Electrical Engineering
https://doi.org/10.1007/s00202-022-01528-1

ORIGINAL PAPER

Urban monthly power load forecasting based on economy-meteorology-gas demand coupling

Bin Ding[1] · Fan Wang[1] · Chen Chen[1] · Shizhao Wang[2]

Received: 28 October 2021 / Accepted: 15 February 2022

Abstract
Accurate power load forecasting is the key foundation and important premise of urban power system planning. Considering that urban power load demand is closely related to economic development and meteorological conditions, and with the application and promotion of various energy coupling equipment and the promotion of electric energy substitution policy, the relationship between gas load and power load demand is increasing. However, most of the existing load forecasting methods only consider a single economic factor or a single meteorological factor, and fail to combine the coupling and complementarity between loads. Based on this, this paper first analyzes the coupling characteristics between power load and urban economic, meteorological, and gas load with Copula theory; Secondly, the key parameters of the least squares support vector machine (LSSVM) are solved by using the salp swarm algorithm (SSA), and the urban monthly power load forecasting model based on SSA-LSSVM is established; Finally, based on the macroeconomic data, meteorological observation data and monthly electrical load of a city in North China, the prediction model is verified. Through comparison with different prediction scenarios, it is confirmed that the built model can effectively improve the prediction accuracy and has good application effects.

Keywords Economic development · Meteorological load · Electrical coupling · Load forecasting

1 Introduction

The construction of a new power system with new energy as the main body, serving carbon peaks and carbon neutral goals means that wind power and photovoltaic power generation will gradually become the main body of the power system in the future. Fundamental changes will take place in the energy and power system, which will not only promote the clean power supply, intelligent power grid and electrification of users, but also change the operation characteristics of power grid [1]. At the same time, reasonable and accurate load forecasting plays an important role in power system production arrangement, economic dispatch, power generation planning, safe operation, and system security evaluation [2]. The monthly power load forecast data is an important input data for formulating the planning and dispatching scheme of the urban power system within one year, which can directly affect the development direction and optimization direction of the future power system and even the city. For the short term, monthly power load forecast is conducive to the safe and stable operation of the power system and effectively reduces operation and maintenance costs. For the medium and long term, it provides a reliable basis for power companies to formulate monthly and annual power generation plans, and has important guiding significance for economic dispatch control of power systems.

The electricity consumption behavior and the load of power users are affected by a variety of factors, such as weather change, seasonal alternation, macroeconomic development, urban expansion, energy structure adjustment, photovoltaic power generation, wind farm, start-up and shut-down of large power users, equipment maintenance, large-scale activities, etc. [3].Studying the relationship between economic development and power demand is the basis of many current power load forecasting methods. The main research methods include elastic coefficient, grey correlation theory, economics, and other theories. In the research of Neera et al. [4], the irregular and noisy behavior in the observed data makes it difficult to achieve better forecast-

✉ Bin Ding
15023602691@163.com

1 State Grid Xiongan New Area Electric Power Supply Company, Hebei, China

2 POWERCHINA Shanghai Electric Power Engineering Co., Ltd., Shanghai, China

Published online: 06 May 2022

ing accuracy. To handle this, we propose a new model, named singular spectrum analysis-long short- term memory (SSA-LSTM). SSA is a signal processing technique used to eliminate the noisy components of a skewed load series. LSTM model uses the outcome of SSA to forecast the final load. Huang et al. [5] analyzed the change of power demand in Liaoning under the new normal, studied the impact of economic situation change on power demand, and forecasted the power demand in Liaoning in the near future. Xu and Zhu [6] took the time series of GDP and regional electricity consumption of four provinces in Central China as the analysis object, and used H-P filtering method, cointegration test and spectral analysis method to analyze the change characteristics of electricity demand in Central China and its relationship with economic development. According to the research by Han et al. [7], a power load forecasting model considering economic transformation index was constructed by constructing an index system to measure the degree of economic transformation. Jin et al. [8] established the index system of endogenous characteristics and electricity demand externality suitable for judging the saturation stage of large regional development, constructed the electricity demand externality model by cointegration method, and predicted the saturation demand.

2 Related works

For a long time, many domestic and foreign experts and scholars have conducted extensive research on the influence of meteorological conditions on the electrical load of different cities. Masoud Sobhani et al. [9] put forward that weather is a key factor affecting electricity demand. Many load forecasting models rely on weather variables. Weather stations provide point measurements of weather conditions in a service area. Gilkeson [10] pointed out that adverse weather caused heavy losses to the power industry, and made a scientific analysis on how to use weather forecast to reduce losses. According to Peng et al. [11], the sensitivity model of meteorological factors to load change was obtained by support vector regression. Chen et al. [12] and Cong et al. [13] put forward the Grey Combination Model Based on the optimal weight combination, and personalized intelligent power consumption package, respectively. Considering the coupling effect and cumulative effect of meteorological factors on daily electricity consumption, Wang et al. [14] analyzed the relationship between meteorological indexes and daily electricity consumption, and established a single prediction model of daily electricity consumption.

In the study of load coupling characteristics, the current analysis mainly focuses on the coupling characteristics between cooling, heat and electricity. Wang et al. [15] proposed an integrated energy system short-term load forecasting method considering the coupling characteristics of cooling, thermal and electrical loads as well as the time dynamics. Chen et al. [16] investigated the spatial and temporal variation of electrical load characteristics due to cold and thermal load variations. Wang et al. [17] considered the coupling characteristic curves between multiple energy sources such as cooling, heat and electricity to reflect the changing characteristics of the load. Chen et al. [18] analyzed the nonlinear relationship between cold and thermal loads and electrical loads, and proposed a multi-energy synergistic electric load forecasting model. Lv et al. [19] consider the coupling characteristics between central energy station, district heating network (DHN) and building loads, which can better cope with the load uncertainty. Qi et al. [20] demonstrated the existence of coupling between loads using Pearson correlation coefficients and extracted the characteristic quantities of the coupling characteristics using a convolutional neural network (CNN) for load prediction.

