Short-term power load forecasting based on gray relational analysis and support vector machine optimized by artificial bee colony algorithm

Xinfu Pang [Corresp., 1] , **Wei Sun** [1] , **Haibo Li** [1] , **Yibao Wang** [1] , **Changfeng Luan** [2]

[1] Key Laboratory of Energy Saving and Controlling in Power System of Liaoning Province, Shenyang Institute of Engineering, Shenyang, Liaoning, China

[2] Yingkou Power Supply Company, State Grid Liaoning Electric Power Co., Ltd., Yingkou, Liaoning, China

Corresponding Author: Xinfu Pang
Email address: pangxf@sie.edu.cn

Short-term power load forecasting is essential in ensuring the safe operation of power systems and a prerequisite in building automated power systems. Short-term power load demonstrates substantial volatility because of the effect of various factors, such as temperature and weather conditions. However, the traditional short-term power load forecasting method ignores the influence of various factors on the load and presents problems of limited nonlinear mapping ability and weak generalization ability to unknown data. Therefore, a short-term power load forecasting method based on GRA and ABC-SVM is proposed in this study. First, the Pearson correlation coefficient method is used to select critical influencing factors. Second, the gray relational analysis (GRA) method is utilized to screen similar days in the history, construct a rough set of similar days, perform K-means clustering on the rough sets of similar days, and construct the set of similar days further. Artificial bee colony (ABC) algorithm is then utilized to optimize penalty coefficient and kernel function parameters of the support vector machine (SVM). Finally, the above method is applied on the basis of actual load data in Nanjing for simulation verification, and the results show the effectiveness of the proposed method.

Short-term power load forecasting based on GRA and ABC-SVM

Xinfu Pang[1, *], Wei Sun[1], Haibo Li[1], Yibao Wang[1], Changfeng Luan[2]

[1] Key Laboratory of Energy Saving and Controlling in Power System of Liaoning Province, Shenyang Institute of Engineering, Shenyang, Liaoning Province, China
[2] Yingkou Power Supply Company, State Grid Liaoning Electric Power Co., Ltd., Yingkou, Liaoning Province, China

Corresponding Author:
Xinfu Pang[1]
Shenyang, Liaoning, 110136, China
Email address: pangxf@sie.edu.cn

ABSTRACT

Short-term power load forecasting is essential in ensuring the safe operation of power systems and a prerequisite in building automated power systems. Short-term power load demonstrates substantial volatility because of the effect of various factors, such as temperature and weather conditions. However, the traditional short-term power load forecasting method ignores the influence of various factors on the load and presents problems of limited nonlinear mapping ability and weak generalization ability to unknown data. Therefore, a short-term power load forecasting method based on GRA and ABC-SVM is proposed in this study. First, the Pearson correlation coefficient method is used to select critical influencing factors. Second, the gray relational analysis (GRA) method is utilized to screen similar days in the history, construct a rough set of similar days, perform K-means clustering on the rough sets of similar days, and construct the set of similar days further. Artificial bee colony (ABC) algorithm is then utilized to optimize penalty coefficient and kernel function parameters of the support vector machine (SVM). Finally, the above method is applied on the basis of actual load data in Nanjing for simulation verification, and the results show the effectiveness of the proposed method.

INTRODUCTION

Short-term power load forecasting is the basis for the safe operation of power systems and the formulation of dispatching plans (*Liao et al., 2011*). The cyclical nature and substantial uncertainty in the load due to the effect of many factors lead to many load forecasting challenges. Short-term power load forecasting based on selecting similar days can provide a reference for the operation and dispatch of regional power systems.

39   Short-term load forecasting approaches are mainly categorized as traditional and machine
40   learning methods. Traditional methods include time series (*C. E. Borges, Y. K. Penya & I.*
41   *Fenandez, 2013; M. Espinoza et al., 2005*), Kalman filter (*Shalini Sharma & Angshul Majumdar,*
42   *2020*), exponential smoothing (*Siddharth Arora & James W, 2013*) methods and so on. Machine
43   learning methods include artificial neural networks (*Deng et al., 2019; J. Liu, Y. Liu & Q. Zhang,*
44   *2022*), support vector machine (*Bo-Juen Chen, Ming-Wei Chang & Chih-Jen Lin, 2001; Qi et*
45   *al., 2021*), long short-term memory networks (*Shafiul Hasan Rafi et al., 2021*) and so on. *Mai et*
46   *al.* (*2015*) presents a short-term load forecasting method based on Kalman filter, but its weak
47   mapping ability to nonlinearity affects the forecasting accuracy. *L.J. Hu, Z.Z. Guo & J.S. Wang*
48   (*2021*) uses sparrow search algorithm to optimize the penalty coefficient C and kernel function
49   parameter g of the least squares support vector machine and improve the prediction accuracy.
50   *Y.C. Wang, Y.K. Dou & R.Q. Meng* (*2021*) utilizes a multicore extreme learning machine with
51   improved particle swarm optimization parameters for prediction. However, global search ability
52   is weak when particle swarm optimization parameters are optimized. *Liu et al.* (*2018*) applies
53   empirical mode decomposition to divide the load into various IMS components and bat
54   algorithm to optimize parameters of the support vector machine. The output results are then
55   corrected using the Kalman filter method to improve the prediction accuracy. However, the bat
56   algorithm demonstrates the limitations of long search time and weak global search ability. To
57   sum up, although the traditional method based on statistics presents advantages of simplicity and
58   speed, it ignores the influence of various external factors on the load and shows poor nonlinear
59   mapping ability. The method based on machine learning is suitable for processing. The nonlinear
60   short-term power load problem exhibits strong generalization ability to anonymous data but
61   difficulty in optimizing parameters.
62   Preprocessing historical data is essential before using them as input data for short-term load
63   forecasting. *Zhu et al. (2021); Ervin Ceperic, Vladimir Ceperic & Adrijan Baric (2013); Chen et*
64   *al. (2021); Behnam Farsi et al. (2021); Manh-Hai Pham, Minh-Ngoc Nguyen & Yuan-Kang Wu*
65   *(2021); Yin et al. (2020)* screen external factors that affect the load but fail to select similar days
66   for historical data; the results showed that massive data unrelated to the day to be forecasted in
67   the input data occupy a large amount of computing resources. The forecasting time is long, and
68   the forecasting model fails to predict the causal relationship between external factors and the
69   load on the day to be forecasted accurately. *P. Zhao & Y.M. Dai (2020)* uses hourly granularity
70   to select similar days in holiday forecasting to enhance the prediction accuracy but only selects
71   similar days once, only considers the time factor, and ignores important factors, such as weather
72   and temperature. *X.Y. Liu & Y.C. Wei (2020)* utilizes an improved gray relational analysis
73   method to select the set of similar days but only considers the geometric similarity between
74   influencing factors and neglects the numerical similarity. Hence, a large error exists in the
75   selection of similar days. *Wu et al. (2018)* adopts an improved gray relational method to select
76   similar days while solving the problem that gray relational analysis only considers the geometric
77   similarity between factors. However, the selection of similar days is not carried out twice, and
78   the collection of similar days is slightly rough. Therefore, the problem of redundant input data

79   after selecting influencing factors and processing a similar day on input data still exists, thereby

80   decreasing the prediction accuracy.

81   To sum up, the traditional forecasting method ignores the influence of various factors on the

82   short-term power load. Meanwhile, the support vector machine can learn the relationship

83   between external factors and the load and rapidly deal with nonlinear problems. The grey

84   relational analysis method screens historical data for geometric similarity, screening daily

85   historical information similar in shape to the load curve of the day to be forecasted. In addition,

86   because this method cannot screen daily historical data identical to the daily load value to be

87   predicted, the set of similar days screened out is relatively rough. However, K-means clustering

88   filters historical information for numerical similarity. Therefore, using the grey relational

89   analysis method and K-means clustering to select two similar days can screen historical data

90   from both geometric and numerical aspects and remove and predict to the greatest extent. A

91   more accurate collection of similar days can be obtained for historical data with low daily

92   correlation. The artificial bee colony algorithm presents the advantages of fast optimization of

93   parameters and difficulty in falling into the local optimum. On this basis, GRA and ABC-SVM

94   short-term power load forecasting methods are proposed. First, the Pearson correlation

95   coefficient method is adopted to screen external factors affecting the load. Second, historical data

96   are screened twice using gray relational analysis and K-means clustering. Third, the ABC-SVM

97   load forecasting model is constructed. Finally, actual electricity consumption data of a city for 1

98   year is verified, and the prediction results are then compared with those of multiple models, such

99   as long short-term memory (LSTM) neural network, to confirm the effectiveness of the proposed

100   method. The proposed method uses grey relational analysis and K-means clustering to select

101   similar days twice, which solves the problem of reducing prediction accuracy caused by

102   excessive input of irrelevant data in short-term load forecasting. The artificial bee colony

103   algorithm is used to optimize the critical parameters of the support vector machine model, which

104   avoids the problem that the prediction accuracy is reduced due to improper parameter selection

105   of the support vector machine in short-term load forecasting.

106

107   **PROBLEM DESCRIPTION OF SHORT-TERM POWER LOAD**

108   **FORESCASTING**

109   Short-term power load forecasting must consider not only the change of load over time and the

110   influence of temperature, weather conditions, wind direction, wind force, and other factors on the

111   load but also external factors to improve the prediction accuracy. Hence, identifying external

112   factors that exert a important influence on the load is necessary. At the same time, historical data

113   irrelevant to or weakly correlated with the date to be forecasted are necessary due to the massive

114   amount of data required for short-term load forecasting. If these data are used as input, then the

115   forecasting accuracy will decrease. Therefore, filtering historical data and finding historical data

116   closely related to the day to be predicted are necessary for training. Traditional methods based on

117   statistics are unsuitable for nonlinear problems. Hence, choosing an appropriate machine

118   learning algorithm for short-term power load forecasting is necessary. The problem description
119   of short-term power load forecasting is shown in Fig. 1. your materials and methods here.
120   **Figure 1 Description of the short-term power load forecasting problem**
121

STRATEGY STRUCTURE OF SHORT-TERM POWER LOAD
123 FORECASTING

124   The process of short-term power load forecasting based on GRA and ABC-SVM is presented as
125   follows. First, the Pearson correlation coefficient analysis method is used to identify critical
126   influencing factors. Second, the gray correlation analysis method is applied to construct a rough
127   set of similar days, and *K*-means clustering is utilized to filter the similarities again. Third, the
128   ABC algorithm is adopted to find the optimal penalty coefficient *C* and kernel function
129   parameter *g* of the SVM. Finally, the similar day set is inputted into the optimized SVM model
130   as input data for training. Mean absolute (MAE), mean absolute percentage (MAPE), and root
131   mean square (RMSE) errors were selected to measure the performance of the forecasting
132   method. The specific implementation strategy of the method is presented in Fig. 2.
133   **Figure 2 Strategy map for short-term power load forecasting methods**
134

Selection of influencing factors based on Pearson correlation coefficient
136 method

137   Short-term power load forecasting requires not only massive load data but also external factor
138   data. External factors include temperature, weather conditions, wind direction, wind force,
139   whether the day is a workday, and whether the day is a holiday. Changes in external factors will
140   cause variations in the power load. Therefore, the relationship between load and external factors
141   can be accurately determined by filtering external factors with a significant impact on the load to
142   improve the prediction accuracy.
143   The Pearson correlation coefficient method is a statistical indicator used to measure the degree
144   of correlation between the influencing factor *X* and the load *Y*, and its value is in the $[-1,1]$
145   interval (*L. Kong & H. Nian, 2021*). The Pearson correlation coefficient is calculated as follows:

146
$$\rho = \frac{\sum_{k=1}^{n}\left(y_k - \bar{y}\right)\left(x_k - \bar{x}\right)}{\sqrt{\sum_{k=1}^{n}\left(y_k - \bar{y}\right)^2 \sum_{k=1}^{n}\left(x_k - \bar{x}\right)^2}} \ . \tag{1}$$

147   The degree of correlation between the two variables is high when the absolute value of ρ is
148   close to 1. By contrast, the degree of correlation between the two variables is low when the
149   absolute value of ρ is close to 0.

150   The standard definition of correlation for the Pearson correlation coefficient analysis method
151   is presented in Table 1.
152   **Table 1 Correlation coefficient and correlation degree table**
153

Construction of similar daily rough sets based on gray relational analysis

Gray relational analysis is a statistical method for mining the similarity between values *(Xiangyu Kong, Chuang Li & Feng Zheng, 2020)*. A large correlation value corresponds to high similarity between the historical day and the day to be predicted in the selection process of similar days. By contrast, a small correlation value indicates low similarity between the historical day and the day to be measured. The input is the characteristic value of the maximum temperature, minimum temperature, whether the day is a workday, and whether the day is a holiday in the historical day and the day to be predicted. The output is the gray correlation value between the historical day and the day to be predicted. The process of GRA is illustrated in Fig. 3.

Figure 3 Flowchart of selecting the similar day rough set through gray relational analysis

(1) Extract the value of each external factor on the *i*th day to form a set Y_i:

$$Y_i = \begin{bmatrix} y_{i1} & y_{i2} & y_{i3} \cdots y_{im} \end{bmatrix}, \tag{2}$$

where $i = 1, 2, ..., n$; n is the number of historical load data samples; y_{im} is the value of the *m*th external factor on the *i*th day; and m is the number of external factors.

Eigenvector of the day to be predicted Y_0 is expressed as follows:

$$Y_0 = \begin{bmatrix} y_{01} & y_{02} & y_{03} \cdots y_{0m} \end{bmatrix}, \tag{3}$$

where y_{0m} is the value of the *m*th external factor on the day to be predicted.

