

Lu, Hongfang; Azimi, Mohammadamin; Iseley, Tom

Article

Short-term load forecasting of urban gas using a hybrid model based on improved fruit fly optimization algorithm and support vector machine

Energy Reports

Provided in Cooperation with:

Elsevier

Suggested Citation: Lu, Hongfang; Azimi, Mohammadamin; Iseley, Tom (2019) : Short-term load forecasting of urban gas using a hybrid model based on improved fruit fly optimization algorithm and support vector machine, Energy Reports, ISSN 2352-4847, Elsevier, Amsterdam, Vol. 5, pp. 666-677,
<https://doi.org/10.1016/j.egyr.2019.06.003>

This Version is available at:

<https://hdl.handle.net/10419/243620>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

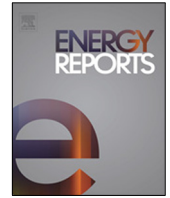
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>



Research paper

Short-term load forecasting of urban gas using a hybrid model based on improved fruit fly optimization algorithm and support vector machine

Hongfang Lu^{*}, Mohammadamin Azimi, Tom Iseley

Trenchless Technology Center, Louisiana Tech University, 599 Dan Reneau Dr., Engineering Annex, Ruston, LA 71270, United States

HIGHLIGHTS

- A hybrid algorithm CF-SA-FFOA-SVM is used to forecast the short-term urban gas load.
- The influence of temperature types in forecasting models on forecasting results is discussed.
- The applicability of the algorithm for forecasting daily, weekly and monthly gas load is analyzed.
- The effect of data grouping-based forecasting method is analyzed.

ARTICLE INFO

Article history:

Received 6 February 2019

Received in revised form 7 May 2019

Accepted 3 June 2019

Available online xxxx

Keywords:

Short-term

Gas load

Forecasting

Urban gas

Fruit fly optimization algorithm

Support vector machine

ABSTRACT

The accurate forecasting of short-term load for urban gas is the premise of gas supply sales, pipe network planning, and energy optimization scheduling. This paper adopts a hybrid model that integrates the fruit fly optimization algorithm (FFOA), simulated annealing algorithm (SA), cross factor (CF) and support vector machine (SVM) to forecast the short-term gas load of urban gas. In the model, SA and CF are used to optimize the FFOA algorithm. This paper takes the urban gas system in Kunming, China as a case study and uses the CF-SA-FFOA-SVM algorithm to predict the gas consumption and compares the results with the other four forecasting methods such as back-propagation neural network (BPNN) and autoregressive integrated moving average model (ARIMA). Besides, this paper analyzes the influence of temperature types (daily maximum temperature and daily mean temperature) in models on forecasting results, and the applicability of the algorithm for forecasting weekly and monthly gas load is analyzed. Moreover, the impact of grouping raw data by the feature on forecasting result is discussed. The following conclusions are drawn: (1) compared with other forecasting models, CF-SA-FFOA-SVM model has higher gas load forecasting accuracy. (2) for Kunming city, if the daily maximum temperature is used as the input variable in the gas load forecasting model, the forecasting accuracy is higher. (3) grouping raw data according to holiday attributes or gas types can effectively improve the accuracy of load forecasting.

© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

As a clean and efficient energy source, natural gas is currently accounting for an increasing proportion of energy structure. According to the 67th edition of the statistical review released by British Petroleum (BP) in June 2018, the global market share of oil has declined slightly, while natural gas accounts for 23.4% of primary energy consumption (British Petroleum, 2018). In the natural gas gathering, transportation and distribution system, the production of gas, the planning of the pipe network and the sale of gas all need to grasp the gas demand accurately. Especially for the urban gas distribution system, more management problems

are involved, such as gas peak-shaving, resource allocation, supply chain management (Lu et al., 2019), and gas storage facilities management, which require higher accuracy of gas load data (Bai, 2016). Therefore, the accurate forecasting of the gas load is of great significance to ensure the normal operation of the industrial chain, reduce economic losses, and ensure the balance between supply and demand.

Short-term gas load refers to the amount of natural gas consumed hourly, daily, weekly or monthly. According to the existing load forecasting methods, it can be roughly divided into two categories: traditional forecasting methods represented by time series method and intelligence forecasting methods represented by artificial neural network (ANN) (Suganthi and Samuel, 2012; Gorucu, 2004; Wang et al., 2018). Table 1 summarizes the methods included in the two classifications.

^{*} Corresponding author.

E-mail address: luhongfang_sci@126.com (H. Lu).

Abbreviations

CF	cross factor
SA	simulated annealing algorithm
FFOA	fruit fly optimization algorithm
SVM	support vector machine
BPNN	back-propagation neural network
PSO	particle swarm optimization
GM(1,1)	gray model with first order and one variable
ARIMA	autoregressive integrated moving average model
BP	British Petroleum
NG	natural gas
ANN	artificial neural network
SARIMA	seasonal ARIMA
ANFIS	adaptive network-based fuzzy inference system
LS-SVM	least squares support vector machine
WLS	weighted least squares
MLP	multilayer perceptron
RBF	radial basis function
OLS	ordinary least squares regression
GA	genetic algorithm
SVRLP	support vector regression local predictor
FNf	false neighbors filtered
ARMA	autoregressive moving average model
KNN	k-nearest neighbors
WT	wavelet transform
FFNN	feed-forward neural network
BAG	bat algorithm with Gaussian perturbations
MGM	rolling metabolic gray model
NMGM	non-linear metabolic gray model
MAPE	mean absolute percentage error
RMSE	root mean square error
MSE	mean squared error
FTL	forecasting time length

Smith et al. (1996) used the expert system method to predict the short-term load of gas. The results show that the expert system method has a high prediction accuracy compared with other traditional methods because the number of days with an error greater than 5% is the least. Gorucu (2004) used the ANN model to forecast the gas demand of Ankara (Turkey), the model considered several factors such as temperature, sales price, number of customers, and exchange rate. The coefficient of determination of non-seasonal gas consumption forecast results reached 0.9833. Pereira et al. (2006) used the SVM method to predict the hourly electrical energy consumption of the Brazilian electric utility located in the Northeast region in Brazil, and compared it with the ANN method, and found that the SVM method has higher prediction accuracy. Also, they analyzed the impact of the kernel function in the SVM model on the prediction results and found that Hyperbolic Tangent and Linear kernels are the most stable. Ediger and Akar (2007) used the ARIMA and seasonal ARIMA (SARIMA) method to forecast Turkey's primary energy demand for 2005–2020. The results show that the ARIMA algorithm's forecast result of total primary energy demand is more accurate than the sum of the individual predictions. Kumar and Jain (2010)

