, Eq. (16) is closed form as Eq. (17):

$$W_{out} = (X^{-1} \times Y)^T = [(X^T X)^{-1} X^T Y]^T. \tag{17}$$

3.3. Basic description of backtracking search optimization algorithm

BSA is a promising evolutionary algorithm (Tsai, 2019). These processes of BSA are presented as follows:

(a) Initialization

Obtain the initial populations using Eqs. (18) and (19).

$$P_{i,j} \sim U_j(low_j, up_j), \tag{18}$$

$$oldP_{i,j} \sim U_j(low_j, up_j), \tag{19}$$

for $i = 1, 2, \ldots, popSize$ and $j = 1, 2, \ldots, D$, where $popSize$ and D are the population size and the problem dimension, respectively; P_i and $oldP_i$ are the target individuals in populations P and $oldP$, respectively; and U_j is the uniform distribution in $[low_j, up_j]$ for the j-dimension.

(b) Selection-I

This operator controls the historical population $oldP$ to be applied for computing the search direction using Eq. (20).

$$if\ a < b\ then\ oldP = P | a, b \sim U(0,1). \tag{20}$$

Then, Eq. (21) is applied to randomly replace the order of the individuals in $oldP$.

$$oldP = permuting(oldP). \tag{21}$$

(c) Mutation

This operator generates the trial population $Mutant$ through Eq. (22).

$$Mutant = P + F \times (oldP - P), \tag{22}$$

where the mutation factor F determines the amplitude of the search-direction matrix $(oldP - P)$.

(d) Crossover

This operator obtains the trial population T through Eq. (23).

$$T_{i,j} = \begin{cases} P_{i,j}, & map_{i,j} = 1 \\ Mutant_{i,j}, & \text{otherwise} \end{cases}, \tag{23}$$

where map is computed by Eq. (24):

$$\begin{cases} map_{i,u_{(1:\lceil mixrate \cdot rand \cdot D \rceil)}} = 0, & a < b | u = permuting\left(\langle 1, 2, \ldots, D \rangle\right) \\ map_{i,randi(D)} = 0, & \text{otherwise} \end{cases}. \tag{24}$$

In Eq. (24), $\lceil\rceil$ is the ceiling function. $rand$, a, and b are randomly generated and $rand, a, b \sim U(0,1)$, where $U(0,1)$ is a standard uniform distribution. $randi(D)$ is used to obtain a pseudorandom integer from a uniform discrete distribution with an interval $(0, D)$. The mix rate parameter ($mixrate$) controls the number of elements of individuals using $\lceil mixrate \cdot rand \cdot D \rceil$. The boundary control mechanism is applied to obtain the trial population in the bound as below:

$$T_{i,j} = rand \times (up_j - low_j) + low_j. \tag{25}$$

(e) Selection-II.

The offspring population is obtained by Eq. (26). f is the fitness value.

$$P_i^{next} = \begin{cases} T_i & if\ f(T_i) \le f(P_i) \\ P_i & otherwise \end{cases}. \tag{26}$$

3.4. Proposed DRESN with improved backtracking search optimization

In this section, the rationality of DRESN optimized by improved BSA is presented. A detailed description of the proposed IBSA–DRESN method for electricity load forecasting is also provided.

3.4.1. Rationality of DRESN optimized by backtracking search optimization

The improved BSA algorithm shows a great advantage in solving numerical optimization problems (Wang et al., 2020). Its unique historical population search structure incorporates the historical information of the previous generation into the new population formation operation, thereby forming a guiding search direction. The mutation and crossover process of the algorithm produces an efficient experimental population. The generation strategy of the adaptive mutation factor F can generate the large and small amplitude values required for the global and local searches, respectively. The proposed IBSA improves the operations of selection-I, mutation, and selection-II, which enhance the local development capability and increase the diversity of the population.

ESN has good nonlinear processing capability and a successful application in time series problems (Bianchi et al., 2015). The two reservoirs included in DRESN are generated independently, and different reservoirs can be selected. The structure of the two reservoirs plays a crucial role in the performance of DRESN, and the reservoir structure is mainly affected by several important parameters, such as SR, SD, IS, and N. N is large, indicating that the network is more complex. SD is the number of non-zero connections that reflect the connection among neurons in the reservoir. IS influences the reconstruction of the input vector. SR is the maximum eigenvalue of the interconnected weight matrix in the reservoir, and it can be set randomly. Several other parameters can also be optimized, but only IS, SR, N, and SD are considered, and the reasons are as follow: (i) According to Zhong et al. (2017) and Wang et al. (2018a) those four hyperparameters have a larger influence on the forecasting capability of DRESN than others; (ii) if all hyperparameters are considered, the optimization results may be disappointing as the search space will be too large and the optimization will be a time-consuming operation. Therefore, setting reasonable reservoir parameters can ensure a DRESN model with improved performance. IBSA has advantages in processing optimization problems and can achieve a good balance between development and exploration capabilities. Thus, IBSA is utilized to form an initial population, narrow the search space, and then optimize the key parameters in the DRESN reservoirs.

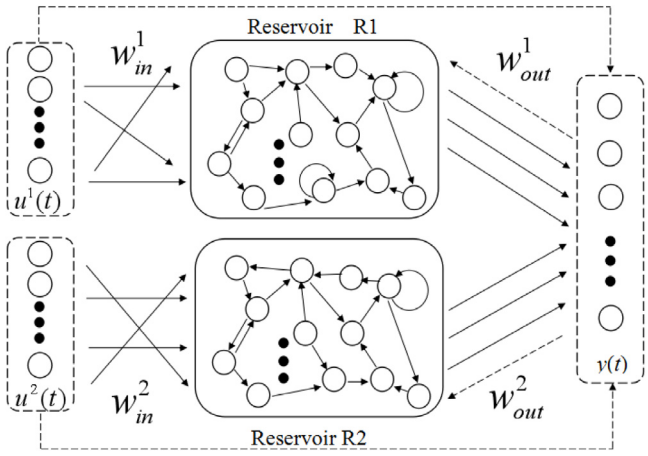


Fig. 3. DRESN structure.

3.4.2. Double-reservoir echo state network

The nonlinear solving capability of ESN is mainly derived from its reservoir network. However, the reservoirs even have the same SD, and the generated internal weight matrix is random. Adding a reservoir based on the standard ESN prediction model will result in a more complex nonlinear processing capability. The combined output of double reservoirs can also alleviate the drawbacks of the randomness of the internal weight matrix. Fig. 3 shows the DRESN structure.

Fig. 3 shows that DRESN has two sets of input vectors, two reservoirs, and one output vector. The two sets of input vectors are represented in Eq. (27). Their lengths can be the same or different. The two sets of input data in multifactor forecasting problems include two sets of different data, which influence the output. Different input vectors are fitted with two various reservoirs, and the internal states are presented in Eq. (28). N_1 and N_2 correspond to the number of internal neurons in reservoirs R_1 and R_2, respectively. The model has only one output vector (Eq. (29)), where L is the number of predicted nodes and $y(t)$ can be calculated by Eq. (30). In this study, the output vector has only one element and only one output neuron for one-step forecasting.

$$\begin{cases} u^1(t) = \left[u_1^1(t), u_2^1(t), \ldots, u_k^1(t)\right]^T \\ u^2(t) = \left[u_1^2(t), u_2^2(t), \ldots, u_k^2(t)\right]^T \end{cases}. \tag{27}$$

$$\begin{cases} x^1(t) = \left[x_1^1(t), x_2^1(t), \ldots, x_{N_1}^1(t)\right]^T \\ x^2(t) = \left[x_1^2(t), x_2^2(t), \ldots, x_{N_2}^2(t)\right]^T \end{cases}. \tag{28}$$

$$y(t) = \left[y_1(t), y_2(t), \ldots, y_L(t)\right]^T. \tag{29}$$

$$y(t) = f_{out}(W_{out}^1 \times (x^1(t), u^1(t)) + W_{out}^2 \times (x^2(t), u^2(t))). \tag{30}$$

3.4.3. Improved backtracking search optimization algorithm

In this study, the operations of selection-I, mutation, and selection-II are improved by adding the selection of roulette, introducing the adaptive mutation factor, and designing the adaptive niching process. Thus, IBSA can increase local development capabilities and enrich population diversity. The specific improvement ideas are presented as follows:

(a) Selection-I operation

The generation of the new population does not use the information of the optimal individual throughout the iteration. The roulette selection process for IBSA is added. The ratio of the average fitness value of the population P to the historical population $oldP$ is computed as the total fitness value. For the minimization problem, the average fitness value of the historical population is large, and the probability of the historical population being replaced is great. As shown in Eqs. (31) and (32), $f_{avg}(P)$ and $f_{avg}(oldP)$ represent the average value of the

fitness values of the population P and the historical population *oldP*, respectively; and *rand* is a random number in (0, 1).

$$P(oldP) = \frac{f_{avg}(oldP)}{f_{avg}(P) + f_{avg}(oldP)}. \tag{31}$$

$$oldP = \begin{cases} P & if\ (rand < P(oldP)) \\ oldP & otherwise \end{cases}. \tag{32}$$

(b) Mutation operation

The mutation factor F controls the search magnitude and affects the convergence speed. If F is very large, the efficiency of the algorithm is reduced. If F is very small, then the premature problem of the algorithm will be aggravated. Thus, the diversity of the population cannot be guaranteed. As shown in Eq. (33), the adaptive F has a large value to promote the diversity of the population at the beginning, and F gradually decreases to ensure convergence. *GenM* is the maximum number of iterations. G is the current number of iterations. r is a random value between (0,0.5). F_{min} and F_{max} are the minimum and maximum mutation factors, respectively.

$$F = inx * ((F_{max} - F_{min}) * e^{1-\frac{GenM}{GenM-G+1}} + F_{min} + r), \tag{33}$$

where $inx = \frac{\sum_i^D (oldP(randi1(NP)) - oldP(randi2(NP)))}{|\sum_i^D (oldP(randi1(NP)) - oldP(randi2(NP)))|}$. *randi*1(*NP*) and *randi*2(*NP*) are different random integers smaller than the population size (*NP*) . D is the individual dimension. *inx* is applied to randomly choose two different individuals in the historical population to subtract and take the sign of the sum of each dimension as the sign of the mutation factor.

(c) Selection-II

The greedy algorithm is used to eliminate individuals with poor fitness value in standard BSA. The one-to-one competition elimination method reduces the diversity of the population. The niching operator for IBSA is designed to select the population before competition selection to improve the diversity of the population. The niching operator has been applied to various evolutionary algorithms to improve its performance (Shir et al., 2010). Such an operation can eliminate similar individuals and supply new ones.

The minimum distance between individuals is calculated by using Eq. (34). The niche radius (*LN*) is calculated by Eq. (35). As iteration increases, the individuals become increasingly similar to each other, and the screening should adapt according to the change. Thus, the adaptive niching factor (*fl*) is designed to make niching screening increasingly strict as the iteration increases. Large *fl* correspond to strict screening. fl_{min} and fl_{max} are the minimum and maximum niching factors, respectively.

$$T_i = \min\left\{S_{i,1}, S_{i,2}, \ldots, S_{i,N-1}\right\}, S_{i,j} = \sqrt{\sum_{K=1}^{D} (P_{i,k} - P_{j,k})^2}. \tag{34}$$

$$LN = \frac{\sum_{i=1}^{N-1} T_i}{fl * (NP - 1)}, fl = fl_{max} - fl_{min} * e^{1-\frac{GenM}{GenM-G+1}}. \tag{35}$$

The distance between the individuals of population P and the individuals of the test population T is calculated by Eq. (36). $P_{i,k}$ and $T_{i,k}$ represent the kth dimension of the ith individual in populations P and T, respectively.

$$R_{i,j} = \sqrt{\sum_{K=1}^{D} (P_{i,k} - T_{i,k})^2}. \tag{36}$$

If the distance between the two individuals is lower than the niche radius (*LN*), then the individuals with larger adaptation values are updated by using Eq. (37).

$$\begin{cases} P_i = rand * (up - low),\ if\ f(P_i) > f(T_i) \\ T_i = rand * (up - low),\ if\ f(P_i) \le f(T_i) \end{cases}, \quad for\ R_{i,j} < LN, \tag{37}$$

where f is the fitness value, *up* and *low* are D-dimensional vectors that represent the maximum and minimum values of the individual population, respectively.

Table 2
The benchmark functions.

No.	Functions	x_i	x^*	$f(x^*)$
F1	$f(x) = \sum_{i=1}^{n} (10^6)^{\frac{i-1}{n-1}} x_i^2$	(−100,100)	0	0
F2	$f(x) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$	(−10,10)	0	0
F3	$f(x) = -20\exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}) - \exp(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)) + 20 + e$	(−32,32)	0	0

Table 3
The results of benchmark test.

Criterion	DE	GA	BSA	IBSA
Mean	1.47×10^{1}	5.33×10^{1}	9.54×10^{1}	9.06×10^{0}
STDE	3.55×10^{1}	8.71×10^{2}	8.23×10^{3}	2.66×10^{1}
Mean	1.61×10^{0}	5.75×10^{-1}	1.42×10^{-2}	6.19×10^{-3}
STDE	1.66×10^{-1}	8.31×10^{-3}	8.09×10^{-6}	9.55×10^{-6}
Mean	2.92×10^{-1}	1.11×10^{-1}	4.83×10^{-2}	3.30×10^{-2}
STDE	8.73×10^{-3}	2.37×10^{-4}	1.04×10^{-3}	1.14×10^{-3}

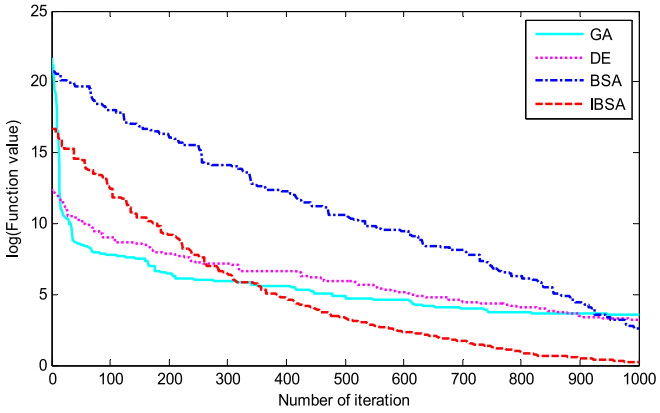


Fig. 4. Convergence results for F1.

Three benchmark functions (Table 2) are used to test the performance of IBSA (Lv et al., 2018). The parameters of DE, genetic algorithm (GA), and BSA are set the same as in Wang et al. (2020). The number of the dimension is set to 30, the size of population is set to 20, and the maximum iteration is set to 1000. For IBSA, the values of *Fmin*, *Fmax*, fl_{min}, fl_{max}, and *Mixrate* are set to 0.1, 0.9, 1, 3, and 1. Table 3 summarizes the mean and standard deviations (STDE) after running 30 times. Convergence results for three functions are shown in Figs. 4–6. Since the range of function value for function F1 is too large, loge(Function value) is used as the ordinate for function F1 . Table 3 and Figs. 4–6 illustrate that the IBSA has better optimization capability than DE, GA, and BSA.

