International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 1906-1923

https://doi.org/10.70517/ijhsa463151

Carbon Emission and Prediction Model for Power System Based on Machine Learning Algorithm

Ximing Zhang[1], Huan Xu[1], Yingqiao Ling[2], Fan Zhang[2] and Yanlu Huang[2,*]
[1] China Southern Power Grid Co., LTD., Guangzhou, Guangdong, 510000, China
[2] Southern Power Grid Artificial Intelligence Technology Co., LTD., Guangzhou, Guangdong, 510000, China
Corresponding authors: (e-mail: huangyanlu_20@163.com).

Abstract Global climate change has become the focus of the international community's attention, and in order to cope with the challenges it brings, China has set the goal of striving to achieve carbon peaking by 2030 and carbon neutrality by 2060. Based on the sparrow search algorithm, this paper proposes a least squares support vector machine method to solve the problems of predictive pattern classification and function estimation, and simplify the complexity of calculation. The carbon emission boundary of the power system is clarified, and the fitting function of the carbon emission of the power industry based on the night lighting data is constructed, taking into account the night lighting problem of the power system, to further improve the carbon emission prediction accuracy of ISSA-LSSVM. The prediction effect of ISSA-LSSVM is validated using ten-fold cross validation, and the experimental results show that the model has the highest fit, with a residual square of 0.08821 and a Pearson of 0.98876, which is better than other models. Predicting the predicted carbon emissions of the subject provinces under four scenarios, namely, green development scenario, low carbon scenario, baseline scenario and high carbon scenario, it is found by analyzing the data in 2030 that under the baseline scenario, the carbon emissions are 118.979Mt, which is an increase of 30.8427Mt compared to 2022, an increase of 34.994%, and the annual growth rate of the carbon emissions is 3.888%, and the baseline scenario dominates in carbon emissions.

Index Terms Sparrow Search Algorithm, Least Squares Support Vector Machine, Carbon Emission Boundary, Nighttime Lighting Data, Carbon Emission Forecasting

I. Introduction

Environmental problems caused by carbon emissions are becoming increasingly serious in today's world, and with the introduction of the dual-carbon target, carbon reduction and carbon measurement in the process industry have become very important [1]. Due to the rising energy demand and the increasingly significant impact of carbon emissions on the environment, the global power industry is accelerating its transformation towards cleaner and lower carbon development [2], [3]. As an important part of the power industry, the power market coordinates and manages power generation, power sales, power transmission, and customer relations in the power system [4]. Among many industries, the power industry is often regarded as the largest user of coal and the most important source of carbon emissions, and the proportion of electricity in the end-use energy consumption has been increasing year by year, in which the power system plays an important role as the main force [5]-[7]. A large number of new energy sources connected to the power system in high proportion will reduce the dependence of the power system on traditional energy sources, reduce the CO2 gas emission of the system, and greatly change the distribution of power system currents, but the distribution of currents has a close relationship with the transmission of carbon emissions from the power system, which in turn generates the problems of carbon measurement, carbon reduction and optimization of the power system brought about by new energy sources connected to the power system [8]-[11].

In response to the above research problems, carbon emission prediction models for power systems with the themes of formulating and implementing low-carbon policy measures, increasing low-carbon constraints to optimize power system operation, and researching and applying low-carbon technologies have been put forward in large numbers. Yu, H. et al. constructed an early warning system for carbon emission in a community based on the Support Vector Regression (SVR) model and Genetic Optimization Algorithm (GA), which helps to assist in the formulation of strategies related to carbon emission reduction in the community by obtaining the dynamic emission coefficient of power system curve, and make reasonable prediction of the carbon emission trend of the community's future electricity consumption, which helps to assist the formulation of strategies related to community carbon emission reduction [12]. Wu, Z. et al. simulated and analyzed carbon emissions under different scenarios based on a system dynamics model, so as to predict the time of the peak carbon value of the electric power industry and the specific peak value, and help the electric power system to achieve the goal of "double carbon" [13]. Luo, J. et al. combined multiverse quantum and acoustic search algorithms and dynamic fuzzy system integration methods to form a composite prediction model to forecast carbon emissions and low-carbon economic development in the energy market, which provides scientific support for future decision-making in energy structure optimization and other aspects [14]. Bao, X., et al. In order to address the problems of large carbon emissions and high energy consumption in the thermal power industry, a system dynamics model was used to simulate the carbon emissions and development trends of the thermal power industry under different scenarios, and to promote the implementation of carbon emission reduction measures in the thermal power industry [15]. Wang, X. et al. introduced methods such as Extreme Gradient Boosting Algorithm (XG-Boost) to design carbon emission influencing factor indicators for coal-fired power plants, and at the same time, utilized Sparrow Search-Long and Short-Term Memory Network Hybrid Algorithm to establish a generalized regression prediction model of carbon emission, which provides a reference to formulate a targeted carbon emission reduction strategy [16]. Chen, Z. et al. showed that accurate load forecasting is an important basis for integrated energy systems to reduce carbon emissions while improving energy efficiency, and proposed the construction of a multi-scale fusion convolutional neural network to extract the multi-scale features of the energy system data so as to guide the energy industry to carry out low-carbon production and storage [17]. Wang, H. et al. proposed the improved stochastic impact regression population, affluence and technology (STIRPAT) model, which was applied to the carbon emission prediction of the electric energy market, effectively guaranteeing the electric power industry to produce electricity in a low-carbon and clean form [18]. It is not difficult to see that the use of machine learning algorithms to construct a carbon emission assessment model of the power system is of practical significance for determining emission reduction strategies, optimizing the energy structure, improving the efficiency of energy use, and promoting energy conservation and emission reduction.

In this paper, three accounting methods for carbon emissions are summarized and generalized according to the different conditions of use. Distinguishing from the traditional accounting methods, this paper proposes a novel machine learning optimization algorithm, which performs local and global search through the behavior of individual sparrows, and demonstrates the steps of using the SSA algorithm. In order to simplify the computational complexity of this algorithm, a least squares support vector machine model is introduced and the boundaries of carbon emission measurement are defined. The influencing factors of the power industry system are identified, the indicator set of carbon emission influencing factors in the power industry is established, and the extended STIRPAT model for carbon emission forecasting in the power industry is constructed. The ten-fold cross-validation method is used to evaluate the performance and interpretability of the model through simulation experiments. Different carbon emission scenarios are set up, and the model is used to predict the carbon emissions under these scenarios. Countermeasures are proposed to address the problems of carbon emissions in terms of resources, system security and economic constraints.

II. Introduction to theories related to carbon emissions

II. A. Theories related to carbon emission projections

II. A. 1) Emission factor approach

Carbon emission accounting can quantify the emission of carbon dioxide, through the carbon emission data can be targeted to find out the possibility of carbon emission reduction, which is of great significance for realizing the dual-carbon target [19]. At present, there is no uniform standard for the accounting method of carbon emissions, according to the different conditions of use, the main methods of measurement are emission factor method, mass balance method and actual measurement method.

Carbon emission factor method is the earliest accounting method and is widely used. The carbon emission factor method is based on the carbon emission factors in the IPCC Guidelines for National Greenhouse Gas Inventories and the consumption of various types of energy to determine carbon emissions. The formula is as follows:

$$E = AD \times EF \tag{1}$$

Where E is the amount of carbon emissions, and AD is the amount of each type of production and consumption activity that generates carbon emissions, such as the consumption of fossil fuels. EF is the carbon emission factor corresponding to each type of energy activity. According to the assumption of IPCC, the carbon emission factor is constant in the process of energy consumption.

Based on the carbon emission factor, the carbon emission of energy can be calculated, indicating the carbon elements contained in each ton of standard coal. The carbon emission in this paper refers to the emission of carbon dioxide, so it is necessary to get the amount of carbon dioxide dispatched per unit of energy after full combustion. The National Development and Reform Commission stipulates that the emission coefficient of carbon dioxide is 2.4567, indicating that each ton of standard coal will produce 2.4567 tons of carbon dioxide after full combustion. Therefore, carbon dioxide emissions can be obtained by combining the carbon emission factors obtained for each type of energy source with the carbon dioxide emission factor.