used three kinds of time series models to predict India's conventional energy consumption, and the results were compared with Indian planning agencies, indicating that the time series method has great potential in energy forecasting. Azadeh et al. (2010) presented an adaptive network-based fuzzy inference system (ANFIS) for short-term gas load forecasting. This method considers new indicators on the basis of the original time series method. The forecasting results show that the forecasting accuracy of this model is higher than that of the ANN and traditional time series method. Zhang et al. (2011) used SVM to predict the daily gas load and analyzed the impact of normalization methods and data grouping methods on prediction. The results show that MAPE is less than 5%. Wang (2012) combined particle swarm optimization with weighted LS-SVM to predict gas load. It is found that the prediction accuracy of PSO-WLS-SVM is improved compared with the WLS-SVM method because the average MAPE is reduced from 3.22% to 3.09%. Taşpınar et al. (2013) used four models to predict the daily gas load in Turkey, including SARIMAX model, ANN-MLP model, ANN-RBF model, and multivariate OLS model. The input variables include air temperature, daily residential gas consumption, average cloud cover, atmospheric pressure, relative humidity, and wind speed. The calculation results show that the SARIMAX model is the best for short-term load forecasting of gas in Sakarya province (Turkey). Uzlu et al. (2014) combined the teaching-learning-based optimization algorithm with the ANN to predict Turkey's energy consumption. They compared the predicted results of the model with BPNN, indicating that the new model has a better prediction effect. Yu and Xu (2014) developed a combinational model to predict the daily gas load, which integrates the advantages of improved BPNN and optimized GA algorithm to avoid partial dinky in the prediction process. They used this model to forecast the daily gas load in Shanghai (China), and the results show that the algorithm can get higher prediction accuracy with fewer iterations. Zhu et al. (2015) presented a hybrid model to predict short-term gas demand. They combined the SVRLP with the FNf method, which optimizes the data of the training samples to obtain more accurate predictions. The case study shows that the prediction accuracy of this new model is higher than the ARMA method, the ANN method, and SVRLP method. Shaikh and Ji (2016) used the logistic modeling method to predict the gas demand of China in 2020 and 2035, and used the Levenberg–Marquardt Algorithm to improve the prediction accuracy. The results show that the logistic and logistic-population modeling methods are effective in long-term gas load forecasting. Bai (2016) predicted the daily gas load of the Sichuan–Chongqing region (China) in winter. First, the data was preprocessed by wavelet denoising, and time series method and SVM were used to forecast the gas in different periods. The results show that the SVM method is better during major holidays, and the frequency-division combination forecasting method is better during non-holidays. Yu (2015) used the k-nearest neighbors (KNN) algorithm to preprocess the gas load data and combined the improved particle swarm algorithm with BPNN and SVM to predict the gas load. Barak and Sadegh (2016) combined the ARIMA method with the ANFIS method to predict the annual energy consumption of Iran. The results show that the combined algorithm has higher prediction accuracy than the single algorithm because the model's MSE criterion decreased from 0.058% to 0.026%. Karadede et al. (2017) proposed a breeder hybrid algorithm that combines a non-learning-based breeder GA with SA. The calculation results show that the MAPE of the predicted value is less than 2%. Panapakidis and Dagoumas (2017) proposed a model that combines WT, GA, ANFIS, and FFNN. Among them, WT is used to decompose the original signals in a set of sub-series, and then use GA-optimized ANFIS to provide predictions for each sub-series. FFNN is used to improve initial predictions and improve overall prediction accuracy.

The results show that the model is suitable for all distribution points that make up a national natural gas pipeline network. Wu and Peng (2017) developed a hybrid model BAG-SA to predict China's energy demand, which combines BA, Gaussian perturbations and SA to improve global search capability and convergence speed. The results show that this hybrid algorithm has higher precision than other single algorithms and has a better fitting ability. Ding (2018) proposed a new self-adaptive gray forecasting model to predict China's future natural gas demand. In this new model, the new initial value has the advantage of adjustable weighting coefficients for each component of the accumulated sequence. The results show that China's demand for natural gas will reach 340 billion cubic meters by 2020, which is consistent with the forecast of other professional institutions. Wang et al. (2018) developed single-linear (MGM model), hybrid-linear (MGM-ARIMA), and non-linear (NMGM) prediction techniques based on gray theory to predict energy demand in China and India. The results show that the three methods have higher prediction accuracy, the MAPE ranges for these three methods are: 1.30–3.08%, 0.80–2.57%, and 2.06–2.19%, respectively. Shu et al. (2018) used the least squares support vector machine (LS-SVM) method to predict the consumption of natural gas in Chengdu, China. The model considered meteorological factors, policy factors, and date factors. They compared the results with ARIMA, BPNN and other methods, and the results show that the LS-SVM method has higher prediction accuracy.

It can be seen from the literature review that the load forecasting method is moving toward a combination of multiple algorithms and intelligent algorithms because a single algorithm has drawbacks. On the other hand, most of the literature is to study the algorithm, and this paper also analyzes the matters other than the algorithm, such as the selection of temperature parameters in the SVM model, and the improvement of gas load forecasting ideas through data grouping.

The rest of the paper is organized as follows: Section 2 introduces the basic theory of natural gas load forecasting, including support vector machine model, improved fruit fly optimization algorithm, forecasting steps, and error evaluation. Section 3 introduces the data sources and parameter settings in the prediction model. Section 4 shows the prediction results of the model used in this paper, as well as the prediction results of the other four reference models. Section 5 discusses the prediction model and method, and considers three factors affecting the prediction accuracy of the model.

2. Methodology

2.1. Support vector machine (SVM)

The support vector machine is a machine learning method widely used in statistical classification and regression analysis. It has a solid theoretical foundation and good generalization performance and is often used to solve nonlinear problems with small amount of data (Vapnik, 2013).

In regression analysis, assume that the training set is:

$$\{(x_i, y_i) | x_i \in R^n, y_i \in R, i = 1, 2, \dots, m\} \quad (1)$$

where x_i is the input vector; y_i is the output vector.

The optimal linear decision function constructed in high-dimensional space is:

$$f(x) = \omega^T \varphi(x) + b \quad (2)$$

where $\varphi(x)$ is nonlinear mapping function; ω is weighted vector; b is deviator.

Based on the principle of structural risk minimization (SRM), introducing the slack variables ξ_i, ξ_i^* , then the linear regression function can be expressed as:

$$\min \left[0.5 \|\omega\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \right] \quad (3)$$

$$\text{s.t.} \begin{cases} y_i - \omega^T \varphi(x_i) - b \leq \varepsilon + \xi_i \\ \omega^T \varphi(x_i) - y_i + b \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (4)$$

where C is penalty factor; ε is insensitive loss function.

The Lagrange function is introduced and converted into dual form:

$$\max \left[-0.5 \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) + \sum_{i=1}^n \alpha_i (y_i - \varepsilon) - \sum_{i=1}^n \alpha_i^* (y_i - \varepsilon) \right] \quad (5)$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^n \alpha_i = \sum_{i=1}^n \alpha_i^* \\ 0 \leq \alpha_i \leq C, 0 \leq \alpha_i^* \leq C \end{cases} \quad (6)$$

where $K(x_i, x_j)$ is kernel function.

Assume that the optimal solutions obtained by Eq. (5) are α_i and α_i^* , then the regression function can be expressed as:

$$f(x) = \sum_{i=1}^n (\alpha_i^* - \alpha_i) K(x_i, x_j) + b \quad (7)$$

The selection of kernel functions is a key issue of the SVM method, and different kernel functions can lead to different generalization and learning ability of prediction models. Three kinds of kernel functions that are used more: q polynomial function, radial basis function (RBF), and Sigmoid function. Because the characteristic space corresponding to the RBF is infinite, it is sure that the sample can be linearly separable under the condition of the finite sample, so the RBF is used in this paper (Ji et al., 2014):

$$K(x_i, x_j) = e^{\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)} \quad (8)$$

where σ is the Gaussian kernel.

2.2. CF-SA-FFOA algorithm

The FFOA was first proposed by Pan (2012). It is an algorithm that seeks global optimization based on the performance of fruit fly foraging behavior. The FFOA has a simple optimization mechanism, only uses fewer input parameters, has information sharing characteristics, and has strong robustness and adaptability. Although this algorithm has only appeared for 6 years, it has been widely concerned and applied in various fields. Its flow chart is shown in Fig. 1.

However, due to the influence of random oscillating, the FFOA algorithm spends much time on the global optimal, making it slow in convergence and low in convergence precision in dealing with complex problems, and will lose the diversity of the population. Song (2015) proposed an algorithm that combines SA, cross factor, FFOA, and SVM, namely CF-SA-FFOA-SVM algorithm to solve this problem. Among them, the SA algorithm can avoid falling into the optimal local solution, and the cross factor

Table 1
Load forecasting technology classification and applicable occasions.