In terms of other factors, the existing literature mainly analyzes and forecasts the urban power load due to special events and changes of cooling and heating load, while there is less research on urban gas and power load forecasting. A forecasting method of coupled thermal and electrical load is proposed by Luo et al. based on Elman neural network and grey neural network [21]. In the study of Quan et al. [22], the correlation between heat and cooling load is considered, and the corresponding load forecasting model is constructed. According to He et al. [23], the risk factors affecting power load fluctuation include land use change, load density per unit area and so on. Secondly, based on the basic principle of cellular automata, the law and model of land use change considering risk factors are established. After comprehensively considering the impact of risk factors on load density fluctuation, an urban power load forecasting model is established based on the risk analysis of land use change and load density. Moreover, Barman M et al. offered a novel method of power system load forecasting (PSLF) for regional special event days (RSEDs) when the load demand is highly prejudiced by societal considerations like cultural or religious rituals. These rituals abruptly change the consumer behaviors (demand variations) and it makes the load profile of such RSEDs more complex and nonlinear than normal holidays. Therefore, during RSEDs, an accurate PSLF method must integrate these consumer behaviors in the forecasting process [24]. In the process of multiple load forecasting, scholars have adopted a variety of research model methods, such as deep learning [25], multi-task learning [26, 27], long-term and short-term memory neural network [28, 29], and Copula theory [30]. The above models and methods based on the multi load of integrated energy system characteristics provide reference value and guidance for further research.

The above-mentioned literature mainly focuses on the coupling characteristics between cold, heat and electric loads in short-term load forecasting and the influence of electric demand by weather factors, etc., while relatively little research has been done on the relationship between monthly electric load and meteorological factors and macroeconomic indicators and on the establishment of monthly electric load forecasting models. The forecast of monthly electricity load is the basis and important indicator for the short- and medium-term operation planning of the power grid. And a single type of load forecasting model is difficult to reflect the coupling relationship among economic development, meteorological changes, and multiple loads. Based on this, this paper comprehensively considers the coupling characteristics between power load and urban economic, meteorological and gas load, and constructs a load correlation calculation model based on Copula theory; Secondly, the RBF function is used as the kernel function of LSSVM model to construct LSSVM urban monthly power load forecasting model, and the SSA method is used to find the optimal parameters; Finally, a case study of a city in North China is carried out to verify the model. The results indicate that the model is able to effectively enhance the accuracy of urban monthly power demand forecasting.

3 Analysis of load coupling characteristics based on Copula theory

Coupling means that two or more elements or systems interact with each other. In the urban power system, the correlation between power load and urban economic, meteorological, and gas load is defined as economic coupling degree, meteorological coupling degree, and gas demand coupling degree, respectively.

Copula function is a tool used to study the correlation between variables. It can establish a certain mathematical relationship between the joint distribution function and its marginal distribution function. The domain defined by Copula function is [0, 1]. By estimating the distribution of the edge of each random variable, the results are obtained respectively. Finally, a model can be built to help calculate the relationship between the two variables. If F is a function containing several n-dimensional random variables and $G_1, G_2, ...G_n$ are the marginal distribution of marginal distribution function C, then there exists Copula function

$$F(x_1, x_2, ..., x_n) = C(G_1(x_1), G_2(x_2), ..., G_n(x_n)). \tag{1}$$

This method is suitable for multivariate model distribution and stochastic simulation and can be used as a model tool for correlation analysis and model research of urban monthly power load forecasting.

Although the joint distribution function can precisely portray the association relationship between variables, the expression of the distribution function is more complex and lacks a certain intuitive. In order to compare and analyze the economic coupling degree, meteorological coupling degree, and gas load coupling degree more intuitively, the Spearman rank correlation coefficient, which can describe the nonlinear correlation among variables, is selected in the correlation measure derived from Copula function ρ as the final evaluation index of nonlinear correlation.

For two random variables X and Y, their corresponding distribution functions are $G(x)$ and $H(y)$. There must be a Copula function $C(a, b)$ that can pass the Spearman rank correlation coefficient ρ. To characterize the nonlinear correlation between X and Y, that is, we can use Copula function $C(a, b)$ to derive ρ. Namely:

$$\begin{aligned}\rho &= 12\int_0^1\int_0^1 abdC(a,b) - 3\\ &= 12\int_0^1\int_0^1 C(a,b)dadb - 3,\end{aligned} \tag{2}$$

where a, b are the marginal distribution functions of two variables X and Y respectively $G(x)$ and $H(y)$.

There are five frequently used Copula Functions, which are t-Copula, Gumbel–Copula, Clayton–Copula, Frank—Copula and normal-Copula [31]. The Spearman rank correlation coefficient derived from the same Copula function is fixed and unique. However, due to the variety of Copula Functions, the correlation coefficient values derived from different Copula Functions are different. Therefore, the accuracy of correlation analysis based on Copula function depends on the selection of the optimal Copula function. In this paper, the empirical Copula function $\hat{C}(a, b)$ is defined, and by comparing the Euclidean distance between each alternative Copula function and the empirical Copula function, the optimization of the Copula function is realized.

The empirical Copula function $\hat{C}(a, b)$ is defined as follows: let $x_i, y_i (i = 1, 2, \cdots, n)$ take samples of variables X and Y, respectively, and $F_x(x)$, $F_y(y)$ take empirical distribution functions of variables X and Y respectively:

$$\hat{C}(a,b) = \frac{1}{n}\sum_{i=1}^{n} I_{[F_x(x_i)\le a]} I_{[F_y(y_i)\le b]}], \tag{3}$$

where $I_{(if)}$ is an indicative function, when the condition if holds, $I_{(if)} = 1$, otherwise it is 0.