II. A. 2) Mass balance method

The mass balance approach is a new method of accounting for carbon emissions that has emerged in recent years. The mass balance method allows for the calculation of the share of new chemical substances consumed to meet the capacity of new equipment or to replace removed gases, based on the new chemical substances and equipment used in the country's productive life each year. The formula is as follows:

$$E_{CO_2} = (I \times C_l - O \times C_o - W \times C_w) \times \frac{44}{12} \quad (2)$$

where E_{CO_2} is the carbon dioxide emission, I is the input of raw materials, O is the output of products, W is the output of wastes, and C_I, C_O and C_W are the carbon content of the raw materials, products and wastes, respectively. The $\frac{44}{12}$ is the coefficient of conversion of carbon to CO_2.

The method needs to take the emission process into account and is prone to errors. It is mainly used to measure carbon emissions from industrial production processes.

II. A. 3) Actual measurement method

The measurement method is a method of measuring carbon emissions by analyzing the summary of measured data of carbon emissions, which is divided into two methods: on-site measurement and off-site measurement. On-site measurements are generally made by using carbon emission monitoring systems such as CEMS (Continuous Emission Monitoring System) to monitor carbon emissions on site. Off-site measurement refers to the analysis and measurement of carbon emissions by testing samples. Off-site measurements are less accurate than on-site measurement methods because problems such as contamination may occur with the samples.

The on-site measurement method has fewer intermediate links and is currently the most accurate method of carbon emission accounting, but the data requirements of the on-site measurement method are high and it is more difficult to obtain data. At present, the application of on-site measurement method is relatively small, and most of them are off-site measurements.

II. B. Related machine algorithms

II. B. 1) Sparrow Search Algorithm

Sparrow Search Algorithm (SSA) is a new optimization algorithm, the principle of SSA is to simulate the foraging and anti-predator behaviors of sparrows for parameter optimization, local and global search through the behaviors of individual sparrows, with stronger optimization ability and faster convergence [20]. There are three types of sparrows in SSA population, which are the discoverer, joiner and scout. Sparrows as discoverers are well adapted and are responsible for discovering food and providing foraging areas for the entire sparrow population. The joiners use the information provided by the discoverers to find food. Scouts provide scouting warning signals to the population and when a predator is detected, the sparrow population will give up the food and make anti-predator behavior. The steps of SSA algorithm are as follows:

The sparrow population X consisting of n sparrows can be described as:

$$X = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^d \\ x_2^1 & x_2^2 & \dots & x_2^d \\ \dots & \dots & \dots & \dots \\ x_n^1 & x_n^2 & \dots & x_n^d \end{bmatrix} \quad (3)$$

where d is the dimension of the variables of the optimization problem.

(1) Initialize the sparrow population position and fitness:

$$F_x = \begin{bmatrix} f & ([x_1^1 & x_1^2 & \dots & x_1^d]) \\ f & ([x_2^1 & x_2^2 & \dots & x_2^d]) \\ & & \dots & & \\ f & ([x_n^1 & x_n^2 & \dots & x_n^d]) \end{bmatrix} \quad (4)$$

where f is the fitness value.

(2) Sort to derive the current optimal individual position and best fitness

(3) Update the discoverer position:

$$X_{i,j}^{t+1}=\begin{cases}x_{i,j}^{t}\exp\left(\dfrac{-i}{\alpha\cdot I_{\max}}\right),R_2<ST\\ x_{i,j}^{t}+Q\cdot L,R_2\geq ST\end{cases}\tag{5}$$

where t denotes the number of iterations currently performed, j represents the dimension, $X_{i,j}^{t}$ denotes the j-dimensional position of the i th sparrow in the population at the t th iteration, α is a uniform random number in the interval (0, 1], $I_{\max}$ denotes the maximum number of iterations, R_2 is the value of alert Kan , ST is the set warning Kan value, Q is the standard normally distributed random number, and L denotes the $1\times d$ -dimensional all-1 matrix.

(4) Update the predator position:

$$X_{i,j}^{t+1}=\begin{cases}Q\cdot\exp\left(\dfrac{X_{worst}^{t}-X_{i,j}^{t}}{i^2}\right), & i>\dfrac{n}{2}\\ X_{p}^{t+1}+\left|X_{i,j}^{t}-X_{p}^{t+i}\right|\cdot A^{+}\cdot L, & i\leq\dfrac{n}{2}\end{cases}\tag{6}$$

where X_{worst}^{t} is the location of the worst adapted individual in the t th iteration, X_{p}^{t+1} is the location of the optimally adapted individual in the $t+1$ th iteration, and A is a $1\times d$ -dimensional matrix where the elements are preconditioned to be either 1 or -1, and $A^{+}=A^{T}(AA^{T})^{-1}$.

(5) Update the scout position:

$$X_{i,j}^{t+1}=\begin{cases}X_{best}^{t}+\beta\cdot\left|X_{i,j}^{t}-X_{best}^{t}\right|,f_i>f_g\\ X_{i,j}^{t}+K\cdot\dfrac{\left|x_{i,j}^{t}-X_{best}^{t}\right|}{\left(f_t-f_w\right)+\varepsilon},f_i=f_g\end{cases}\tag{7}$$

where X_{best}^{t} denotes the global optimal position in the t th iteration, β is a random number obeying a normal distribution representing the control step, f_i is the fitness value of the current individual, f_g is the fitness value of the current global optimal individual, f_w is the current global fitness value of the worst individual, K is a random number in the interval [-1, 1], and ε is set as a constant to prevent the denominator from taking the value 0.

(6) Calculate the fitness and update the sparrow position.

(7) Determine whether the iteration satisfies the stopping condition, output the optimal sparrow position after satisfying the stopping condition, and return to step 2 if the stopping condition is not satisfied.

Sparrow search algorithm has fast iteration speed and strong optimization ability, and is often used in path planning and image processing, etc. Sparrow search algorithm takes into account all the factors within the population, so that the sparrows of the population move to the optimum, and ultimately converge near the optimal value. However, the sparrow search algorithm tends to fall into the local optimum at the late convergence stage, leading to premature maturation of the algorithm, which results in poor accuracy and stability of the algorithm. Therefore, if the initial selection of the population can be optimized and the position of the population can be perturbed when the population gets the optimal value, the accuracy of the sparrow search algorithm will be improved.

II. B. 2) Least squares vectors

Least Squares Support Vector Machine (LSSVM) is a support vector machine method that can be used to solve problems such as pattern classification and function estimation [21]. Traditional support vector machines use quadratic programming methods, which are difficult to implement for large-scale samples. The LSSVM method, on the other hand, chooses the least squares linear system in terms of loss function, which can simplify the complexity of calculation.

The computational steps of LSSVM are as follows:

(1) Given a training sample set $\{(x_i, y_i), i = 1, 2, ..., n\}$, x_i is the input vector and y_i is the prediction value:

$$y_i = x_i \varphi(xi) + b \tag{8}$$

where ω is the weight vector and b is the bias.

(2) The LSSVM optimization function is solved as:

$$\begin{cases} \min_{\omega,b,e} J(\omega,e) = \frac{1}{2}\|\omega\|^2 + c\frac{1}{2}\sum_{i=1}^{n} e_i^2 \\ \omega\varphi(x_i) + b + e_i = y_i \end{cases} \tag{9}$$

where $\varphi(x_i)$ is the kernel function, c is the penalty parameter, and e_i is the error vector.

(3) Introduce the Lagrangian vector:

$$L(\omega,b,e,\xi) = J(\omega,e) - \sum_{i=1}^{n} \xi_i \left[\omega\varphi(x_i) + \mathrm{b} + e_i - y_i\right] \tag{10}$$

where ξ is the Lagrange multiplier.

(4) Take the partial derivation of the above equation, eliminate ω and e, and use the least squares method to solve b and ξ_i, and finally get the optimization function of LSSVM:

$$y(x) = \sum_{i=1}^{n} \xi_i K(x, x_i) + b \tag{11}$$

where $K(x, x_i)$ is the kernel function.