Classification	Technique	Reference(s)
Traditional method	Exponential smoothing	Akkurt et al. (2010), Akpınar and Yumusak (2016)
	Regression	Herbert et al. (1987)
	ARIMA	Erdogdu (2010), Melikoglu (2013)
	Gray	Zeng and Li (2016), Ma et al. (2019)
Intelligence method	Fuzzy logic	Azadeh et al. (2010)
	ANN	Szoplik (2015), Yu and Xu (2014)
	SVM	Zhu et al. (2015), Iranmanesh et al. (2011)
	Expert system	Smith et al. (1996)

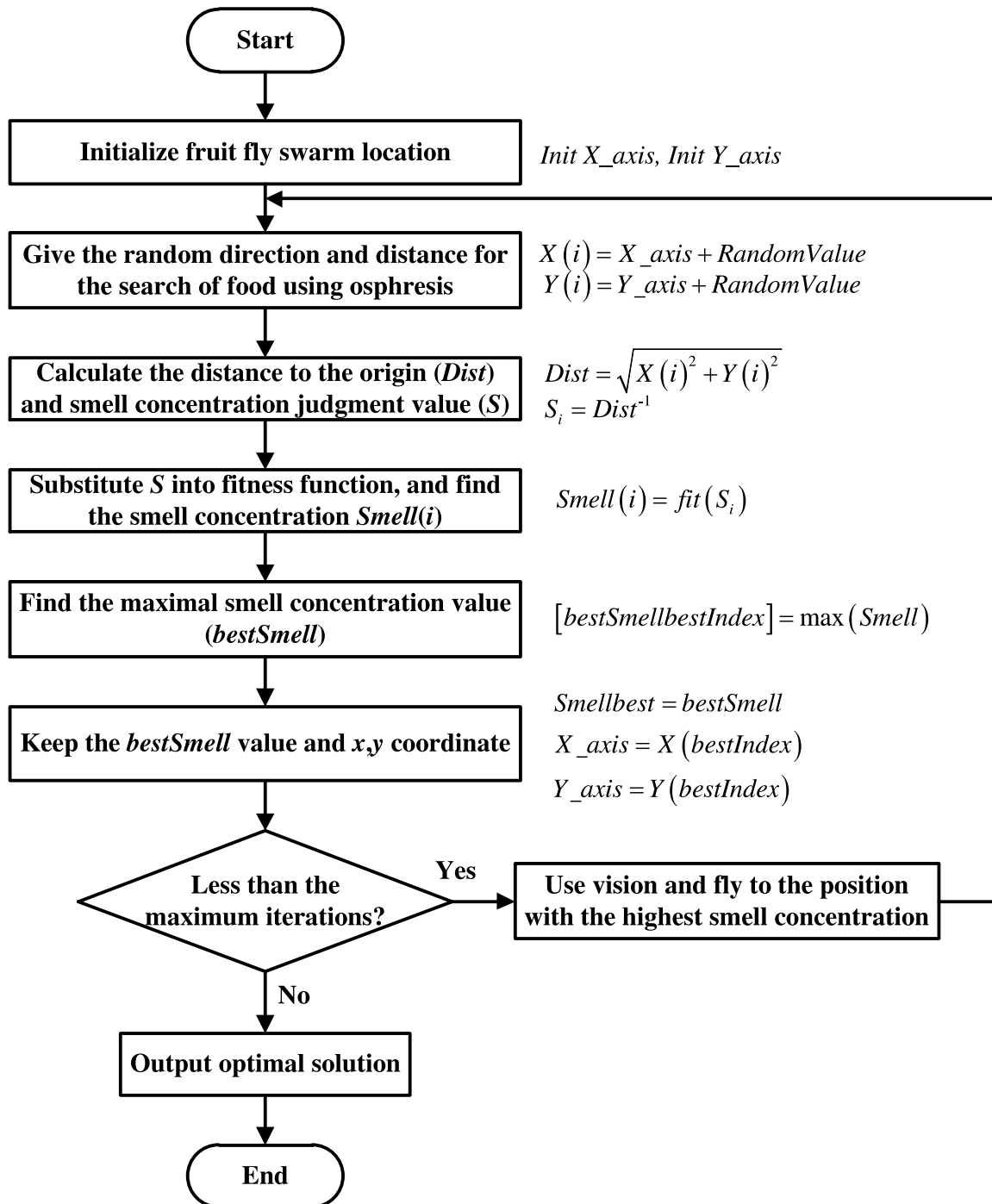


Fig. 1. Flow chart of FFOA (Vapnik, 2013).

significantly affects the convergence of the algorithm. The new population produced by cross factor will be able to adapt to the changes of environment and find the optimal global solution just like the offspring produced by natural evolution. Therefore, the steps of the CF-SA-FFOA algorithm are:

Step 1: Initialize the parameters: The maximum number of iterative evolutions is Gen_max , and the population size is set to Pop_size . The position of the population of the initial randomly distributed fruit fly is X_axis , Y_axis .

Step 2: Assigning each fruit fly to any direction ($RandomValue$) that seeks food through the sense of smell.

$$\begin{aligned} X(i) &= X_axis + RandomValue \\ Y(i) &= Y_axis + RandomValue \end{aligned} \quad (9)$$

where $RandomValue$ is a random number in the interval $[0,1]$.

Step 3: At first, the position of the food source cannot be accurately determined. Therefore, it is necessary to first estimate the starting distance $Dist$ between the individual and the initial set point, and then bring it into the calculation value S_i for calculating the concentration, which is the reciprocal of $Dist$:

$$Dist = \sqrt{X(i)^2 + Y(i)^2} \quad (10)$$

$$S_i = \frac{1}{Dist} \quad (11)$$

Step 4: By substituting S_i into the evaluation function of the taste concentration (also called the fitness function), the concentration value $Smell(i)$ above each fruit fly position can be calculated:

$$Smell(i) = fit(S_i) \quad (12)$$

Step 5: The optimal taste concentration judgment value $bestSmell$ is memorized and stored in its X , Y coordinate position.

Step 6: Cross-operations to create a new generation of populations:

$$\begin{aligned} child_1(x) &= a \times parent_1(x) + (1.0 - a) \times parent_2(x) \\ child_1(y) &= a \times parent_1(y) + (1.0 - a) \times parent_2(y) \\ child_2(x) &= a \times parent_2(x) + (1.0 - a) \times parent_1(x) \\ child_2(y) &= a \times parent_2(y) + (1.0 - a) \times parent_1(y) \end{aligned} \quad (13)$$

where a is a random weighting factor.

Step 7: The new individuals after the intersection are retained by the simulated annealing (SA) algorithm.

Step 8: The stored new individual position is placed in steps 2–5, and a value of the best taste concentration in the fruit fly population is calculated, and the concentration and position coordinates of the fruit fly are recorded. It is judged whether the value of the optimal taste concentration is better than the value after the last iteration.

Step 9: If the number of iterations is less than the set maximum number of iterations Gen_max or the target precision of the pair is not achieved, the steps 6–8 are performed again.

$$\begin{cases} Smellbest = bestSmell \\ X_b = X(bestindex) \\ Y_b = Y(bestindex) \end{cases} \quad (14)$$

The flow chart of the CF-SA-FFOA algorithm is shown in Fig. 2.

2.3. Forecasting process

The specific steps of using CF-SA-FFOA-SVM algorithm to forecast gas load are shown in Table 2 (Song, 2015):

Table 2

Forecasting process (Song and Pan, 2017).

Step	Explanation
1	Collect gas load data for a period.
2	Divide the data into training and test sets.
3	Use the CF-SA-FFOA algorithm to optimize the parameter C of the SVM model and the variance of the radial basis kernel function γ .
4	Train the data in the training set.
5	Forecast the data of the test set.
6	Output forecasting results and conduct error analysis.