The Euclidean distance between the empirical Copula function and each alternative empirical Copula function is as follows:

$$d_j = \sum_{i=1}^{n}\left|\hat{C}(a_i, b_i) - \hat{C}_j(a_i, b_i)\right|, \tag{4}$$

where a_i is the empirical distribution function value $F_x(x_i)$ corresponding to x_i; b_i is the empirical distribution function value $F_y(y_i)$ corresponding to y_i; $C_j(\cdot)$ is an alternative empirical Copula function.

The steps of coupling analysis using Copula theory are as follows: firstly, the kernel density estimation method is used to obtain the marginal distribution of the variable, and the Maximum Likelihood Method (MLM) is used to calculate the unknown parameters of each candidate Copula function according to the marginal distribution. The empirical distribution function of the variable is calculated according to the sample point data, and then the spline interpolation method is used to obtain the empirical distribution function value corresponding to each sample point, and calculate the corresponding empirical Copula function value and the candidate Copula function value from the empirical distribution function value according to formula (3) and the candidate Copula function. According to formula (4), calculate the Euclidean distance between the alternative copula function and the empirical copula function, and select the minimum Euclidean distance as the optimal copula function. on the basis of formula (2), the Spearman correlation coefficient can be derived through the corresponding integral calculation of the optimal Copula function.

4 Urban monthly power load forecasting model based on SSA-LSSVM

The time series of actual data can show the development trend and law of the research object in a certain period, so we can find out the characteristics, trend and development law of variable change from the time series, so as to effectively predict the future change of variables. Based on the analysis of historical data, it is found that urban power load has two peak periods in winter and summer, and its change shows obvious periodic fluctuation. Therefore, this paper selects the time series forecasting model based on least squares support vector machine for urban power load forecasting.

It should be pointed out that the time series prediction method has the defect of prediction error because it highlights that the time series does not consider the influence of external factors temporarily. When there are large changes in the outside world, there will often be large deviations. The effect of time series prediction method on medium and short-term prediction is better than that of long-term prediction. Therefore, this paper comprehensively considers the impact of external factors such as economic development, meteorological changes and natural gas demand on urban power load, and establishes an urban monthly power load forecasting model based on the coupling of economy, meteorology and gas demand.

4.1 Building LSSVM prediction model

Least squares support vector machine (LSSVM) is a computing model of support vector machine under quadratic probability loss function, which can help solve optimization problems by solving linear model. The main principles are as follows.

Assuming that the sample set $S = (x_i, y_i)_{i=1}^{L}$, $x_i \in X \in R^n$ is input vector and $y_i \in R$ is the corresponding output value of the sample, the decision function can be constructed as follows:

$$f(x) = \omega\varphi(x) + b, \tag{5}$$

where $\varphi(x)$ is a nonlinear high dimensional mapping of $x_i \in X \in R^n$; ω is the weight; b is the offset value.

The structural risk function is as follows:

$$R = \frac{1}{2}\|\omega\|^2 + \frac{1}{2}c \cdot R_{\text{emp}}, \tag{6}$$

where $\|\omega\|^2$ is the accuracy and complexity of the model; c is the regularization parameter;R_{emp} is an empirical risk.

In the process of modeling, $R_{\text{emp}} = \sum_{i=1}^{L} \zeta_i^2$, the decomposition formula is as follows:

$$\min R = \frac{1}{2}\boldsymbol{\omega}^T \boldsymbol{\omega} + \frac{1}{2}c \cdot \sum_{i=1}^{L} \zeta_i^2, \tag{7}$$

$$s.t.\ y_i = \omega_i\varphi(x_i) + b + \zeta_i\ , \tag{8}$$

where ζ_i is the error relaxation variable, $i = 1, 2, \ldots, L$.

Then the optimization problem is shown in Eqs. (7)—(8) can be expressed as:

$$\begin{bmatrix} 0 & \boldsymbol{I_v^T} \\ \boldsymbol{I_v} & \Omega + c^{-1}\boldsymbol{I} \end{bmatrix}\begin{bmatrix} b \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \tag{9}$$

where $I_v = [1, 1, \ldots, 1]^T$, and there are l elements in total, $\Omega_{ij} = K\left(x_i, y_j\right)i, j = 1, 2, 3, \ldots, l$.

The kernel function itself is the internal machine of the mapping relationship. In LSSVM model, the adaptability of the RBF kernel function is that the number of setting parameters is small. Therefore, the RBF function is used as the kernel function in this paper.

$$K(x_i, y_j) = \exp\left(-\frac{\left\|x_i - y_j\right\|^2}{\sigma^2}\right). \tag{10}$$

By solving the above equation, the decision function is obtained directly:

$$f(x) = \sum_{i=1}^{L} \lambda_i K(x_i, x) + b. \tag{11}$$

4.2 Finding optimal parameters by SSA method

On the premise of determining the regularization parameters and kernel function parameters, this paper improves the traditional LSSVM method using the network cross validation method for parameter selection, and uses the bottle ascidian group optimization algorithm to find the best parameter value and reduce the error automatically.

Styela Corylifolia is a group of deep-sea organisms. In daily predatory and other ecological activities, Styela Corylifolia aggregation group formed by group movement is called ascidian chain. Australian researcher Mirjalili s proposed an algorithm for optimization of Salp swarm algorithm (SSA) [32]. In SSA, the target group can be divided into leaders and followers. The role of the leader is to guide the SALP group and each follower follows the former. It is assumed that the target source exists in the planning scope. The specific optimization methods based on the change of behavior are as follows.