(2) Construct the gray relational matrix F:

$$F = \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1m} \\ F_{21} & F_{22} & \cdots & F_{2m} \\ . & . & & . \\ F_{n1} & F_{n2} & \cdots & F_{nm} \end{bmatrix}. \tag{4}$$

Each element in the matrix is calculated as follows:

$$F_{ik} = \frac{\min\limits_{i=1,\dots,n} \min\limits_{k=1,\dots,m} |y_{0k} - y_{ik}| + \rho \max\limits_{i=1,\dots,n} \max\limits_{k=1,\dots,m} |y_{0k} - y_{ik}|}{|y_{0k} - y_{ik}| + \rho \max\limits_{i=1,\dots,n} \max\limits_{k=1,\dots,m} |y_{0k} - y_{ik}|}, \tag{5}$$

where $F_{ik} (i = 1, 2, ..., n; k = 1, 2, ..., m)$ is the correlation coefficient corresponding to the *k*th external factor of the *i*th historical load sample, $y_{ik} (i = 1, 2, ..., n; k = 1, 2, ..., m)$ is the value of the *k*th external factor of the *i*th historical load sample, $y_{0k} (k = 1, 2, ..., m)$ is the value of the *i*th external factor on the day to be forecasted, and ρ is the resolution coefficient typically set to ρ =0.5. The Pearson correlation coefficient method is applied to determine the weight of each external factor as follows:

$$W = \begin{bmatrix} \omega_1, & \omega_2, & \cdots & ,\omega_m \end{bmatrix}, \quad \omega_k = \frac{p_k}{\sum\limits_{k=1}^{m} p_k}, \tag{6}$$

182   where p_k is the absolute value of the Pearson correlation coefficient of the kth external factor;

183   ω_k is the weight occupied by the kth external factor; and $k = 1, 2, ..., m$.

184   (3) Use the weights of external factors to weigh the gray relational matrix F and construct the

185   gray relational decision matrix F_1:

186   $$F_1 = FW^T = \begin{bmatrix} \omega_1 F_{11} & \omega_2 F_{12} & ... & \omega_m F_{1m} \\ \omega_1 F_{21} & \omega_2 F_{22} & ... & \omega_m F_{2m} \\ . & . & & . \\ \omega_1 F_{n1} & \omega_2 F_{n2} & ... & \omega_m F_{nm} \end{bmatrix}. \tag{7}$$

187   (4) Add the element values of each row in the matrix to obtain the correlation value D_i of each

188   historical sample.

189   $$D_i = \sum_{k=1}^{m} \omega_k F_{ik}, \tag{8}$$

190   where D_i is the gray correlation value between the ith historical day and the day to be predicted.

191

192   **Secondary selection of similar day sets based on *K*-means clustering**

193   *K*-means clustering *(Xi et al., 2019)* is used to build a set of similar days to reduce historical data

194   in the set with low correlation and days to be predicted further given that the gray relational

195   analysis method only builds a rough set of similar days for geometric similarity. First, the

196   number of cluster centers is set by the silhouette coefficient. Second, the Euclidean distance

197   between influencing factors and the cluster center in each historical day sample in the rough set

198   of similar days is calculated using Formula (9), and the daily samples are classified. Finally, the

199   distance between each cluster center and influencing factors of the day to be predicted is

200   calculated, the cluster center with the minimum distance is selected, and the historical daily

201   samples included in the cluster center are considered the final set of similar days. This process is

202   shown in Fig. 4.

203   $$d_j = \sqrt{\sum_{k=1}^{m} \left(x_k - p_{jk} \right)^2}, \tag{9}$$

204   where $x_k \left(k = 1, 2, ..., m \right)$ is the eigenvalue of the kth factor of influencing factors on the day to be

205   predicted and $p_{jk} \left(j = 1, 2, ..., l; k = 1, 2, ..., m \right)$ is the value of the kth influence factor of the jth

206   group of cluster centers.

207   **Figure 4 *K*-means clustering flowchart for selecting similar day sets**

208

209   **ABC-SVM short-term power load forecasting model**

210   *Mathematical model of ABC algorithm to optimize SVM parameters*

211   (1) Objective function

212   The minimum mean square error J between predicted and actual load values of the support

213   vector machine is used as the objective function in this study.

214    $$\min J = \frac{1}{N}\sum_{i=1}^{N}(y_i' - y_i)^2,$$ (10)

215   where N is the time point, y_i' is the predicted load value at the ith time, and y_i is the actual load

216   value at the ith time.

217   (2) Constraints

218   Setting upper and lower bound constraints on parameters to be optimized is necessary to

219   improve the accuracy of the algorithm, find optimal parameters of the support vector machine,

220   and reduce the optimization time of the algorithm.

221    $$\begin{cases} C \in [0.01, 50] \\ g \in [0.01, 50] \end{cases},$$ (11)

222   where C is the penalty coefficient of the support vector machine and g is the kernel function

223   parameter of the support vector machine.

224

225   ***Design of artificial bee colony optimization algorithm***

226   Karaboga *(Dervis Karaboga & Bahriye Akay, 2007)* proposed an intelligent optimization

227   algorithm that imitates the honey-collecting behavior of bees called the artificial bee colony

228   (ABC) algorithm in 2005. ABC consists of four parts, namely, food source, lead bee, watcher

229   bee, and scout bee, and demonstrates the advantages of fast optimization and difficulty in falling

230   into the local optimum.

231   The relationship between the honeybee collection behavior and the optimization problem in

232   the ABC algorithm is presented in Table 2.

233   **Table 2 Correspondence between honeybee collecting behavior and parameter**

234   **optimization problem**

235   The process of finding optimal parameters of the SVM is presented as follows:

236   (1) Initialize the honey source

237   First, set the population size, number of feasible solutions SN, maximum number of

238   iterations MCN, control parameter *limit*, upper and lower bounds of the penalty coefficient C,

239   and kernel function parameter g. Second, randomly generate two-dimensional SN vector

240   solutions $X_i^t \ (i = 1, 2, ..., SN)$ according to Formula (12), where t is the number of iterations.

241    $$x_{i,j} = x_{\min,j} + rand(0,1)(x_{\max,j} - x_{\min,j}),$$ (12)

242   where $j \in (1,2)$ and $x_{\max,j}$ and $x_{\min,j}$ are the upper and lower bounds of the penalty coefficient C

243   and the kernel function parameter g, respectively. These feasible solutions will be randomly

244   assigned to SN-employed bees, and the fitness of these solutions will be calculated. The

245   relationship between the objective function value of the solution and the fitness value is

246   expressed as follows:

247    $$fit_i = \begin{cases} 1/(1+f_i), & f_i \geq 0 \\ 1 + \mathrm{abs}(f_i), & otherwise \end{cases},$$ (13)

248   where f_i is the objective function value.

249   The objective function value takes the feasible solution as the penalty coefficient C and the

250   kernel function parameter g of the SVM. The mean square error between predicted and actual

251   load values of the predicted day is obtained after inputting historical data in a similar day set for

252   training.

253   (2) Leading bee stage

254   Each lead bee uses Formula (14) to search around its feasible solutions and find new feasible

255   solutions. The mean square error between predicted and actual load values of the support vector

256   machine corresponding to the new solution will be calculated when a new solution is found. If

257   the mean square error corresponding to the new solution is smaller than the old solution, then the

258   lead bee will give up the old solution. A new solution is selected and the penalty

259   coefficient C and kernel function parameter g corresponding to the new solution are recorded.

260   Conversely, the lead bee retains the old solution.

261     $$v_{i,j} = x_{i,j} + \varphi_{i,j}\left(x_{i,j} - x_{k,j}\right), \tag{14}$$

262   where $k \in \{1, 2, ..., SN\}$ and $j \in (1, 2)$ are randomly selected values in the corresponding interval,

263     $k \neq i$, and $\varphi_{i,j}$ is a random number in the range $[-1, 1]$.

264   (3) Calculate the probability that the follower bee follows the leader bee

265   Leading bees will dance in the recruitment area to share the solution information with the

266   follower bees after all the leading bees complete the search. The follower bees calculate the

267   selection probability of each solution as follows:

268     $$p_i = \frac{fit_i}{\sum\limits_{i=1}^{SN} fit_i}, \tag{15}$$

269   where fit_i is the fitness value of the ith solution.

270   (4) Scouting bee search stage

271   The scout bee stage prevents the ABC algorithm from falling into the local optimum. If a

272   feasible solution X_i^t has not been updated, then the number of times *the trail* exceeds the *limit*.

273   The hired bee corresponding to the solution will become a scout bee, abandon the solution, and

274   find a new solution in the neighborhood of the solution. The new solution X_i^{t+1} is determined as

275   follows:

276     $$X_i^{t+1} = \begin{cases} x_{min,j} + rand \cdot \left(x_{max,j} - x_{min,j}\right), trail \geq limit \\ X_i^t, trail_i < limit \end{cases}. \tag{16}$$

277   Figure 5 shows the flowchart of the ABC algorithm to express the specific steps of the ABC

278   algorithm clearly for optimizing the parameters of the support vector machine.

279   **Figure 5 Flowchart of ABC algorithm in optimizing SVM parameters**

280

281   ***SVM short-term power load forecasting model***

282 SVM is a machine learning algorithm proposed by Vapnik et al. based on the structural risk
283 minimization criterion in statistical learning theory, which has the functions of classification and
284 regression *(Mithun Madhukumar et al., 2022; Li et al., 2020; Tan et al., 2020)*. Short-term load
285 forecasting uses the regression function of the support vector machine. The short-term power
286 load presents nonlinear characteristics because of the effect of weather, temperature, and time.
287 Hence, finding a general rule is difficult. SVM can map the short-term power load that shows
288 nonlinear laws from the low-dimensional space to the high-dimensional space through the kernel
289 function and then predicts the load.
290   Specifically, the training set is a collection of m historical daily sample data and expressed as
291 follows:

292 $$T = \{(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)\} \in (X \times Y)^m, \tag{17}$$

293 where x is the normalized value of the highest and lowest temperatures in the historical day,
294 which may a workday or a holiday; y is the load value of the historical day; and m is the number
295 of samples in the training set.
296   SVM regression is ideal for solving the load forecasting problem and obtaining a regression
297 model (Formula [18]). Note that the predicted value $f(x)$ is close to the actual load value y.

298 $$f(x) = \omega^T \varphi(x) + \hat{b}, \tag{18}$$

299 where ω is the weight vector, b is a constant, and $\varphi(x)$ is the nonlinear mapping function.

300   SVM can be expressed after adding the slack variable $\xi_i, \hat{\xi}_i$ and penalty coefficient C to
301 Equation (18) as follows:

302 $$\begin{cases} \min \dfrac{1}{2}\|\omega\|^2 + C\displaystyle\sum_{i=1}^{m}\left(\xi_i - \hat{\xi}_i\right) \\[2mm] \text{s.t. } f(x_i) - y_i \le \varepsilon + \hat{\xi}_i \\[2mm] \quad y_i - f(x_i) \le \varepsilon + \hat{\xi}_i \\[2mm] \xi_i \ge 0, \hat{\xi} \ge 0, i = 1, 2, ..., m \end{cases}, \tag{19}$$

303 where ε is an insensitive loss factor. The insensitive loss function determines the acceptable
304 error size during SVM training. The larger ε is, the greater the acceptable error during SVM
305 training, and the lower the prediction accuracy. The smaller ε, the smaller the acceptable error
306 during SVM training, and the lower the prediction accuracy. If ε is too large, it will lead to
307 under-fitting in training, and if ε is too small, it will lead to over-fitting in training, so the
308 default value of $\varepsilon = 0.1$ is usually taken. The dual problem of SVM can be obtained by
309 introducing Lagrange multipliers.

310 $$\max_{\alpha, \hat{\alpha}} \sum_{i=1}^{m}\left(\hat{\alpha}_i - \alpha_i\right)y_i - \sum_{j=1}^{m}\left(\hat{\alpha}_j + \alpha_j\right)\varepsilon - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\hat{\alpha}_i - \alpha_i\right)\left(\hat{\alpha}_j - \alpha_j\right)K(x_i, x_j), \tag{20}$$

311 where $\alpha_i, \hat{\alpha}_i, \alpha_j, \hat{\alpha}_j$ is the Lagrange multiplier and $K\left(x_i, x_j\right)$ is the kernel function. *Y. Liu, X.X.*
312 *Peng & S.D. Zheng (2021)* selects the RBF kernel function and obtains better prediction results.
313 So the RBF kernel function used in this work is expressed as follows:

314 $$K\left(x_i, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2g^2}\right), \tag{21}$$

315 where g is the kernel function parameter. The following final regression model of SVM can be
316 obtained when Formula (20) is solved:

317 $$f\left(x, \hat{\alpha}_i, \alpha_i\right) = \sum_{i=1}^{m}\left(\hat{\alpha}_i - \alpha_i\right)K\left(x_i, x_j\right) + \hat{b}. \tag{22}$$

318   Among the variables, ω and b are obtained during the operation. Values of the penalty
319 coefficient C and the kernel function parameter g exert the maximum impact on the training
320 accuracy of the support vector machine. The artificial bee colony algorithm will be used to
321 optimize it, and the loss function uses the default value of 0.1.
322

323 **EXAMPLE ANALYSIS**
324 **Dataset description**
325 The study uses a one-year power load dataset in Nanjing, China. The data set includes power
326 load, wind direction, wind power, maximum temperature, minimum temperature, weather
327 conditions, whether the day is a holiday, and whether the day is a workday. The sampling
328 interval of the influencing factor data is 1 day. The sampling interval of load data is 15 minutes.
329

330 **Data preprocessing and performance evaluation indicators**
331 *Data preprocessing*
332 (1) Raw data missing value padding
333   In the process of raw data collection, individual data is inevitably missing. To reduce the
334 impact of missing data on the accuracy of short-term load forecasting, it is necessary to fill in the
335 missing raw information before predicting. There are many filling methods, such as 0 filling,
336 mean filling, median filling, etc. In this study, the average value is used to serve. That is, the
337 average value of the load before and after the missing value is used as the filling data to make it
338 as close to the actual value as possible.
339   (2) Data normalization
340   Historical data are normalized as follows to avoid errors due to different unit dimensions
341 between influencing factors and loads:

342 $$X' = \frac{X - X_{min}}{X_{max} - X_{min}}, \tag{23}$$

343   where X' is the normalized value of historical data, X is the historical data value, $X_{\max}$ is the
344   maximum value of the load and each influencing factor in the historical data, and $X_{\min}$ is the
345   minimum value of the load and each influencing factor in the historical data.