3.4.4. IBSA– DRESN model

IBSA–DRESN is a two-stage forecasting model that combines IBSA and DRESN. The first phase uses the IBSA algorithm to optimize the reservoir parameters in DRESN, including SR, SD, IS, and N. The decision variable of IBSA consists of these parameters, and the fitness function is the mean absolute percentage error (MAPE) obtained by DRESN. The second phase is forecasting. The corresponding decision variables of an optimal individual obtained by IBSA are used as the initial parameters of DRESN. The training set is used to obtain the output weight matrix of DRESN, and finally, the test set data is predicted.

(a) Coding of individuals

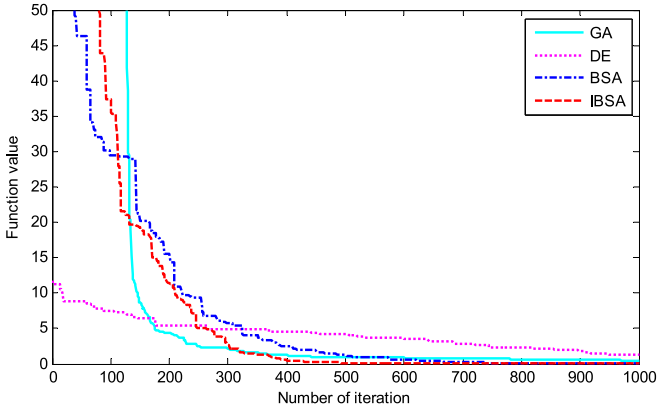


Fig. 5. Convergence results for F2.

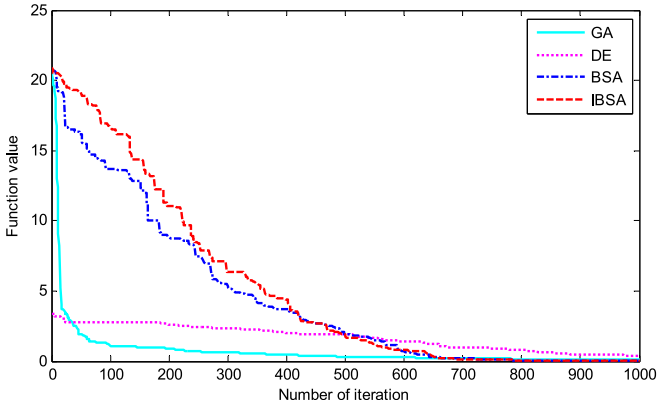


Fig. 6. Convergence results for F3.

In the optimization phase of IBSA–DRESN, the IBSA individuals have eight decision variables. For every individual, a vector of (6+ p_1+ p_2) dimensions exists, where p_1 and p_2 are the number of input variables of reservoirs R_1 and R_2, respectively. The individual represents the eight parameters of the two reservoirs that need to be optimized by IBSA. Fig. 7 shows the individual's structure. N_1 and N_2 represent the number of neurons in reservoirs R_1 and R_2, respectively. SR_1 and SD_1 are the SR and SD of reservoir R_1. SR_2 and SD_2 are the SR and SD of reservoir R_2. $IS_1^1, IS_1^2, \ldots, IS_1^{p_1}$ are the IS of reservoir R_1; $IS_2^1, IS_2^2, \ldots, IS_2^{p_2}$ are the IS of reservoir R_2.

(b) Data preprocessing

To prevent the occurrence of large value dominance of input data, the original data must be preprocessed to ensure that the model has good performance. If there are zero load values in the dataset, the average value of several numbers before and after the specified load values will be used. The data preprocessing is shown in Eq. (38). x_{max} and x_{min} are the maximum and minimum values, respectively.

$$x' = \frac{\log(x) - \log(x_{min})}{\log(x_{max}) - \log(x_{min})}. \tag{38}$$

(c) Niching adjusting operator

Calculating the niche radius using the total distance of the individuals will make two individuals who are not similar to also be selected for updating with a high probability because the range of different decision variables varies. Therefore, the niching operator is adjusted, and two conditions must be satisfied for the individual to perform the niching adjusting operator. First, following the niching operation method represented in Section 3.4.3, it is updated such that the dimension of individuals must be updated by the niching operator. Second, the calculated number (the distance between the first six dimensions of two individuals is less than the niche radius of the first six dimensions) is greater than the scale (*fl*). The niche radius of the first six dimensions of the individual is calculated independently. The larger the scale is, the harsher the condition is. The calculation of the niche radius of the first six dimensions is shown in Eq. (39).

$$\begin{aligned} &S_{i,j,g} = (P_{i,g} - P_{j,g})^2. \\ &L_g = \frac{\sum_{i=1}^{N-1} T_{i,g}}{NP-1}, T_{i,g} = \min\left\{S_{s,1,g}, S_{i,2,g}, \ldots, S_{i,NP-1,g}\right\}, \end{aligned} \tag{39}$$

where $g = 1, 2, \ldots, 6$; $i = 1, 2, \ldots, NP$; $j = 1, 2, \ldots, i-1, i+1, \ldots, NP$, and NP is the population size.

Fig. 8 shows the IBSA–DRESN model structure. The process is presented by eight steps:

Step 1: Data division. The data are divided into an optimization set, validation set, training set, and test set. The optimization and validation sets are combined to obtain excellent parameter sets. The training set is used to obtain the output weight matrix of the DRESN model, and the prediction result of the test set is the final predicted value.

Step 2: Initialize. For IBSA, the maximum iteration number (*maxGen*), population size (*NP*), and the bound of individuals are set. Set G=0. Generate an initial population.

Step 3: Calculate the fitness value of the initial population. The MAPE of the validation set is computed as the fitness value. The initial optimal value is recorded.

Step 4: Selection-I, mutation, crossover, and niching operations are used to obtain new populations. The fitness value of new populations is calculated, and the new optimal fitness value is updated.

Step 5: Selection-II operation. The optimal individual and optimal fitness values are obtained. Set G=G+1.

Step 6: If G>*maxGen*, the optimal individual obtained by Step 5 is decoded as the corresponding DRESN parameter. Otherwise, repeat Step 4–5.

Step 7: The network structure of DRESN is trained using the training set.

Step 8: The test set is used to obtain the final predicted result by using the trained DRESN network.

3.5. Forecasting strategy

For one-step ahead forecasting, $\{y_1, y_2, \ldots, y_t\}$ is the input and $\hat{y}_{t+1}$ is the output. Then , $\hat{y}_{t+1} = \hat{f}(y_1, y_2, \ldots, y_t)$. $\hat{f}$ is the forecasting model.

For multi-step ahead forecasting, $\{y_1, y_2, \ldots, y_t\}$ is the input and $\left\{\hat{y}_{t+1}, \hat{y}_{t+2}, \ldots, \hat{y}_{t+h}\right\}$ is the output. h is the length of output. In this study, the recursive forecasting strategy is used as shown in Eq. (40).

$$\begin{cases} \hat{y}_{t+h} = \hat{f}(y_h, \ldots, y_t, \hat{y}_{t+1}, \ldots, \hat{y}_{t+h-1}), 1 < h \le t \\ \hat{y}_{t+h} = \hat{f}(\hat{y}_h, \ldots, \hat{y}_{t+h-2}, \hat{y}_{t+h-1}), h > t \end{cases}. \tag{40}$$

For example, if $t = 3$, $h = 2$, then $\begin{cases} \hat{y}_4 = \hat{f}(y_1, y_2, y_3) \\ \hat{y}_5 = \hat{f}(y_2, y_3, \hat{y}_4) \end{cases}$. if $t = 3$, $h = 4$, then

$$\begin{cases} \hat{y}_4 = \hat{f}(y_1, y_2, y_3) \\ \hat{y}_5 = \hat{f}(y_2, y_3, \hat{y}_4) \\ \hat{y}_6 = \hat{f}(y_3, \hat{y}_4, \hat{y}_5) \\ \hat{y}_7 = \hat{f}(\hat{y}_4, \hat{y}_5, \hat{y}_6) \end{cases}.$$

As one day has 24 h, for the one-day ahead forecasting, h is equal to 24 in this study.

4. Experiments and analysis

In this section, IBSA–DRESN is tested by two real-life data sets. Data description is presented in Section 4.1. Benchmark methods and forecasting accuracy measurements are given in Sections 4.2 and 4.3. The results and analysis of two real-life data sets are given in Sections 4.4 and 4.5.

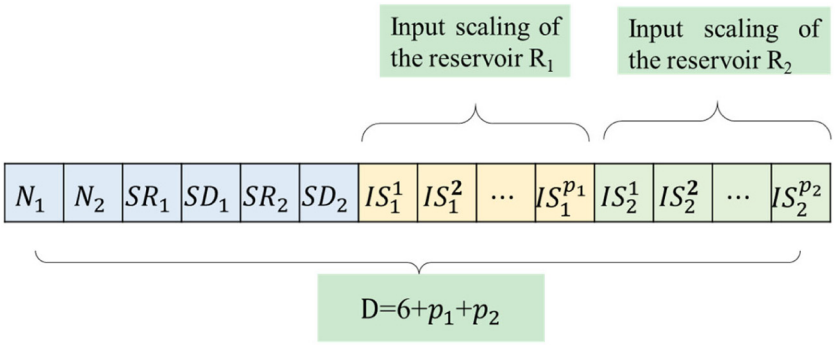


Fig. 7. Structure of the individuals.

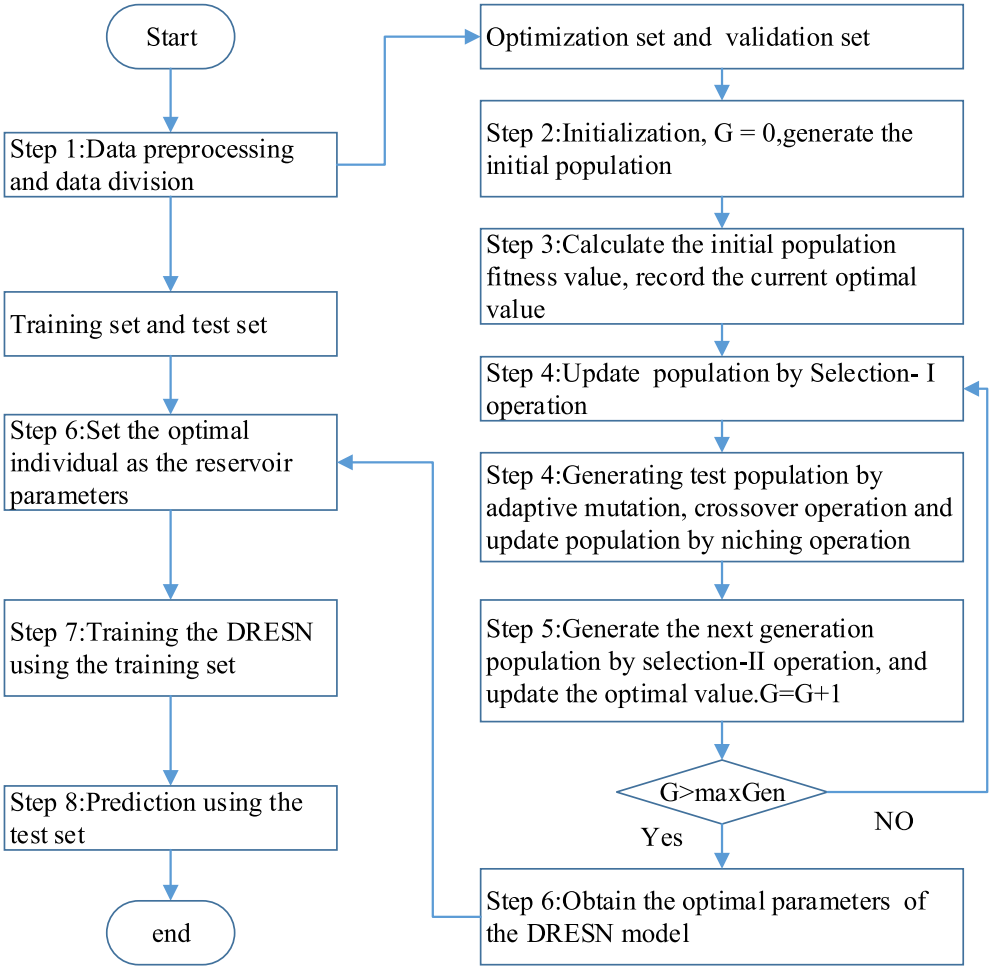


Fig. 8. Structure of IBSA–DRESN model.

4.1. Data description

The IBSA–DRESN model is tested by using two electricity load datasets, namely, North-America and PJM, for one-step ahead forecasting and one-day ahead forecasting. As the data is the hourly load, the one-step-ahead forecasting is the one-hour ahead forecasting. The recursive forecasting is adopted for one-day-ahead forecasting.

The hourly load, time information, and temperature data from November 1, 1989, to June 30, 1990 and from November 1, 1990, to June 30, 1991, in a North American electric utility are used in experiment 1. The data is also used in Hu et al. (2015). The sampling period is one hour. Raw datasets were processed using Eq. (38) at the beginning of the experiment. The dataset was set as optimization, validation, training, and test sets (Table 4). The statistical value of load and temperature is shown in Table 5.

The hourly load and time information of PJM from January 1, 2011 to December 31, 2011 in ROT are used in experiment 2 (data source: PJM Interconnection, 2020). The dataset was divided into optimization, validation, training, and test sets as shown in Table 6 . The data division of February for experiment 2 is shown in Fig. 9. The statistical value of load is shown in Table 7.

4.2. Benchmark methods

Eight popular methods, including SVR, ESN, BPNN, Random Forest (RF), LSTM, GA–DRESN, DE–DRESN, and BSA–DRESN, are selected

Table 4
Data division of experiment 1.

Month	Stage	Dataset division	Start date	End date	Number of samples
January	Optimization	Optimization set	1989.11.1	1989.12.31	1464
		Validation set	1990.1.1	1990.1.31	744
	Prediction	Training set	1990.11.1	1990.12.31	1464
		Test set	**1991.1.1**	**1991.1.31**	**744**
February	Optimization	Optimization set	1989.12.1	1990.1.31	1488
		Validation set	1990.2.1	1990.2.28	672
	Prediction	Training set	1990.12.1	1991.1.31	1488
		Test set	**1991.2.1**	**1991.2.28**	**672**
March	Optimization	Optimization set	1990.1.1	1990.2.28	1416
		Validation set	1990.3.1	1990.3.31	744
	Prediction	Training set	1991.1.1	1991.2.28	1416
		Test set	**1991.3.1**	**1991.3.31**	**744**
April	Optimization	Optimization set	1990.2.1	1990.3.31	1416
		Validation set	1990.4.1	1990.4.30	720
	Prediction	Training set	1991.2.1	1991.3.31	1416
		Test set	**1991.4.1**	**1991.4.30**	**720**
May	Optimization	Optimization set	1990.3.1	1990.4.30	1464
		Validation set	1990.5.1	1990.5.31	744
	Prediction	Training set	1991.3.1	1991.4.30	1464
		Test set	**1991.5.1**	**1991.5.31**	**744**
June	Optimization	Optimization set	1990.4.1	1990.5.31	1464
		Validation set	1990.6.1	1990.6.30	720
	Prediction	Training set	1991.4.1	1991.5.31	1464
		Test set	**1991.6.1**	**1991.6.30**	**720**

Table 5
Statistical value of Load and Temperature used in experiment 1.