In this paper, the radial basis kernel function RBF with wide convergence domain and strong generalization ability is selected, and the expression is:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{12}$$

where σ is the width factor of the kernel function.

Least squares vector machine is an improved vector machine method based on statistical theory, characterized by the ability to transform quadratic optimization problems into the solution of linear systems of equations, simplifying the computational process, and is also applicable to the classification and regression tasks in high-dimensional input spaces, but the choice of regularization coefficients and parameters of the kernel function in the least squares vector machine affects the prediction results of the model, and the parameter values of which are too large will lead to the phenomenon of overfitting, while the value of the too small parameter values can lead to model underfitting. Therefore, if the two parameters of the least squares vector machine can be optimized, the accuracy of the algorithm can be improved.

II. C.Carbon Emission Prediction Model Construction of Electric Power System Based on Multivariate Data

II. C. 1) Power System Carbon Emission Boundary and Measurement

Before measuring and analyzing the carbon emissions of the power industry, it is necessary to define its boundaries. Carbon emissions from the power industry generally refer to the carbon emissions generated during the life cycle of energy and electricity, which includes four stages: power generation, power supply, transmission and distribution, and electricity consumption by users, of which the carbon emissions generated by the consumption of fossil fuels, such as coal, oil, and natural gas, are the main ones in the power generation process. In this paper, only the power generation side is considered when measuring carbon emissions from the electric power industry, and the processes of power supply, transmission and distribution, and electricity consumption are considered when studying the factors affecting carbon emissions from the electric power industry.

Secondary energy sources such as electricity are usually calculated using the following conversion formula:

$$CE_e = \sum_{j=1}^{m} AD_{ej} \cdot \alpha_{ej} \cdot EF_{ej} \tag{13}$$

where CE_e denotes the total amount of carbon emissions from electric power, AD_{ej} and EF_{ej} denote the consumption of the i th energy source or material and the carbon emission factor, respectively, and α_{ej} denotes the ratio of the first energy source or material in the electric power production process to the total energy source or material.

In this paper, the updated carbon emission factor is adopted, and the updated carbon emission measurement formula of the electric power industry is adopted as shown below:

$$CE_{ue} = \sum_{k=1}^{K} AD_{ek} \cdot \alpha_{ek} \cdot EF_{uk} + \sum_{l=1}^{L} AD_{el} \cdot \alpha_{el} \cdot \lambda_{el} \cdot \mu \tag{14}$$

where K denotes the number of updated carbon emission factors, L denotes the number of missing carbon emission factors, EF_{uk} is the updated carbon emission factor, λ_{el} denotes the coefficient for converting the l th energy source or material to standard coal, and μ is the carbon emission factor for standard coal.

II. C. 2) Carbon Emission Prediction Model for Power Systems Based on Nighttime Lighting

As the new satellite NPP adopts a more advanced detector VIRS, there are big differences between the acquired data and DMSP/OLS in terms of precision and resolution, etc. In order to unify the two nighttime lighting data formats, it is necessary to preprocess the two kinds of nighttime lighting data in the following steps:

Step (1): Reproject, resample and recrop the DMSP/OLS data from 1992-2013 to get the light data within the target area. For the missing unstable pixel values of some images, their DN values are replaced with 0, and the overly bright noise points present in the image data are removed.

Step (2): re-project, re-sample and re-crop the NPP/VIIRS data after 2012, and adjust the spatial resolution of this data to match the DMSP/OLS data, replace the DN values of the missing unstable pixels of the image to 0, and eliminate the noise points.

Step (3): Construct the regression model of DN values of DMSP/OLS and NPP/VIIRS data in 2012 and 2013, and the model expression is shown in equation (15):

$$f_d(n) = an^2 + bn + c \tag{15}$$

where $f_d(n)$ is the annual DN value of DMSP/OLS nighttime lighting from 2012 to 2013, n is the annual DN value of NPP/VIIRS nighttime lighting from 2012 to 2013, and a,b and c are the model parameters, the values of which are determined by the lighting data of the specific area.

Step (4): time series correction of the two kinds of nighttime lighting data by fitting the model, and finally get the fused lighting dataset.

II. D.Predictive modeling

Based on the measured carbon emissions from the electric power industry CE_{ue}, a regression prediction model is constructed between the total luminance values of nighttime lighting data DN obtained by fusing and correcting the DMSP/OLS and NPP/VIIRS data, where the corrected DN values range from 0 to 63. By analyzing the trend of CE_{ue} and DN values in the study area over the historical time period, different expressions of the fitting function are established, which can be used as the expression of the carbon emissions prediction model for electric power. Be used as the model expression for the prediction of carbon emissions from electricity. The fitting function for the carbon emissions of the electric power industry based on the nighttime lighting data is constructed as follows:

$$y_n = \begin{cases} A \cdot (r_1 \cdot s_1 \cdot DN)^a + B \cdot (r_1 \cdot s_1 \cdot DN)^b + Cor \\ A \cdot (r_1 \cdot s_1 \cdot DN) + Bor \\ A \cdot e^{(r_1 \cdot s_1 \cdot DN)} + Bor \\ A \cdot \ln(r_1 \cdot s_1 \cdot DN) + B \end{cases} \tag{16}$$

where y_n denotes the measured value of carbon emissions from the electric power industry in millions of tons. A,B and C are the parameters of the prediction model, respectively. r_1 and s_1 are the coefficients of the DN value, because the nighttime light data reflect the consumption of electricity, and the electricity consumption of a certain region may partly come from the power supply of the external region, and secondly, only the part of the electricity consumption of the thermal power generates carbon emissions, so when constructing the fit function of a certain region with the DN value and the introduction of and the two parameters, which represent the proportion of the regional power supply and the thermal power proportion of the electricity consumption, respectively, the proportion of internal power supply and the proportion of thermal power. Therefore, when constructing the fitting function with DN value for a certain region, two parameters are introduced and they represent the proportion of regional power supply and the proportion of thermal power in electricity consumption, respectively, and the specific value is determined by the trend of the change with DN value.

II. D. 1) Factors affecting carbon emissions in the power sector

The traditional STIRPAT model includes only three indicators of population, wealth and technology, while the actual carbon emissions are also constrained by many other influencing factors, thus it is necessary to analyze and screen the influencing factors before establishing the extended STIRPAT model of carbon emissions in the electric power industry [22]. Considering the processes that may generate carbon emissions in the stages of power supply, transmission and distribution, and electricity consumption, a set of factors influencing carbon emissions in the power sector is established, as shown in Table 1.

Table 1: Influencing factors of carbon emissions in the power industry

Influencing factor	Symbol	Definition
Population scale (Ten thousand)	P	The number of permanent residents in the region
Per capita GDP (yuan/per)	A	Regional GDP and permanent resident population e
Urbanization rate	P_s	The number of permanent urban residents in a region/the number of permanent residents
Industrial structure	I_s	The added value of the secondary industry in the region and the regional GDP
Electricity consumption (billion kilowatt-hours)	E_l	The total electricity consumption of the entire society in the region
The proportion of new energy power generation	E_s	Regional new energy power generation/power generation
Standard coal consumption for power generation (grams per kilowatt-hour)	E_c	The consumption of each 1kwh of electricity generated by the regional power enterprise Standard coal quantity
The proportion of new energy installed capacity	E_n	The installed capacity of new energy power generation in the region is high
Energy consumption: 10,000 tons of standard coal.	E_e	Total regional energy consumption

II. D. 2) Predictive modeling

Based on the set of influencing factors in Table 1, the carbon emission prediction model of the power industry can be constructed, and based on the traditional model, the quadratic term of GDP per capita is introduced, and the extended STIRPAT model for carbon emission prediction of the power industry is constructed, with the following expressions:

$$\begin{aligned}\ln y_c = \alpha_1 \ln P + \alpha_2 \ln P_s + \alpha_3 \ln A + \alpha_4 \left(\ln A\right)^2 + \alpha_5 \ln I_s + \alpha_6 \ln E_l \\ + \alpha_7 \ln E_s + \alpha_8 \ln E_c + \alpha_9 \ln E_n + \alpha_{10} \ln E_e + \ln e\end{aligned} \tag{17}$$

where y_c denotes the measured value of carbon emissions from the electric power industry in millions of tons, α_1 to α_9 are the elasticity coefficients of the influencing factors, and e is the model error term.