(1) Parameter setting. Parameter settings. Including the number of groups, the number of influencing factors, the maximum number of iterations and the upper and lower limits of the variables.
(2) Species initialization. The following matrix is shown.

$$S = [s_{ij}]_{n \times d}, \tag{12}$$

where s_{ij} represents the j-th variable value of the i-th bottle ascidian,$i = 1, 2, \ldots, n$,$j = 1, 2, \ldots, d$.

$$s_{ij} = \text{rand}(i, j) \times [ub(i) - lb(i)] + lb(i), \tag{13}$$

where rand(i, j) is a random matrix whose element domain is [0,1]. $ub(i)$ and $lb(i)$ represent the maximum and minimum values of the i-th bottle ascidian, respectively.

(3) Construct fitness function. The fitness function is used to calculate, and the matrix OS is set to store the value:

$$OS = \begin{bmatrix} OS_1 \\ OS_2 \\ \vdots \\ OS_n \end{bmatrix} = \begin{Bmatrix} f\left[\left(s_{11}\ s_{12}\ \cdots\ s_{1d}\right)\right] \\ f\left[\left(s_{21}\ s_{22}\ \cdots\ s_{2d}\right)\right] \\ \vdots \\ f\left[\left(s_{n1}\ s_{n2}\ \cdots\ s_{nd}\right)\right] \end{Bmatrix}. \tag{14}$$

In matrix OS, the target source F is the one with the best fitness, and its position is determined by the influence of the bottle ascidian chain. Therefore, the optimal value can be solved by changing the position of the target source F.

(4) Determine the number of iterations. In order to avoid local optimal solution, all elements need to be updated and iterated by function operation. Among them, the formula of leader's positioning and updating to target source is as follows:

$$x_j^1 = \begin{cases} F_j + c_1\left[\left(ub_j - lb_j\right)c_2 + lb_j\right], c_3 \geq 0 \\ F_j - c_1\left[\left(ub_j - lb_j\right)c_2 + lb_j\right], c_3 < 0 \end{cases}, \tag{15}$$

where x_j^1 is the position of the first leader, namely the bottle ascidian, in the j-th dimension, and F_j is the coordinate of the target source. ub_j and lb_j are upper and lower limits, respectively. c_1, c_2, c_3 are random numbers generated by setting. c_2 and c_3 are in the range of [0,1] to determine the position change. c_1 is

$$c_1 = 2e^{-\left(\frac{4l}{L}\right)^2}, \tag{16}$$

where L and l are the maximum and current iterations, respectively. The location of followers is updated as follows:

$$x_j^i = \frac{1}{2}\left(x_j^i + x_j^{i-1}\right). \tag{17}$$

All steps are iterated until the end.

4.3 Model checking

In this paper, posterior tests were performed after the calculation to verify the accuracy of the prediction model using the small error probability (P) and the posterior error ratio (c). The posterior difference test is a method of testing the model based on the statistics between the predicted and actual values of the model, which is transposed from the probabilistic prediction method. It is based on the residuals (absolute error) ε and according to the absolute value of the residual in each period, the probability of the occurrence of points with smaller residuals and the magnitude of the indicators related to the variance of the prediction error are examined. The specific steps are as follows:

Set the historical load sequence.

$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}. \tag{18}$$

Set the sequence of predicted values.

$$\hat{x}^{(0)} = \{\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \cdots, \hat{x}^{(0)}(n)\}. \tag{19}$$

The difference between the actual value $x^{(0)}(k)$ and the calculated value (predicted value) $\hat{x}^{(0)}(k)$ at time k is $\varepsilon(k)$, which is called the time k residual.

$$\varepsilon(k) = \left| x^{(0)}(k) - \hat{x}^{(0)}(k) \right| \quad (k = 1, 2, \cdots, n). \tag{20}$$

Let the average of the actual value $x^{(0)}(k)k = 1, 2, \cdots, n$ be. That is

$$\overline{x} = \frac{1}{n} \sum_{k=1}^{n} x^{(0)}(k). \tag{21}$$

Let the mean value of residual $\varepsilon(k)k = 1, 2, \cdots, m$ be $\overline{\varepsilon}$, where

$$\overline{\varepsilon} = \frac{1}{m} \sum_{k=1}^{m} \varepsilon(k), \tag{22}$$

where m is the number of predicted residual data, generally with $m \leq n$.

Note that the variance of the historical data (actual values) is S_1^2.

$$S_1^2 = \frac{1}{n} \sum_{k=1}^{n} (x^{(0)}(k) - \overline{x})^2. \tag{23}$$

Let that the residual variance is S_2^2, and there is

$$S_2^2 = \frac{1}{m} \sum_{k=1}^{m} (\varepsilon(k) - \overline{\varepsilon})^2. \tag{24}$$

Then the posterior error ratio C and the small error probability P are obtained.

$$\begin{aligned} C &= \frac{S_2}{S_1} \\ P &= P\{|\varepsilon(K) - \overline{\varepsilon}| < 0.6745 S_1\}. \end{aligned} \tag{25}$$

The value of Index C is greater than 0, and the value of index p is [0,1].

The smaller the indicator C, the larger the S1 and the smaller the S2. a large S1 indicates a large variance in the historical data and a large degree of dispersion in the historical data. a small S2 indicates a small variance in the residuals and a small degree of dispersion in the residuals. a small C indicates that although the historical data are discrete, the difference between the predicted and actual values obtained by the model is not too discrete. The larger the indicator P, the greater the number of points where the difference between the residuals and the mean of the residuals is less than the given value of 0.6745S1. Based on C and P, the accuracy of the prediction model can be evaluated comprehensively, and the evaluation criteria are shown in Table 1.