	Minimum	Maximum	Average	Standard deviation	Coefficient of variation	Skewness	Kurtosis
Load	1228	4635	2350.72	558.99	0.24	0.34	−0.20
Temperature	12	84	47.15	9.10	0.19	−0.04	1.14

Table 6
Data division of experiment 2.

Month	Stage	Dataset division	Start date	End date	Number of samples
February	Optimization	Optimization set	2011.1.1	2011.1.21	504
		Validation set	2011.1.22	2011.1.31	240
	Prediction	Training set	2011.2.1	2011.2.21	504
		Test set	**2011.2.22**	**2011.2.28**	**168**
April	Optimization	Optimization set	2011.3.1	2011.3.21	504
		Validation set	2011.3.22	2011.3.31	240
	Prediction	Training set	2011.4.1	2011.4.21	504
		Test set	**2011.4.22**	**2011.4.30**	**216**
June	Optimization	Optimization set	2011.5.1	2011.5.21	504
		Validation set	2011.5.22	2011.5.31	240
	Prediction	Training set	2011.6.1	2011.6.21	504
		Test set	**2011.6.22**	**2011.6.30**	**216**
August	Optimization	Optimization set	2011.7.1	2011.7.21	504
		Validation set	2011.7.22	2011.7.31	240
	Prediction	Training set	2011.8.1	2011.8.21	504
		Test set	**2011.8.22**	**2011.8.31**	**240**
October	Optimization	Optimization set	2011.9.1	2011.9.21	504
		Validation set	2011.9.22	2011.9.30	216
	Prediction	Training set	2011.10.1	2011.10.21	504
		Test set	**2011.10.22**	**2011.10.31**	**240**
December	Optimization	Optimization set	2011.11.1	2011.11.21	504
		Validation set	2011.11.22	2011.11.30	216
	Prediction	Training set	2011.12.1	2011.12.21	504
		Test set	**2011.12.22**	**2011.12.31**	**240**

as the benchmark methods. SVR is a popular model used to solve nonlinear problems (Che and Wang, 2014). ESN and BPNN are popular neural networks (Lv et al., 2018; Hu et al., 2019). GA, DE, and BSA are advanced intelligent optimization algorithms (Cheng et al., 2018;

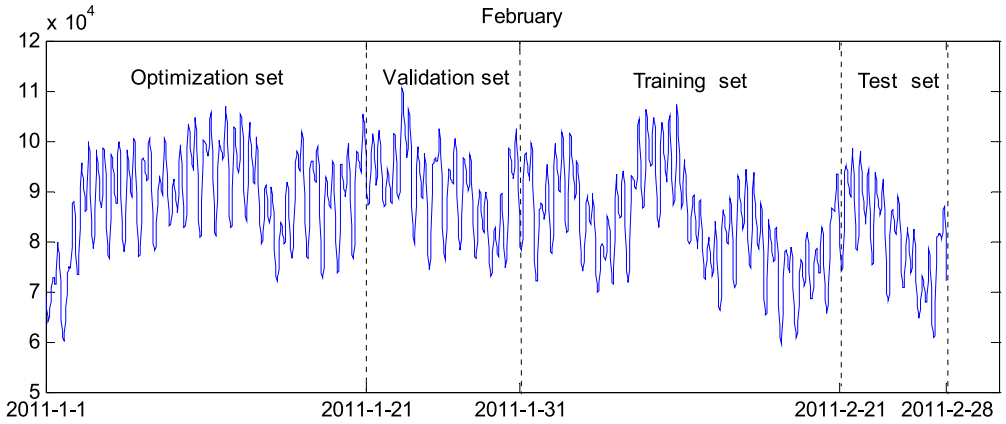


Fig. 9. Data division of February for experiment 2.

Table 7
Statistical value of Load used in experiment 2.

	Minimum	Maximum	Average	Standard deviation	Coefficient of variation	Skewness	Kurtosis
Load	50644.68	158042.51	84385.03	16574.17	0.20	0.91	1.21

Table 8
Parameter setting of different methods.

Methods	Parameters	Value
SVR (grid-search)	C	[−10.10]
	Gamma	[−10,10]
ESN	N	100
	SD	0.05
	SR	0.8
	IS	1
BPNN	Hidden neurons	30
	Epochs	2000
	Learning rate	0.005
	Tolerance	0.0005
RF	Depth of the tree	5
	Number of trees	80
LSTM	Batch size	100
	Hidden neurons	30
	Epochs	1000
GA–DRESN	Crossover factor	0.8
	Mutation factor	0.3
DE–DRESN	Crossover factor	0.5
	Mutation factor	0.2
BSA–DRESN	*Mixrate*	1.5
	Mutation factor	3·randn
IBSA–DRESN	*Mixrate*	1.5
	F_{min}	0.1
	F_{max}	3
	fl_{min}	1
	fl_{max}	3

Hu and Chen, 2018). To further compare the effect of IBSA, GA–DRESN, DE–DRESN, and BSA–DRESN are added as benchmark models. Also, Random Forest (RF) and LSTM (Stéfano Frizzo Stefenon et al., 2020) are used as the benchmark models. According to Ribeiro and Coelho (2020) and Tian et al. (2020), cross-validation can be used to increase the reliability of the forecasting model. The effect of the prediction model is evaluated by 5 fold cross-validation, which is often used by researchers(lv et al. 2018). The training data is randomly partitioned into 5 equally sized subsamples. Four subsamples are the training data, and the remaining subsample is used for validation. After repeating 5 times, the minimum root mean square error (RMSE) corresponding model is the final forecast model. All algorithms are implemented in MATLAB (2019a) with an Intel(R) Core(TM) i5-4210U CPU @4.00 GHz, 8 GB RAM, and Windows 8.1 operating system.

Table 9
Selected inputs between the candidate input factors and the output of experiment 1.

Input	Rank	Standard mutual information
L($t-24$)	1	0.892
L($t-48$)	2	0.808
H	3	0.758
L($t-72$)	4	0.752
L($t-168$)	5	0.704
T($t-12$)	6	0.688
T($t-11$)	7	0.682
T($t-1$)	8	0.677
T($t-10$)	9	0.676
T($t-13$)	10	0.671
T(t)	11	0.671
T($t-9$)	12	0.662
T($t-2$)	13	0.647
T($t-8$)	14	0.639
T($t-14$)	15	0.635
T($t-7$)	16	0.621
T($t-3$)	17	0.610
T($t-6$)	18	0.601
T($t-24$)	19	0.597
Mean		0.595

Based on several trials and similar works (Wang et al., 2019), the parameters of all methods for two experiments are presented in Table 8. For GA–DRESN, DE–DRESN, BSA–DRESN, and IBSA–DRESN, the lower and upper bounds of N are set to 1 and 100, respectively. The lower bounds of SR, SD, and IS are set as 0.01, 0.006, and 0.0001, respectively. The upper bounds of SR, SD, and IS are all set as 1. The maximum number of iterations is 50, and the population size is 20.

4.3. Forecasting accuracy measurements

Similar to Jiang et al. (2017) & Wang et al. (2018a,b), mean absolute error (MAE), MAPE, and RMSE are used as error criteria in the electricity load forecasting and are expressed as Eq. (41).

$$\begin{cases} \mathrm{MAE} = \dfrac{1}{N_t}\sum_{n=1}^{N_t}\left|\widetilde{y}_n - y_n\right|, \\ \mathrm{MAPE} = \dfrac{1}{N_t}\sum_{n=1}^{N_t}\dfrac{\left|\widetilde{y}_n - y_n\right|}{y_n}, \\ \mathrm{RMSE} = \sqrt{\dfrac{1}{N_t}\sum_{n=1}^{N_t}\left(\widetilde{y}_n - y_n\right)^2}, \end{cases} \tag{41}$$

Table 10
Final selected inputs of experiment 1.

Reservoir	One-step ahead	One-day ahead
R1	L($t-1$), L($t-2$),…, L($t-12$)	L($t-1$), L($t-2$),…, L($t-24$)
R2	T($t-1$), T($t-12$), L($t-24$), L($t-168$), H	T($t-1$), T($t-12$), L($t-24$), L($t-168$), H

Table 11(a)
The selected hyper-parameters of one-step ahead for experiment 1.

Month	N1	N2	SR1	SD1	SR2	SD2
January	99	80	0.67	0.45	0.69	0.15
February	95	97	0.93	0.68	0.93	0.62
March	99	52	0.81	0.37	0.87	0.96
April	37	100	0.9	0.28	0.85	0.91
May	100	87	0.15	0.39	0.93	0.29
June	89	80	0.64	0.16	0.77	0.59

where N_t is the number of test set samples, y_n is the actual value of nth prediction point, and $\widetilde{y}_n$ is the corresponding forecasting value.

4.4. Experiment 1: Electricity load in North-America

In this section, for the electricity load in North-America, the result of feature selection using MI is present in Section 4.4.1. The forecasting result and statistical testing are given in Sections 4.4.2 and 4.4.3.

4.4.1. Feature selection using MI

To enhance the generalization capability of the prediction model, the low-significance input features must be eliminated, and the key input features must be retained. Some researchers have proven the effectiveness of MI in feature selection (Rahmaninia and Moradi, 2018). Similar to Hu et al. (2015), the hourly load values of the previous 1–12, 24, 48, 72, and 168 h are the input factors. For the temperature variables, the temperature at the time of the predicted point, the temperature of the previous 1–24 h, and the temperature of the previous 48, 72, and 168 h are used as input factors. Time information is also used as a candidate input factor.

The candidate input factors shown in Eq. (42) has 46 candidate input factors. L(t) and T(t) are the electric load and temperature at time t, respectively. $H(t)$ is the number of hours at time t, and $H(t) \in \{1, 2, \ldots, 24\}$. $D(t) \in \{0.75, 0.25\}$ as the working day is set as 0.75, and the weekend is set as 0.25.

$$\begin{aligned} Input\ factors(t) = \{ & L(t-1), L(t-2), \ldots, L(t-12), L(t-24) \\ & L(t-48), L(t-72), L(t-168), T(t), \\ & T(t-1), T(t-2), \ldots, T(t-24), T(t-48), \\ & T(t-72), T(t-168), H(t), D(t)\}. \end{aligned} \tag{42}$$

MI is used to further select candidate input sets. The specific strategies are presented as follows:

For the reservoir R1 of IBSA–DRESN, the load of the previous 1–12 h ($\{L(t-1), L(t-2), \ldots, L(t-12)\}$) are directly used as the input data for one-step ahead forecasting. For one-day ahead forecasting, the load of 1–24 h ($\{L(t-1), L(t-2), \ldots, L(t-24)\}$) in the previous day are directly used as the input data as the load in the prediction day is unknown.

For the reservoir R2 of IBSA–DRESN, the load of the previous 24, 48, 72, and 168 h, temperature, and time information are used as input factors. The MI between these candidate input factors and the output is calculated, and the variables with strong MI correlation are retained. In this study, the average standard MI of all features is selected as the selection criterion, and the feature that the MI is larger than the average standard MI is retained. The reserved input factors are shown in Table 9, and 19 features are reserved as alternative input factors.

The MI among the 19 retained input factors is further calculated to eliminate the variables that contain duplicate information and to reduce computational cost. The selection criteria is 0.8. If the standard MI between the two features is greater than the selection criteria, then the two features are considered to have more repetition values, and

Table 11(b)
The selected hyper-parameters of one-step ahead for experiment 1.

Month	January	February	March	April	May	June
IS2	0.55	0.40	0.65	0.83	0.07	0.01
	0.08	0.10	0.40	0.10	0.22	0.01
	0.42	0.68	0.77	0.93	0.64	0.50
	0.97	0.93	0.83	0.95	0.84	0.73
	0.22	0.03	0.07	0.04	0.04	0.46
IS1	0.03	0.68	0.22	0.06	0.63	0.17
	0.97	0.25	0.41	0.64	0.22	0.52
	0.21	0.59	0.10	0.49	0.49	0.47
	0.01	0.46	0.05	0.19	0.99	0.42
	0.28	0.09	0.52	0.72	0.77	0.92
	0.51	0.53	0.69	0.29	0.66	0.54
	0.51	0.21	0.18	0.14	0.72	0.63
	0.36	0.36	0.50	0.41	0.68	0.23
	0.68	0.38	0.94	0.42	0.80	0.20
	0.24	0.49	0.32	0.31	0.14	0.52
	0.07	0.60	0.28	0.24	0.45	0.83
	0.91	0.94	0.98	0.84	0.80	0.99

Table 12(a)
The selected hyper-parameters of one-day ahead for experiment 1.

Month	N1	N2	SR1	SD1	SR2	SD2
January	35	47	0.16	0.54	0.13	0.08
February	91	71	0.17	0.85	0.64	0.64
March	87	36	0.03	0.40	0.25	0.06
April	13	77	0.65	0.03	0.72	0.17
May	53	75	0.08	0.28	0.05	0.71
June	5	52	0.33	0.75	0.06	0.78

Table 12(b)
The selected hyper-parameters of one-day ahead for experiment 1.

Month	January	February	March	April	May	June
IS2	0.23	0.83	0.88	0.71	0.31	0.02
	0.8	0.92	0.19	0.99	0.66	0.02
	0.11	0.55	0.25	0.38	0.39	0.41
	0.64	0.04	0.3	0.13	0.34	0.48
	0.73	0.98	0.84	0.77	0.59	0.13
IS1	0.12	0.44	0.62	0.44	0.75	0.93
	0.84	0.9	0.71	0.91	0.65	0.33
	0.7	0.52	0.49	0.51	0.31	0.13
	0.63	0.13	0.99	0.94	0.87	0.76
	0.38	0.16	0.34	0.56	0.66	0.55
	0.66	0.08	0.75	0.4	0.41	0.59
	0.24	0.85	0.55	0.22	0.57	0.06
	0.89	0.03	0.22	0.34	0.37	0.49
	0.36	0.98	0.55	0.75	0.41	0.05
	0.4	0.38	0.55	0.11	0.67	0.88
	0.23	0.08	0.4	0.19	0.51	0.32
	0.76	0.02	0.46	0.84	0.77	0.09
	0.59	0.36	0.98	0.52	0.6	0.49
	0.03	0.96	0.79	0.63	0.43	0.01
	0.53	0.9	0.64	0.26	0.22	0.07
	0.22	0.81	0.68	0.69	0.4	0.56
	0.91	0.81	0.52	0.83	0.86	0.08
	0.15	0.69	0.73	0.17	0.66	0.62
	0.92	0.49	0.39	0.06	0.47	0.69
	0.18	0.56	0.66	0.62	0.71	0.59
	0.8	0.62	0.77	0.23	0.6	0.84
	0.38	0.01	0.2	0.37	0.26	0.45
	0.3	0.15	0.03	0.04	0.09	0.08
	0.18	0.08	0.05	0.31	0.11	0.01

Table 13
Results of one-step ahead for experiment 1.