346   *Performance evaluation index*

347   The study used mean absolute error (MAE), mean absolute percentage error (MAPE), and root
348   mean square error (RMSE) as the evaluation indicators of prediction methods. MAE, MAPE, and
349   RMSE are calculated as follows:

350   $$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i' - y_i\right|, \tag{24}$$

351   $$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i' - y_i}{y_i}\right| \times 100,$$

352   (25)

353   $$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i' - y_i\right)^2}, \tag{26}$$

354   where y_i' is the predicted load value and y_i is the actual load value.

355

356   ## Identification of important influencing factors

357   Weather conditions, wind direction, and holidays in influencing factors are usually expressed in
358   Chinese characters and unrelated to the load. Therefore, quantization processing is required.
359   According to the quantitative standard of *Zou et al. (2022)*, Monday to Friday are set as
360   workdays and quantified as 1, and Saturday and Sunday are set as nonworkdays and quantified
361   as 0. Similarly, the quantitative representation of wind direction, weather conditions, and
362   whether the day is a holiday is listed in Table 3.

363   **Table 3 Weather condition, wind direction, and date type quantitative code**

364   Let the influencing factor data be X, and set the matrix $X = \left[X_1, X_2, X_3, X_4, X_5, X_6\right]$, where X_1
365   is the wind direction, X_2 is the minimum temperature, X_3 is the maximum temperature, X_4 is a
366   workday, X_5 is a holiday, and X_6 is the weather condition. Let the load data be Y. Python is used
367   to analyze the Pearson correlation coefficient between each influencing factor and the load and
368   obtain the Pearson correlation coefficient value between each influencing factor and the load in
369   Table 4.

370   **Table 4 Pearson correlation coefficient of influencing factors**

371   As shown in Table 4, the absolute value of the Pearson correlation coefficient is the lowest
372   temperature among the six influencing factors at 0.786. The minimum temperature exerts the
373   maximum influence on the load, and the absolute value of the Pearson correlation coefficient is
374   the weather condition at 0.026. Hence, weather conditions exert the minimum effect on the load.
375   The influence of wind direction on the load is less than that of other factors. Therefore, wind
376   direction and weather conditions are considered weak influencing factors that will be ignored in
377   this study. The four external factors of minimum temperature, maximum temperature, whether

378   the day is a workday, and whether the day is a holiday are important influencing factors for
379   selecting similar days.
380
Similar day selection
382   August 31, 2003 is used as the date to be forecasted in this section. Data of influencing factors
383   between the date to be forecasted and each historic day are extracted, and Python is utilized for
384   gray correlation analysis and calculation of the gray correlation degree between each historical
385   day and the day to be forecasted (Fig. 6). Because the Pearson correlation coefficient between
386   the lowest temperature and the highest temperature is the largest, and it occupies the largest
387   weight in the selection of similar days, so from the time scale, the closer the historical day to the
388   day to be predicted, the greater the gray correlation value. *Huang et al. (2021)* sets the threshold
389   to 0.7 and achieves better prediction results. Therefore, the threshold is set at 0.7, and the
390   historical days greater than 0.7 are taken to form a rough set of similar days, including a total of
391   134 historical days.
392   **Figure 6 Gray correlation between days to be predicted and historical days**
393   Influencing factor data of each historic day in the rough set of the day to be predicted and
394   similar days are extracted as the input feature dimension according to the rough set of similar
395   days. Python is applied to perform *K*-means clustering. *K*-means clustering divides the historical
396   days in the rough set of similar days into different categories according to the number of cluster
397   centres. It obtains the coordinate values of the centre points of each type. Calculate the Euclidean
398   distance between each influencing factor data and the coordinate value of each centre point on
399   the day to be predicted, and the historical day included in the category with the smallest
400   Euclidean distance from the day to be expected is the final set of similar days. The number of
401   cluster centers will affect the clustering effect, and the silhouette coefficient (SIL) is an essential
402   indicator for judging the number of cluster centers. The relationship between the number of
403   cluster centers and the silhouette coefficient is shown in Fig. 7. The clustering effect enhances as
404   the silhouette coefficient approaches 1. Therefore, the number of cluster centers is determined as
405   2.
406   **Figure 7 Relationship between the number of cluster centers and the silhouette**
407   **coefficient**
408   The clustering effect of each historic day in the rough set of similar days is presented in Fig. 8.
409   A total of 134 data points exist, and some data points are covered due to overlapping. Table 5
410   shows the Euclidean distance between the influencing factors on the day to be predicted and each
411   cluster center. It can be seen from Table 5 that the Euclidean distance between the day to be
412   predicted and the cluster center of the first type is small, and the cluster where the cluster center
413   is located contains a total of 88 historical days. Therefore, the set of these 88 historical days is
414   taken as the final set of similar days.
415   **Table 5 Euclidean distance between the day to be predicted and the cluster center**
416   **Figure 8 Clustering effect diagram of each historical day in the rough set of similar days**
417

418   **Validity analysis of the selection of similar days**
419   A total of 296 historical days in the original data, 134 historical days in the rough set of similar
420   days, and 88 historical days in the collection of similar days are used as input data to verify the
421   validity of the selection of similar days. The top 80% of the historical data were selected. The
422   daily data is used as the training set, and 20% of the historical daily data is used as the test set,
423   which is input into the SVM prediction model after the parameters are optimized by the grid
424   search method *(Jiang et al., 2018)* for training and testing. The penalty coefficient C of SVM is
425   0.8, and the kernel function parameter g is 100.
426   Three methods are programmed and case analysis is conducted in this study. (1) Method 1:
427   Utilize Pearson correlation coefficient analysis to screen the influencing factors, use the original
428   data as input data, apply the grid search method to optimize SVM parameters, establish the
429   optimized SVM model, and record the method as SVM. (2) Method 2: Utilize Pearson
430   correlation coefficient analysis to filter influencing factors, adopt the gray correlation analysis
431   method to construct the similar daily rough set, use the similar daily rough set as the input data,
432   apply the grid search method to optimize SVM parameters, establish the optimized SVM model,
433   and denote the method as GSVM. (3) Method 3: Use Pearson correlation coefficient analysis to
434   filter influencing factors, utilize the gray correlation analysis method to construct a rough set of
435   similar days, apply K-means clustering to construct a set of similar days further, use the set of
436   similar days as input data, use the grid search method, optimize SVM parameters, establish the
437   optimized SVM model, and record the method as GKSVM.
438   Errors of the three methods are illustrated in Fig. 9. MAE, MAPE, and RMSE values for the
439   three methods are listed in Table 6. SVM shows the maximum improvement in prediction
440   accuracy when similar days are selected twice, and the MAE, MAPE, and RMSE values are the
441   minimum. The prediction accuracy of SVM without a similar date and time is the minimum, and
442   the MAE, MAPE and RMSE values are the maximum. Therefore, the prediction accuracy will
443   reduce when the original data contains a large amount of daily historical data unrelated to the
444   date to be predicted. This finding proves the effectiveness of the proposed method. The load
445   prediction values of the three methods are shown in Fig. 10. The results showed that the
446   proposed method achieves the maximum degree of coincidence with the actual value.
447   **Table 6 MAE, MAPE, and RMSE values for the three methods**
448   **Figure 9 Predicted load error value at each moment for the three methods**
449   **Figure 10 Comparison chart of predicted and actual values of 96 sampling points for the**
450   **three methods**
451
452   **Artificial bee colony algorithm for optimizing SVM parameters**
453   The artificial bee colony algorithm is used to optimize parameters C and g because the grid
454   search method fails to find optimal parameters C and g of the support vector machine to improve
455   the prediction accuracy of the short-term power load. The set of similar days is used as input
456   data, the population size of the ABC algorithm is 20, the number of nectar sources is SN=10, the
457   maximum number of iterations is MCN=100, the maximum number of cycles of nectar sources is

458   limited to 100, the SVM penalty coefficient is $C \in [0.01, 50]$, and the kernel function parameter

459   is $g \in [0.01, 50]$. The mean square error (MSE) between the predicted load value of SVM and

460   the actual load value is utilized as the objective function *J*.

461   Ninety-six support vector machine prediction models are designed given that the relationship

462   between load values corresponding to 96 sampling points in a day and influencing factors is

463   different. Meanwhile, the ABC algorithm is used to optimize the penalty coefficient *C* and kernel

464   function parameters *g* of 96 vector machines. Optimal parameters *C* and *g* of the SVM at each

465   sampling point are listed in Table 7. Iterative curves of the artificial bee colony, PSO (particle

466   swarm optimization) algorithms and GWO (grey wolf optimization) algorithm are shown in Fig.

467   11. Convergence times of the artificial bee colony algorithm are about 22, and the objective

468   function value finally converges to 36. Convergence times of the particle swarm algorithm are

469   approximately 60, and the objective function value finally converges to 50. The number of

470   convergence times of the grey wolf optimization algorithm is about 35 times, and the objective

471   function value finally converges to 40 times. Convergence times and the final convergence value

472   of the artificial bee colony algorithm are better than those of the particle swarm algorithm and

473   the grey wolf optimization algorithm.

474   **Table 7 Optimal parameters *C* and *g* of the support vector machine at each moment**

475   **Figure 11 ABC, PSO and GWO iteration curves**

476   Optimal solutions of the ABC, PSO and GWO algorithms were collected after running them

477   20 times to compare the performance of the three algorithms further. Relative percentage growth

478   rate (*RPI*) is used to evaluate the performance of the three algorithms as follows:

479    $$RPI(f) = \frac{(f - f^*)}{f^*} \times 100, \tag{27}$$

480   where f represents the minimum mean square error between predicted and actual values of the

481   support vector machine after a single operation of each algorithm and f^* represents the

482   minimum mean square error among all the minimum mean square errors. The *RPI* values of the

483   three algorithms ABC, PSO and GWO are shown in Table 8.

484   **Table 8 RPI values of ABC, PSO and GWO**

485   Optimal parameters obtained by the artificial bee colony algorithm demonstrate a more

486   significant improvement in the prediction accuracy of the support vector machine compared with

487   those of the particle swarm optimization algorithm and the grey wolf optimization algorithm.

488   Multivariate analysis of variance (ANOVA) was performed, in which ABC, PSO and GWO

489   were defined as factors, to show the performance gap between ABC, PSO and GWO intuitively.

490   The results with a confidence level of 95% are shown in Fig. 12.

491   **Figure 12 Multivariate ANOVA of ABC, PSO and GWO**

492

493   ## Analysis of prediction results

494   The five methods are programmed and numerical example analysis is conducted in this study. (1)

495   Method 1: Use the set of similar days as input data, optimize SVM parameters via the grid search

496 method, build the optimized SVM model, and record the method as GKSVM. (2) Method 2: Use
497 the set of similar days as input data, optimize SVM parameters through the particle swarm
498 algorithm, build the optimized SVM model, and record the method as PSO-GKSVM. (3) Method
499 3: Take the set of similar days as the input data, use the grey wolf optimization algorithm to
500 optimize the SVM parameters, build the optimized SVM model, and record the method as GWO-
501 GKSVM. (4) Method 4: Use the set of similar days as input data to build an LSTM model, and
502 this method is recorded as LSTM. (5) The proposed method in this study: Use the set of similar
503 days as input data, utilize the ABC algorithm to optimize SVM parameters, establish the SVM
504 model, and record the method as ABC-GKSVM.
505   The MAE, MAPE, and RMSE values of the five methods are listed in Table 9. The error
506 between predicted and actual load values at 96 time points on August 31, 2003 is shown in Fig.
507 13. Figure 14 illustrates the comparison of predicted and actual values of the five methods at 96
508 time points. The scatter plot in Fig. 15 depicts the degree of agreement between predicted values
509 of the five methods and actual values. The diagonal line of $y=x$ in the figure denoted that the
510 predicted value is equal to the actual value. The point in the figure is close to the diagonal line
511 when the difference between predicted and actual values is negligible.
512   **Table 9 MAE, MAPE, and RMSE values for the five methods**
513   **Figure 13 Comparison chart of error values at each moment of the five methods**
514   **Figure 14 Comparison of predicted and actual values for each method**
515   **Figure 15 Scatter plot of predicted values versus actual values for the five methods**
516   Figures. 13, 14, and 15 and Table 9 showed that the points of ABC-GKSVM are generally the
517 closest to the diagonal among the four methods when two similar days are selected; abnormal
518 points are absent; and the MAE, MAPE, and RMSE values are all less than those of the support
519 vector machine model optimized via the grid search method and particle swarm optimization
520 algorithm. The consistency between the curve trend and actual values verified the effectiveness
521 of the ABC algorithm in optimizing SVM parameters. Meanwhile, compared with the LSTM
522 prediction model, the proposed method reduces the MAE, MAPE, and RMSE indicators by
523 48.44MW, 2.69%, and 41.2MW, respectively, thereby confirming the effectiveness of the
524 proposed method further.
525

526 ## CONCLUSIONS

527 Gray relational analysis and K-means clustering are used in this study to screen 298 historical
528 data. A set of similar days is formed as input data after eliminating historical data that are
529 irrelevant or slightly relevant to the day to be predicted. The ABC algorithm optimizes
530 parameters C and g of the SVM, establishes the ABC-SVM short-term power load forecasting
531 model, and finally obtains forecast values 96 times for the day to be forecasted. Furthermore, the
532 effectiveness of the proposed method is proven using examples.
533   (1) The Pearson correlation coefficient method is used to calculate the Pearson coefficient
534 value between each external factor and the load. External factors that significantly influence the
535 load are screened according to the coefficient value.

536   (2) The gray correlation degree between the day to be predicted and each historical day is

537 calculated through gray correlation analysis, and rough sets of similar days are filtered by setting

538 the threshold. *K*-means clustering is utilized to classify the rough sets of similar days. The cluster

539 center with the minimum Euclidean distance to each influencing factor in the days to be

540 predicted is set as the set of similar days to reduce the input of irrelevant data.

541   (3) The artificial bee colony algorithm is applied to find the optimal penalty coefficient *C* and

542 kernel function parameter *g* of the support vector machine suitable for the data of this study and

543 address the problem of decreasing prediction accuracy of the support vector machine due to

544 improper parameter selection.