However, due to the differences in socio-economic and energy data of each country or region, the influencing indicators involved in the above model may not always have the greatest impact on carbon emissions, and thus it is necessary to screen the importance of the selected indicators, this paper utilizes the degree of importance of the indicators and screens them, and the screened indicators are used as the independent variables in the final extended STIRPAT model for carbon emission prediction.

In the process of ordinary regression of the extended STIRPAT model, there may be multicollinearity between multiple independent variables, which is easy to interfere with the relationship between the independent variables and the dependent variable, thus leading to errors in the prediction results. Thus, this paper adopts the ridge regression method to analyze the extended STIRPAT model. The main idea of the ridge regression method is to first convert the log-linear equation of Eq. (17) into matrix form as follows:

$$\ln y_c = \ln X \cdot a + \ln e \tag{18}$$

where $\ln X$ is the independent variable matrix, a is the coefficient matrix, and $\ln e$ is the constant matrix. Usually linear regression is estimated by ordinary least squares method with the following formula:

$$a = ((\ln X)^{'} \ln X)^{-1} (\ln X)^{'} \ln y_c \quad (19)$$

When there is covariance between the independent variables, the change in the values of $\ln X$ tends to converge resulting in the rank of the matrix $(\ln X)^{'} \ln X$ being less than n, making α meaningless. To solve this problem, the ridge regression method adds a constant matrix Kc to this matrix, where K is positive, so that the independent variable covariance problem can be handled. The estimation formula based on ridge regression is as follows:

$$a(K) = ((\ln X)^{'} \ln X + Kc)^{-1} (\ln X)^{'} \ln y_c \quad (20)$$

where, where is the ridge regression parameter, by adjusting the K value can be obtained under different ridge regression estimation of a, when adjusting the value to no longer change tends to stabilize, you can choose this value as the final coefficient matrix, can be largely eliminated ordinary least squares estimation of the regression equation caused by the covariance between the independent variables.

II. D. 3) Carbon Emission Forecasting Model for the Electricity Sector

In order to further improve the accuracy of carbon emission prediction, this section constructs a carbon emission prediction model for the power industry based on multi-source data. Based on the advantage of the power demand forecasting model in dealing with high-dimensional multi-source data, this section also adopts the improved LSSVM model to forecast carbon emissions from the power industry.

The main idea of the model is to input the nighttime lighting data and the filtered indicator data as well as the residual value (the difference between the real value and the predicted value) in the respective prediction model into the prediction model as the new carbon emission impact indicator of the power industry. On the one hand, the model in this section takes into account the prediction residuals of the models in the previous two sections, which can be used to correct the prediction model, thus reducing the error of the prediction process, and on the other hand, it takes into account the influence of multi-source data on the carbon emissions of the electric power industry, which allows the model to learn more comprehensive and comprehensive sample information, thus improving the prediction accuracy of the model. The main flow chart of the model is shown in Figure 1, and the main steps are as follows:

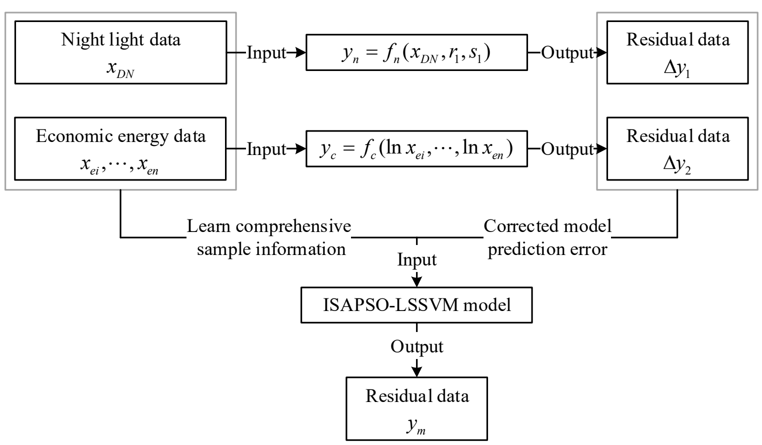


Figure 1: Main flowchart

The residual values of the predicted and true values of the prediction model are calculated separately, and the two columns of values are used as the new feature data sources, followed by the socio-economic data sources with the nighttime light brightness values DN and those screened by the random forest algorithm as the model data inputs, and all the input data sources of the model are normalized. The main prediction model expression is:

$$\Delta y_1 = y_d - y_{real} \quad (21)$$

$$\Delta y_2 = y_c - y_{real} \quad (22)$$

$$y_m = f_{ILSSVM}(\Delta y_1, \Delta y_2, x_{DN}, x_{ei}, \cdots, x_{en}) \tag{23}$$

Among them, y_m represents the estimated value of carbon emissions in the power industry in million tons, $f_{ILSSVM}(x)$ represents the ISSA-LSSVM model proposed in this paper, Δy_1 and Δy_2 represent the difference between the predicted value and the actual measured value of carbon emissions in the power industry, x_{DN} is the nighttime light brightness value, and x_{ei} is the difference between the predicted value of carbon emissions in the power industry and the actual measured value, x_{ei} to x_{en} indicates the filtered socio-economic and energy statistical indicators.

III. Prediction of carbon emissions from power systems based on machine learning algorithms

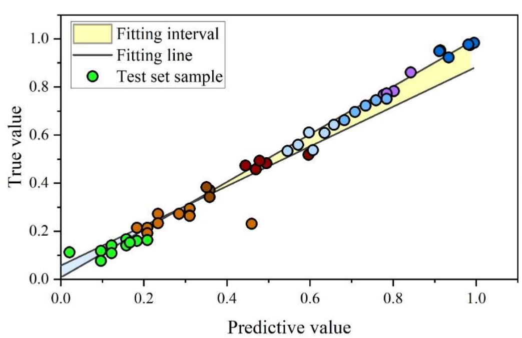
III. A. Analysis of model prediction results

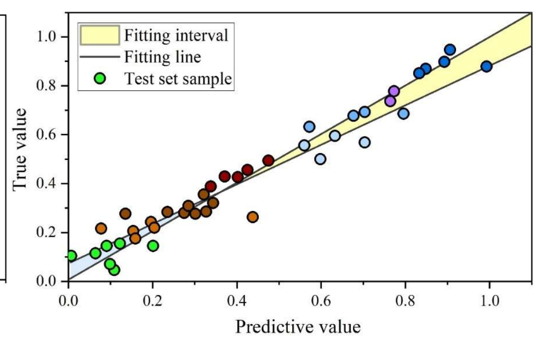
III. A. 1) Analysis of machine learning model prediction results

Cross-validation is suitable for training optimization models when there is insufficient data, and in the paper, ten-fold cross-validation is used to assess the prediction effect of each machine learning model using R^2 as the evaluation index. The entire sample is divided into 10 equal-sized sample subsets, 9 copies are randomly selected as the training set each time, and the remaining 1 copy is used as the test set, which is repeated for ten times, and the results of these 10 independent tests are averaged as an indicator for evaluating the model performance.

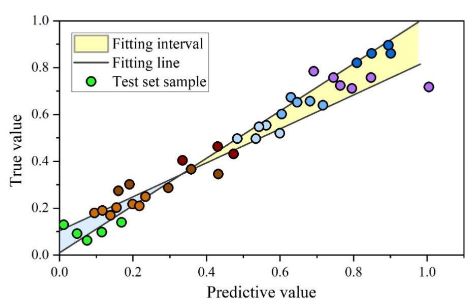
The hyperparameters of a machine learning model have a great impact on its regression performance, and manual parameter tuning is time-consuming and ineffective, usually with the help of optimization algorithms. Bayesian tuning is a hyperparameter tuning method based on the Bayesian optimization algorithm, which, through the Gaussian process, takes into account the previous prior information when selecting hyperparameters, constitutes new combinations for the next round of attempts, and gradually reduces the search space until the optimal hyperparameter combination is reached.