Month		SVR	ESN	BPNN	RF	LSTM	GA–DRESN	DE–DRESN	BSA–DRESN	IBSA–DRESN
January	RMSE	60.67	76.10	66.37	127.99	65.48	71.04	64.93	66.67	**57.95**
	MAE	44.22	57.73	52.90	95.34	47.52	52.12	47.26	48.21	**41.89**
	MAPE	1.55%	2.02%	1.89%	3.44%	1.62%	1.81%	1.62%	1.65%	**1.43%**
February	RMSE	61.63	63.92	60.16	106.14	56.65	62.35	55.73	55.46	**51.70**
	MAE	49.17	49.77	46.92	84.06	42.10	48.09	40.97	41.40	**37.72**
	MAPE	2.01%	2.03%	1.90%	3.52%	1.72%	1.92%	1.63%	1.65%	**1.50%**
March	RMSE	47.85	58.94	57.40	113.66	66.51	54.65	49.31	50.34	**47.71**
	MAE	36.20	45.46	44.07	82.84	51.03	42.58	37.33	38.55	**36.19**
	MAPE	1.45%	1.82%	1.77%	3.37%	2.07%	1.69%	1.48%	1.53%	**1.44%**
April	RMSE	70.81	67.16	54.71	97.13	57.91	55.94	51.32	54.19	**50.23**
	MAE	51.56	52.56	40.61	72.43	43.31	42.29	38.63	40.21	**38.03**
	MAPE	2.37%	2.38%	1.82%	3.41%	1.91%	1.86%	1.70%	1.77%	**1.69%**
May	RMSE	45.41	52.49	42.36	81.27	38.24	43.54	36.62	37.88	**36.38**
	MAE	33.21	40.20	32.03	64.55	28.35	33.18	27.17	27.79	**26.50**
	MAPE	1.67%	1.99%	1.62%	3.34%	1.43%	1.64%	1.35%	1.37%	**1.31%**
June	RMSE	38.19	49.67	39.22	62.52	41.84	35.94	36.59	40.84	**34.44**
	MAE	28.43	37.08	30.33	43.37	28.62	26.32	24.45	28.38	**23.86**
	MAPE	1.44%	1.91%	1.63%	2.35%	1.49%	1.37%	1.27%	1.47%	**1.23%**

Table 14
Results of one-day ahead for experiment 1.

Month		SVR	ESN	BPNN	RF	LSTM	GA–DRESN	DE–DRESN	BSA–DRESN	IBSA–DRESN
January	RMSE	157.62	206.92	200.01	211.35	**152.82**	194.49	162.16	172.95	162.99
	MAE	116.52	157.80	153.99	161.90	**112.75**	150.35	122.03	131.98	123.55
	MAPE	4.09%	5.53%	5.27%	5.78%	**3.84%**	5.39%	4.33%	4.73%	4.38%
February	RMSE	142.02	191.59	148.86	156.42	153.74	174.47	173.82	158.49	**142.90**
	MAE	107.48	151.95	108.56	119.14	116.90	128.69	127.76	118.80	**103.93**
	MAPE	4.37%	6.30%	4.48%	4.91%	4.79%	5.16%	5.15%	4.95%	**4.23%**
March	RMSE	149.59	204.77	225.90	194.21	**133.28**	189.21	150.49	147.22	153.33
	MAE	113.88	164.07	164.80	144.71	**104.79**	146.21	116.43	114.97	119.21
	MAPE	4.53%	6.67%	6.66%	5.75%	**4.25%**	5.88%	4.73%	4.71%	4.88%
April	RMSE	172.15	197.39	243.15	181.03	152.11	155.37	150.58	174.56	**137.56**
	MAE	127.55	148.15	179.00	131.55	112.84	113.75	108.47	131.74	**100.76**
	MAPE	5.62%	6.50%	8.00%	5.80%	4.96%	5.00%	4.75%	5.73%	**4.48%**
May	RMSE	145.37	193.11	130.43	136.49	**110.72**	139.37	143.52	144.53	122.60
	MAE	113.60	153.54	101.67	106.01	**86.68**	107.89	111.10	109.43	95.16
	MAPE	5.46%	7.36%	4.95%	5.22%	**4.34%**	5.23%	5.34%	5.23%	4.65%
June	RMSE	112.43	211.70	147.09	122.48	106.35	136.91	138.51	134.13	**116.88**
	MAE	80.28	163.41	107.40	88.05	81.04	93.50	96.58	90.11	**78.64**
	MAPE	4.07%	8.27%	5.55%	4.48%	4.12%	4.75%	4.84%	4.53%	**4.03%**

the features with high correlation with the output are retained. The feature that ranks higher in Table 9 is retained. The final selected characteristics are shown in Table 10.

4.4.2. Result analysis

The selected inputs of reservoir R1 and R2 are used as input data for SVR, ESN, BPNN, RF and LSTM. For GA–DRESN, DE–DRESN, BSA–DRESN, and IBSA–DRESN, the input data are the same. The selected hyper-parameters for IBSA–DRESN are shown in 11 and 12. The forecasting results of one-step ahead and one-day ahead forecasting by using different forecasting models are shown in Tables 13 and 14. The partial forecasting and actual values for the one-step ahead forecasting are shown in Fig. 10. Fig. 11 shows the scatter plots of different models for one-step ahead in June. If correlation coefficient is equal to 0, it means that the forecasting data and the actual data have no relationship. If correlation coefficient is equal to 1, it means that the forecasting data and the actual data have a linear relationship.

In one-step ahead forecasting, compared with the SVR, ESN, BPNN, RF, LSTM, GA–DRESN, DE–DRESN, and BSA–DRESN models, IBSA—DRESN obtains better prediction performance according to RMSE, MAE, and MAPE. In one-day ahead forecasting, IBSA–DRESN obtains better prediction performance in most cases. Hence, the proposed IBSA–DRESN has better prediction accuracy than SVR, ESN, BPNN, RF, LSTM, GA–DRESN, DE–DRESN, and BSA–DRESN models for the electricity load forecasting in the North-America dataset. Compared with GA, DE, and BSA, IBSA shows its good capability in the optimization of the DRESN parameters.

4.4.3. Statistical testing of experiment 1

According to Civicioglu (2013), the Wilcoxon signed-rank test at the 0.05 level has been usually adopted for pairwise comparisons. The null hypothesis H0 for the Wilcoxon signed-rank test is: 'There is no difference between the median of the results obtained by method A and the median of the results obtained by method B for the same benchmark problem'. If the p-value is less than 0.05, method A significantly differs from method B.

Wilcoxon signed-rank test at the 0.05 level using the forecasting values is performed to check if the performance of the IBSA–DRESN significantly differs from those of SVR, ESN, BPNN, RF, LSTM, GA–DRESN, DE–DRESN, and BSA-DRESN models. For one-step ahead forecasting, the results in Table 15 show that the IBSA–DRESN is statistically superior to SVR, ESN, and BPNN for all months. For one-day ahead forecasting, the results in Table 16 show that the IBSA–DRESN is statistically better than ESN, DE–DRESN, and BSA–DRESN for all months.

Table 15
Wilcoxon signed-rank test results of one-step ahead for experiment 1.

Month		IBSA–DRESN versus SVR	IBSA–DRESN versus ESN	IBSA–DRESN versus BPNN	IBSA–DRESN versus RF	IBSA–DRESN versus LSTM	IBSA–DRESN versus GA–DRESN	IBSA–DRESN versus DE–DRESN	IBSA–DRESN versus BSA–DRESN
January	p value	0.0021	0	0	0.0134	0.0506	0	0.0793	0
	Different?	Yes	Yes	Yes	Yes	No	Yes	No	Yes
February	p value	0	0	0	0	0	0	0	0
	Different?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
March	p value	0	0.0001	0	0	0	0.9873	0	0.0003
	Different?	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
April	p value	0.0279	0	0.0108	0	0.0012	0	0.9736	0.0552
	Different?	Yes	Yes	Yes	Yes	Yes	Yes	No	No
May	p value	0.0134	0	0	0.0024	0.0468	0	0.4132	0.0004
	Different?	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
June	p value	0	0	0	0.0678	0	0	0	0
	Different?	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes

Table 16
Wilcoxon signed-rank test results of one-day ahead for experiment 1.

Month		IBSA–DRESN versus SVR	IBSA–DRESN versus ESN	IBSA–DRESN versus BPNN	IBSA–DRESN versus RF	IBSA–DRESN versus LSTM	IBSA–DRESN versus GA–DRESN	IBSA–DRESN versus DE–DRESN	IBSA–DRESN versus BSA–DRESN
January	p value	0.3316	0.0012	0	0.0418	0	0.0003	0.0210	0
	Different?	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
February	p value	0	0	0.0052	0	0	0.9151	0	0.0002
	Different?	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
March	p value	0.2220	0	0	0	0	0.2069	0	0
	Different?	No	Yes	Yes	Yes	Yes	No	Yes	Yes
April	p value	0	0	0	0.2376	0.6755	0	0	0
	Different?	Yes	Yes	Yes	No	No	Yes	Yes	Yes
May	p value	0	0	0	0.3795	0	0	0	0
	Different?	Yes	Yes	No	No	Yes	Yes	Yes	Yes
June	p value	0	0	0	0	0	0.1328	0	0
	Different?	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes

4.5. Experiment 2: electricity load of PJM

In this section, for electricity load of PJM, the result of feature selection using MI is present in Section 4.5.1. The forecasting result and statistical testing are given in Sections 4.5.2 and 4.5.3.

4.5.1. Feature selection by MI

Similar to experiment 1, the hourly load values of the previous 1–12, 24, 48, 72, and 168 h are the input factors. The time information is also used as the candidate input factor.

The candidate input factors are shown in Eq. (43), which has 18 candidate input factors. L(t) is the electric load at time t. $D(t) \in \{0.75, 0.25\}$ as the working day is set as 0.75, and the weekend is set as 0.25.

$$\begin{aligned} Input\ factors(t) = \{&L(t-1), L(t-2), \ldots, L(t-12), L(t-24) \\ &L(t-48), L(t-72), L(t-168), H(t), D(t)\}. \end{aligned} \tag{43}$$

MI is used to further select candidate input sets. The specific strategies are presented as follows:

For the reservoir R1 of IBSA–DRESN, we directly take the load of the previous 1–12 h ($\{L(t-1), L(t-2), \ldots, L(t-12)\}$) as the input data. For one-day ahead forecasting, the load of 1–24 h ($\{L(t-1), L(t-2), \ldots, L(t-24)\}$) in the previous day are directly used as the input data as the load in the prediction day is unknown.

For the reservoir R2 of IBSA–DRESN, the load of the previous 24, 48, 72, and 168 h and time information are used as the input factors. The MI between these candidate input factors and the output is calculated, and the variables with strong correlation by MI are retained. In this study, the average standard MI of all features is selected as the selection criterion, and the feature that the MI is larger than the average standard MI is retained. The reserved input factors are shown in Table 17, and 5 features are reserved as alternative input factors. The final selected characteristics are shown in Table 18.

Table 17
Selected inputs between the candidate input factors and the output of experiment 2.

Input	Rank	Standard mutual information
L($t-24$)	1	0.798
D	2	0.681
L($t-72$)	3	0.676
L($t-168$)	4	0.666
L($t-48$)	5	0.666
Mean		0.641

4.5.2. Result analysis

The selected inputs of reservoir R1 and R2 are used as input data for SVR, ESN, BPNN, RF and LSTM. For GA–DRESN, DE–DRESN, BSA–DRESN, and IBSA–DRESN, the input data are the same. The selected hyper-parameters for IBSA–DRESN are shown in 19 and 20 The forecasting results of one-step ahead and one-day ahead forecasting by using different forecasting models are shown in Tables 21 and 22. The partial forecasting and actual values for the one-step ahead and one-day ahead forecasting are shown in Figs. 12 and 13, respectively.

In one-step ahead forecasting, compared with the SVR, ESN, BPNN, RF, LSTM, GA–DRESN, DE–DRESN, and BSA–DRESN models, IBSA–DRESN obtains better prediction performance according to RMSE, MAE, and MAPE. In one-day ahead forecasting, IBSA–DRESN obtains better prediction performance in most cases. Hence, the proposed IBSA–DRESN has better prediction accuracy than SVR, ESN, BPNN, RF, LSTM, GA–DRESN, DE–DRESN, and BSA–DRESN models for the electricity load forecasting of PJM.

In addition, the one-day ahead forecasting results of GA–DRESN, DE–DRESN, BSA–DRESN, and IBSA–DRESN models are obviously better than that of SVR, ESN, BPNN. Thus, the hybrid models show better forecasting ability than these single models in solving the one-day ahead forecasting problem of PJM.