545   The selection of external factors with a significant impact on the load and an hourly

546 granularity can be the focus of future investigations to improve the prediction accuracy further

547 given that the influence of temperature and weather conditions on the load contains hysteresis

548 and cumulative effects.

549

REFERENCES

551 **Behnam Farsi, Manar Amayri, Nizar Bouguila, Ursula Eicker. 2021.** On short-term power load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach, *IEEE Access* **9**: 31191-31212.

553 **Bo-Juen Chen, Ming-Wei Chang, Chih-Jen Lin. 2004.** Load forecasting using support vector machines: a study on EUNITE competition 2001, *IEEE Transactions on Power Systems* **19(4)**: 1821-1830.

555 **C. E. Borges, Y. K. Penya, I. Fenandez. 2013.** Evaluating combined load forecasting in large power system and smart grids, *IEEE Transactions on Industrial Informatics* **9(3)**: 1570-1577.

557 **Chen KJ, Chen KL, Wang Q, He ZY, Hu J, He JL. 2021.** Short-term power load forecasting with deep residual networks, *IEEE Transactions on Smart Grid* **10(4)**: 3943-3952.

559 **Deng ZF, Wang BB, Xu YL, Xu TT, Liu CX, Zhu ZL. 2019.** Multi-scale convolutional neural network with time-cognition for multi-step short-term power load forecasting, *IEEE Access* **7**: 88058-88071.

561 **Dervis Karaboga, Bahriye Akay. 2007.** Artificial bee colony (ABC) algorithm, on training artificial neural networks, *2007 IEEE 15th Signal Processing and Communications Applications*, June 11-13, Eskisehir, Turkey, 10286583.

563 **Ervin Ceperic, Vladimir Ceperic, Adrijan Baric. 2013.** A strategy for short-term power load forecasting by support vector regression machines, *IEEE Transactions on Power Systems* **28(4)**: 4356-4364.

565 **Hu LJ, Guo ZZ, Wang JS. 2021.** Short-term power load forecasting based on ISSA-LSSVM model, *Science Technology and Engineering* **21(23)**: 9916-9922.

567 **Huang DM, Zhuang XK, Hu AD, Sun JZ, Shi S, Sun Y, Tang Z. 2021.** Short-term load forecasting based on similar-day selection with GRA-K-means, *Electric Power Construction* **42(7)**: 110-117.

569 **Jiang HG, Zhang YC, Eduard Muljadi, Jun Jason Zhang, David Wenzhong Gao. 2018.** A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization, *IEEE Transactions on Smart Grid* **9(4)**: 3341-3350.

572 **Kong L, Nian H. 2021.** Fault detection and location method for mesh-type DC microgrid using Pearson correlation coefficient, *IEEE Transactions on Power Delivery* **36(3)**: 1428-1439.

574 **Kong XY, Li C, Zheng F. 2020.** Improved deep belief network for short-term power load forecasting considering demand-side management, *IEEE Transactions on Power Systems* **35(2)**: 1531-1538.

576 **Li WW, Shi Q, Muhammad Sibtain, Li D, Daniel Eliote Mbanze. 2022.** A hybrid forecasting model for short-term power load based on sample entropy, two-phase decomposition and whale algorithm optimized support vector regression, *IEEE Access* **8**: 166907-166921.

578 **Liao NH, Hu ZH, Ma YY, Lu WY. 2011.** Review of the short-term load forecasting methods of electric power system, *Power System Protection and Control* **39(1)**: 147-152.

580 **Liu J, Liu Y, Zhang Q. 2022.** A weight initialization method based on neural network with asymmetric activation function. *Neurocomputing* **484**: 171-182.

582 **Liu QZ, Shen YB, Wu L, Li J, Zhuang LR, Wang SF. 2018.** A hybrid FCW-EMD and KF-BA-SVM based model for short-term load forecasting, *CSEE Journal of Power and Energy Systems* **4(2)**: 226-237.

584 **Liu XY, Wei YC 2020.** Short-term load forecasting of BA-BP neural network based on improved grey relational analysis, *Science Technology and Engineering* **20(1)**: 223-227.

586 **Liu Y, Peng XX, Zheng XD. 2021.** Research on short-term power load forecasting method based on improved LS-SVM, *Electrical Measurement & Instrumentation* **58(5):** 176-181.

588 **Manh-Hai Pham, Minh-Ngoc Nguyen, Yuan-Kang Wu. 2021.** A novel short-term power load forecasting method by combining the deep learning with singular spectrum analysis, *IEEE Access* **9**: 73736-73746.

590 **Mai KH, Xiao JH, Wu XC, Chen C. 2015.** Research on ARIMA model parallelization in load prediction based on R language, *Power System*

591   *Technology* **39(11)**: 3216-3220.

592   **M. Espinoza, C. Joye, R. Belmans, B.De Moor. 2005.** Short-term power load forecasting, profile identification, and customer segmentation: a
593   methodology based on periodic time series, *IEEE Transactions on Power Systems* **20(3)**: 1622-1630.

594   **Mithun Madhukumar, Albino Sebastian, Xiaodong Liang, Mohsin Jamil, Md Nasmus Sakib Khan Shabbir. 2022.** Regression model-based
595   short-term load forecasting for university campus load, *IEEE Access* **10**: 8891-8905.

596   **Qi XY, Liu HJ, Hou QH, Liu XY, Tan YC, Wang LC. 2021.** Short-term load forecasting of iron and steel industry area based on combination
597   model of SVM and LSTM, *Journal of Shandong University(Engineering Science)* **51**: 91-98.

598   **Shafiul Hasan Rafi, Nahid-Al-Masood, Shohana Rahman Deeba, Eklas Hossain. 2021.** A short-term power load forecasting method using
599   integrated CNN and LSTM network, *IEEE Access* **51**: 32436-32448.

600   **Shalini Sharma, Angshul Majumdar. 2020.** Víctor Elvira, and Emilie Chouzenoux, Blind kalman filtering for short-term power load forecasting,
601   *IEEE Transactions on Power Systems* **35(6)**: 4916-4919.

602   **Siddharth Arora, James W. 2013.** Taylor, Short-term forecasting of anomalous load using rule-based triple seasonal methods, *IEEE Transactions*
603   *on Power Systems* **28(3)**: 3235-3242.

604   **Tan ZQ, Zhang J, He Y, Xiong GJ, Liu Y. 2020.** Short-term load forecasting based on integration of SVR and stacking, *IEEE Access* **8**: 227719-
605   227728.

606   **Wang YC, Dou YK, Meng RQ. 2021.** Multicore neural network short-term load forecasting model based on fuzzy C-mean clustering-variational
607   model Decomposition and Chaotic swarm intelligence optimization, *High Voltage Engineering* https://doi.org/10.13336/j.1003-
608   6520.hve.20210664.

609   **Wu Y, Lei JW, Bao LS, Li CZ. 2018.** Short-term load forecasting based on improved grey relational analysis and neural network optimized by
610   bat algorithm, *Automation of Electric Power Systems* **42(20)**: 67-72.

611   **Xi YW, Wu JY, Shi C, Zhu XW, Cai R. 2019.** A refined load forecasting based on historical data and real-time influencing factors, *Power System*
612   *Protection and Control* **47(1)**: 80-87.

613   **Yin LF, Sun ZX, Gao F, Liu H. 2020.** Deep forest regression for short-term power load forecasting of power systems, *IEEE Access* **9**: 49090-
614   49099.

615   **Zhao P, Dai YM 2020.**, Power load forecasting of SVM based on real-time price and weighted grey relational projection algorithm, *Power System*
616   *Technology* **44(4)**: 1325-1332.

617   **Zhu LJ, Xun ZH, Wang YX, Cui Q, Chen WY, Lou JC. 2021.** Short-term power load forecasting based on CNN-BILSTM, *Power System Technology*
618   **45(11)**: 4532-4539.

619   **Zou Z, Wu TZ, Zhang XX, Zhang ZM. 2022.** Short-term load forecasting based on Bayesian optimized CNN-BiGRU hybrid neural networks, *High*
620   *Voltage Engineering* https://doi.org/10.13336/j.1003-6520.hve.20220168.

621

Figure 1

Description of the short-term power load forecasting problem

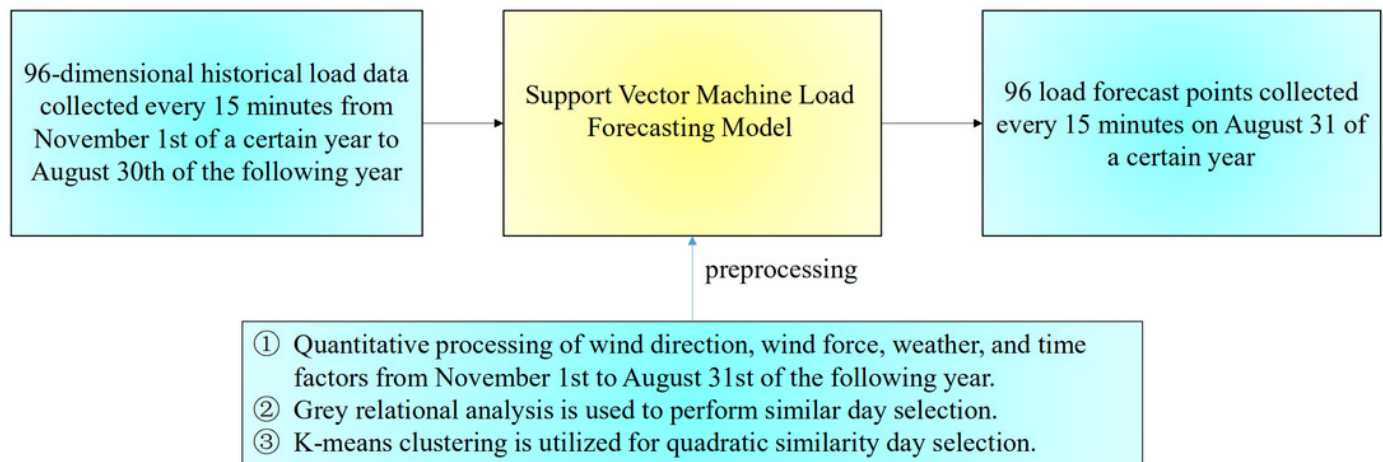



Figure 2

Strategy map for short-term power load forecasting methods

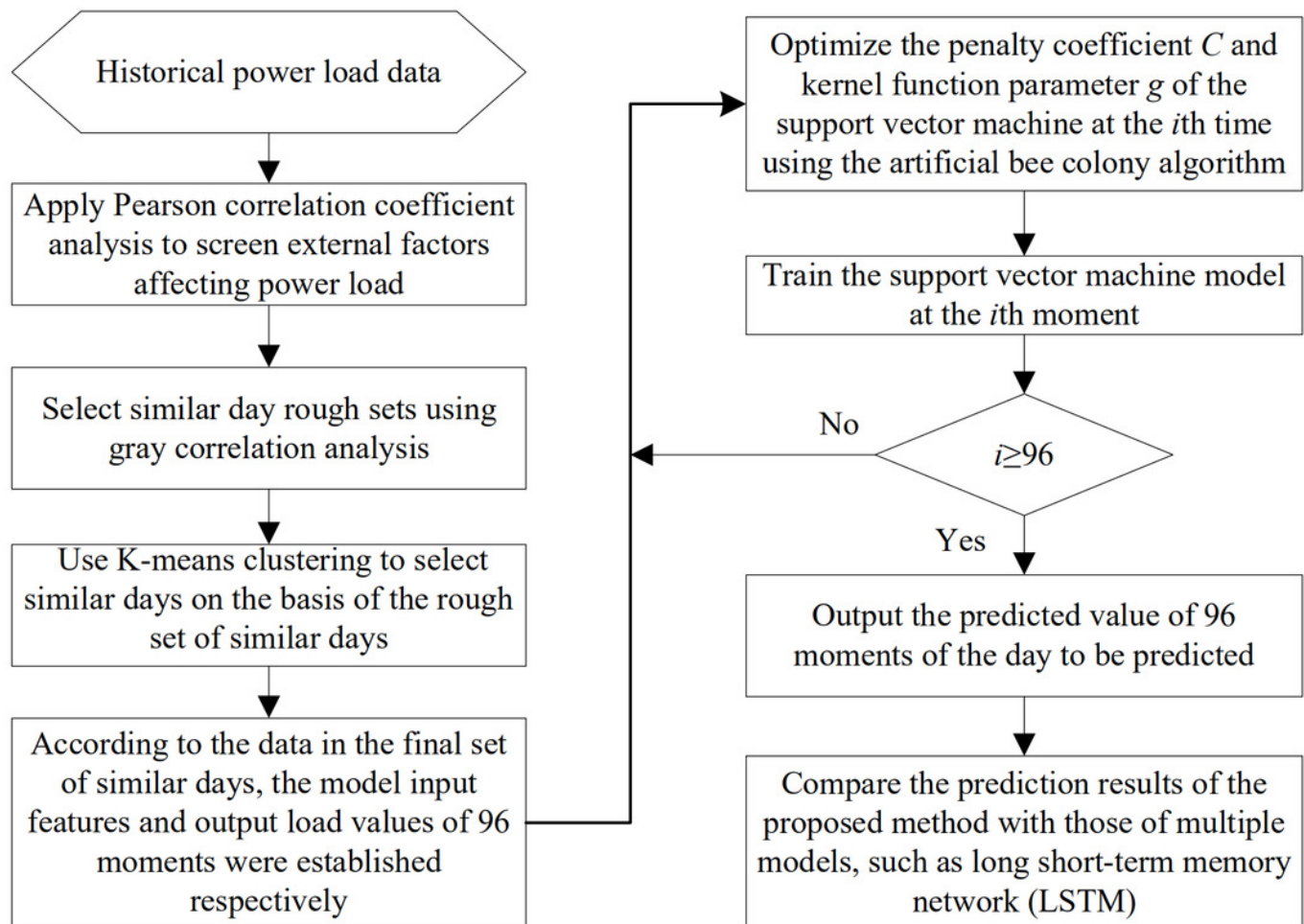



Figure 3

Flowchart of selecting the similar day rough set through gray relational analysis