2.4. Error evaluation

To evaluate a forecasting result, it is usually necessary to perform error analysis on the model. Currently, three error methods are commonly used as performance indicators: MAPE, RMSE, and MSE. The smaller their values, the higher the forecasting accuracy (Zhang et al., 2011; Shaikh and Ji, 2016).

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A - F}{A} \right| \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (A - F)^2} \quad (16)$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (A - F)^2 \quad (17)$$

where A represents actual value, F represents forecasted value.

Among them, MAPE is the most intuitive forecasting accuracy judgment criterion, and it can generally be considered that the forecasting accuracy is higher when the MAPE is less than 10% (Lewis, 1982).

3. Case study

3.1. Basic data

The case study is based on the gas data of Kunming, Yunnan, China, provided by PetroChina Kunlun Gas Ltd. From January 1, 2012 to November 30, 2013, a total of 700 sets of data are used as the research object. According to the FTL, the gas loads of the next 10 days to 30 days are forecasted. Their training set and test set are shown in Table 3. In addition, the daily temperature data of Kunming city was collected from the timeanddate.com website.

3.2. Parameter setting

When using SVM to forecast the gas load, the input vector has a great impact on the prediction model. When considering variables, on the one hand, essential factors should not be missed, on the other hand, the number of variables should be as small as possible, because when the number of variables is large, the complexity of the model increases, the prediction accuracy will be reduced. Therefore, this paper selects the daily mean temperature, weather, date, holiday, and historical value of related parameters as the variables of the input vector, and historical data considers the impact of the previous seven days (Gu et al., 2017; Soldo, 2012). Then the forecasting model can be expressed as:

$$L_{fd} = f(T_{dm}, W_d, D, H, HIST_{(T,W,D,H,L),(7d)}) \quad (18)$$

where L represents gas load; T represents temperature variable; W represents weather variable; D represents date variable, H

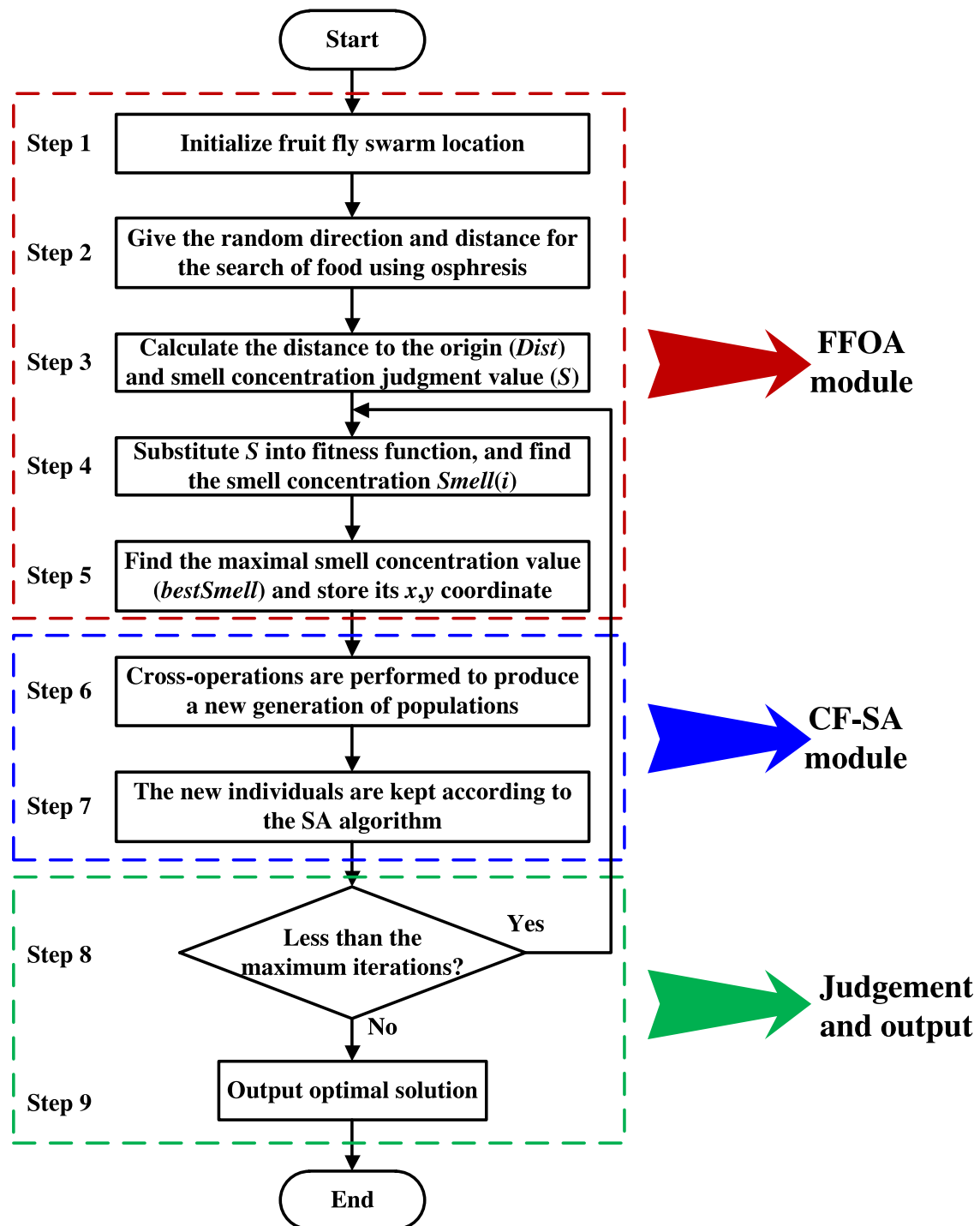


Fig. 2. Flow chart of CF-SA-FFOA (Song, 2015).

Table 3

Training sets and test sets with different forecasting time lengths.

FTL (Day)	Training set (Number of data)	Test set (Number of data)
10	Jan 1, 2012–Nov 20, 2013 (690)	Nov 21, 2013–Nov 30, 2013 (10)
20	Jan 1, 2012–Nov 10, 2013 (680)	Nov 11, 2013–Nov 30, 2013 (20)
30	Jan 1, 2012–Oct 31, 2013 (670)	Nov 1, 2013–Nov 30, 2013 (30)

represents holiday variable, $HIST$ represents historical variables; fd represents forecasted day value; dm represents daily mean value; d represents daily value; $7d$ represents seven days.

Eq. (18) is explained as follows: the gas load of the day needs to consider the daily mean temperature of the day, the weather of the day, the date attribute of the day, the holiday attribute of the

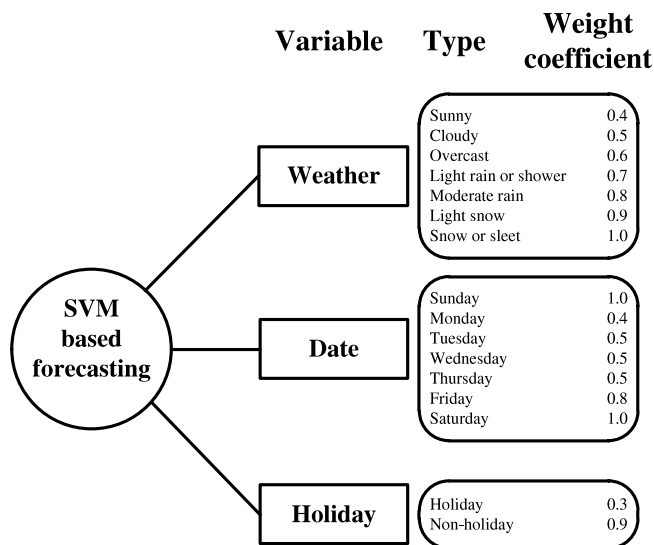


Fig. 3. Weight coefficient corresponding to different variables.