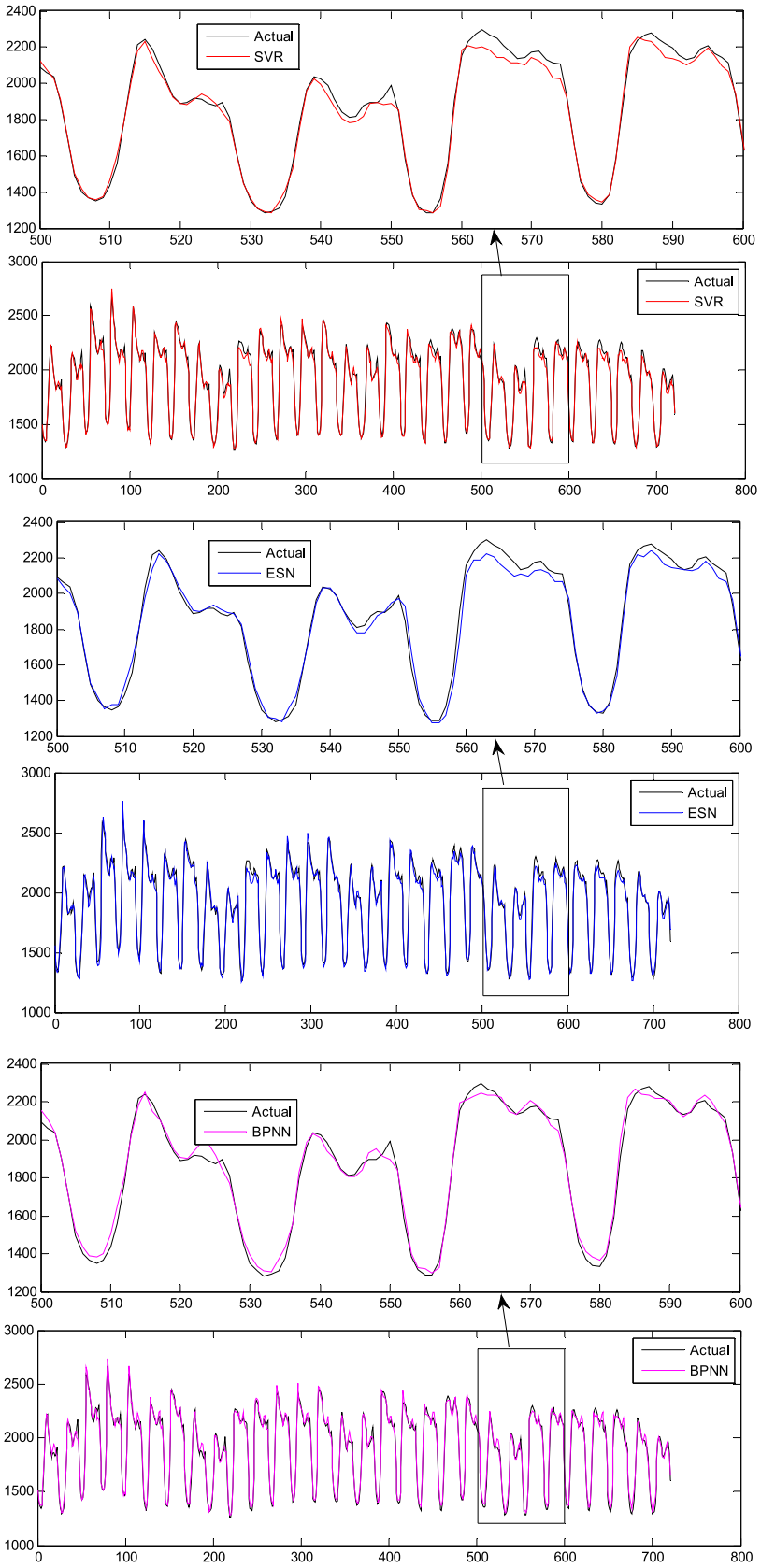


Fig. 10. Forecasting results of one-step ahead in June for experiment 1.

According to Tian (2020), the Taylor diagram is also used to specifically describe the forecasting results of different models. Figs. 14 and 15 are the prediction results of February and April for one-day ahead forecasting, respectively. The root mean square deviation (RMSD), standard deviations (STDE), and the correlation coefficient of forecasting

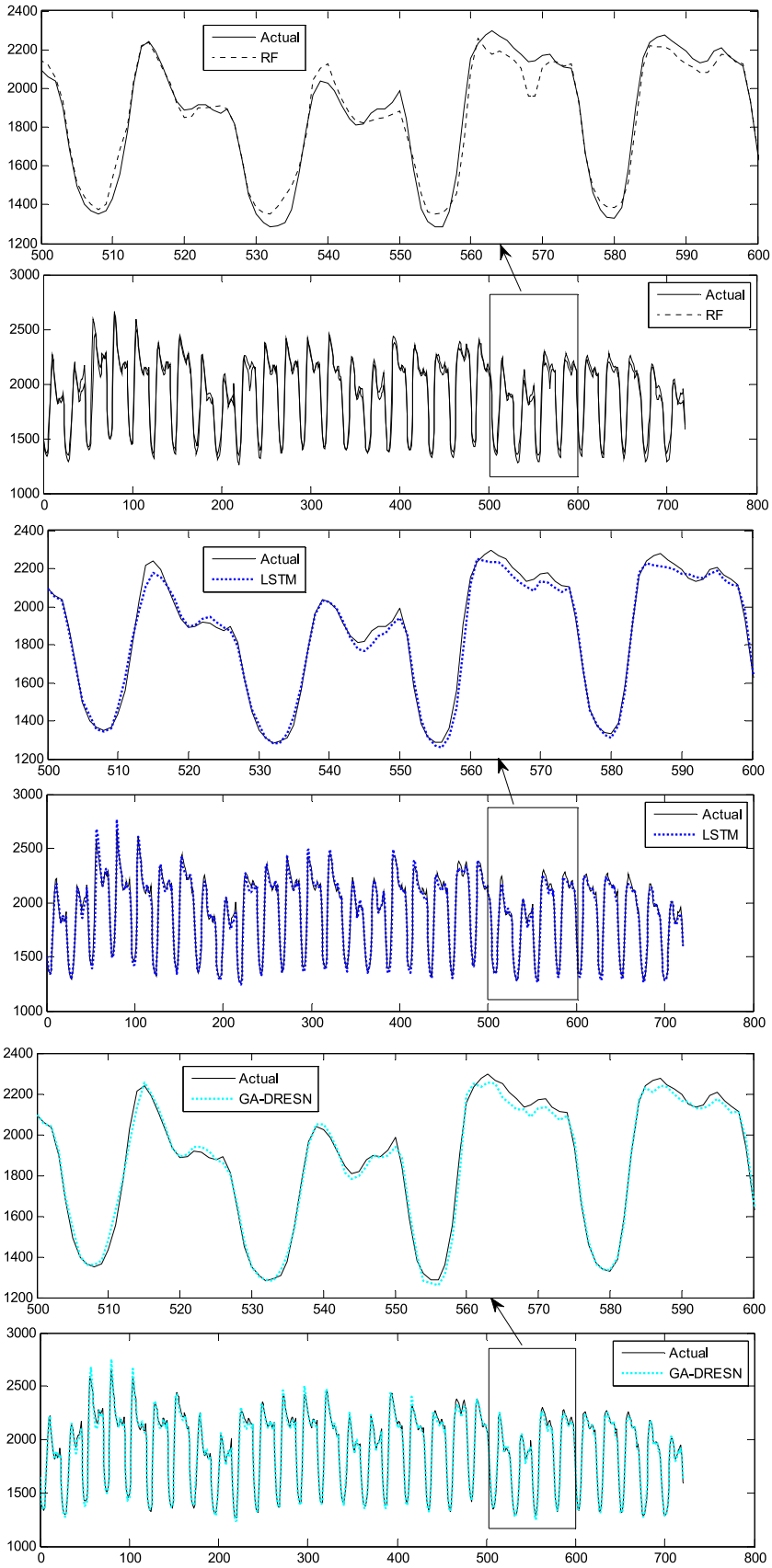


Fig. 10. (*continued*).

value between every prediction method and actual value are clearly presented.

In these figures, the sector curve is the correlation coefficient. The correlation coefficient is larger, the forecast value of the model and

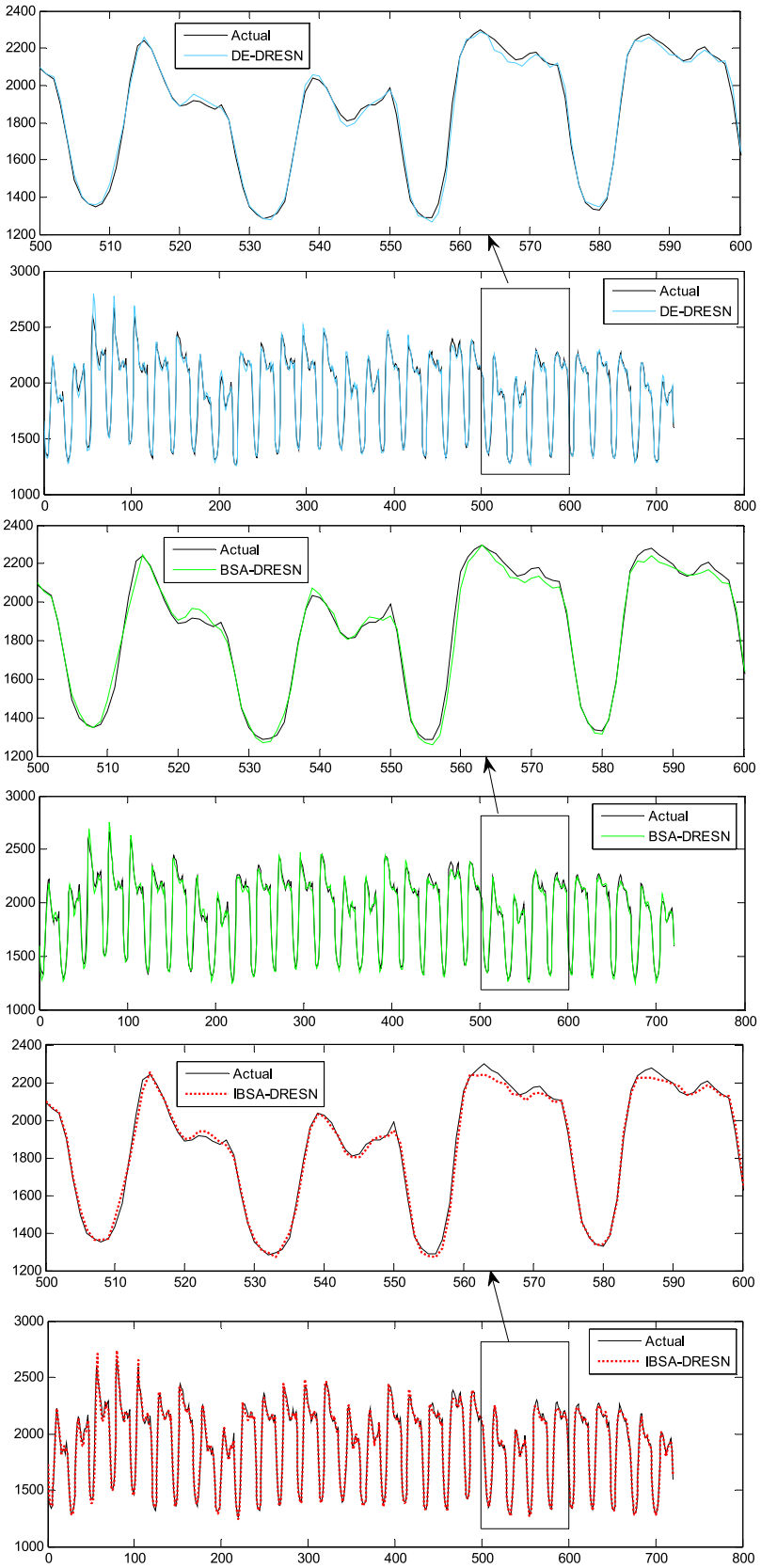


Fig. 10. (*continued*).

the actual value fit better. The horizontal and vertical coordinates are standard deviation. The STDE of the forecast series is proportional to the radial distance between the origin and the model point. The virtual curve is RMSD. Dot A is the actual value of electricity load. The other dots are the STDE, RMSD, and correlation coefficient of IBSA–DRESN and other benchmark models. Fig. 14 shows that IBSA–DRESN and BSA–DRESN are more close to dot A than other models, the correlation coefficient of the IBSA–DRESN and BSA–DRESN are larger than other

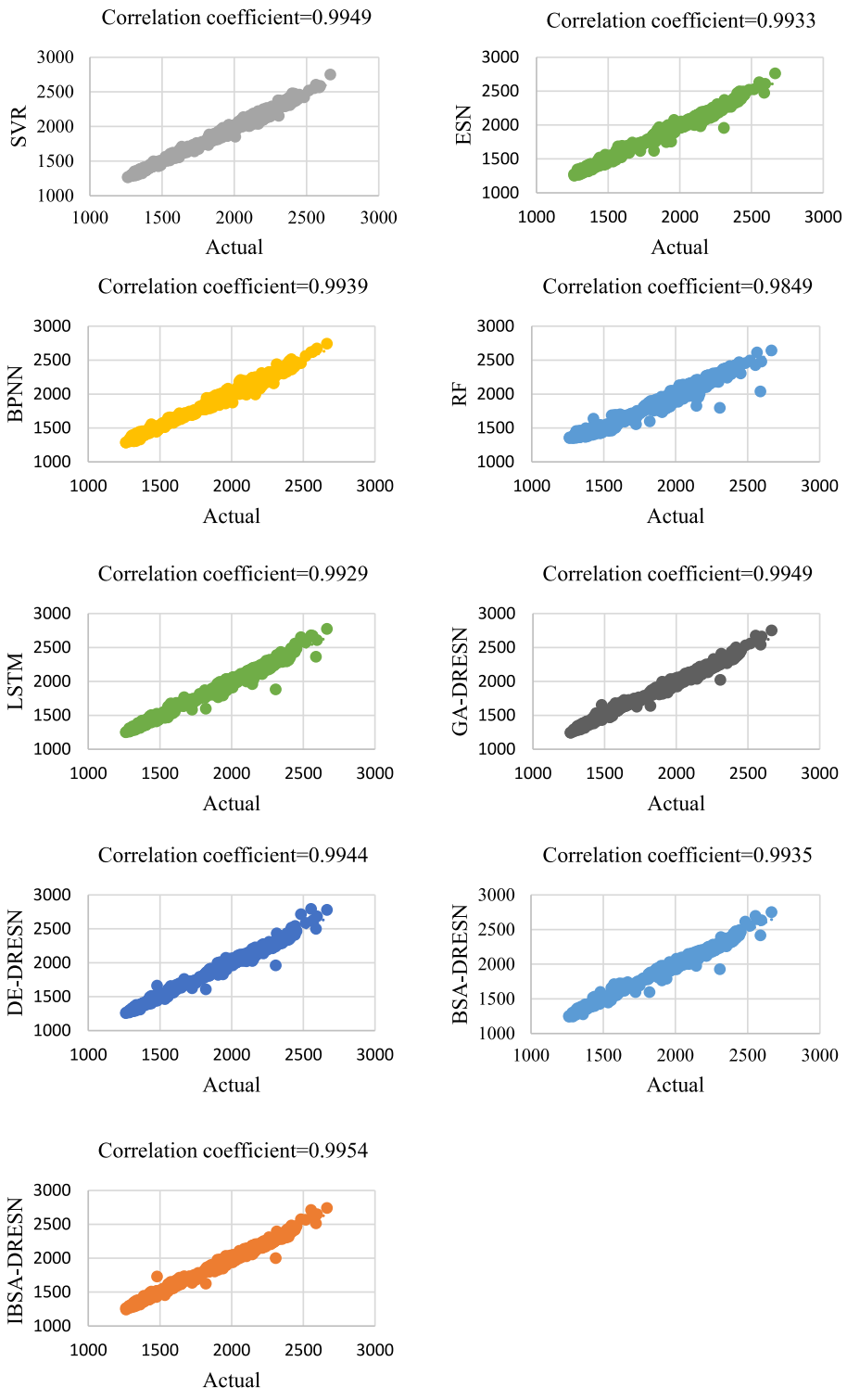
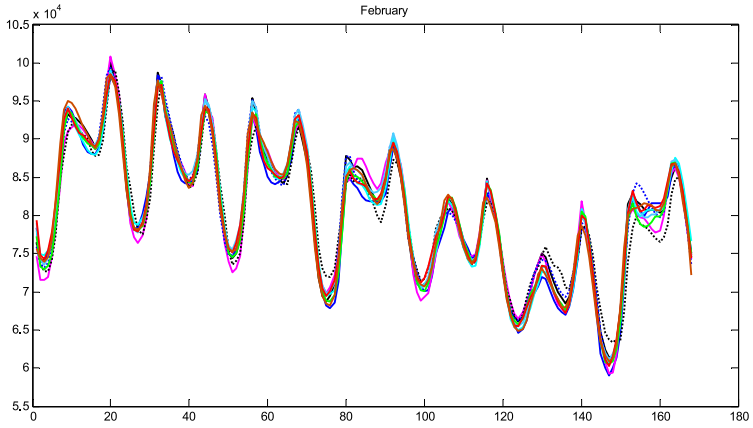


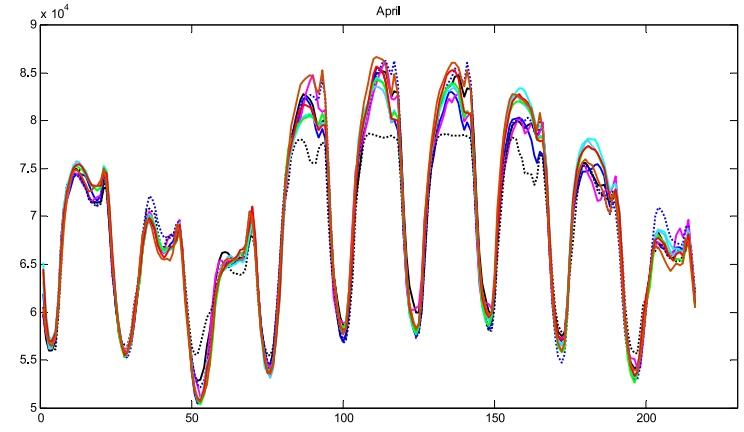
Fig. 11. Scatter plots of different models for one-step ahead in June.