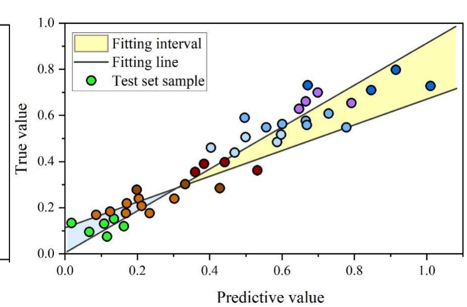
Fig. 2 shows the machine learning fitting graph, Fig. (a) shows ISSA-LSSVM model, Fig. (b) shows XGboost model, Fig. (c) shows LightGBM model, Fig. (d) shows KNN model, Fig. (e) shows Adaboost model, ISSA-LSSVM model has the highest fit, the sample of the test set is close to the line of fit, and the residuals squared is 0.08821, the Pearson of 0.98876 and adjusted R^2 = 0.97718, which is better than other models.



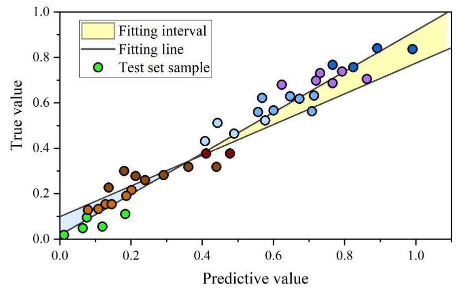


(a)ISSA-LSSVM (b)XGboost





(c)LightGBM (d)KNN



(e)Adaboost

Figure 2: Machine learning fitting

Performance evaluation

In order to compare the prediction level of each model, mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2) are used as the evaluation indexes, and the calculations are shown in Eqs. (24) to (26). MAE is the average of the absolute values of the errors of the true value and the predicted value, which indicates the size of the actual prediction error, and the closer the value is to 0 the better it is, and RMSE is the standard deviation of the difference between the true and predicted values. RMSE is the standard deviation of the difference between the true value and the predicted value, which indicates the accuracy of the model, and the closer its value is to 0, the better it is, R^2 indicates the goodness-of-fit of the model, and the closer its value is to 1, the better it is, and in general, when R^2 is greater than 0.7, it can be assumed that the model is a better fit to the data:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - y_i^{'} | \tag{24}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - y_i^{'})^2} \tag{25}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - y_i^{'})^2}{\sum_{i=1}^{n} (y_i - y_{mean})^2} \tag{26}$$

where: n is the number of samples, y is the true value, $y^{'}$ is the predicted value, and y_{mean} is the mean value of the true value.

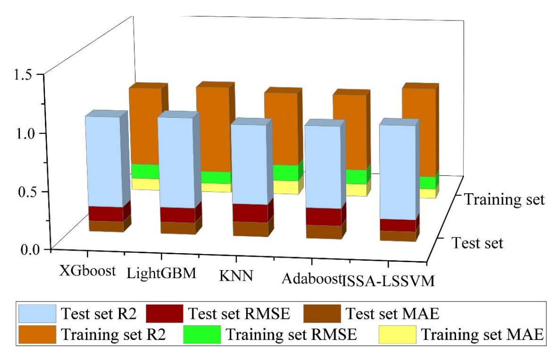


Figure 3: Performance evaluation index of machine learning model

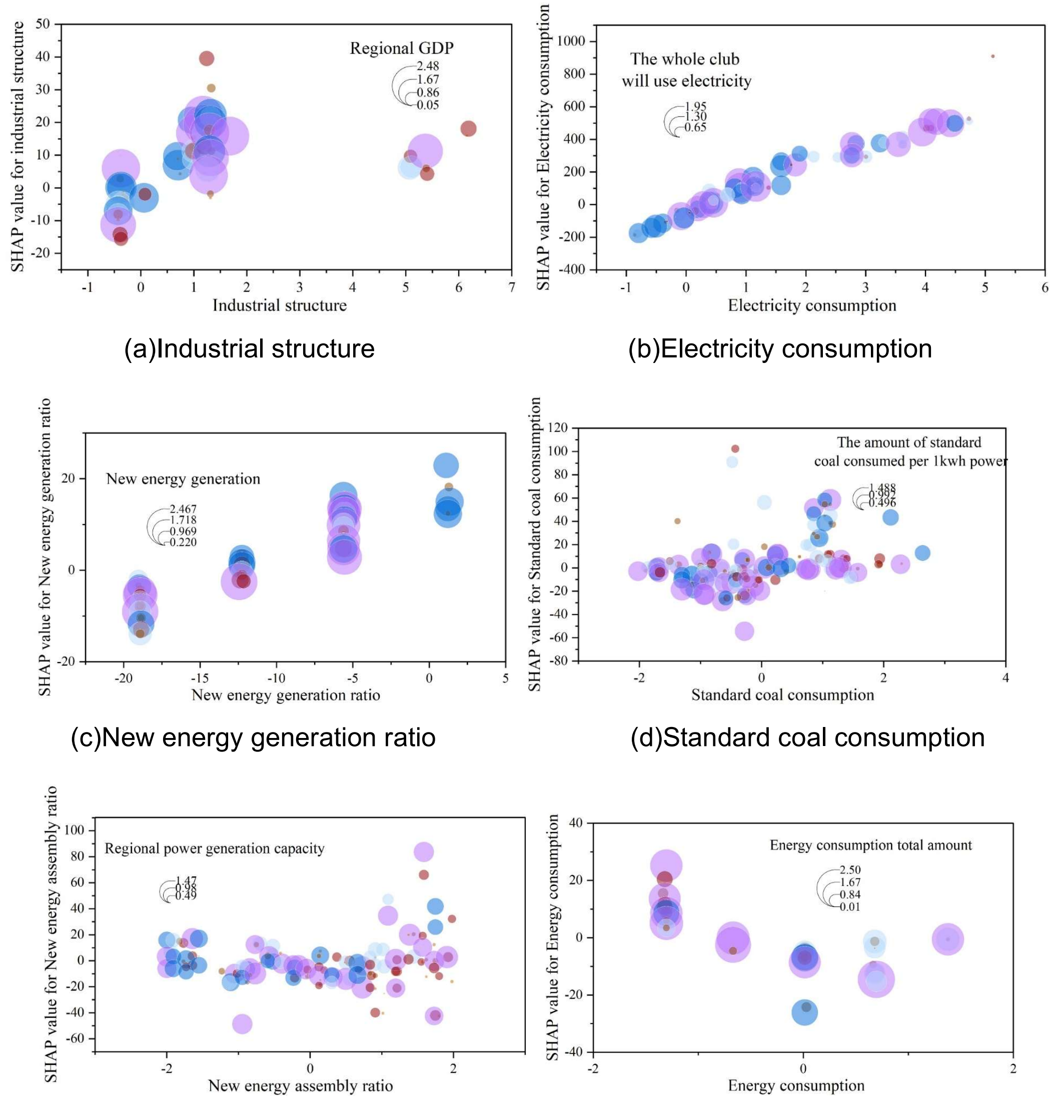
Five machine learning models are used to predict the carbon emissions of the power system, and the performance evaluation indexes of the test set and the training set are obtained as shown in Fig. 3. Analyzing from a single model, the ISSA-LSSVM model has the best prediction effect, and it can predict the test set data more accurately after learning from the training set, and the data of MAE, RMSE and R^2 of the ISSA-LSSVM model's performance indexes in the test set are 0.0848, 0.1054, and 0.8165, respectively, and the MAE and RMSE are in the 5 kinds of models closest to 0, and R^2 is greater than 0.7 and closest to 1. The reason for this is analyzed because the ISSA-LSSVM model is good at predicting with less training data and relies on random selection of dividing points, and has good generalization performance, and in terms of the overall effect analysis, the machine learning model is more outstanding, and because it is an integration of a single model, it usually has high accuracy and stability, can avoid overfitting problems and reduce variance.