Table 4
The initialization parameters of CF-SA-FFOA-SVM.

Parameter	Value
Population	30
Maximum iterations	30
Optimization dimension	2
Fruit fly initialization position	[0,10]
Initialization flight direction and distance	[−10,10]

day, and attributes related to temperature, weather, date, holiday, and gas load for the previous 7 days.

According to the index quantization rule proposed in the literature (Zhang et al., 2011, 2014), the qualitative factors such as weather and date are converted into quantitative data, which are mapped to the interval [0,1], and the weight coefficients are shown in Fig. 3. The initialization parameters of CF-SA-FFOA-SVM are shown in Table 4.

4. Results

To evaluate the performance of the CF-SA-FFOA-SVM method, comparison with PSO-SVM, BPNN, GM (1,1) and ARIMA are conducted. The comparative experiments are divided into three groups, corresponding to three different FTLs (according to Table 2). The error analysis results are shown in Fig. 4. Since there is only one square operation between RMSE and MSE, the error analysis only shows MAPE and RMSE.

As can be seen from Fig. 4, the CF-SA-FFOA-SVM method performs best in the five models. In the three sets of experiments, the MAPE of CF-SA-FFOA-SVM is only 3.13%–3.54%. Taking the FTL of 20 days as an example, the MAPE of the CF-SA-FFOA-SVM method is reduced by 6.23%, 8.93%, 41.04%, and 41.70%, respectively, compared with the PSO-SVM method, the BPNN method, the GM (1,1) method, and the ARIMA method. Matlab 2018 is used to program and generate convergence curves, taking the FTL for 30 days as an example, the convergence curve of the CF-SA-FFOA-SVM algorithm is compared with the PSO-SVM algorithm, as shown in Fig. 5. It can be seen that the CF-SA-FFOA-SVM algorithm converges after the 10th iteration, while the PSO-SVM algorithm converges after the 39th iteration. And the convergence accuracy of the CF-SA-FFOA-SVM algorithm is higher due to its RMSE is 0.02274, while the RMSE of PSO-SVM is 0.02275. In summary, the CF-SA-FFOA-SVM algorithm has higher

accuracy in gas load forecasting and has a faster convergence speed, which has particular advantages.

5. Discussions

5.1. Temperature type

According to the investigation, the daily mean temperature is always used as one of the input variables for gas forecasting in most studies, and the daily maximum temperature is seldom selected. In 2018, Shu et al. (2018) took the maximum temperature as the input variable but did not compare the two parameters. Therefore, the effect of daily mean temperature T_m and daily maximum temperature T_{max} on forecasting results are discussed in this paper. If the daily maximum temperature T_{dmax} is used as the input variable, the revised forecasting model is:

$$L_{fd} = f(T_{dmax}, W_d, D, H, HIST_{(T,W,D,H,L),(7d)}) \quad (19)$$

Six sets of experiments are designed according to the FTL (10–60 days). Fig. 6 shows the results when the FTL is 30 days.

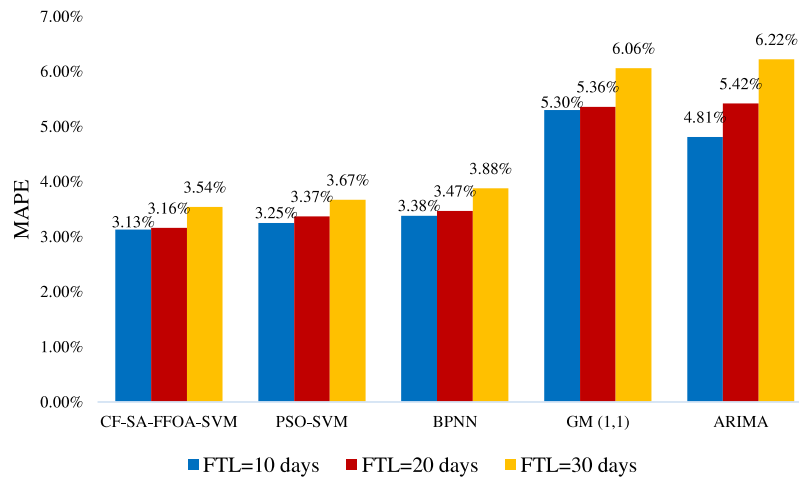
It can be seen from Table 5 that the MAPEs of the basic model (Eq. (18)) are larger than that of the revised model (Eq. (19)) except that the FTL is 10 days. Considering that the 10-day time is short, it is difficult to avoid the occurrence of large errors. Therefore, it is reasonable to use the average value of multiple sets of experiments. From Table 5, it can be concluded that in the six sets of experiments, the average MAPE of the basic model is 3.56%, which is 6.56% lower than the average MAPE of the revised model, indicating that for Kunming city, the demand for gas occurs mostly in the higher temperature period every day. Therefore, when establishing the gas load forecasting model, it is necessary to analyze the gas usage law of the city in advance based on historical data. Based on this, the discussion of other factors in this paper adopts the revised forecasting model (Eq. (19)).

5.2. Short-term time type

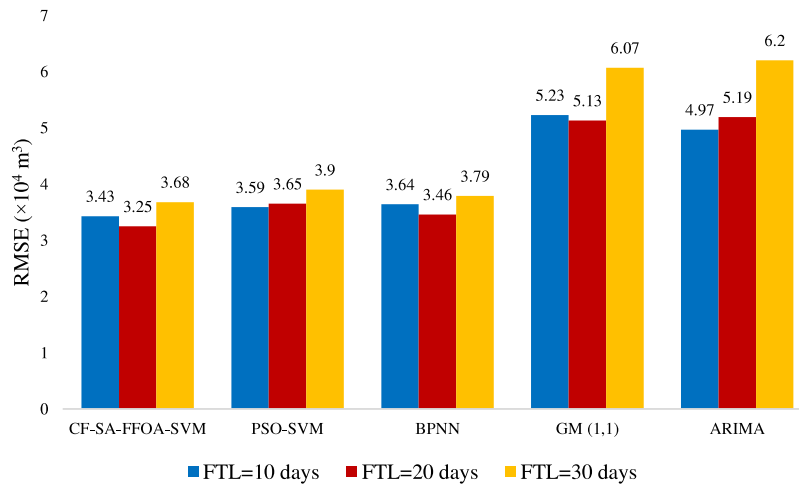
As mentioned earlier, short-term load forecasting includes multiple time types, ranging from hourly to monthly. This paper analyzes the adaptability of the CF-SA-FFOA-SVM for other time types: weekly and monthly. Table 6 lists the training and test sets for the forecasting models corresponding to different time types. It should be noted that in China, Monday is the beginning of a week, and Sunday is the last day of the week. Since January 1, 2012 is Sunday and New Year's Day in China, to facilitate data collation, the data of January 1, 2012 and November 25–30, 2013 are removed in the study of the weekly load forecasting model. Table 7 lists the forecasting errors of models based on different time types. It can be seen that the MAPEs of the forecasting results of the three models are all below 5%, and the overall forecasting results are good, and the accuracy is high. It can also be seen that CF-SA-FFOA-SVM has the highest forecasting accuracy for daily load, followed by weekly load and monthly load. Even so, the accuracy of forecasting can still be met.

5.3. Data grouping-based forecasting

Since the gas load will be affected by factors such as holidays and seasons, and under different main control factors, the gas load will change to varying degrees. If the data with similar features are classified, the forecast result will be more regular (Barman et al., 2018). Thus, this paper uses data grouping method to forecast gas load. As mentioned earlier, the input variables in the SVM forecasting in this paper include the daily maximum temperature, weather, date, and holiday. The two factors, date



(a)



(b)

Fig. 4. Error analysis results for various forecasting methods. (a) MAPE error. (b) RMSE error.**Table 5**

The error of models corresponding to two temperature types.