Table 18
Final selected inputs of experiment 2.

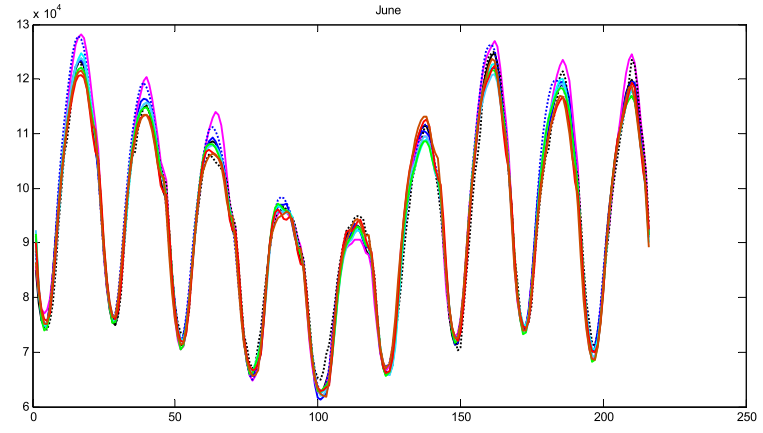
Reservoir	One-step ahead	One-day ahead
R1	$L(t-1), L(t-2),\ldots, L(t-12)$	$L(t-1), L(t-2),\ldots, L(t-24)$
R2	$L(t-24), L(t-48), L(t-72), L(t-168)$, D	$L(t-24), L(t-48), L(t-72), L(t-168)$, D



(a) February



(b) April



(c)June

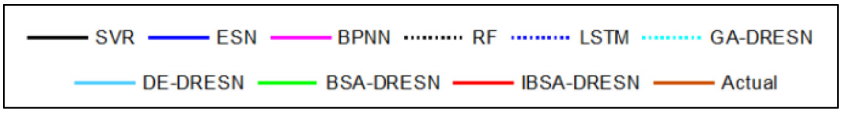


Fig. 12. Forecasting results of one-step ahead for experiment 2.

Table 19(a)
The selected hyperparameters of one-step ahead for experiment 2.

Month	N1	N2	SR1	SD1	SR2	SD2
February	95	85	0.20	0.75	0.88	0.17
April	90	95	0.54	0.53	0.97	0.43
June	73	19	0.92	0.81	0.71	0.49
August	93	69	0.51	0.22	0.8	0.28
October	84	85	0.46	0.21	0.99	0.46
December	87	52	0.18	0.54	0.41	0.88

Table 19(b)
The selected hyperparameters of one-step ahead for experiment 2.

Month	February	April	June	August	October	December
IS2	0.01	0.72	0.06	0.61	0.32	0.43
	0.57	0.87	0.16	0.45	0.70	0.40
	0.40	0.56	0.22	0.56	0.43	0.61
	0.06	0.77	0.07	0.17	0.37	0.84
	0.68	0.14	0.21	0.48	0.71	0.09
IS1	0.51	0.79	0.19	0.22	0.24	0.21
	0.19	0.61	0.95	0.29	0.59	0.15
	0.80	0.38	0.67	0.51	0.44	0.23
	0.20	0.04	0.59	0.16	0.50	0.32
	0.19	0.42	0.14	0.74	0.93	0.56
	0.33	0.45	0.26	0.47	0.34	0.15
	0.80	0.18	0.37	0.22	0.50	0.69
	0.42	0.49	0.20	0.38	0.10	0.79
	0.61	0.19	0.11	0.20	0.18	0.19
	0.51	0.26	0.40	0.56	0.91	0.74
	0.56	0.37	0.16	0.41	0.53	0.30
	0.93	0.92	0.97	0.95	0.95	0.97

Table 20(a)
The selected hyperparameters of one-day ahead for experiment 2.

Month	N1	N2	SR1	SD1	SR2	SD2
February	97	9	0.44	0.75	0.23	0.19
April	83	14	0.11	0.27	0.6	0.16
June	6	70	0.49	0.58	0.35	0.39
August	18	84	0.13	0.14	0.33	0.58
October	3	47	0.70	0.81	0.03	0.31
December	43	64	0.79	0.98	0.18	0.87

benchmark methods. Fig. 15 shows that IBSA–DRESN has the smallest RMSD and dot J is more close to dot A, the correlation coefficient of the IBSA–DRESN is larger than other benchmark methods. Therefore, the forecast value of IBSA–DRESN and the actual value fit better. In general, the IBSA–DRESN shows good forecasting ability.

Table 20(b)
The selected hyperparameters of one-day ahead for experiment 2.

Month	February	April	June	August	October	December
IS2	0.64	0.51	0.04	0.90	0.18	0.01
	0.92	0.25	0.62	0.84	0.55	0.78
	0.17	0.62	0.65	0.88	0.97	0.31
	0.15	0.12	0.32	0.61	0.03	0.08
	0.09	0.01	0.13	0.12	0.35	0.64
IS1	0.51	0.43	0.22	0.40	0.59	0.39
	0.32	0.63	0.65	0.37	0.96	0.20
	0.86	0.08	0.33	0.92	0.26	0.44
	0.42	0.29	0.89	0.96	0.44	0.05
	0.98	0.47	0.31	0.89	0.48	0.22
	0.50	0.97	0.46	0.27	0.31	0.75
	0.09	0.76	0.89	0.06	0.26	0.28
	0.61	0.44	0.20	0.07	0.98	0.80
	0.20	0.10	0.60	0.64	0.82	0.74
	0.36	0.13	0.79	0.50	0.99	0.70
	0.43	0.61	0.87	0.23	0.23	0.01
	0.42	0.72	0.34	0.44	0.92	0.75
	0.79	0.57	0.67	0.76	0.76	0.70
	0.72	0.84	0.89	0.35	0.81	0.40
	0.12	0.62	0.36	0.20	0.57	0.30
	0.61	0.59	0.42	0.29	0.70	0.28
	0.03	0.22	0.81	0.89	0.59	0.06
	0.02	0.15	0.67	0.37	0.78	0.04
	0.56	0.75	0.63	0.12	0.71	0.18
	0.11	0.13	0.71	0.24	0.97	0.19
	0.79	0.65	0.38	0.07	0.29	0.12
	0.54	0.23	0.33	0.05	0.31	0.97
	0.07	0.73	0.62	0.39	0.17	0.77
	0.98	0.96	0.40	0.50	0.03	0.18

4.5.3. Statistical testing of experiment 2

Wilcoxon signed-rank test at the 0.05 level is performed to determine whether the performance of the IBSA–DRESN model differs greatly from that of SVR, ESN, BPNN, RF, LSTM, GA–DRESN, DE–DRESN, and BSA–DRESN models. For one-step ahead forecasting, the results in Table 23 shows that the IBSA–DRESN model is statistically better than ESN and BPNN for all months. For one-day ahead forecasting, the results in Table 24 show that the IBSA–DRESN model is statistically better than SVR for all months.

5. Discussions

In this section, the correlation analysis and stability analysis are provided to prove the advantage of IBSA–DRESN.

Table 21
Results of one-step ahead for experiment 2.

Month		SVR	ESN	BPNN	RF	LSTM	GA–DRESN	DE–DRESN	BSA–DRESN	IBSA–DRESN
February	RMSE	1164.77	1499.34	1743.54	2580.50	1340.97	1486.85	1141.72	1405.20	**1118.63**
	MAE	902.48	1204.61	1404.99	2076.60	1081.42	1169.78	900.01	1094.67	**867.67**
	MAPE	1.12%	1.49%	1.77%	2.62%	1.35%	1.43%	1.11%	1.35%	**1.08%**
April	RMSE	1567.08	1880.48	1780.99	3502.15	1506.30	1709.86	1659.58	1501.91	**1451.63**
	MAE	1232.54	1431.45	1358.62	2620.89	1191.90	1336.58	1256.86	1144.28	**1091.67**
	MAPE	1.75%	1.94%	1.88%	3.59%	1.72%	1.85%	1.72%	1.57%	**1.52%**
June	RMSE	1653.84	1959.97	3222.58	2884.08	2730.49	1933.70	1821.39	1814.06	**1562.54**
	MAE	1367.33	1573.88	2381.01	2270.36	2093.18	1500.29	1443.38	1410.84	**1239.89**
	MAPE	1.48%	1.68%	2.35%	2.53%	2.14%	1.60%	1.52%	1.52%	**1.39%**
August	RMSE	1659.71	1760.28	2667.94	4186.48	2159.60	2301.18	2165.04	1654.81	**1615.73**
	MAE	1295.00	1423.51	2078.65	3030.67	1724.33	1812.12	1734.27	1309.54	**1269.84**
	MAPE	1.53%	1.65%	2.30%	3.76%	1.95%	2.14%	2.01%	1.53%	**1.45%**
October	RMSE	2204.52	1981.27	2445.30	2611.16	1527.19	1970.46	2196.86	1541.92	**1364.95**
	MAE	1686.73	1520.66	1816.20	1978.69	1151.46	1511.83	1655.05	1183.09	**1056.51**
	MAPE	2.20%	1.94%	2.37%	2.55%	1.50%	1.96%	2.14%	1.52%	**1.36%**
December	RMSE	2331.30	1843.44	2246.38	4108.67	2006.17	2089.50	1852.27	2038.17	**1776.25**
	MAE	1681.38	1476.12	1744.90	3459.99	1516.43	1658.62	1497.28	1777.48	**1438.26**
	MAPE	2.07%	1.76%	2.14%	4.39%	1.87%	1.99%	1.80%	2.16%	**1.72%**

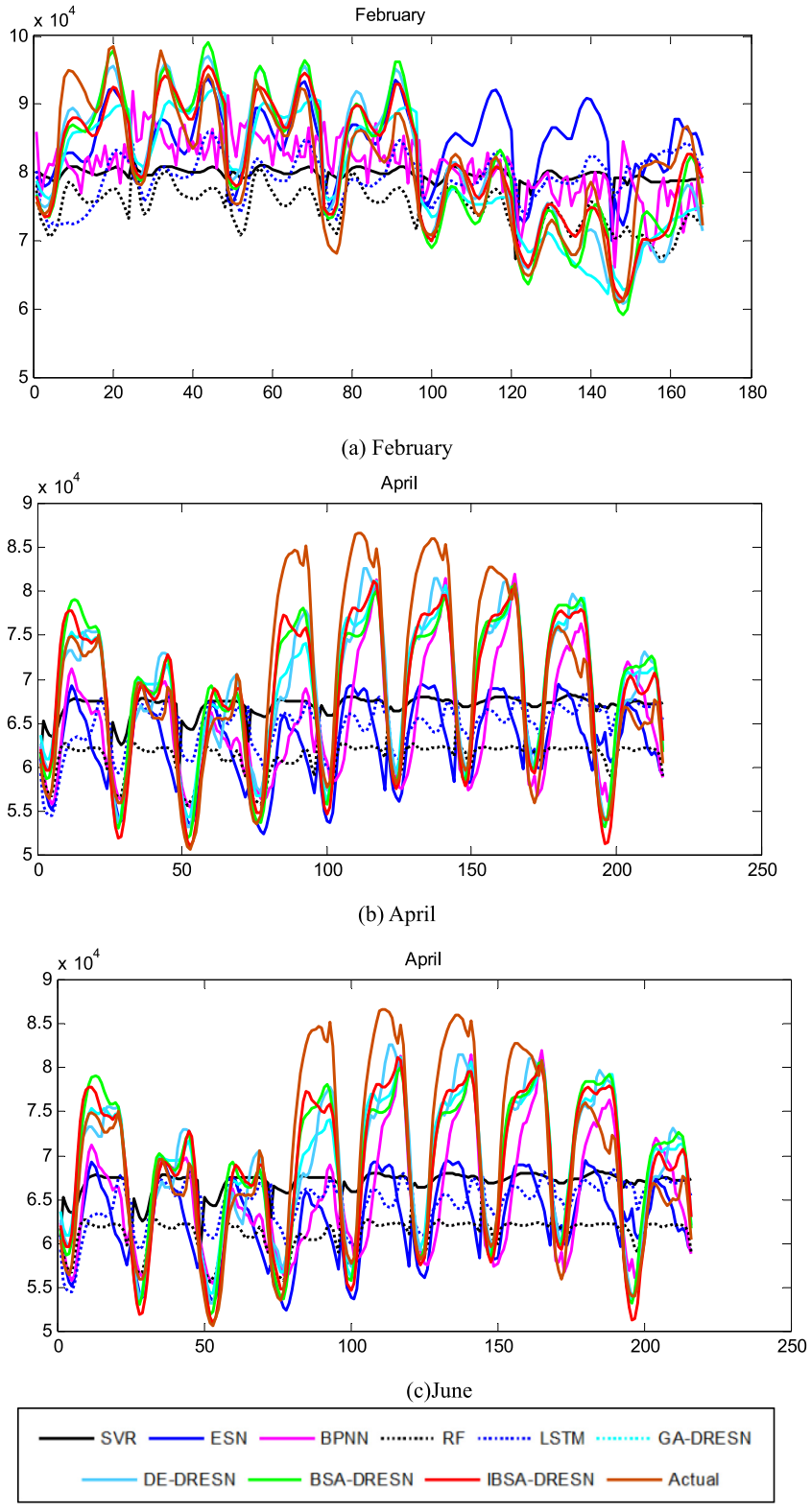


Fig. 13. Forecasting results of one-day ahead for experiment 2.