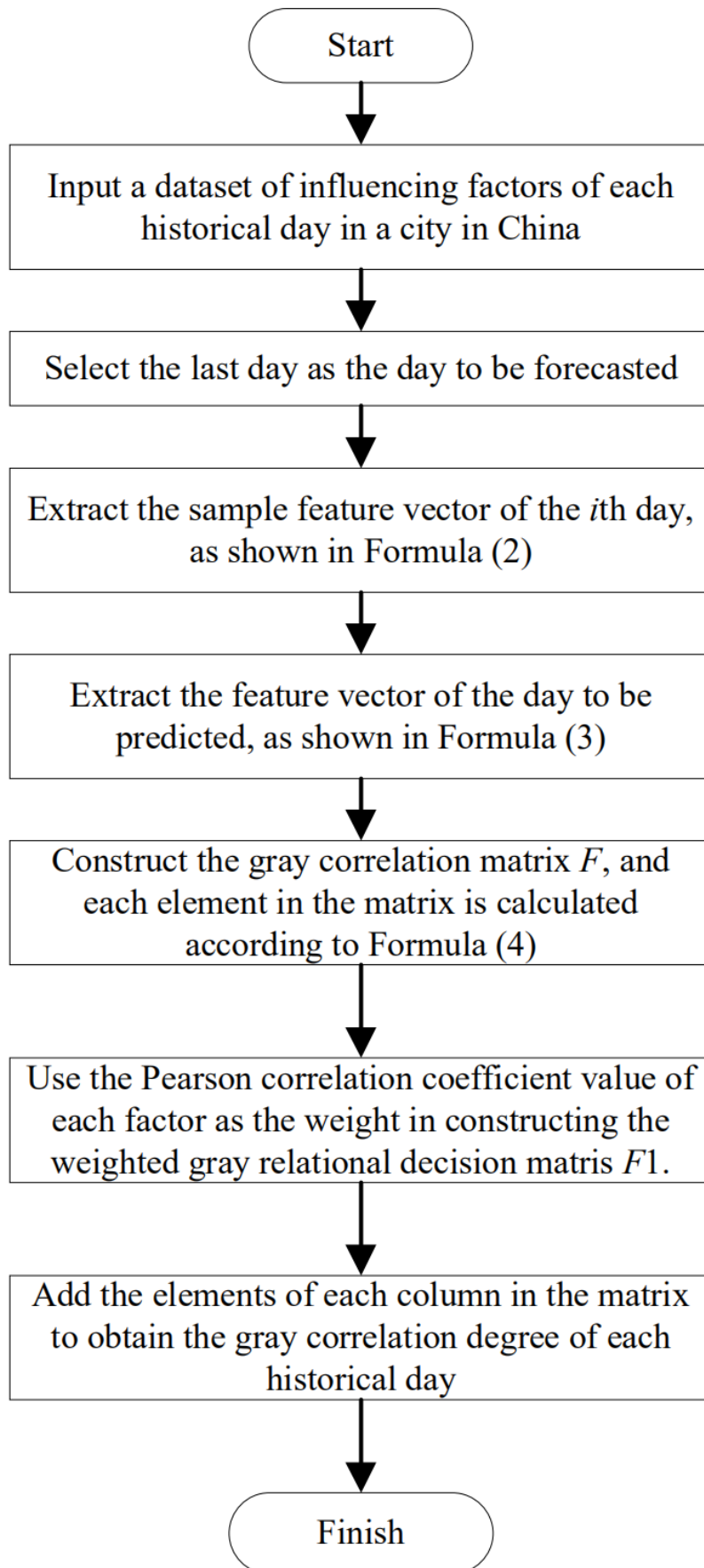



Figure 4

K-means clustering flowchart for selecting similar day sets

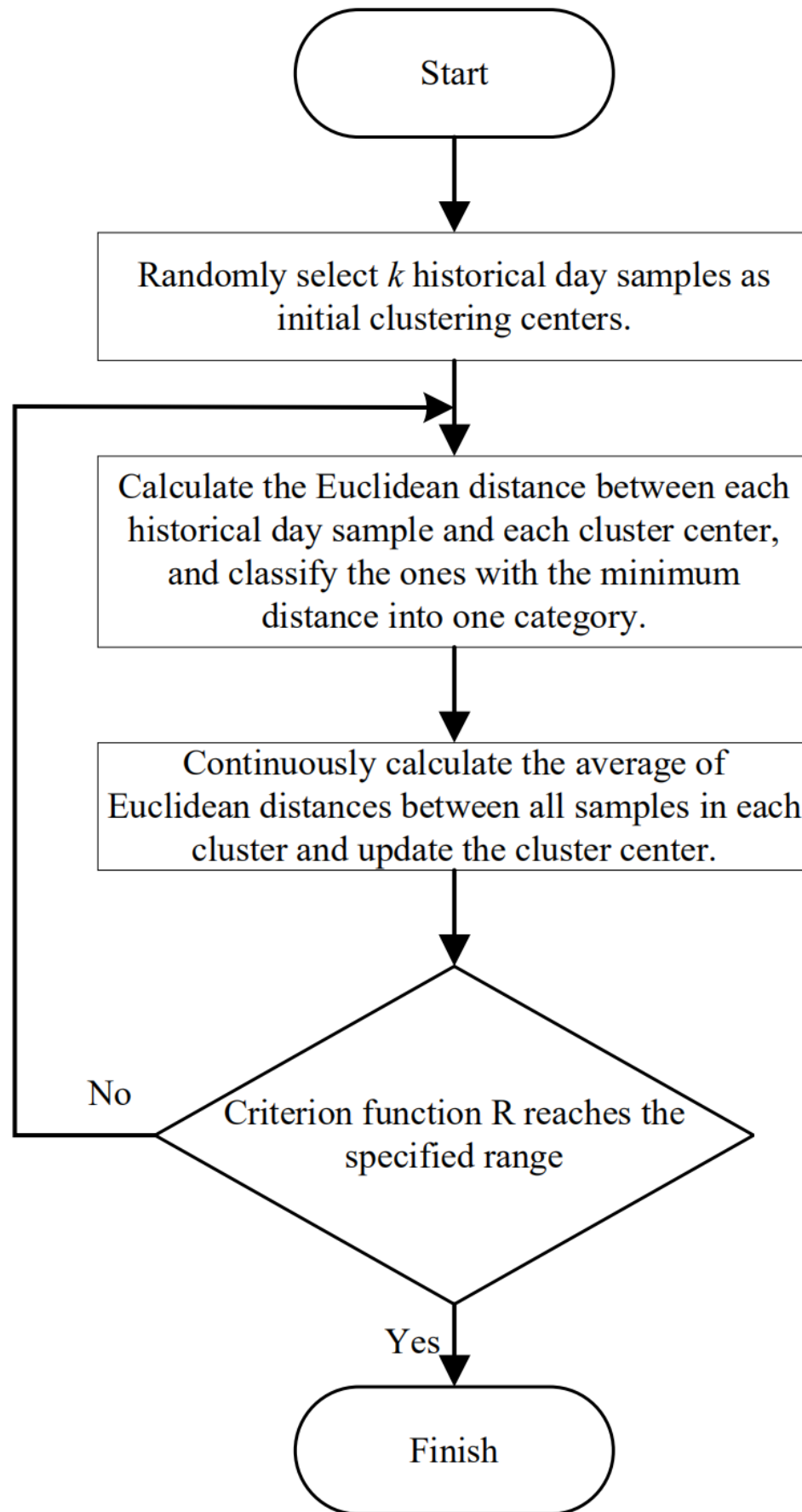



Figure 5

Flowchart of ABC algorithm in optimizing SVM parameters

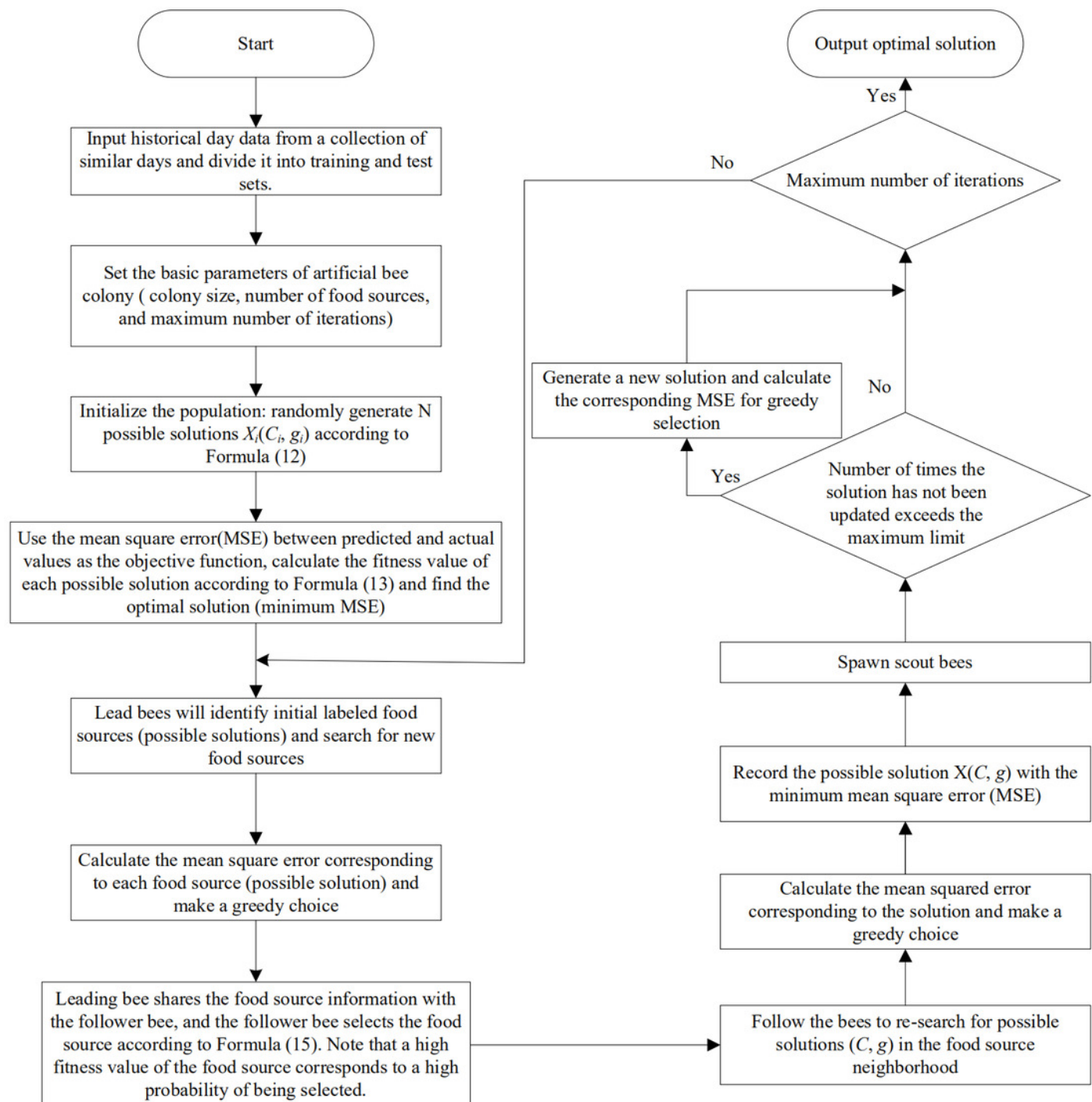



Figure 6

Gray correlation between days to be predicted and historical days

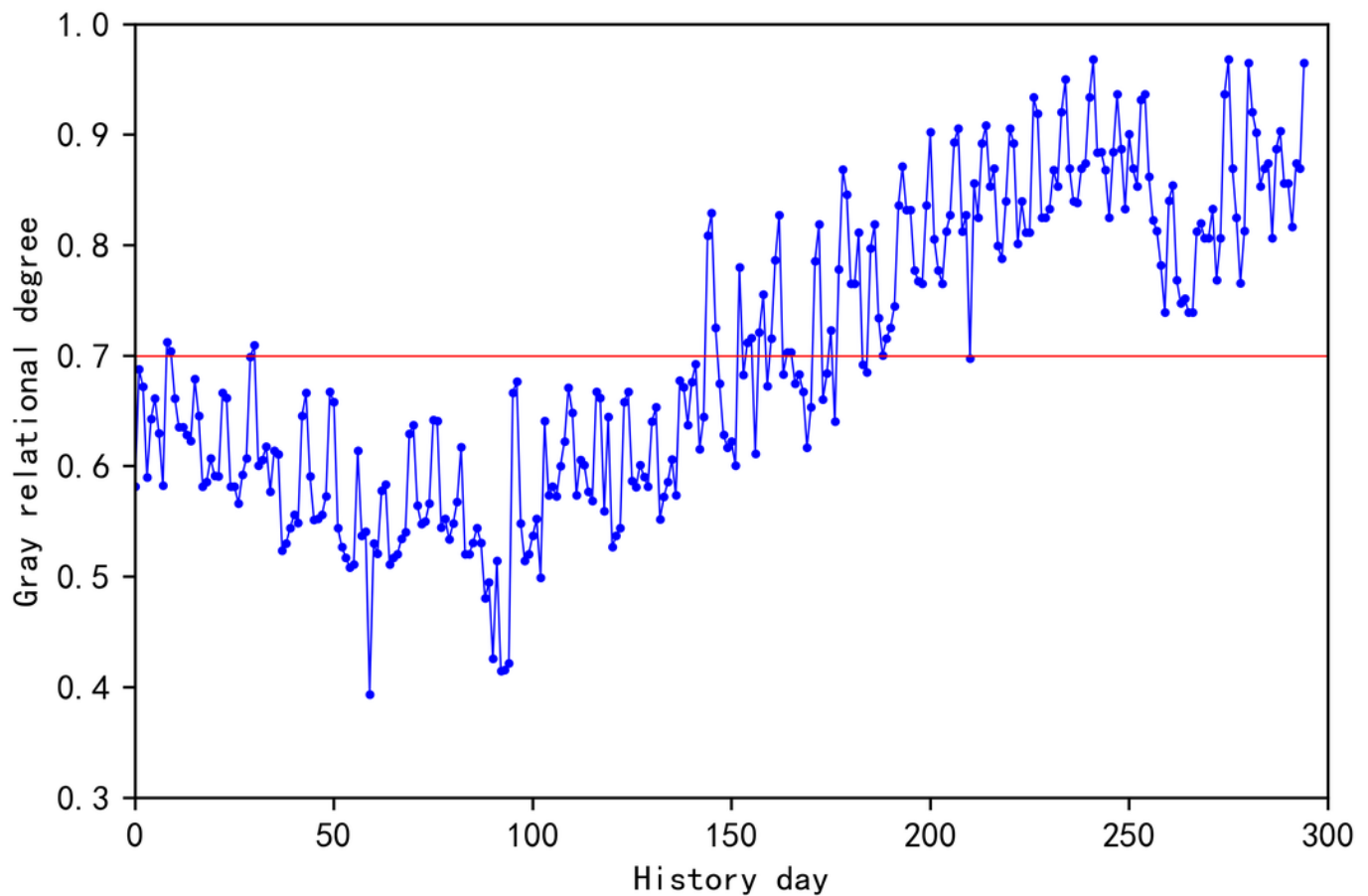



Figure 7

Relationship between the number of cluster centers and the silhouette coefficient