Model number	Temperature type	Forecasting model	Forecasting time length (days)	Error type		
				MAPE	MSE (m^3)	RMSE (m^3)
Model 1	Daily mean temperature	$L_{fd} = f(T_{dm}, W_d, D, H, HIST_{(T,W,D,H,L),(7d)})$	10	3.13%	1.18×10^9	3.43×10^4
			20	3.16%	1.06×10^9	3.25×10^4
			30	3.54%	1.36×10^9	3.68×10^4
			40	3.57%	1.41×10^9	3.75×10^4
			50	3.89%	1.73×10^9	4.15×10^4
			60	5.56%	3.21×10^9	5.66×10^4
Model 2	Daily maximum temperature	$L_{fd} = f(T_{dmax}, W_d, D, H, HIST_{(T,W,D,H,L),(7d)})$	10	3.34%	1.31×10^9	3.62×10^4
			20	3.07%	1.07×10^9	3.28×10^4
			30	3.35%	1.14×10^9	3.38×10^4
			40	3.19%	1.07×10^9	3.27×10^4
			50	3.32%	1.31×10^9	3.61×10^4
			60	5.10%	2.94×10^9	5.42×10^4

and holiday, are fixed, while temperature and weather are difficult to control. In fact, it is still possible to classify according to the season (temperatures in different seasons are significantly different). However, the temperature changes in the four seasons in Kunming city are not noticeable. Moreover, the date does not

have a very significant effect on the gas load. Therefore, this paper considers data classification by the holiday attribute.

According to the weight coefficient shown in Fig. 3, data can be divided into two groups: holiday (weight coefficient is 0.3), non-holiday (weight coefficient is 0.9). Considering that there were

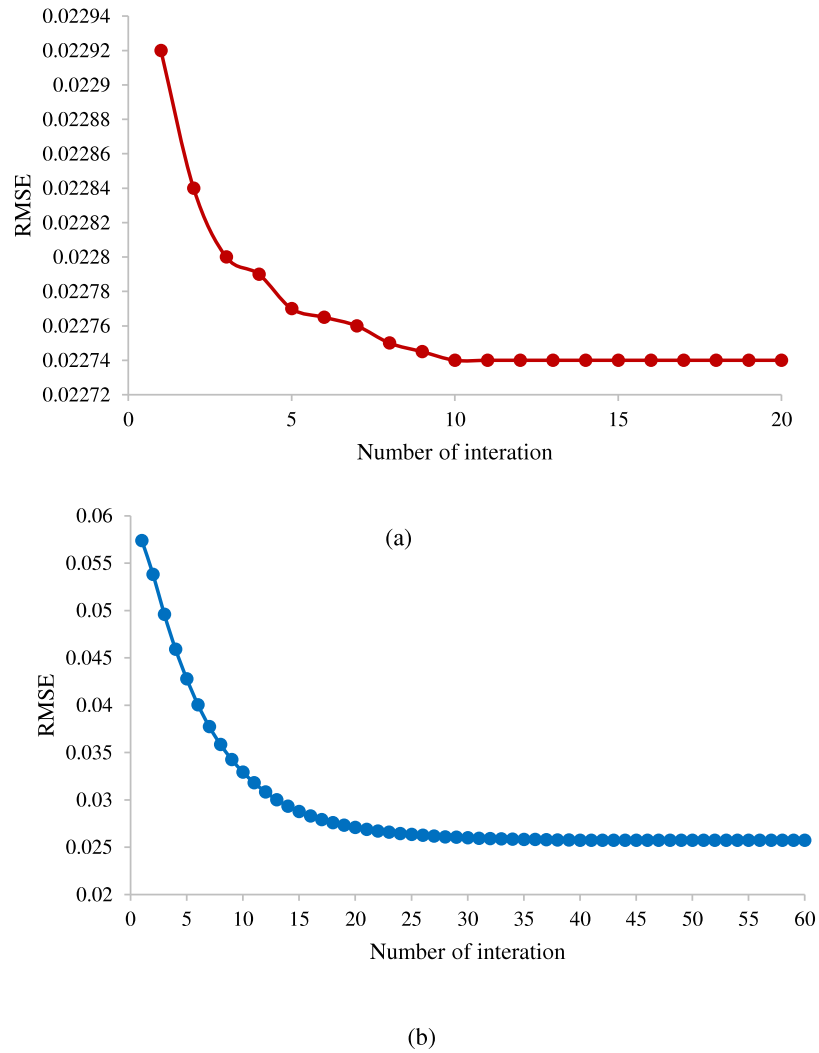


Fig. 5. Convergence curve of the algorithm. (a) CF-SA-FFOA-SVM. (b) PSO-SVM.

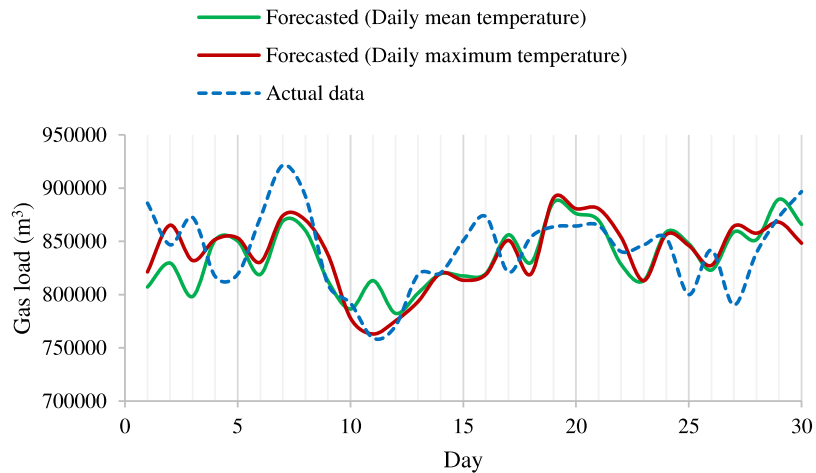


Fig. 6. Comparison between forecasted value and the real value of gas load (FTL is 30 days).

no major holidays in November 2013, the data in October and November was selected as the test set, that is, the data from January 1, 2012 to September 30, 2013 is used as the training set, and the data from October 1, 2013 to November 30, 2013 is used as the test set. Table 8 lists the number of two kinds of data in the training set and test set. Then the new forecasting model

is:

$$L_{fd} = f(T_{d\max}, W_d, D, HIST_{(T, W, D, L), (7d)}) \quad (20)$$

From Fig. 7, the MAPE of the overall forecasting results using the data grouping-based method is 4.52%, which is 11.2% lower

Table 6

Training and test sets for the forecasting models corresponding to different time types.

Time type	Forecasting model	Training set (Number of data)	Test set (Number of data)
Daily	$L_{fd} = f(T_{dmax}, W_d, D, H, HIST_{(T,W,D,H,L),(7d)})$	Jan 1, 2012–Oct 31, 2013 (670)	Nov 1, 2013–Nov 30, 2013 (30)
Weekly	$L_{fw} = f(T_{wamax}, W_{wa}, D, H, HIST_{(T,W,H,L),(4w)})$	Jan 2, 2012–Oct 27, 2013 (95)	Oct 28, 2013–Nov 24, 2013 (4)
Monthly	$L_f = f(T_{mamax}, W_{ma}, D, H, HIST_{(T,W,D,H,L),(4m)})$	Jan 1, 2012–Oct 31, 2013 (22)	Nov 1, 2013–Nov 30, 2013 (1)

Note: fw represents forecasted week value; wamax represents the average daily maximum value of that week; wa represents the average daily parameters of the week; 4w represents four weeks; mamax represents the average daily maximum value of that month; 4m represents four months.

Table 7

The error of models corresponding to different time types.