5.1. Correlation analysis

According to Tian (2020), Pearson's test is added to further prove the advantage of IBSA–DRESN. If Pearson's correlation coefficient is equal to 0, it means that the forecasting data and the actual data have no relationship. If Pearson's correlation coefficient is equal to 1, it means that the forecasting data and the actual data have a linear relationship. The results of Pearson's test are shown in Tables 25–28. As the results of IBSA–DRESN are higher than those of the other benchmark models in most cases, IBSA–DRESN is superior to other models in most cases.

Table 22
Results of one-day ahead for experiment 2.

Month		SVR	ESN	BPNN	RF	LSTM	GA–DRESN	DE–DRESN	BSA–DRESN	IBSA–DRESN
February	RMSE	8281.78	7786.57	8368.02	9298.29	8520.28	5790.76	5070.66	4607.71	**4451.61**
	MAE	6916.59	6054.03	7071.44	7595.71	6776.72	4550.11	3791.66	3579.85	**3451.80**
	MAPE	8.66%	7.94%	8.97%	8.95%	8.42%	5.63%	4.68%	4.40%	**4.27%**
April	RMSE	9462.91	10190.98	8306.95	12222.55	10227.72	4826.42	4770.19	4737.60	**4149.77**
	MAE	7750.52	7907.55	6062.43	9457.81	7993.71	3711.50	3603.26	3835.95	**3412.42**
	MAPE	10.93%	10.54%	8.16%	12.47%	10.89%	5.16%	5.01%	5.29%	**4.73%**
June	RMSE	12743.25	9762.75	12322.90	17176.60	11629.13	6281.37	8855.98	6903.91	**6287.97**
	MAE	10668.89	7643.57	9638.96	13841.27	9651.74	4934.33	6892.59	5554.38	**4812.75**
	MAPE	11.73%	8.37%	9.87%	13.47%	10.06%	5.27%	7.47%	5.86%	**5.17%**
August	RMSE	12923.83	9745.61	9644.37	14077.43	9834.65	**6079.35**	6140.18	6360.46	6205.81
	MAE	10986.73	7452.32	7433.74	11625.50	8097.94	**4708.10**	4988.28	5230.71	5035.33
	MAPE	12.31%	8.17%	8.49%	12.39%	9.57%	**5.48%**	5.83%	6.15%	5.87%
October	RMSE	8233.01	10647.00	10058.36	15034.02	8440.38	3641.79	4100.98	4033.42	**3539.39**
	MAE	6611.36	8527.41	7289.27	13069.55	6537.94	2793.83	3134.19	3079.25	**2730.61**
	MAPE	8.04%	10.38%	8.97%	16.03%	8.07%	3.59%	4.08%	3.98%	**3.53%**
December	RMSE	9819.13	8767.15	10353.43	7743.95	10950.93	6235.91	6444.78	5978.39	**5987.84**
	MAE	8135.33	7436.48	7927.40	6519.30	8972.02	5287.98	5363.25	4997.40	**4864.93**
	MAPE	10.46%	9.21%	9.24%	8.05%	10.66%	6.40%	6.63%	6.10%	**5.95%**

Table 23
Wilcoxon signed-rank test results of one-step ahead for experiment 2.

Month		IBSA–DRESN versus SVR	IBSA–DRESN versus ESN	IBSA–DRESN versus BPNN	IBSA–DRESN versus RF	IBSA–DRESN versus LSTM	IBSA–DRESN versus GA–DRESN	IBSA–DRESN versus DE–DRESN	IBSA–DRESN versus BSA–DRESN
February	p value	0.5505	0	0.0448	0.0809	0.4809	0.0007	0	0.0001
	Different?	No	Yes	Yes	No	No	Yes	Yes	Yes
April	p value	0.3656	0	0.0043	0	0.7173	0.9532	0.0800	0
	Different?	No	Yes	Yes	Yes	No	No	No	Yes
June	p value	0.0007	0.0001	0	0.0380	0	0.1741	0.1908	0.0650
	Different?	Yes	Yes	Yes	Yes	Yes	No	No	No
August	p value	0.8178	0.0175	0	0.9511	0.3453	0	0.6958	0.0178
	Different?	No	Yes	Yes	No	No	Yes	No	Yes
October	p value	0	0	0	0	0.0665	0	0	0.0127
	Different?	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
December	p value	0.5114	0.0152	0	0	0.0010	0	0.1317	0.0015
	Different?	No	Yes	Yes	Yes	Yes	Yes	No	Yes

Table 24
Wilcoxon signed-rank test results of one-day ahead for experiment 2.

Month		IBSA–DRESN versus SVR	IBSA–DRESN versus ESN	IBSA–DRESN versus BPNN	IBSA–DRESN versus RF	IBSA–DRESN versus LSTM	IBSA–DRESN versus GA–DRESN	IBSA–DRESN versus DE–DRESN	IBSA–DRESN versus BSA–DRESN
February	p value	0.0067	0	0.7133	0	0.0030	0	0.2438	0.2257
	Different?	Yes	Yes	No	Yes	Yes	Yes	No	No
April	p value	0.0010	0	0	0	0	0.0006	0.0031	0.0038
	Different?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
June	p value	0.0364	0.0001	0	0	0	0.2567	0.0004	0.9100
	Different?	Yes	Yes	Yes	Yes	Yes	No	Yes	No
August	p value	0	0	0	0	0.1037	0.0568	0.0313	0
	Different?	Yes	Yes	Yes	Yes	No	No	Yes	Yes
October	p value	0	0	0	0	0	0.0063	0.8063	0.0001
	Different?	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
December	p value	0	0.4169	0	0	0	0	0	0
	Different?	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes

5.2. Stability analysis

The randomness noise signal with an amplitude of 5% is added to measure the stability of the hybrid IBSA–DRESN model. The results of North-America for one-step ahead forecasting are shown in Table 29. As the RMSE, MAE, and MAPE have little change, IBSA–DRESN can adapt to the new and noisy data set. Therefore, the hybrid IBSA–DRESN model has good stability.

5.3. The advantages of IBSA-DRESN

As shown in Section 3.4.3, the results of the performance test using three benchmark functions show that the IBSA is superior to GA, DE, and BSA. Tables 13 and 21 shows that MAPE of the IBSA–DRESN is below 2% for one-step ahead forecasting. Tables 14 and 22 shows that MAPE of the IBSA–DRESN is below 6% for one-day ahead forecasting.

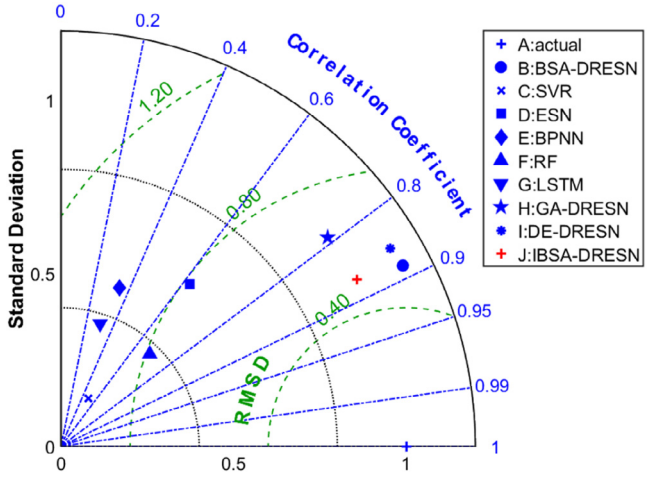


Fig. 14. Taylor diagram of February for one-day ahead forecasting.

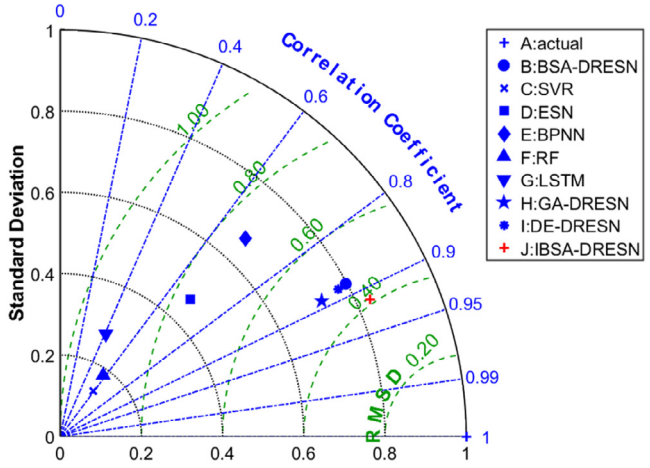


Fig. 15. Taylor diagram of April for one-day ahead forecasting.

Table 25
Pearson's test of one-step ahead for experiment 1.

	January	February	March	April	May	June
SVR	0.9932	0.9914	0.9948	0.9853	0.9931	0.9949
ESN	0.9890	0.9929	0.9907	0.9903	0.9915	0.9933
BPNN	0.9922	0.9915	0.9932	0.9914	0.9934	0.9939
RF	0.9708	0.9731	0.9688	0.9736	0.9803	0.9849
LSTM	0.9923	0.9918	0.9897	0.9903	0.9945	0.9929
GA–DRESN	0.9914	0.9912	0.9935	0.9916	0.9939	0.9949
DE–DRESN	0.9929	0.9929	0.9945	0.9928	0.9951	0.9944
BSA–DRESN	0.9924	0.9938	0.9944	0.9928	0.9951	0.9935
IBSA–DRESN	0.9939	0.9945	0.9954	0.9934	0.9951	0.9954

Table 26
Pearson's test of one-day ahead for experiment 1.

	January	February	March	April	May	June
SVR	0.9527	0.9539	0.9549	0.9154	0.9424	0.9571
ESN	0.9128	0.9076	0.9129	0.8610	0.8665	0.8780
BPNN	0.9373	0.9390	0.8703	0.8400	0.9485	0.9237
RF	0.9181	0.9362	0.9018	0.9039	0.9495	0.9460
LSTM	0.9624	0.9634	0.9591	0.9374	0.9558	0.9634
GA–DRESN	0.9440	0.9280	0.9320	0.9342	0.9489	0.9205
DE–DRESN	0.9523	0.9342	0.9479	0.9381	0.9515	0.9280
BSA–DRESN	0.9498	0.9318	0.9490	0.9280	0.9462	0.9299
IBSA–DRESN	0.9510	0.9439	0.9464	0.9460	0.9545	0.9418

Table 27
Pearson's test of one-step ahead for experiment 2.

	February	April	June	August	October	December
SVR	0.9917	0.9919	0.9955	0.9956	0.9807	0.9720
ESN	0.9871	0.9897	0.9940	0.9942	0.9792	0.9822
BPNN	0.9805	0.9893	0.9905	0.9894	0.9712	0.9699
RF	0.9586	0.9718	0.9847	0.9700	0.9682	0.9390
LSTM	0.9890	0.9889	0.9918	0.9914	0.9883	0.9834
GA–DRESN	0.9858	0.9861	0.9933	0.9916	0.9824	0.9826
DE–DRESN	0.9925	0.9885	0.9941	0.9914	0.9787	0.9836
BSA–DRESN	0.9877	0.9907	0.9940	0.9951	0.9882	0.9815
IBSA–DRESN	0.9925	0.9906	0.9955	0.9952	0.9906	0.9846

Table 28
Pearson's test of one-day ahead for experiment 2.

	February	April	June	August	October	December
SVR	0.4989	0.5884	0.8473	0.6526	0.7315	0.3915
ESN	0.6227	0.6894	0.8543	0.8140	0.6417	0.4302
BPNN	0.3461	0.6839	0.8438	0.8056	0.4903	0.3806
RF	0.6940	0.5757	0.8430	0.7827	0.6786	0.6629
LSTM	0.3049	0.4082	0.8199	0.8118	0.5836	0.0806
GA–DRESN	0.7883	0.8884	0.9265	0.9332	0.9370	0.7625
DE–DRESN	0.8579	0.8843	0.8616	0.9289	0.9023	0.8150
BSA–DRESN	0.8847	0.8825	0.9228	0.9286	0.9210	0.7758
IBSA–DRESN	0.8718	0.9151	0.9271	0.9297	0.9367	0.8035

Table 29
The results of North-America for one-step ahead forecasting after adding noise.

Month		No noise	Adding noise	Month		No noise	Adding noise
January	RMSE	57.95	59.21	April	RMSE	50.23	48.12
	MAE	41.89	42.66		MAE	38.03	36.71
	MAPE	1.43%	1.46%		MAPE	1.69%	1.64%
February	RMSE	51.70	52.57	May	RMSE	36.38	38.54
	MAE	37.72	39.02		MAE	26.50	28.12
	MAPE	1.50%	1.56%		MAPE	1.31%	1.39%
March	RMSE	47.71	49.48	June	RMSE	34.44	35.23
	MAE	36.19	36.98		MAE	23.86	23.80
	MAPE	1.44%	1.46%		MAPE	1.23%	1.25%

Tables 15–16 and Tables 23–24 show the performance of the IBSA–DRESN significantly differs from other models in most cases using the Wilcoxon signed-rank test.

Thus, according to the discussions of two electricity load examples, the proposed IBSA–DRESN model shows a powerful prediction ability for one-step ahead and one-day ahead forecasting. The main advantages of the proposed hybrid forecasting model are as follows: (i) MI is used to eliminate the low-significance input features and select the key input features. The forecasting difficulty is reduced. (ii) The DRESN structure can increase the diversity of the network and alleviate the drawbacks of the randomness of the internal weight matrix. The learning ability of the network is improved. (iii) Roulette strategy, adaptive mutation operator, and niche operator are introduced to enhance the capability of standard BSA. The optimization capability is increased. (iv) IBSA is utilized to optimize the key parameters in the DRESN reservoirs. The forecasting performance of DRESN is increased.

6. Conclusions and future research

This study aims to design a hybrid model based on ESN to forecast electricity load. MI is used to eliminate the low-significance input features and select the key input features. The DRESN structure is designed to increase the diversity of the network and improve its generalization capability. Roulette strategy, adaptive mutation operator, and niche operator are introduced to enhance the performance of the standard BSA algorithm. IBSA is applied to optimize several important parameters in the DRESN neural network.