4.4 Overall process of the model

The main factors that affect urban power load demand include temperature, humidity and other meteorological factors, as well as economic development, industrial structure, power substitution and other social environmental factors. In urban power load forecasting, it is necessary to deeply consider the coupling relationship between power load and urban economy, meteorology, and gas demand. Based on this, this paper establishes a Copula based load correlation analysis model to calculate the economic coupling, meteorological coupling, and gas coupling characteristics of the urban power system. Secondly, the bottle ascidian algorithm is used to solve the optimization problem of regularization parameters and kernel function parameters in LSSVM. In the prediction model, the factors such as temperature and economic development are considered, and the coupling relationship between gas and power load is taken as the influencing factor to build a multiple load prediction model based on SSA-LSSVM, as shown in Fig. 1.

5 Case analysis

5.1 Basic data

In the case analysis, this paper selects a city in North China for medium and long-term power load forecasting. By forecasting and analyzing the monthly load of the city in a year, the rationality and accuracy of the model are verified. This city is a major economic center and power load center in North China, and its data is relatively complete and representative. At the same time, the data used mainly comes from the relied fund projects and the electricity consumption information collection system of the State Grid. By forecasting the monthly electricity load, the city's grid planning scheme can be optimized, thereby helping to build a smart city (Figs. 2, 3, 4, 5).

The data used in this paper can be divided into four categories: meteorological, power load, macroeconomic and natural gas load. The data from 2010 to 2018 is the basic data set, and the data from 2019 is the verification set. The monthly average values of average high temperatures, average low temperatures and average temperatures are extracted from the meteorological data of the city.

Taking the historical economic, meteorological, power load and gas load data as the input data of the model can realize the quantitative transformation of energy policy, economic policy and other factors, so as to consider the impact of social and scientific factors affecting urban power load demand.

Macroeconomic data are extracted from the monthly statistical reports published by the Municipal Bureau of

Table 1 Small error probability (P) and posterior error ratio (c) of comprehensive evaluation prediction model

Prediction accuracy level	*P*	*C*	Prediction accuracy level	*P*	*C*
Level I	>0.95	<0.35	Level III	$0.5\leq C<0.7$	$0.5\leq C<0.65$
Level II	>0.8	$0.35\leq C<0.5$	Level IV	≥ 0.7	≥ 0.65