Time type	Error type		
	MAPE	MSE (m ³)	RMSE (m ³)
Daily	3.35%	1.14×10^9	3.38×10^4
Weekly	3.59%	1.07×10^{11}	3.26×10^5
Monthly	4.33%	1.20×10^{12}	1.09×10^6

Table 8

The number of two types of data in the training set and test set.

Training set		Test set	
Holiday	Non-holiday	Holiday	Non-holiday
49	590	7	54

than the original method. Among them, the MAPE of the forecasted gas load of non-holidays is 3.43%, which is 19.7% lower than that of the original method. This is because separating non-holiday data from holiday data can better determine the impact of other factors on gas load. It can be obtained that the data grouping method has a good effect in improving the overall forecasting accuracy, which is equivalent to removing a variable in the input vector of the model. On the other hand, the holiday load forecasting error has increased slightly. Therefore, the load of a specific holiday should be predicted based on historical data for many years to achieve higher forecasting accuracy.

Urban gas can be divided into domestic, industrial gas and so on according to the different supply objects. Fig. 8 shows the load of different types of gases in 2012 (366 days), it can be seen that different types of gas have different load characteristics, that is, the impact of various external factors on them is different. Thus, data grouping can be performed according to the type of gas. According to statistics, the types of gas in Kunming city can be divided into domestic gas, industrial gas, public service gas, gas

for transfer and other (refers to the self-use gas at the station). CF-SA-FFOA-SVM method is used to forecast five kinds of gas separately and then summarize them.

Five sets of experiments are conducted according to the FTL (10–50 days), Fig. 9 shows the MAPE of the forecasted results of the original method and this data grouping-based method. It can be seen that the MAPE of several forecasting results by the data grouping-based method (by gas type) is lower than the original method, which is reduced by 2.69%–10.18%, indicating that grouping forecast according to gas type has a good effect on improving forecasting accuracy.

6. Conclusions

Based on the historical data of Kunming, China, this paper uses CF-SA-FFOA-SVM algorithm to forecast the gas load. By comparing the results of other forecasting methods, it proves that CF-SA-FFOA-SVM algorithm has superiority. This paper also discusses the influence of temperature types on the forecasting results in the gas load forecasting model and analyzes the applicability of CF-SA-FFOA-SVM algorithm to other types of short-term load forecasting (weekly and monthly). In addition, the data are grouped according to holiday attributes and gas types, and the load is forecasted in groups. The following main conclusions are drawn:

(a) compared with PSO-SVM, BPNN, GM (1,1) and ARIMA, CF-SA-FFOA-SVM algorithm has higher gas load forecasting accuracy (when the FTL is 20 days, MAPE decreased by 6.23%, 8.93%, 41.04%, and 41.70%, respectively), and its convergence speed and precision are higher than the PSO-SVM algorithm.

(b) for Kunming city, if the input vector of the forecasting model adopts the daily maximum temperature, the MAPE of the forecasting result is 6.56% lower than that of the model using the daily mean temperature as the input vector.

(c) in the short-term gas load forecasting, the CF-SA-FFOA-SVM algorithm has the highest forecasting accuracy for the daily

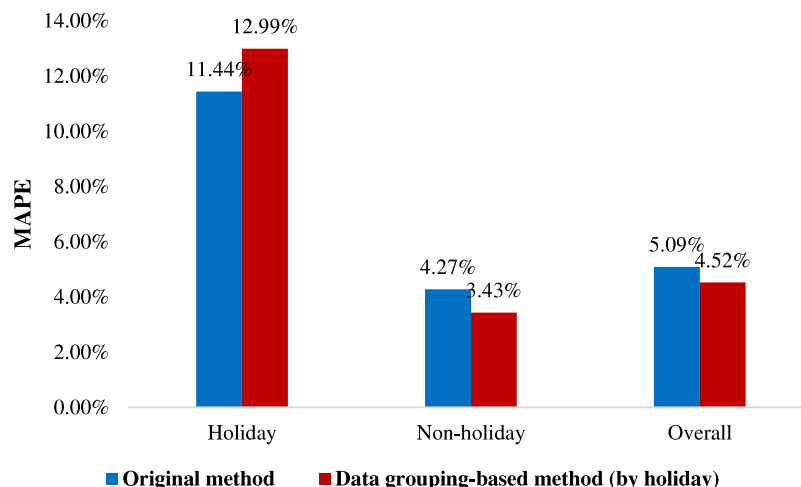


Fig. 7. The MAPE of the original method and data grouping-based method (according to holiday attribute).

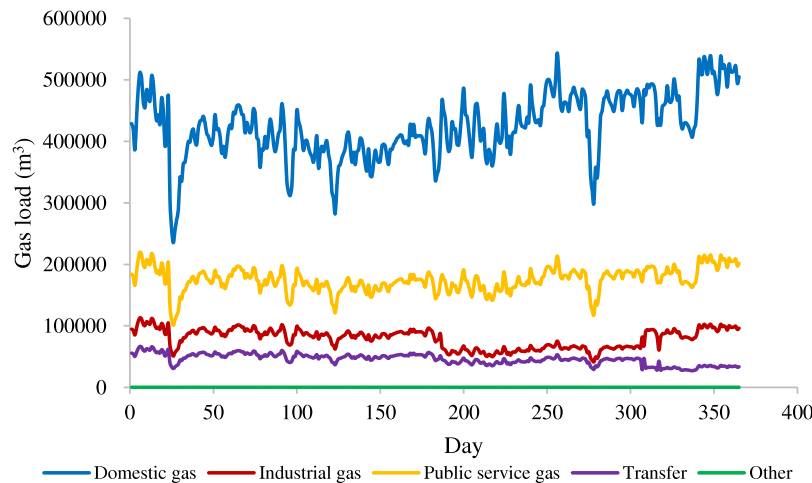


Fig. 8. The gas load of Kunming city in 2012.

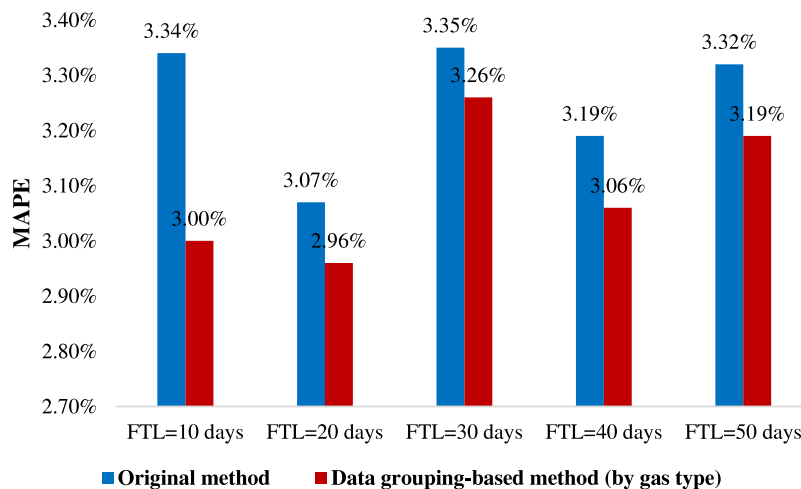


Fig. 9. The MAPE of the original method and data grouping-based method (gas type).

load (MAPE is 3.35%), but the forecasting accuracy of weekly and monthly loads is also satisfactory (MAPE is less than 5%).

(d) If the original data is grouped, the final forecasting accuracy can be improved, because the data grouping method integrates data with similar characteristics, which can better eliminate the interference of some important factors.

Acknowledgments

This article is funded by China Scholarship Council (201708030006). Special thanks to PetroChina Kunlun Gas Ltd. for providing gas data for the study.