The proposed IBSA–DRESN method is evaluated using two electricity load datasets, namely, North-America and PJM. Eight benchmark

methods are compared to verify the effectiveness of the IBSA–DRESN method: SVR, ESN, BPNN, RF, LSTM, GA–DRESN, DE–DRESN, and BSA–DRESN. Three error measures (e.g., RMSE, MAE, and MAPE) are applied to test the performance of these forecasting models. The prediction results show that IBSA–DRESN obtains a better forecasting effect compared with the eight models used for one-step ahead and one-day ahead forecasting. Besides, Wilcoxon signed-rank test at the 0.05 level is performed to show the performance of the IBSA–DRESN model differs greatly from that of SVR, ESN, BPNN, RF, LSTM, GA–DRESN, DE–DRESN, and BSA–DRESN models in most cases. Thus, the proposed hybrid IBSA–DRESN model shows better forecasting ability than these eight models in solving the electricity load forecasting problem.

For future work, the proposed IBSA–DRESN model can be used to predict other complicated prediction problems (Zhang et al., 2017; Wu et al., 2021), such as Long-term bank failure prediction (Behbood et al., 2013), or traffic flow prediction (Zhao et al., 2019). Moreover, other feature candidates, such as human activities and big events, can be used as input factors. The data seasonality can also be considered in future. Other popular deep learning methods supported by intelligent optimization technologies can also be used to solve these electricity load forecasting problems (Peng et al., 2020; Qu et al., 2020).

CRediT authorship contribution statement

Lu Peng: Methodology, Software, Investigation, Writing - original draft. **Sheng-Xiang Lv:** Visualization, Validation, Writing - original draft. **Lin Wang:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition. **Zi-Yun Wang:** Investigation, Software, Validation.

Declaration of competing interest

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

Acknowledgment

This research is partially supported by National Natural Science Foundation of China (Nos: 71771095; 71810107003).

References

Battiti, R., 1994. Using mutual information for selecting features in supervised neural net learning. IEEE Trans. Neural Netw. 5 (4), 537–550. http://dx.doi.org/10.1109/72.298224.

Behbood, V., Lu, J., Zhang, G., 2013. Fuzzy refinement domain adaptation for long term prediction in banking ecosystem. IEEE Trans. Ind. Inf. 10 (2), 1637–1646. http://dx.doi.org/10.1109/TII.2012.2232935.

Bianchi, F.M., Scardapane, S., Uncini, A., Rizzi, A., Sadeghian, A., 2015. Prediction of telephone calls load using echo state network with exogenous variables. Neural Netw. 71, 204–213. http://dx.doi.org/10.1016/j.neunet.2015.08.010.

Bouzgou, H., Gueymard, C.A., 2019. Fast short-term global solar irradiance forecasting with wrapper mutual information. Renew. Energy 133, 1055–1065. http://dx.doi.org/10.1016/j.renene.2018.10.096.

Chaib, A.E., Bouchekara, H.R.E.H., Mehasni, R., Abido, M.A., 2016. Optimal power flow with emission and non-smooth cost functions using backtracking search optimization algorithm. Int. J. Electr. Power Energy Syst. 81, 64–77. http://dx.doi.org/10.1016/j.ijepes.2016.02.004.

Che, J., Wang, J., 2014. Short-term load forecasting using a kernel-based support vector regression combination model. Appl. Energy 132, 602–609. http://dx.doi.org/10.1016/j.apenergy.2014.07.064.

Cheng, F., Fan, T., Fan, D., Li, S., 2018. The prediction of oil price turning points with log-periodic power law and multi-population genetic algorithm. Energy Econ. 72, 341–355. http://dx.doi.org/10.1016/j.eneco.2018.03.038.

Chitsazan, M.A., Fadali, M.S., Trzynadlowski, A.M., 2019. Wind speed and wind direction forecasting using echo state network with nonlinear functions. Renew. Energy 131, 879–889. http://dx.doi.org/10.1016/j.renene.2018.07.060.

Chouikhi, N., Ammar, B., Rokbani, N., Alimi, A.M., 2017. PSO-based analysis of echo state network parameters for time series forecasting. Appl. Soft Comput. 55, 211–225. http://dx.doi.org/10.1016/j.asoc.2017.01.049.

Civicioglu, P., 2013. Backtracking search optimization algorithm for numerical optimization problems. Appl. Math. Comput. 219 (15), 8121–8144. http://dx.doi.org/10.1016/j.amc.2013.02.017.

Elhendawi, M., Wang, Z., 2020. An ensemble method of full wavelet packet transform and neural network for short term electrical load forecasting. Electr. Power Syst. Res. 182, 106265. http://dx.doi.org/10.1016/j.epsr.2020.106265.

Hu, Z., Bao, Y., Xiong, T., Chiong, R., 2015. Hybrid filter–wrapper feature selection for short-term load forecasting. Eng. Appl. Artif. Intell. 40, 17–27. http://dx.doi.org/10.1016/j.engappai.2014.12.014.

Hu, Y.L., Chen, L., 2018. A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and differential evolution algorithm. Energy Convers. Manage. 173, 123–142. http://dx.doi.org/10.1016/j.enconman.2018.07.070.

Hu, Y., Li, J., Hong, M., Ren, J., Lin, R., Liu, Y., Liu, M., Man, Y., 2019. Short term electric load forecasting model and its verification for process industrial enterprises based on hybrid GA-PSO-BPNN algorithm—A case study of papermaking process. Energy 170, 1215–1227. http://dx.doi.org/10.1016/j.energy.2018.12.208.

Hu, H.L., Wang, L., Lv, S.X., 2020a. Forecasting energy consumption and wind power generation using deep echo state network. Renew. Energy 154, 598–613. http://dx.doi.org/10.1016/j.renene.2020.03.042.

Hu, H.L., Wang, L., Peng, L., Zeng, Y.R., 2020b. Effective energy consumption forecasting using enhanced bagged echo state network. Energy 193, 116778. http://dx.doi.org/10.1016/j.energy.2019.116778.

Jiang, P., Liu, F., Song, Y., 2017. A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting. Energy 119, 694–709. http://dx.doi.org/10.1016/j.energy.2016.11.034.

Jiang, H., Zhang, Y., Muljadi, E., Zhang, J.J., Gao, D.W., 2018. A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization. IEEE Trans. Smart Grid 9 (4), 3341–3350. http://dx.doi.org/10.1109/TSG.2016.2628061.

Khwaja, A.S., Anpalagan, A., Naeem, M., Venkatesh, B., 2020. Joint bagged-boosted artificial neural networks: Using ensemble machine learning to improve short-term electricity load forecasting. Electr. Power Syst. Res. 179, 106080. http://dx.doi.org/10.1016/j.epsr.2019.106080.

Kim, J., Moon, J., Hwang, E., Kang, P., 2019. Recurrent inception convolution neural network for multi short-term load forecasting. Energy Build. 194, 328–341. http://dx.doi.org/10.1016/j.enbuild.2019.04.034.

Lusis, P., Khalilpour, K.R., Andrew, L., Liebman, A., 2017. Short-term residential load forecasting: Impact of calendar effects and forecast granularity. Appl. Energy 205, 654–669. http://dx.doi.org/10.1016/j.apenergy.2017.07.114.

Lv, S.X., Peng, L., Wang, L., 2018. Stacked autoencoder with echo-state regression for tourism demand forecasting using search query data. Appl. Soft Comput. 73, 119–133. http://dx.doi.org/10.1016/j.asoc.2018.08.024.

Madasu, S.D., Kumar, M.S., Singh, A.K., 2017. Comparable investigation of backtracking search algorithm in automatic generation control for two area reheat interconnected thermal power system. Appl. Soft Comput. 55, 197–210. http://dx.doi.org/10.1016/j.asoc.2017.01.018.

Maldonado, S., Gonzalez, A., Crone, S., 2019. Automatic time series analysis for electric load forecasting via support vector regression. Appl. Soft Comput. 83, 105616. http://dx.doi.org/10.1016/j.asoc.2019.105616.

Moon, J., Jung, S., Rew, J., Rho, S., Hwang, E., 2020. Combination of short-term load forecasting models based on a stacking ensemble approach. Energy Build. 216, 109921. http://dx.doi.org/10.1016/j.enbuild.2020.109921.

Peng, L., Liu, S., Liu, R., Wang, L., 2018. Effective long short-term memory with differential evolution algorithm for electricity price prediction. Energy 162, 1301–1314. http://dx.doi.org/10.1016/j.energy.2018.05.052.

Peng, L., Zhu, Q., Lv, S.X., Wang, L., 2020. Effective long short-term memory with fruit fly optimization algorithm for time series forecasting. Soft Comput. http://dx.doi.org/10.1007/s00500-020-04855-2.

PJM Interconnection, 2020. http://dataminer2.pjm.com/feed/hrl_load_metered (Accessed 10 October 2020).

Qian, W., Shu, W., 2015. Mutual information criterion for feature selection from incomplete data. Neurocomputing 168, 210–220. http://dx.doi.org/10.1016/j.neucom.2015.05.105.

Qu, H., Ai, X.Y., Wang, L., 2020. Optimizing an integrated inventory-routing system for multi-item joint replenishment and coordinated outbound delivery using differential evolution algorithm. Appl. Soft Comput. 86, 105863. http://dx.doi.org/10.1016/j.asoc.2019.105863.

Rahmaninia, M., Moradi, P., 2018. OSFSMI: Online stream feature selection method based on mutual information. Appl. Soft Comput. 68, 733–746. http://dx.doi.org/10.1016/j.asoc.2017.08.034.

Ribeiro, M.H., Coelho, L.D., 2020. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. Appl. Soft Comput. 86, 105837. http://dx.doi.org/10.1016/j.asoc.2019.105837.

Ribeiro, G.T., Mariani, V.C., dos Santos Coelho, L., 2019. Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. Eng. Appl. Artif. Intell. 82, 272–281. http://dx.doi.org/10.1016/j.engappai.2019.03.012.

Shi, H., Xu, M., Li, R., 2018. Deep learning for household load forecasting—a novel pooling deep RNN. IEEE Trans. Smart Grid 9 (5), 5271–5280. http://dx.doi.org/10.1109/TSG.2017.2686012.

Shir, O.M., Emmerich, M., Bäck, T., 2010. Adaptive niche radii and niche shapes approaches for niching with the CMA-ES. Evol. Comput. 18 (1), 97–126. http://dx.doi.org/10.1162/evco.2010.18.1.18104.

Singh, P., Dwivedi, P., 2018. Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem. Appl. Energy 217, 537–549. http://dx.doi.org/10.1016/j.apenergy.2018.02.131.

Singh, P., Dwivedi, P., Kant, V., 2019. A hybrid method based on neural network and improved environmental adaptation method using Controlled Gaussian Mutation with real parameter for short-term load forecasting. Energy 174, 460–477. http://dx.doi.org/10.1016/j.energy.2019.02.141.

Srivastava, V.K., Fahim, A., 2007. An optimization method for solving mixed discrete-continuous programming problems. Comput. Math. Appl. 53 (10), 1481–1491. http://dx.doi.org/10.1016/j.camwa.2007.01.006.

Stéfano Frizzo Stefenon, Ribeiro, M.H.D.M., Nied, A., Mariani, V.C., Ruano, A., 2020. Wavelet group method of data handling for fault prediction in electrical power insulators. Int. J. Electr. Power Energy Syst. 123, 103573,106269. http://dx.doi.org/10.1016/j.ijepes.2020.106269.

Tascikaraoglu, A., Sanandaji, B.M., 2016. Short-term residential electric load forecasting: A compressive spatio-temporal approach. Energy Build. 111, 380–392. http://dx.doi.org/10.1016/j.enbuild.2015.11.068.

Tian, Z., 2020. Short-term wind speed prediction based on LMD and improved FA optimized combined kernel function LSSVM. Eng. Appl. Artif. Intell. 91, 103573. http://dx.doi.org/10.1016/j.engappai.2020.103573.

Tian, Z., Li, S., Wang, Y., 2020. A prediction approach using ensemble empirical mode decomposition-permutation entropy and regularized extreme learning machine for short-term wind speed. Wind Energy 23 (2), 177–206. http://dx.doi.org/10.1002/we.2422.

Tsai, H.C., 2019. Improving backtracking search algorithm with variable search strategies for continuous optimization. Appl. Soft Comput. 80, 567–578. http://dx.doi.org/10.1016/j.asoc.2019.04.032.

Vaghefi, A., Jafari, M.A., Bisse, E., Lu, Y., Brouwer, J., 2014. Modeling and forecasting of cooling and electricity load demand. Appl. Energy 136, 186–196. http://dx.doi.org/10.1016/j.apenergy.2014.09.004.

Wang, L., Hu, H., Ai, X.Y., Liu, H., 2018a. Effective electricity energy consumption forecasting using echo state network improved by differential evolution algorithm. Energy 153, 801–815. http://dx.doi.org/10.1016/j.energy.2018.04.078.

Wang, L., Lv, S.X., Zeng, Y.R., 2018b. Effective sparse adaboost method with ESN and FOA for industrial electricity consumption forecasting in China. Energy 155, 1013–1031. http://dx.doi.org/10.1016/j.energy.2018.04.175.

Wang, L., Peng, L., Wang, S.R., Liu, S., 2020. Advanced backtracking search optimization algorithm for a new joint replenishment problem under trade credits with grouping constraint. Appl. Soft Comput. 2020 (86), 105953. http://dx.doi.org/10.1016/j.asoc.2019.105953.

Wang, Z., Zeng, Y.R., Wang, S., Wang, L., 2019. Optimizing echo state network with backtracking search optimization algorithm for time series forecasting. Eng. Appl. Artif. Intell. 81, 117–132. http://dx.doi.org/10.1016/j.engappai.2019.02.009.

Wu, B.R., Wang, L., Lv, S.X., Zeng, Y.R., 2021. Effective crude oil price forecasting using new text-based and big-data-driven model. Measurement 168, 108468. http://dx.doi.org/10.1016/j.measurement.2020.108468.

Xu, M., Han, M., 2016. Adaptive elastic echo state network for multivariate time series prediction. IEEE Trans. Cybern. 46 (10), 2173–2183. http://dx.doi.org/10.1109/TCYB.2015.2467167.

Zeng, Y.R., Zeng, Y., Choi, B., Wang, L., 2017. Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. Energy 127, 381–396. http://dx.doi.org/10.1016/j.energy.2017.03.094.

Zhang, Y., Chen, H., Lu, J., Zhang, G., 2017. Detecting and predicting the topic change of Knowledge-based Systems: A topic-based bibliometric analysis from 1991 to 2016. Knowl.-Based Syst. 133, 255–268. http://dx.doi.org/10.1016/j.knosys.2017.07.011.

Zhang, J., Wei, Y.M., Li, D., Tan, Z., Zhou, J., 2018. Short term electricity load forecasting using a hybrid model. Energy 158, 774–781. http://dx.doi.org/10.1016/j.energy.2018.06.012.

Zhao, L., Zhou, Y., Lu, H., Fujita, H., 2019. Parallel computing method of deep belief networks and its application to traffic flow prediction. Knowl.-Based Syst. 163, 972–987. http://dx.doi.org/10.1016/j.knosys.2018.10.025.

Zhong, S., Xie, X., Lin, L., Wang, F., 2017. Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction. Neurocomputing 238, 191–204. http://dx.doi.org/10.1016/j.neucom.2017.01.053.