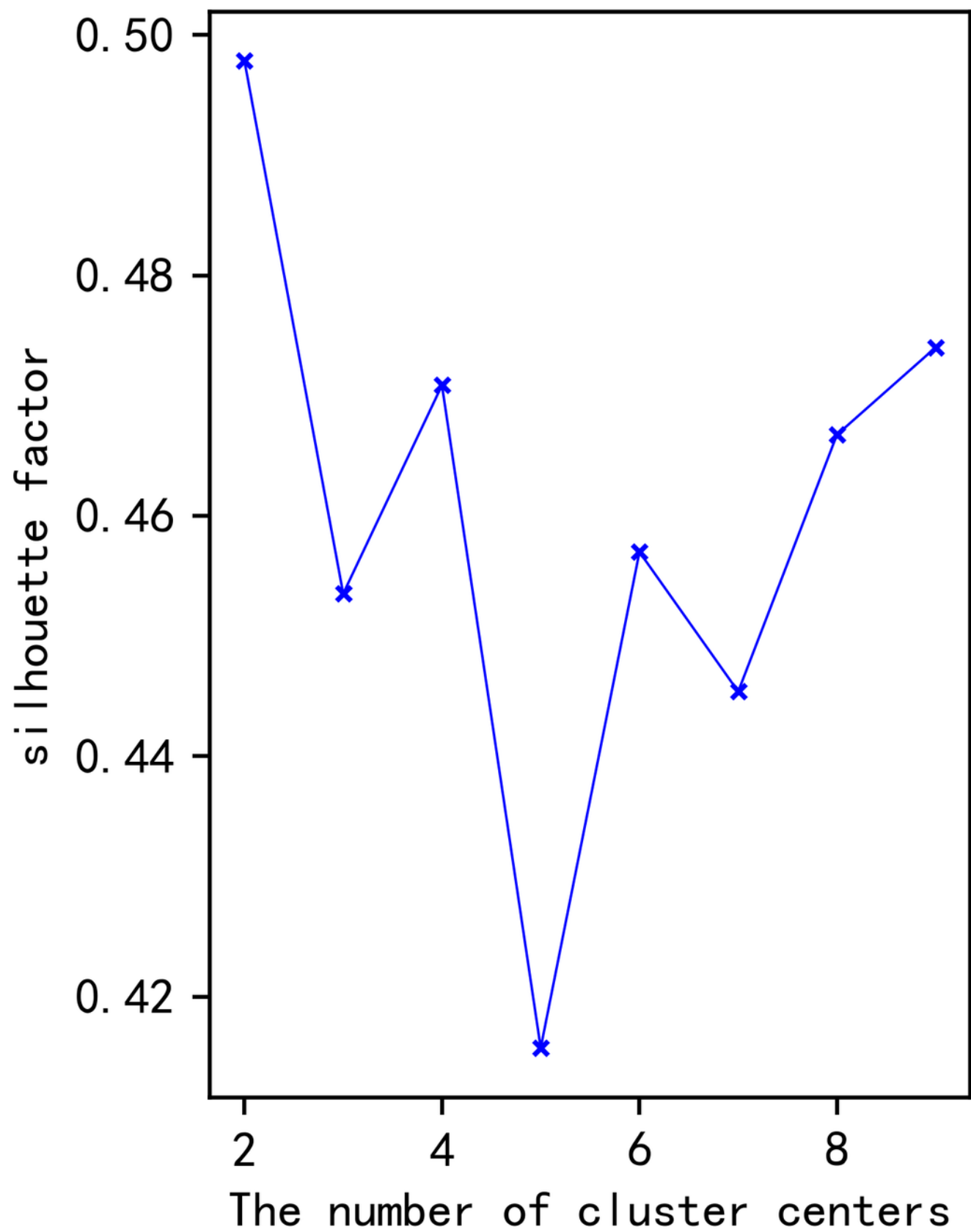



Figure 8

Clustering effect diagram of each historical day in the rough set of similar days

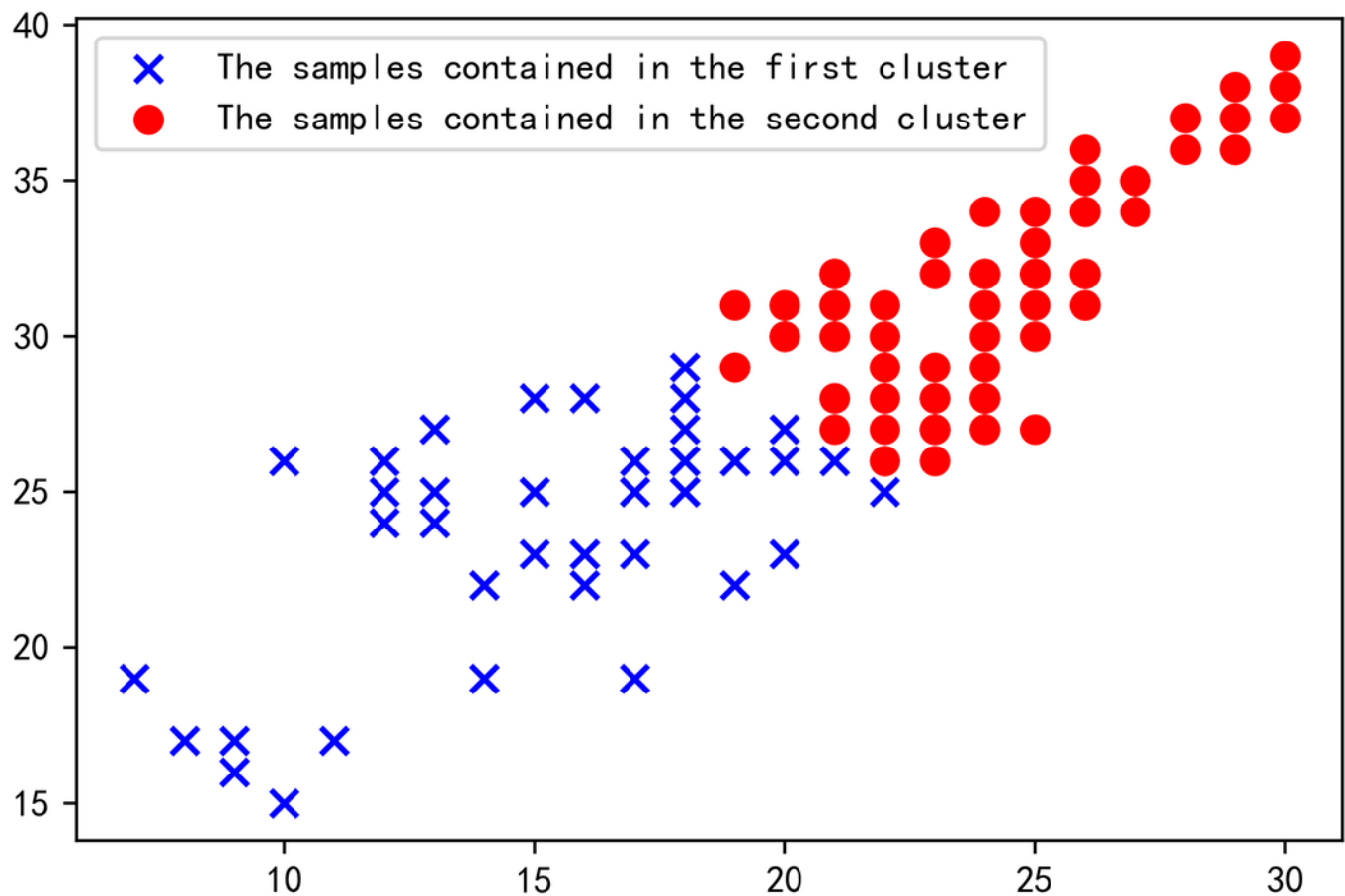



Figure 9

Predicted load error value at each moment for the three methods

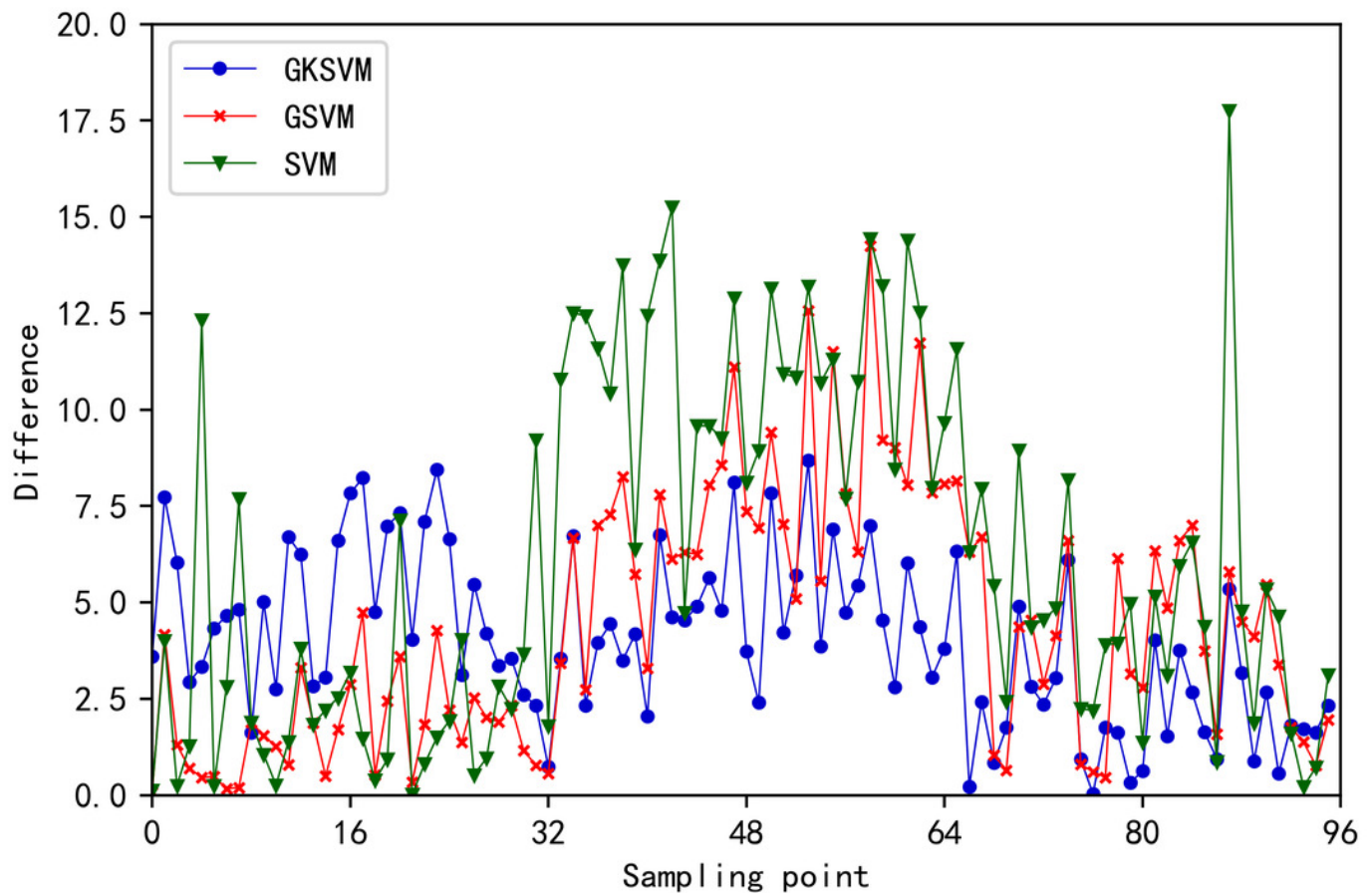



Figure 10

Comparison chart of predicted and actual values of 96 sampling points for the three methods