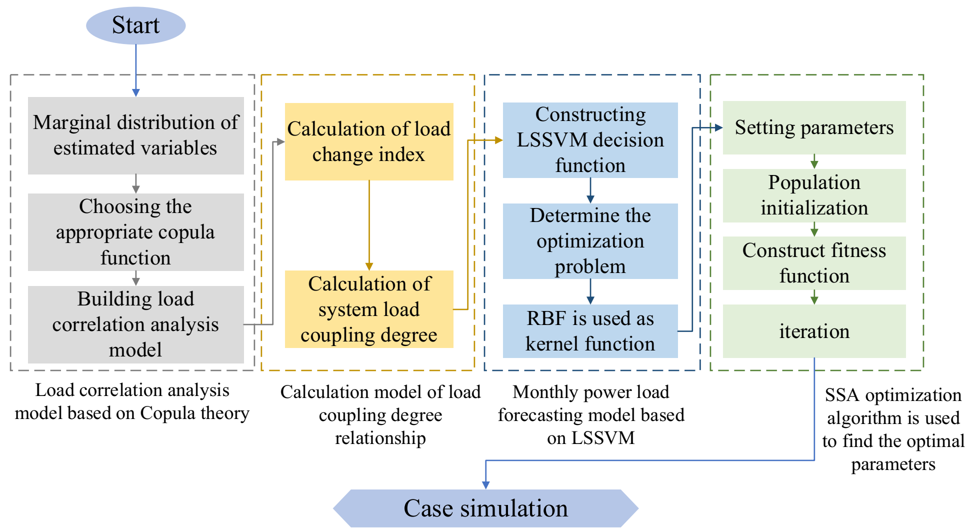


Fig. 1 Monthly urban power load forecasting process based on SSA-LSSVM model

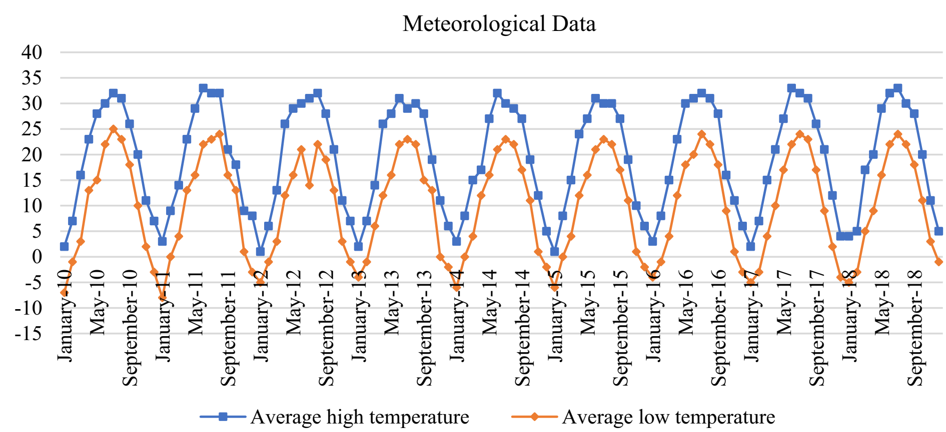


Fig. 2 Meteorological data of S city from 2010 to 2019

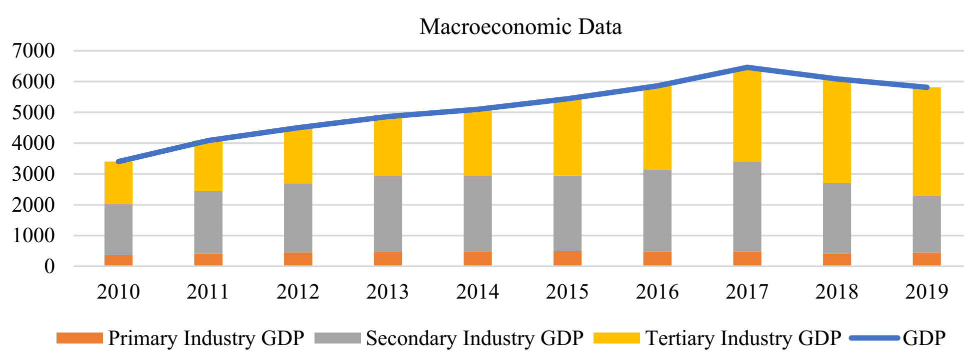


Fig. 3 Macroeconomic data of S city

Statistics, including monthly GDP, value added of primary, secondary and tertiary industry.

Total electricity consumption, monthly average load, monthly average peak value, and monthly average valley value are extracted from the database of power supply company and natural gas company in the city.

Fig. 4 Electricity load data of S city

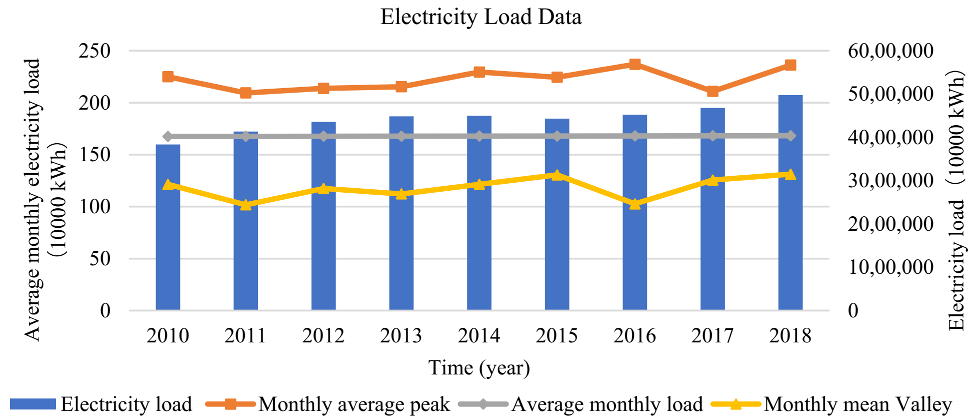
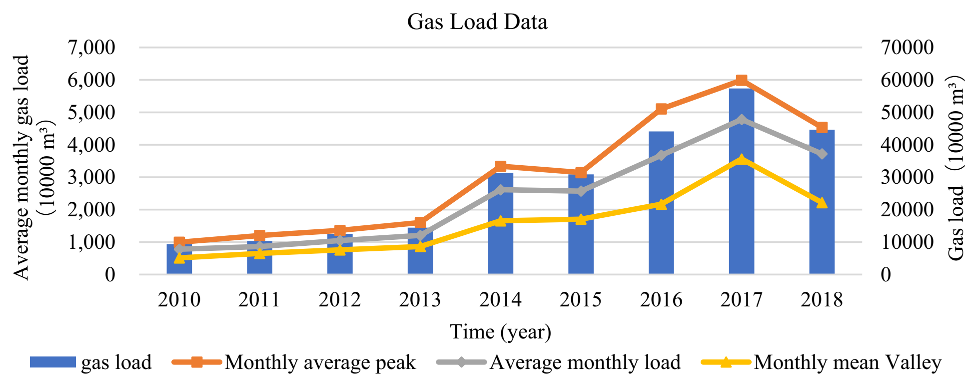


Fig. 5 Gas load data of S city



5.2 Load coupling analysis

Spearman rank correlation coefficients can be derived from Frank-Copula, t-Copula, and normal-Copula, and the selection of the optimal Copula function is shown in Table 2. It can be seen that the Euclidean distances between the three alternative Copula Functions and the empirical Copula function are different. The smaller the Euclidean distance, the closer the alternative Copula function is to the empirical Copula function, and the more accurately the correlation coefficient describes the correlation between the variables. Bold values indicate that this result is better than other Copula functions.

The results of the correlation analysis derived from the optimal Copula function are shown in Table 3. From Table 3, it can be seen that there is a high degree of nonlinear correlation between electricity load and GDP, GDP of primary industry, GDP of secondary industry, GDP of tertiary industry, average high temperature, average low temperature, average temperature, average monthly gas load, average monthly gas peak, and average monthly gas valley, but the degree of correlation varies. The correlation between the city power load and Secondary Industry GDP, Tertiary Industry GDP, average temperature average high temperature, average low temperature, monthly average gas load, and the monthly average gas peak value is higher than 0.5. The correlation between s city power load and GDP of secondary industry, average high temperature, and monthly average gas load is even higher than 0.8. The analysis results of the coupling characteristics between power load and macro-economy, meteorological conditions, and urban gas load show the systematicness of urban economy, meteorology and energy and the indivisibility of the prediction process. The influence of the above factors on the prediction results can not be ignored, and it also offers a theory basis for the construction of the prediction data set. Therefore, in purpose of improving the accuracy of urban power load forecasting, this paper takes the GDP of secondary industry, GDP of tertiary industry, the average high temperature, the average low temperature, the average temperature, the average monthly gas load, the average monthly gas peak and other factors with high correlation as the influencing factors of the forecasting model, and forms the input sample set of urban power load forecasting model together with the power load data.

5.3 Forecast results and analysis

To verify the effectiveness of the load forecasting method of the SSA-LSSVM model proposed in this paper, we set up the following three scenarios.