III. A. 2) Model interpretability

In order to further obtain the degree and effect of the respective variables, the paper utilizes the SHAP interpretability analysis method to conduct an interpretability analysis of the ET model with the optimal predictive effect.SHAP is a game theory-based method that links game theory to local interpretation by calculating the contribution of an individual in the cooperation to determine the degree of importance of that individual.Let the i th sample be x_i , the j th feature of the i th sample is x_i , the predicted value of the model for this sample is y_i , the mean value of the target variable of all the samples is y_{mean} , and $f(x_{ij})$ is the SHAP value of the current feature, which is the predicted value of the feature for the i th sample contribution level, when $f(x_i) > 0$, the feature has a positive impact on the predicted value: conversely, the feature has a negative impact on the predicted value, and the SHAP value is denoted as:

$$y_i^{'} = y_{mean} + f(x_{i1}) + f(x_{i2}) + f(x_{i3}) + \cdots + f(x_{ik}) \tag{27}$$

Fig. 4 shows the SHAP relationship dependency diagram, Fig. (a) shows the industrial structure, Fig. (b) shows the electricity consumption, Fig. (c) shows the proportion of new energy power generation, Fig. (d) shows the standard coal consumption of power generation, Fig. (e) the proportion of new energy installed capacity, and Fig. (f) shows energy consumption, which show the carbon emission influencing factors of the electric power industry, respectively. In Figures (a) and (b), the output of the ISSA-LSSVM model is positively affected by the industrial structure and electricity consumption, and the larger these eigenvalues are, the higher the prediction value of the ISSA-LSSVM model is, and there is a clear relationship between the carbon emission level and the model output, especially for the indicator of electricity consumption, and the highest value of the SHAP value reaches about 900 with the increase of electricity consumption. In Fig. (c), as the proportion of new energy generation increases, the carbon emission in the operation phase of the power system also increases, and the highest SHAP value is over 20 when the proportion is 1. The distribution of data points in Figs. (d)~(f) is relatively more dispersed, and the regularity is weaker than that in Figs. (a)~(c), indicating that the standard coal consumption for power generation, the proportion of new energy installed capacity, and the energy consumption have less influence on the carbon emission prediction output of the ISSA-LSSVM model.



(a)Industrial structure (b)Electricity consumption

(c)New energy generation ratio (d)Standard coal consumption

(e)New energy assembly ratio (f)Energy consumption

Figure 4: SHAP dependencies

III. B. Carbon emission projections

III. B. 1) Carbon Emission Projections under Different Scenarios

In this paper, four carbon emission scenarios are modeled as low carbon, baseline, high carbon, and green development. The low-carbon scenario focuses on reducing carbon emissions and assumes that a series of low-carbon emission policies and measures are adopted to achieve the goal of reducing carbon emissions. The baseline scenario represents the trend in carbon emissions under current policy planning and technology levels and serves as a reference for the other scenarios. The high-carbon scenario assumes that the constraints on energy consumption and high-carbon industries are weakened, and that the policies and measures adopted are not effective enough to control carbon emissions, resulting in a higher level of carbon emissions. The green development scenario, on the other hand, further integrates green technology and sustainable development policies on the basis of the low-carbon scenario, and considers the impact of technological progress, such as the improvement of carbon capture technology and clean and efficient combustion technology, so as to reduce carbon emissions to realize the effect of emission reduction.

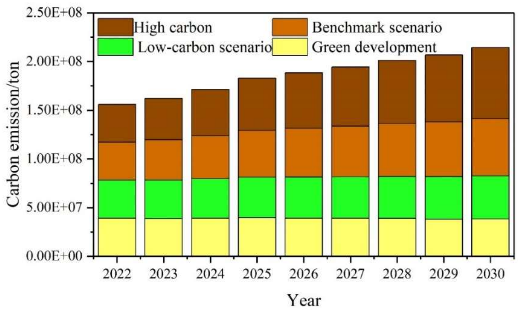
In this paper, based on the four scenarios of low carbon type, baseline type, high carbon type and green development type, the public factor industrial structure and electricity consumption of conventional coal-fired units above 300MW class, conventional coal-fired units of 300MW class and below, and non-conventional coal-fired units from 2023 to 2030 are set, and the optimal model for carbon emission prediction of thermal power generating units (ISSA-LSSVM) is adopted by the Matlab software to run the model, and the public factor industrial structure and electricity consumption were used as model input data to get the predicted value of carbon emission of each coal-fired unit. Figure 5 shows the results of the scenario prediction, Figure (a) is the prediction of carbon emissions from generator group A, Figure (b) is the prediction of carbon emissions from generator group B, Figure (c) is the prediction of carbon emissions from generator group C, and Figure (d) is the overall prediction of the system of electric power enterprises in a certain place.

The carbon emissions of generator group A under the green development scenario show a trend of slowly increasing and then decreasing year by year, and the carbon emissions under this scenario peak in 2025 at 39.548 Mt. From 2022 to 2030, the carbon emissions decrease year by year from 39.106 Mt to 38.337 Mt, a decrease of 0.769 Mt. Compared with the baseline scenario, the carbon emissions under the high carbon type scenario in 2030 increase by an additional 14%, and the carbon emissions under the high carbon type scenario increase by an additional 14%, and the carbon emissions under the high carbon type scenario increase by an additional 14%. Compared with the baseline scenario, the carbon emissions of the high-carbon scenario in 2030 increase by 14.333 Mt. On the contrary, the carbon emissions of the green development scenario and the low-carbon scenario decrease significantly by 20.464 Mt and 14.661 Mt, respectively, which indicates that the impacts of different carbon emission scenarios on the carbon emissions of conventional coal-fired power generation units of 300 MW and above are significantly different, and such power generation units in the green development scenario and the low-carbon scenario show significant advantages in emission reduction. This shows that there are significant differences in the impacts of different carbon emission scenarios on the carbon emissions of conventional coal-fired units above 300MW class, and that such generating units in the green development scenario and the low carbon scenario show significant advantages in emission reduction.

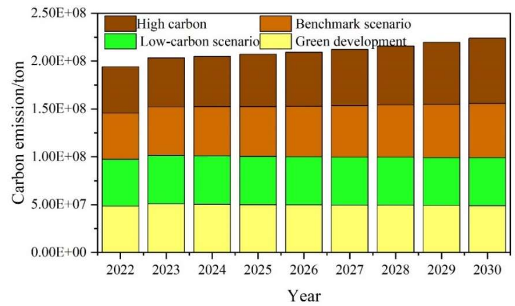
Under both the Green Development Scenario and the Low Carbon Scenario, Genset B shows a positive trend in emission reduction, and both reach their peak carbon emissions in 2023, at 50.783 Mt and 50.642 Mt, respectively. The carbon emissions of these two scenarios start to decrease slightly year by year after reaching the peak, in which the carbon emissions of the green development scenario decrease to 49.002Mt in 2030, and the carbon emissions of the low carbon scenario decrease to 50.027Mt.

Among the carbon emission projections under the four scenarios of non-conventional coal-fired units (Group C), the carbon emission under the green development scenario starts to decrease year by year after reaching the peak of 2.321Mt in 2022, and the carbon emission in 2030 is 1.645Mt, which is a decrease of 0.676Mt compared with that in 2022, with an average annual growth rate of -3.236%. In addition, due to the small number of non-conventional coal-fired units (Group C) with only 2 units, the total carbon emissions are significantly lower than those of Group A and Group B generating units, and its Green Development Scenario and Low Carbon Scenario only provide carbon emission reductions of 0.736Mt and 0.414Mt in 2030 compared with the Baseline Scenario.