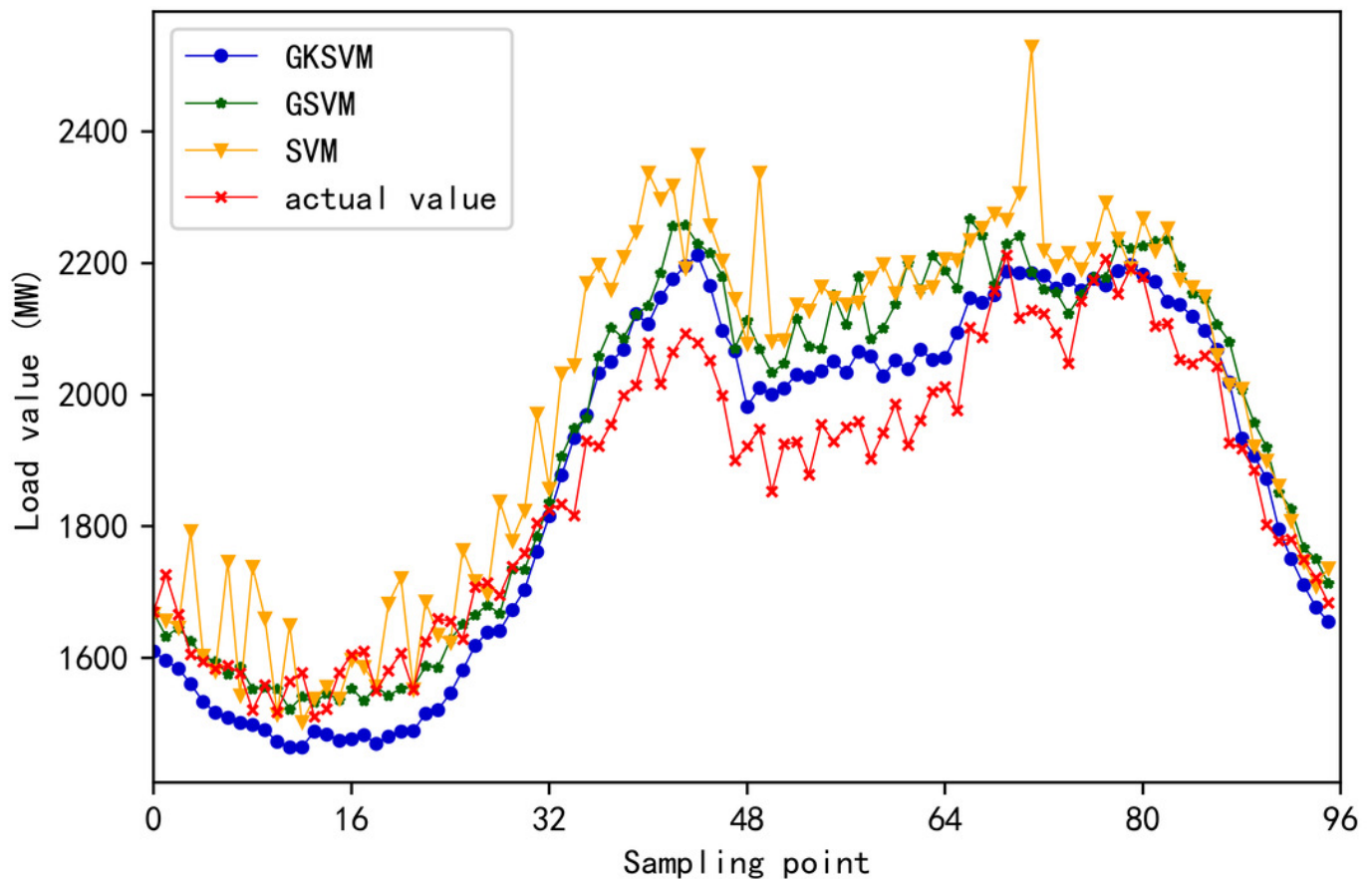



Figure 11

ABC, PSO and GWO iteration curves

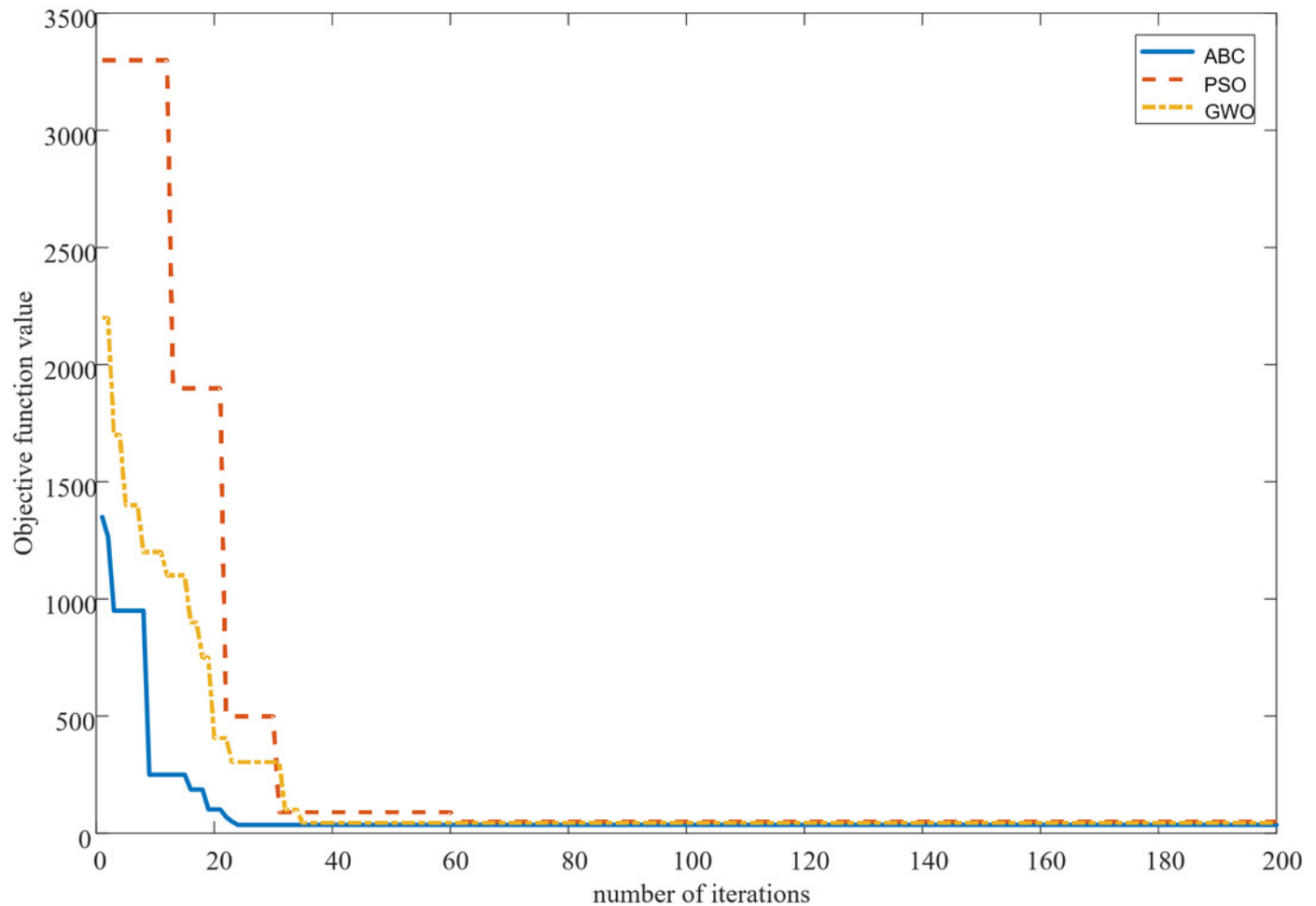



Figure 12

Multivariate ANOVA of ABC, PSO and GWO

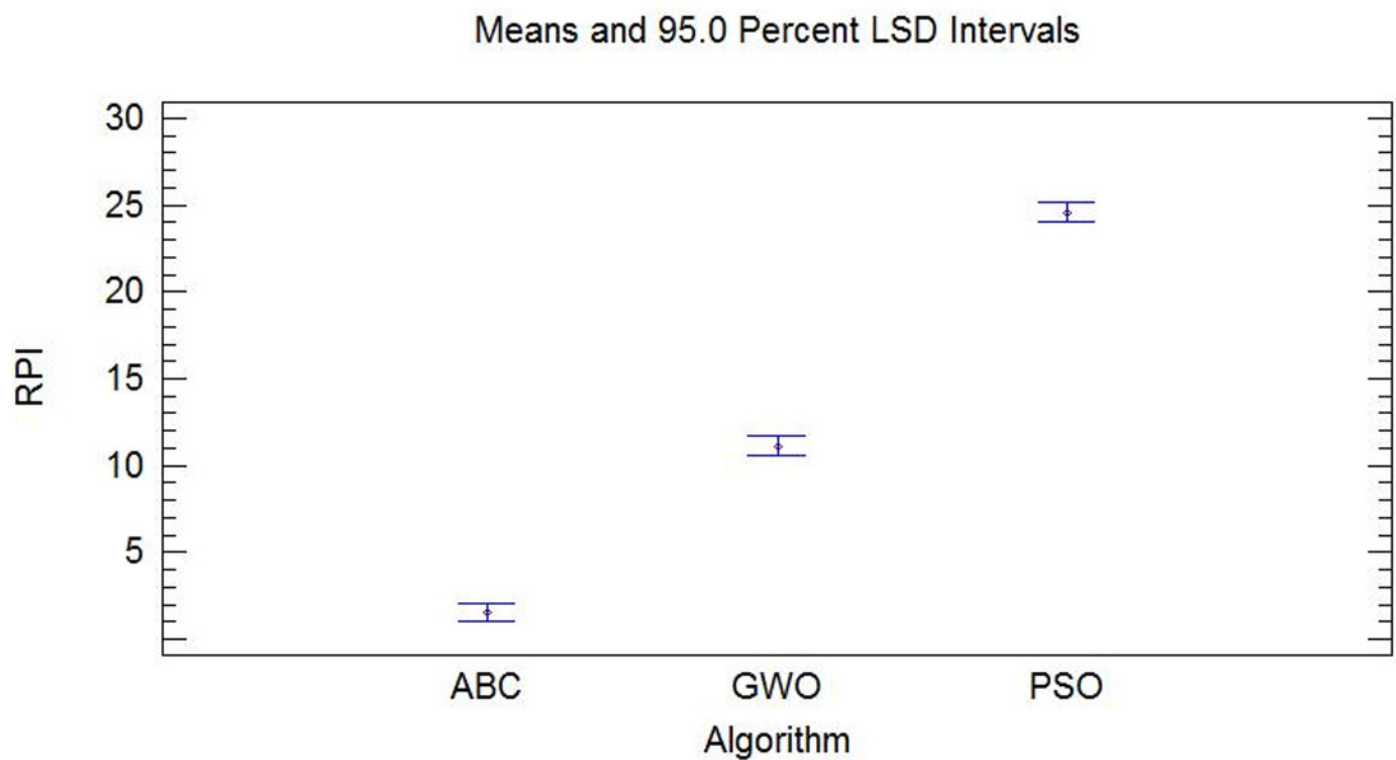



Figure 13

Comparison chart of error values at each moment of the five methods

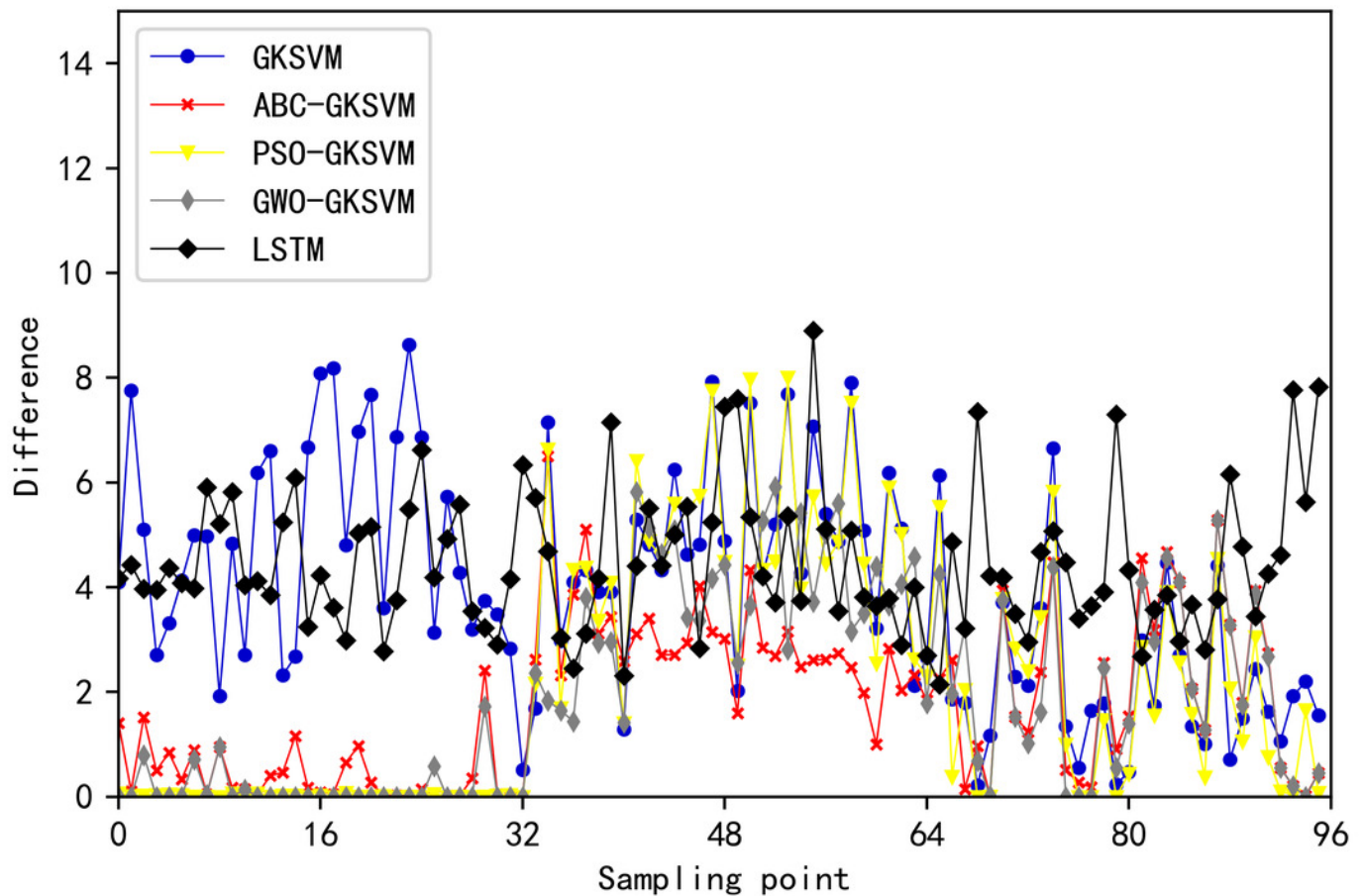



Figure 14

Comparison of predicted and actual values for each method

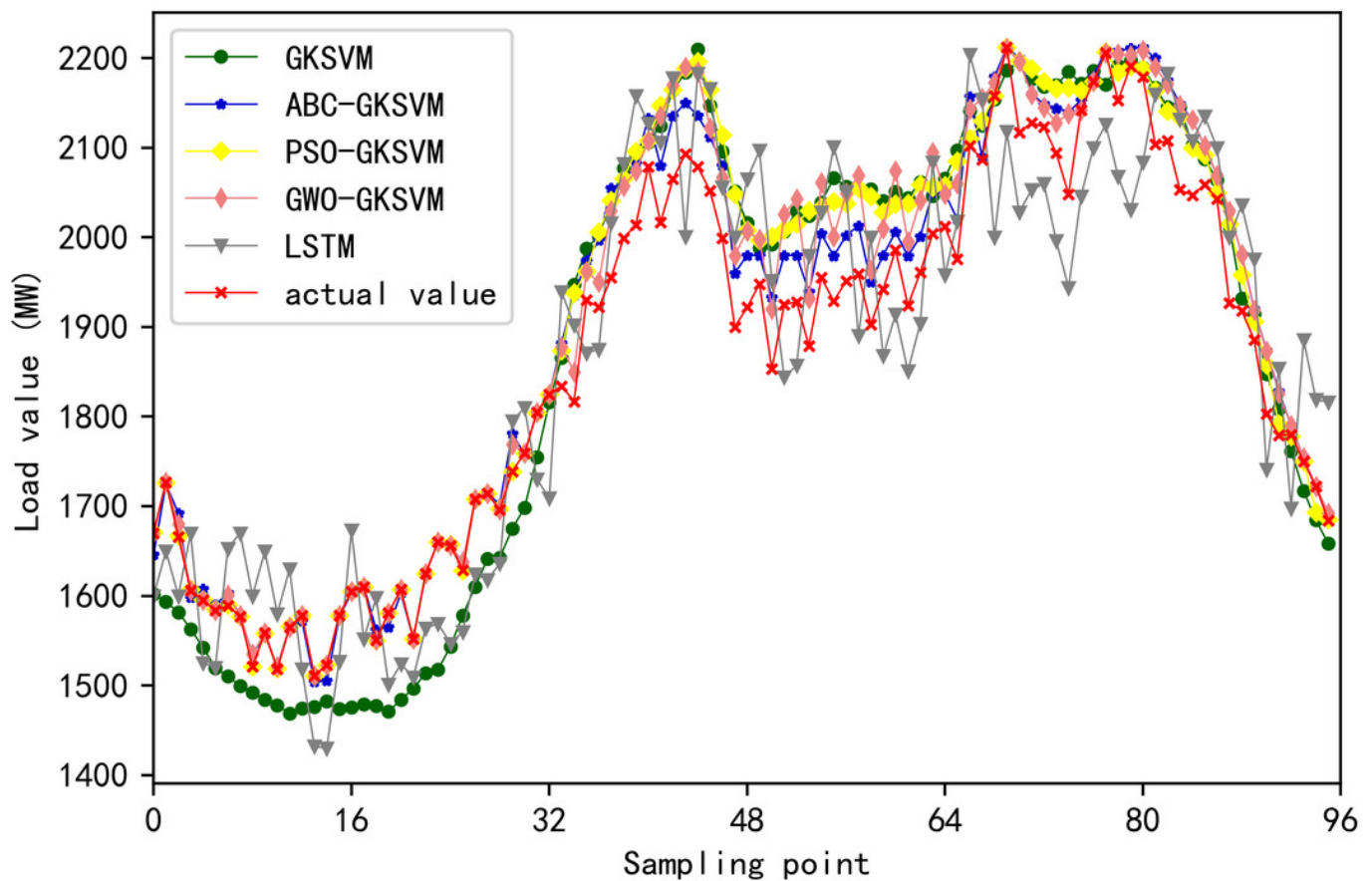
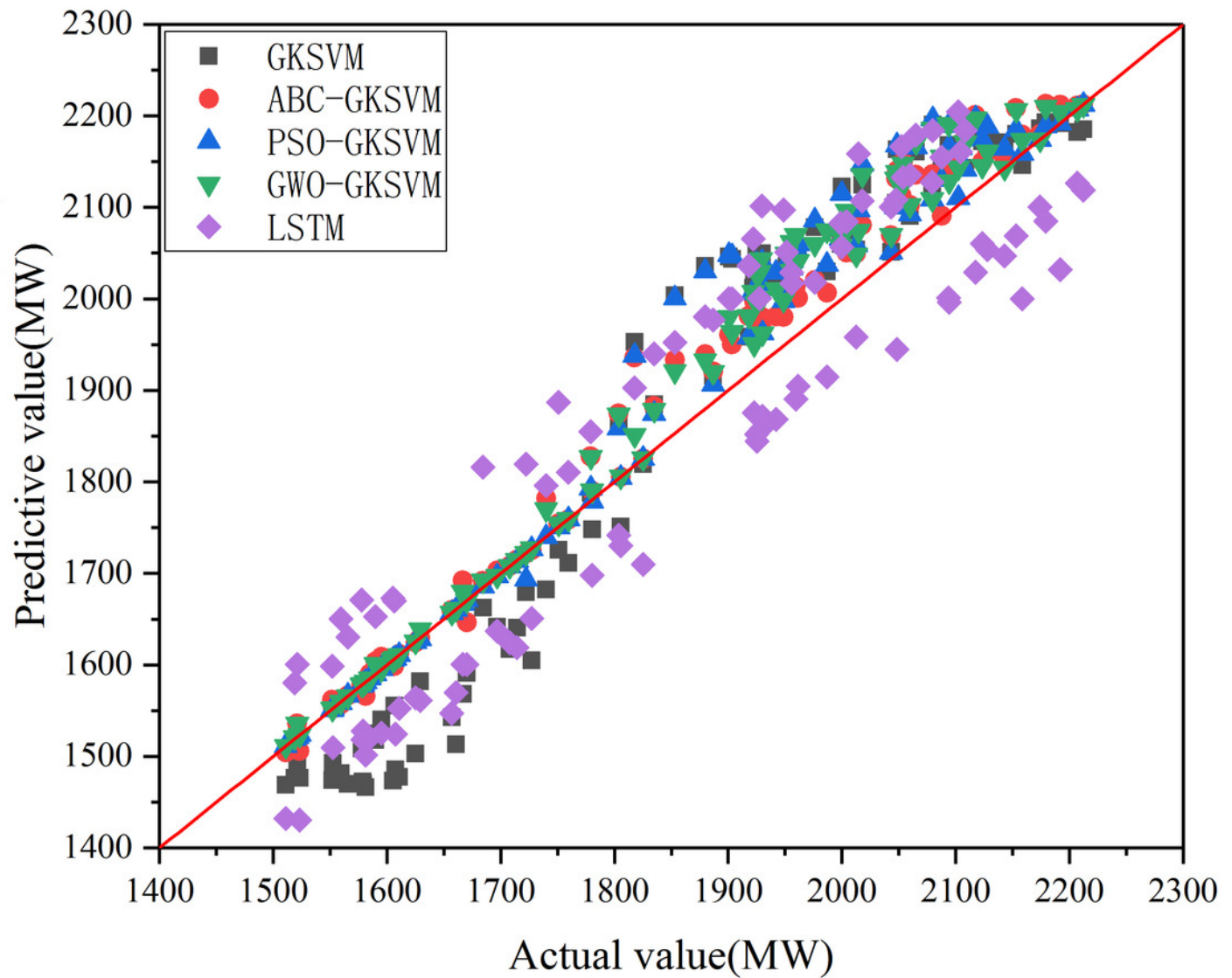



Figure 15

Scatter plot of predicted values versus actual values for the five methods

Correlation coefficient and correlation degree table

1   Table 1 Correlation coefficient and correlation degree table

Correlation coefficient interval	Relevance
$0.2 \leq \lvert \rho \rvert < 0.4$	Very weak correlation
$0.2 \leq \lvert \rho \rvert < 0.4$	Weak correlation
$0.4 \leq \lvert \rho \rvert < 0.6$	Moderately relevant
$0.6 \leq \lvert \rho \rvert < 0.8$	Strong correlation
$0.8 \leq \lvert \rho \rvert < 1.0$	Very strong correlation

Correspondence between honeybee collecting behavior and parameter

1   Table 2 Correspondence between honeybee collecting behavior and parameter optimization problem

Honeybee collection behavior	Specific optimization problems
Food source	Two-dimensional vector composed of SVM parameters C and g
Amount of nectar from food source	Fitness value in the problem
Process of finding and gathering food sources	SVM prediction process
Food source for the maximum amount of nectar	Optimal parameters C and g of SVM

Weather condition, wind direction, and date type quantitative code

1   Table 3 Weather condition, wind direction, and date type quantitative code

Influencing factors	Coding	Influencing factors	Coding
Holiday	1 if the day is a holiday and 0 if otherwise	West wind	1.75
Weekend	0	Northwest wind	1.875
Workday	1	Sunny	2
North wind	1	Partly cloudy	1
Northeasterly wind	1.125	Cloudy day	0
East wind	1.25	Light rain or snow	-1
Southeast wind	1.375	Moderate rain or moderate snow	-2
South wind	1.5	Heavy rain or snow	-3
Southwesterly wind	1.625	Heavy rain or blizzard	-4

Pearson correlation coefficient of influencing factors

1 Table 4 Pearson correlation coefficient of influencing factors

Influencing factors	Coefficients	Influencing factors	Coefficients
Wind direction	0.131	Is the day a workday?	0.537
Minimum temperature	0.786	Is the day a holiday?	−0.584
Maximum temperature	0.756	Weather conditions	0.026

Euclidean distance between the day to be predicted and the cluster center

1 Table 5   Euclidean distance between the day to be predicted and the cluster center