Scenario 1: considering the coupling of economy and meteorology, LSSVM model is selected;

Table 2 Calculation results of Euclidean distance

	Frank–Copula	t-Copula	Normal-Copula
Electricity load—GDP	0.6995	**0.2226**	0.2560
Electricity load—primary Industry GDP	0.7359	**0.6393**	0.6940
Electricity load—secondary Industry GDP	0.9389	**0.1090**	0.8294
Electricity load—tertiary Industry GDP	0.9396	0.7188	**0.6384**
Electricity load—average high temperature	0.9391	0.6005	**0.0664**
Electricity load—average low temperature	0.1675	**0.0349**	0.4809
Electricity load—average temperature	**0.1749**	0.2305	0.7630
Electricity load—monthly average gas load	0.7905	**0.3054**	0.3427
Electricity load—monthly average gas peak value	0.6457	**0.4006**	0.4294
Electricity load—monthly average gas valley value	0.4527	**0.0853**	0.9136

Table 3 Results of correlation analysis with electricity load

Factors	Relevance	Factors	Relevance
GDP	0.26132	Average low temperature	0.53026
GDP of primary industry	0.39566	Average temperature	0.78064
GDP of secondary industry	0.84917	Monthly average gas load	0.94390
GDP of tertiary industry	0.67433	Monthly average gas peak	0.78339
Average high temperature	0.85710	Monthly average gas valley	0.43785

Scenario 2: considering the coupling of economy and meteorology, SSA-LSSVM model is selected;

Scenario 3: considering the coupling of economy, meteorology and gas demand, the prediction model combining SSA-LSSVM is selected.

Under scenario 1, scenario 2 and scenario 3, the monthly power load forecasting curve and forecasting error percentage of s city are shown in Fig. 6. Compared to the actual load, the percentage of forecast error for each scenario is 1.93%, 1.43% and 0.52%, respectively. In Scenario 3, the percentage of monthly load forecast error less than 2% is 50%, while Scenario 1 and Scenario 2 account for 0% and 33.33%, respectively.

Table 4 Posterior error test and small error probability of each scenario

Scenario	Scenario 1		Scenario 2		Scenario 3	
	C	P	C	P	C	P
Value	0.2203	95.64%	0.1914	96.22%	0.1563	98.36%

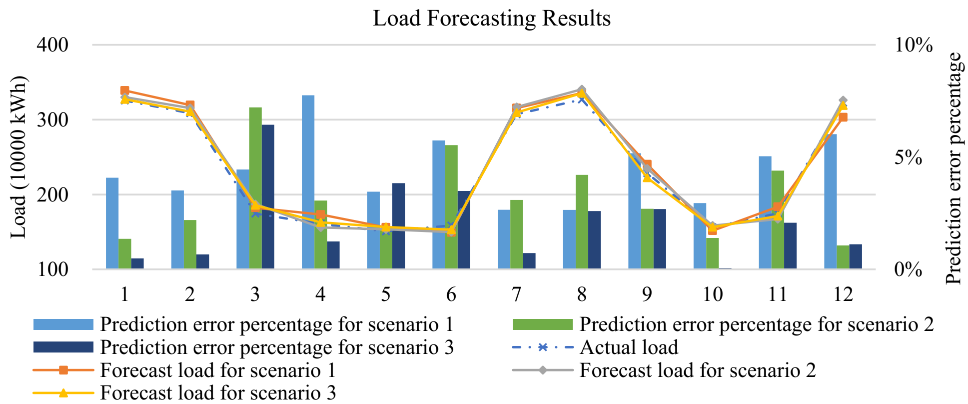
It can be seen from Table 4 that when using SSA-LSSVM model for load forecasting, the posterior error test of scenario 3 is 0.1563, less than 0.35, and the small error probability is 0.9836. The forecasting result belongs to the first level, and the overall forecasting accuracy is high. Therefore, the SSA-LSSVM model constructed in this paper has good prediction accuracy in the urban monthly power load forecasting model. This shows that considering the coupling of economy, meteorology, and gas demand can effectively improve the effectiveness of monthly electricity load forecasting.

6 Conclusion

This paper mainly establishes the urban monthly power load collaborative forecasting model based on Copula theory and SSA-LSSVM. Based on Copula theory, the model focuses on the interaction between economic development, meteorological conditions, natural gas demand and power load, which makes up for the deficiency of the existing load forecasting models in the interaction of internal and external factors. Secondly, SSA algorithm is used to optimize the key parameters of LSSVM to further improve the accuracy of load forecasting. The prediction accuracy of the prediction model is verified by an example. Take the research results as a new service means to provide more professional and quantitative service products for the power industry, provide scientific basis for power grid operation and dispatching, and enhance the vitality of professional meteorological services.

However, there are many factors affecting urban power load, and the relationship between the factors is very complex. The model established in this paper has not considered the influence of power substitution policy, energy conservation and carbon reduction policy and other factors, and has less consideration of air humidity, sunshine conditions and other factors in terms of Meteorological factors. In the next research, we can consider how to further refine the economic, meteorological, gas and other factors, and consider the empirical mode decomposition of urban power load to

Fig. 6 Load forecasting results



stabilize each component and further improve the accuracy of load forecasting.

Funding This research was funded by the science and technology project of State Grid Corporation of China: research and application of key technologies of BIM based smart city power grid planning information fusion project No.: 5400-202013121a-0-0-00.

Data availability The data used to support the findings of this study are included within the article.

Declarations

Conflict of interest The authors declare that they have no competing interest.