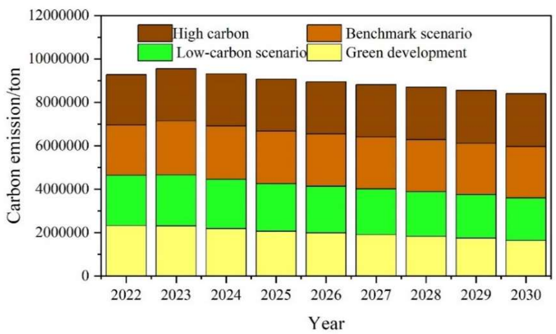
In the previous paper, carbon emission projections have been made for the three groups of generating units A, B and C. The projections are summarized in the following section. Next, this paper obtains the total carbon emissions of the power system as a whole of 18 thermal power generating enterprises in a province during the forecast period by summing up the carbon emission projections of these three groups of generating units under four scenarios: the green development scenario, the low carbon scenario, the baseline scenario and the high carbon scenario. Under the baseline scenario, carbon emissions show a continuous growth. Specifically, the average annual growth rate of carbon emissions under the baseline scenario is 3.888%, and it is found by analyzing the data in 2030 that the baseline scenario carbon emissions are 118.979 Mt, which is an increase of 30.8427 Mt, or 34.994%, compared with 2022. Analyzing the carbon emission share of different unit types, in 2030, the carbon emission share of conventional coal-fired units above 300MW class (Group A) is 48.93%, the share of conventional coal-fired units of 300MW class and below (Group B) is the largest at 49.541%, and the carbon emission share of non-conventional coal-fired units (Group C) is the smallest at 2.346%. Conventional coal-fired generating units Group A and Group B are the main sources of emissions, with a total share of 97.936%, dominating the carbon emissions of the baseline scenario of thermal power generation enterprises in a province.



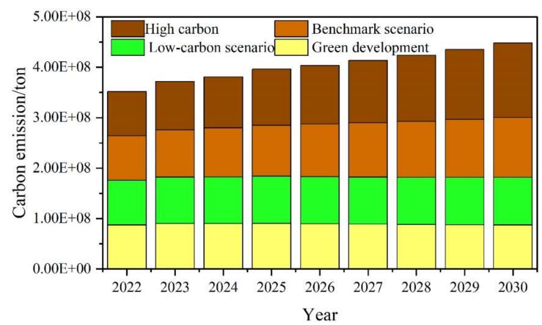
(a)Generator A group carbon emission prediction



(b)Generator B group carbon emission prediction



(c)Generator C group carbon emission prediction



(d)The overall prediction of a local electric enterprise system

Figure 5: Scenario prediction

III. B. 2) Social Costs of Carbon Emissions in the Power Sector in a Peak Carbon Scenario

A total of 18 simulation scenarios were set up by integrating three indicators, namely GDP growth rate, power technology breakthrough time, and change in power consumption intensity per unit of GDP. As shown in Table 2, L, M, and H represent three kinds of GDP growth rates, low, medium, and high, respectively; EC1, EC2, and EC3 represent three kinds of scenarios, namely, increase in power consumption intensity per unit of GDP, unchanged power consumption intensity per unit of GDP, and decrease in power consumption intensity per unit of GDP, respectively; and 2035 and 2050 represent the breakthrough of power technology in 2035 and 2050, respectively.

Table 2: Simulation scenario setting and code

GDP growth	Technology breakthrough time	Increased power consumption	Power consumption intensity is constant	Reduced power consumption
Low GDP	2035	LEC12035	LEC22035	LEC32035
	2050	LEC12050	LEC22050	LEC32050
Mid-GDP velocity	2035	MEC12030	MEC22035	MEC32035
	2050	MEC12050	MEC22050	MEC32050
GDP High speed	2035	HEC12035	HEC22035	HEC32035
	2050	HEC12050	HEC22050	HEC32050

Electricity demand determines the amount of electricity supply, and the amount of electricity supply in the representative nine scenarios is shown in Table 3, and analyzing the amount of electricity supply shows that its size is consistent with the amount of electricity demand. The rule of change is that in the scenario of higher economic growth rate and higher intensity of power consumption, the higher the power demand is, and therefore the larger the power supply is. In the scenario of lower economic growth rate and lower intensity of power consumption, the power demand is lower, so the power supply is smaller, and the power supply in the scenario of LEC32035 in 2060 is 8.1485 trillion kWh.

Table 3: Power Supply Volume in Representative Scenarios(Trillion KWh)

Situation	2020	2030	2040	2050	2060
HEC12035	3.7854	7.6248	12.6148	20.7988	34.7165
HEC22035	3.7854	5.8536	8.3548	12.3157	19.0654
HEC32035	3.7854	5.3215	6.4154	8.4652	10.4264
MEC12035	3.7854	7.1326	11.4987	19.1325	32.7484
MEC22035	3.7854	5.6487	7.6485	10.8484	16.9487
MEC32035	3.7854	5.0698	5.8315	7.6187	9.3487
LEC12035	3.7854	6.0978	10.3264	18.0648	31.1648
LEC22035	3.7854	5.5487	7.4688	10.7488	16.2448
LEC32035	3.7854	4.7987	5.6487	6.8265	8.1485

In the simulation scenarios that can reach the peak during the study period, the peak year and peak data of CO2 emissions are shown in Table 4. A total of five simulation scenarios reach the peak of CO2 emissions in 2030, which are LEC32035, LEC32050, MEC32035, MEC32050, and HEC32035. In the peak scenarios, the peak CO2 emissions of LEC32035 and LEC32050 scenarios are 2,465.5785 Mt and 2,465.5854 Mt, and the peak CO2 emissions of MEC32035 and MEC32050 scenarios are 2548.6695 Mt and 2548.5854 Mt, respectively. Mt, 2548.6695Mt and 2549.4853Mt for MEC32035 and MEC32050 scenarios, and 2618.4854Mt for HEC32035 scenario. It can be seen that a decrease in the intensity of electricity consumption and early breakthroughs in electricity technology will contribute to the realization of the goal of peak carbon attainment. The carbon peak time of the remaining reachable scenarios is between 2031 and 2041, and the peak CO2 emissions are between 2628.1554 and 3426.4151Mt. The earlier the peaking time, the smaller the carbon peak. The earlier the peak is reached when the economic growth rate is smaller and the intensity of electricity consumption is lower. When the economic growth rate and electricity consumption intensity are the same, the early breakthrough of electricity technology promotes the realization of carbon peak.

Table 4: Carbon peak year and carbon dioxide emission peak of 15 simulation scenarios (Mt)

Situation	Peak time	Carbon peak
LEC32035	2030	2465.5785
LEC32050	2030	2465.5854
MEC32035	2030	2548.6695
MEC32050	2030	2549.4853
HEC32035	2030	2618.4854
Situation	Peak time	Carbon peak
HEC32050	2031	2628.1554
LEC22035	2031	2718.6485
MEC22035	2032	2768.6154
HEC12035	2033	3248.4464
HEC22035	2034	2865.4621
Situation	Peak time	Carbon peak
LEC12035	2035	2878.4563
MEC12035	2035	3084.6165
LEC22050	2040	3148.6364
MEC22050	2041	3214.3152
HEC22050	2041	3426.4151

The social cost of power system carbon emissions is analyzed to assess the external costs of different scenarios. The social cost includes the capital cost of power generation equipment as well as fixed and variable maintenance costs, fuel costs, and the externality cost of CO2 emissions, so that the social cost provides a comprehensive picture of the combined cost of carbon emissions from the power system in the specified scenarios. Table 5 shows the social cost of carbon emissions for 2020-2060 in each scenario. Among all the simulated scenarios, the HEC12050 scenario has the highest social cost of 7.996 trillion yuan. The LEC32035 scenario has a social cost of 3.945 trillion yuan, which is the lowest cost for the electric power industry in all the scenarios. The later a scenario peaks or fails to peak, the faster its economic growth rate and the higher its electricity consumption intensity, and therefore the higher the social cost to its electricity generation sector. The earlier the scenario peaks, the lower the social cost. Therefore, promoting the power sector to peak carbon as early as possible has a positive significance in reducing the social cost of the power sector. And under the same economic growth rate and power consumption intensity, the earlier the power technology breakthrough is realized, the lower its total social cost. For example, LEC32035, LEC32050 scenarios, LEC32050 scenarios in 2050 to achieve breakthroughs in power technology, the commercialization of CCS technology, renewable energy equipped with energy storage costs down, then its social cost of 4.426 trillion yuan. LEC32035 scenarios in 2035 to achieve breakthroughs in power technology, for 3.945 trillion yuan. Therefore, the government should vigorously support power technology breakthroughs, for CCS technology, renewable energy generation and energy storage equipment research and development, as well as the cost of technology applications to subsidize, which for the power industry to achieve carbon neutrality as early as possible, and to save the social cost of playing a positive role.