Category	Category 1	Category 2
Euclidean distance	2.85	10.78

Table 6 MAE, MAPE, and RMSE values for the three methods

1   Table 6 MAE, MAPE, and RMSE values for the three methods

Method of prediction	MAE(MW)	MAPE(%)	RMSE(MW)
GSVM	84.46	4.38	106.69
GKSVM	73.97	4.01	83.52
SVM	119.25	6.12	152.47

Optimal parameters C and g of the support vector machine at each moment

1   Table 7 Optimal parameters C and g of the support vector machine at each moment

Point in time	C	g	Point in time	C	g	Point in time	C	g	Point in time	C	g
1	0.01	0.01	25	2.05	0.01	49	0.43	0.01	73	1.71	0.01
2	2.74	0.01	26	0.16	0.01	50	0.01	0.42	74	0.74	0.01
3	0.01	0.24	27	25	0.01	51	0.27	0.01	75	0.42	0.01
4	0.01	0.01	28	6.69	0.01	52	0.24	0.01	76	1.04	0.01
5	0.01	0.25	29	2.86	0.01	53	0.68	0.01	77	0.01	0.25
6	0.13	0.01	30	0.01	37.04	54	1.77	0.01	78	0.01	16.9
7	0.01	50	31	3.7	0.01	55	0.77	0.01	79	0.01	0.01
8	0.23	0.01	32	0.01	46.23	56	0.01	0.16	80	0.01	0.01
9	0.01	0.01	33	0.03	0.01	57	0.01	0.33	81	46.37	25.86
10	0.01	0.56	34	0.9	0.01	58	0.1	0.01	82	0.12	0.01
11	0.01	0.01	35	4.18	0.01	59	0.37	0.01	83	0.01	0.01
12	0.01	12.29	36	1.81	0.01	60	0.01	0.87	84	0.01	0.01
13	0.01	15.43	37	1.49	0.01	61	0.01	0.64	85	0.01	0.01
14	0.01	0.1	38	9.55	0.01	62	0.48	0.01	86	0.01	0.01
15	0.04	0.01	39	1.19	0.01	63	0.01	0.34	87	2.05	0.01
16	0.01	9.66	40	0.77	0.01	64	0.05	0.01	88	0.01	0.01
17	3.02	0.01	41	1.25	0.01	65	0.01	0.45	89	0.01	0.01
18	2.86	0.01	42	0.31	0.01	66	0.01	0.34	90	0.01	0.01
19	0.01	0.57	43	0.01	0.33	67	0.01	0.01	91	0.01	0.03
20	0.01	22.33	44	0.27	0.01	68	0.01	0.07	92	0.01	0.01
21	4.2	0.01	45	2.37	0.01	69	0.14	0.01	93	0.01	0.01
22	0.01	50	46	0.28	0.01	70	0.01	0.01	94	0.01	0.01
23	8.21	0.01	47	0.61	0.01	71	0.21	0.01	95	0.01	0.01
24	14.88	0.01	48	0.01	32.57	72	0.56	0.01	96	0.01	26.21

RPI values of ABC, PSO and GWO

1 Table 8 RPI values of ABC, PSO and GWO

Number of runs	ABC	PSO	GWO	Number of runs	ABC	PSO	GWO
1	1.772	25.221	12.292	11	0.332	27.575	12.569
2	0.249	24.197	9.275	12	2.187	20.903	11.157
3	3.848	23.976	10.382	13	1.467	22.841	9.192
4	1.301	28.820	11.822	14	1.800	23.007	8.195
5	0.554	26.495	10.410	15	0.305	28.239	12.209
6	4.845	22.149	13.372	16	2.298	27.658	12.514
7	3.3225	22.34	12.348	17	0.581	23.062	10.354
8	0.471	24.391	12.431	18	0.720	23.643	10.105
9	1.080	24.668	8.306	19	0	21.733	11.988
10	2.132	25.831	11.101	20	1.772	24.972	12.652

MAE, MAPE, and RMSE values for the five methods

1   Table 9 MAE, MAPE, and RMSE values for the five methods

Method of prediction	MAE(MW)	MAPE(%)	RMSE(MW)
GKSVM	73.23	4.01	83.20
PSO-GKSVM	42.27	2.15	62.84
GWO-GKSVM	38.79	1.96	54.32
LSTM	83.42	4.48	87.58
ABC-GKSVM	34.98	1.79	46.38