References

1. Boicea VA (2021) Energy management: big data in power load forecasting. CRC Press, Boca Raton
2. Grzegorz D, Pawel P, Slawek S (2020) A hybrid residual dilated LSTM and exponential smoothing model for midterm electric load forecasting. IEEE Transact Neural Networks Learn Syst 99:1–13
3. Wang S, Wang S, Wang D (2019) Combined probability density model for medium term load forecasting based on quantile regression and kernel density estimation. Energy Proc 158:6446–6451
4. Neeraj N, Mathew J, Agarwal M et al (2020) Long short-term memory-singular spectrum analysis-based model for electric load forecasting. Electrical Eng 1:1–16
5. Jin H, Xiaowen W, Xingchen P (2019) Analysis of power demand in Liaoning Province under the new economic normal. J Shenyang Institut of Technol (Social Science Edition) 15(03):344–349
6. Ardakani FJ, Ardehali MM (2014) Long-term electrical energy consumption forecasting for developing and developed economies based on different optimized models and historical data types. Energy 65:452–461
7. Han Haian D, Mengke CH, Zonglin Y (2018) Research on power load forecasting considering economic transformation and development. Renew Energy 36(08):1197–1203
8. Wei J, Chen Y, Fei G, Na Y, Jing M, Bao W (2018) Research on power demand of Anhui Province based on the characteristics of socio-economic saturation development. Power Demand Side Manag 20(02):16
9. Sobhani M, Campbell A, Sangamwar S, Li C, Hong T (2019) Combining weather stations for electric load forecasting. Energies 12(8):1510
10. Gilkeson WR (1939) The use of weather forecasts in electric power system operation. Bull Am Meteorol Soc 20(8):338–340
11. Kun P, Xingying C, Bin L, Yingchen L, Kun Y (2016) Analysis on the influence factors of meteorological environment on power grid load. Power Demand Side Manag 18(01):8–13
12. Yiwei C, Qiang G, Limei W, Xiulei L, Lixin L (2019) Research on load coordinated forecasting method based on information re correction. Power Eng Technol 38(02):93–98
13. Cong Xiaohan Su, Huiling LH, Beibei W (2019) Research on personalized intelligent power consumption package based on data mining and demand response. Power Demand Side Manag 21(05):21–25
14. Ruimiao W, Xiaojuan Z, Yue S, Xinyi Le, Zheng L, Wei H, Cheng Xin Hu, Bo XK (2021) Combined forecasting method of daily electricity consumption in summer based on meteorological mapping function. New Technol Electrical Energy 40(03):55–62
15. Wang S, Wang S, Chen H, Qiang G (2020) Multi-energy load forecasting for regional integrated energy systems considering temporal dynamic and coupling characteristics. Energy 195:116964
16. Chen S, Zhang X, Wei S, Yang T, Guan J, Yang W, Lijuan Q, Yunqing X (2019) An energy planning oriented method for analyzing spatial-temporal characteristics of electric loads for heating/cooling in district buildings with a case study of one university campus[J]. Sustain Cities Soc 51:101629
17. Shouxiang W, Kaixin W, Qianyu Z, Shaomin W, Liang F, Zhijie Z, Guo W (2021) Multienergy load forecasting for regional Integrated Energy Systems Considering Multienergy coupling of variation characteristic curves. Front Energy Res
18. Chen B, Wang Y (2021) Short-term electric load forecasting of integrated energy system considering nonlinear synergy between different loads. IEEE Access 99:1–1
19. Lv C, Hao Y, Li P, Wang C, Xiandong X, Li S, Jianzhong W (2019) Model predictive control based robust scheduling of community integrated energy system with operational flexibility. Appl Energy 243:250
20. Qi X, Zheng X, Chen Q (2020) A short term load forecasting of integrated energy system based on CNN-LSTM. Web Confer 185:01032
21. Xiaoman L, Fucheng H, Jiangjun R et al (2019) Heat and power load forecasting of cogeneration unit based on neural network. Thermal Power Gener 48(09):46–50
22. Chao Q, Xiaofeng D, Tong J (2018) Optimization planning of power and natural gas regional integrated energy system based on CCHP coupling. Power Grid Technol 42(08):2456–2466

23. He YX, Zhang JX, Xu Y et al (2015) Forecasting the urban power load in China based on the risk analysis of land-use change and load density. Int J Electr Power Energy Syst 73(10):71–79
24. Barman M, Choudhury N (2020) A similarity based hybrid GWO-SVM method of power system load forecasting for regional special event days in anomalous load situations in Assam, India. Sustain Cities Soc 61(3):102311
25. Yunxuan D, Xuejiao M, Tonglin F (2021) Electrical load forecasting: a deep learning approach based on K-nearest neighbors. Appl Soft Comput J 99:106900
26. Tan Z, De G, Li M, Lin H, Yang S, Huang L, Tan Q (2020) Combined electricity-heat-cooling-gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine. J Clean Prod 248:119252
27. Rahman A, Srikumar V et al (2018) Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl Eenergy 212:372–385
28. Jieyang P, Andreas K, Jiahai W, Xiufeng L, Zhibin N, Jivka O (2021) Dual-stage attention-based long-short-term memory neural networks for energy demand prediction. Energy Build 249:111211
29. Yuntian C, Dongxiao Z (2021) Theory-guided deep-learning for electrical load forecasting (TgDLF) via ensemble long short-term memory. Adv Appl Energy 1:100004
30. Yongli W, Minhan Z, Suhang Y, Mengju W, Yang Y, Zhao L, Mengjin H Multi load collaborative forecasting model of integrated energy system based on multi energy coupling mechanism. J North China Electric Power Univ (Natural Science Edition), {3}, {4}, {5}: 1-10
31. Zhou X, Zhang G, Shaohua Hu, Li J, Ji J (2019) Optimal estimation of shear strength parameters based on copula theory coupling information diffusion technique. Adv Civil Eng 2019:8738969
32. Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM (2017) Salp swarm algorithm: a bio-inspired optimizer for engineering design problems. Adv Eng Softw 114(12):163–191