References

- Akkurt, M., Demirel, O.F., Zaim, S., 2010. Forecasting turkey's natural gas consumption by using time series methods. *Eur. J. Econ. Politi. Stud.* 3 (2), 1–21.
- Akpınar, M., Yumusak, N., 2016. Year ahead demand forecast of city natural gas using seasonal time series methods. *Energies* 9 (9), 727.
- Azadeh, A., Asadzadeh, S.M., Ghanbari, A., 2010. An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: uncertain and complex environments. *Energy Policy* 38 (3), 1529–1536.
- Bai, X., 2016. Characteristic analysis and forecasting method research of daily gas load in winter of towns in sichuan-chongqing area (Master dissertation). Southwest Petroleum University, China, Chengdu.
- Barak, S., Sadeh, S.S., 2016. Forecasting energy consumption using ensemble arima-anfis hybrid algorithm. *Int. J. Electr. Power Energy Syst.* 82, 92–104.
- Barman, M., Choudhury, N.D., Sutradhar, S., 2018. A regional hybrid goa-svm model based on similar day approach for short-term load forecasting in assam, india. *Energy* 145, 710–720.
- British Petroleum (BP), 2018. BP statistical review of world energy, 67th ed. British Petroleum Co, London.
- Ding, S., 2018. A novel self-adapting intelligent grey model for forecasting china's natural-gas demand. *Energy* 162, 393–407.
- Ediger, V.S., Akar, S., 2007. Arima forecasting of primary energy demand by fuel in turkey. *Energy Policy* 35 (3), 1701–1708.
- Erdogdu, E., 2010. Natural gas demand in turkey. *Appl. Energy* 87 (1), 211–219.
- Gorucu, F.B., 2004. Artificial neural network modeling for forecasting gas consumption. *Energy Sources* 26 (3), 299–307.
- Gu, T.T., Luo, Y.Z., Zhang, Q., Zhu, Z.Y., 2017. Study on gas load forecast in hangzhou based on meteorological factors. *J. Meteorol. Environ.* 33 (3), 107–112.
- Herbert, J.H., Sitzler, S., Eades-Pryor, Y., 1987. A statistical evaluation of aggregate monthly industrial demand for natural gas in the usa. *Energy* 12 (12), 1233–1238.
- Iranmanesh, H., Abdollahzade, M., Miranian, A., 2011. Forecasting natural gas consumption using pso optimized least squares support vector machines. *Int. J. Artif. Intell. Appl.* 2 (4), 49.
- Ji, G., Wang, J., Ge, Y., Liu, H., 2014. Urban water demand forecasting by LS-SVM with tuning based on elitist teaching-learning-based optimization. In: *The 26th Chinese Control and Decision Conference (2014 CCDC)*, Changsha, 2014, pp. 3997–4002. <http://dx.doi.org/10.1109/CCDC.2014.6852880>.
- Karadede, Y., Ozdemir, G., Aydemir, E., 2017. Breeder hybrid algorithm approach for natural gas demand forecasting model. *Energy* 141, 1269–1284.
- Kumar, U., Jain, V.K., 2010. Time series models (grey-markov, grey model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in india. *Energy* 35 (4), 1709–1716.

- Lewis, S., 1982. *International and Business Forecasting Methods*. Butter Worths, London.
- Lu, H., Huang, K., Azimi, M., Guo, L., 2019. Blockchain technology in the oil and gas industry: a review of applications, opportunities, challenges, and risks. *IEEE Access* 7, 41426–41444.
- Ma, Xin, Xie, Mei, Wu, Wenqing, Zeng, Bo, Wang, Yong, Wu, Xinxing, 2019. The novel fractional discrete multivariate grey system model and its applications. *Appl. Math. Model.* 70, 402–424. <http://dx.doi.org/10.1016/j.apm.2019.01.039>.
- Melikoglu, M., 2013. Vision 2023: forecasting turkey's natural gas demand between 2013 and 2030. *Renew. Sustain. Energy Rev.* 22, 393–400.
- Pan, W.T., 2012. A new fruit fly optimization algorithm: taking the financial distress model as an example. *Knowl.-Based Syst.* 26, 69–74.
- Panapakidis, I.P., Dagoumas, A.S., 2017. Day-ahead natural gas demand forecasting based on the combination of wavelet transform and anfis/genetic algorithm/neural network model. *Energy* 118, 231–245.
- Pereira, J.J.F., Salgado, R.M., Ohishi, T., Ballini, R., 2006. Short-term load forecasting using support vector machines. In: *IEEE PES Transmission and Distribution Conference*.
- Shaikh, F., Ji, Q., 2016. Forecasting natural gas demand in china: logistic modelling analysis. *Int. J. Electr. Power Energy Syst.* 77, 25–32.
- Shu, M., Liu, X.L., Xu, T., Xie, W.J., He, B., 2018. A new model for forecasting the short-term daily demand of urban natural gas. *Nat. Gas Ind.* 38 (6), 128–132.
- Smith, P., Husein, S., Leonard, D.T., 1996. Forecasting short term regional gas demand using an expert system. *Expert Syst. Appl.* 10 (2), 265–273.
- Soldo, B., 2012. Forecasting natural gas consumption. *Appl. Energy* 92, 26–37.
- Song, J.G., 2015. Research on improvement of fruit fly optimization algorithm (Master dissertation). Jiangxi University of Science and Technology, China, Nanchang.
- Song, J., Pan, H., 2017. Gas daily load forecasting based on fruit fly optimization algorithm and svm. *Control Eng. China* 24 (10), 1995–2002.
- Suganthi, L., Samuel, A.A., 2012. Energy models for demand forecasting—a review. *Renew. Sust. Energ. Rev.* 16 (2), 1223–1240.
- Szoplik, J., 2015. Forecasting of natural gas consumption with artificial neural networks. *Energy* 85, 208–220.
- Taşpınar, F., Celebi, N., Tutkun, N., 2013. Forecasting of daily natural gas consumption on regional basis in turkey using various computational methods. *Energy Build.* 56, 23–31.
- Uzlu, E., Kankal, M., Akpınar, A., Dede, T., 2014. Estimates of energy consumption in turkey using neural networks with the teaching–learning-based optimization algorithm. *Energy* 75, 295–303.
- Vapnik, V., 2013. *The Nature of Statistical Learning Theory*. Springer science & business media.
- Wang, L.J., 2012. Research on forecasting methods of natural gas load based on particle swarm optimization-least squares support vector machine (Master dissertation). East China University of Science and Technology, China, Shanghai.
- Wang, Q., Li, S., Li, R., 2018. Forecasting energy demand in china and india: using single-linear, hybrid-linear, and non-linear time series forecast techniques. *Energy* 161, 821–831.
- Wu, Q., Peng, C., 2017. A hybrid bag-sa optimal approach to estimate energy demand of china. *Energy* 120, 985–995.
- Yu, F., 2015. The research on short-term gas load combination forecasting model based on residual error and similar day load correction (Master Dissertation). Shanghai Normal University, Shanghai.
- Yu, F., Xu, X., 2014. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved bp neural network. *Appl. Energy* 134, 102–113.
- Zeng, B., Li, C., 2016. Forecasting the natural gas demand in china using a self-adapting intelligent grey model. *Energy* 112, 810–825.
- Zhang, C., Liu, Y., Zhang, H., Huang, H., 2011. Research on the daily gas load forecasting method based on support vector machine. In: *2011 Fourth International Conference on Intelligent Computation Technology and Automation*, Shenzhen, Guangdong, 2011, pp. 222–227.
- Zhang, C., Liu, Y., Zhang, H., Huang, H., 2014. Study on urban short-term gas load forecasting based on support vector machine model. *Journal of Tsinghua University(Science and Technology)* 54 (3), 320–325.
- Zhu, L., Li, M.S., Wu, Q.H., Jiang, L., 2015. Short-term natural gas demand prediction based on support vector regression with false neighbours filtered. *Energy* 80, 428–436.