Table 5: Social costs of carbon emissions from electricity in 2020-2060(billion RMB)

Situation	Social cost	Situation	Social cost	Situation	Social cost
HECI2035	7064.4956	MEC12035	6812.4524	LEC12035	6434.6487
HEC12050	7996.4524	MEC12050	7708.6485	LEC12050	7351.6432
HEC22035	5862.3452	MEC22035	5532.4894	LEC22035	5463.9848
HEC22050	6615.4869	MEC22050	6248.3615	LEC22050	6189.6487
HEC32035	4321.6548	MEC32035	4136.6215	LEC32035	3945.4245
HEC32050	4839.9548	MEC32050	4563.9887	LEC32050	4426.6487

IV. Analysis of constraints and responses

In order for the power sector to achieve the goal of carbon neutrality, 75% of the electricity needs to be supplied by new energy generation. The following constraints exist in terms of resource development, system security, and economics in order to create a new power system with greater capacity to consume new energy in the future.

IV. A. Resource development constraints and responses

The objectives, tasks and measures of the development plans of different industries have imposed constraints on new energy development. For example, the spatial planning and land use control of land resources, some local development plans set too many restrictions on the development of the red line, new energy development can be utilized to land resources are very limited, the environmental planning on the one hand requires carbon emission reduction, but on the other hand, do not allow the development. What's more, some people believe that as long as the development is to destroy the environment, environmental protection and resource development in the concept of the formation of the antithesis.

Should further carry out the national wind, solar and other new energy resources survey and assessment, to ensure the accurate and efficient use of resources, maximize solar, photovoltaic, biomass, geothermal and other new energy forms of power generation of the overall power generation, to meet the future growth of the total amount of power generation needs. To form a development pattern combining the scale layout of centralized multi-energy complementary renewable energy power generation bases and the deployment of distributed production and consumption power forms according to local conditions. Promote new energy power generation technology progress, improve the efficiency of new energy resource utilization and power economy, and do a good job of key technology research and development and major engineering layout. At the same time as developing new energy, optimizing the planning of electrochemical energy storage, pumped storage, gas power stations and other flexible power sources to support the continuous improvement of new energy consumption capacity.

IV. B. System security constraints and countermeasures

The high proportion of new energy access poses a great challenge to grid security and operational stability. With the large-scale access of new energy sources, conventional power sources have been replaced in large quantities, and the rotational inertia of the system and its frequency and voltage regulation capabilities have been continuously reduced, so that the risk of large-scale, wide-band and interlocking failures of the power grid has continued to accumulate. In addition, a large number of distributed new energy access to the distribution network may cause system power imbalance, line overload, node voltage overruns and other issues, the reliability of the power supply has brought great challenges, which shows that China's current power system is not ready for large-scale access to renewable energy transformation, more unable to adapt to the future of renewable energy installed capacity proportion of the year-on-year increase in the development trend. And China's power system flexible adjustment capacity there are still short boards, pumped storage power plant, gas power generation and other flexible adjustment power supply proportion of only 6%, large hydropower, additional carbon capture and storage technology of coal power flexibility to transform the capacity benefit of the lack of safeguard policy, constraints on the development of renewable energy.

Fundamentally solve the problem of mismatch between renewable energy development and power system, we should fully explore the potential of flexibility resources from the power supply side, grid side and user side, and guide the coordinated development of various types of flexibility resources of “source network and load reserve” with new planning concepts.

On the power supply side, it is necessary to increase the proportion of flexible power supply in areas rich in new energy resources, plan and build large hydropower stations, pumped storage power stations and gas power stations with good regulation performance, use existing hydropower bases, plan and build “water, wind, light and storage” clean energy bases, give full play to the advantages of the regulation performance of the group of terraced power stations, and optimize the ratio of wind and light resources, on the power supply side, bundle power transmission, and optimize the ratio of wind and light resources, on the power supply side. The power supply side bundles the power to send out, reducing the impact on the power grid.

On the grid side, with the help of artificial intelligence, big data and other new technologies, we can improve the grid's operation and management capability of new energy power generation equipment, establish a high-precision and high-confidence new energy power prediction system, and lay a solid foundation for the optimal operation of a high proportion of new energy resources on the grid. On the user side, energy storage technology can solve the problem of new energy consumption and volatility, regulate the load, and greatly improve the security and stability of the power system. Energy storage system can peak and valley arbitrage, participate in grid demand-side response, provide emergency power backup and other functions, in the field of industrial and commercial energy storage, demand-side response, distributed photovoltaic optimization, charging station expansion, home energy storage and many other user-side has a greater application value.

IV. C. Economic constraints and countermeasures

From the point of view of the overall development of the power system, it is expected that in the “14th Five-Year Plan” and “15th Five-Year Plan” period, the new energy “parity” utilization challenges, but reasonable control of the pace of development, will help to Slow down the cost of new energy utilization. At the same time, it is also necessary to promote the whole society to share the cost of green development through the market competition mechanism. To focus on economic and social benefits synergistic, not only to calculate the “economic accounts”, but also to calculate the “people's livelihood accounts”. First, in the consideration of new energy auxiliary investment based on accelerating technological progress to reduce the cost of new energy power generation further, the second is to take into account the fairness to meet the bottom, to protect the supply of basic public services, to properly deal with the cross-subsidization of electricity prices, to ensure that the residents, agriculture, important public utilities and public welfare services and other electricity prices are relatively stable, the third is to popularize the cost of low-carbon green transformation to the public, and to enhance the understanding and support for the social parties to the price reform. Third, popularize the cost of low-carbon green transformation among the public, enhance the understanding and support of all social parties for the price reform, and form a society-wide synergy to jointly promote the realization of carbon peaking and carbon neutralization.

V. Conclusion

This paper applies the emission factor method, mass balance method and actual measurement method in turn to account for the carbon emission data of the electric power system, based on which the sparrow search algorithm and the least squares support vector machine algorithm are proposed to construct the ISSA-LSSVM model. Between the carbon emission measurement and prediction of the electric power system, the carbon emission boundary is defined, the STIRPAT model is expanded, the set of carbon emission influencing factors of the electric power industry is established, and the prediction research of the carbon emission of the electric power system is launched. The research results of this paper are as follows:

(1) The performance of the model ISSA-LSSVM model constructed in this paper in the test set of MAE, RMSE and R^2 index data were 0.0848, 0.1054, 0.8165, respectively, all of which are the best performance among the five models, which can avoid the overfitting problem and reduce the variance.

(2) Four carbon emission scenario models are set, namely, low carbon, baseline, high carbon and green development, based on which carbon emissions are predicted for different power generation groups from 2023 to 2030. The carbon emission of generator group A under the green development scenario peaks at 39.548 Mt in 2025, and from 2022 to 2030, the carbon emission decreases from 39.106 Mt to 38.337 Mt year by year, a

decrease of 0.769 Mt. This type of generator group under the green development scenario and the low carbon scenario shows a significant advantage in emission reduction.

(3) Integrating the three indicators of GDP growth rate, power technology breakthrough time, and changes in power consumption intensity per unit of GDP, the scenario setting is expanded, and a total of 18 simulation scenarios are set up. In the scenarios with lower economic growth rate and lower power consumption intensity, the demand for electricity is lower, and therefore the supply of electricity is the smaller, and the supply of electricity in the scenario of LEC 32035 in 2060 is 8.1485 trillion kilowatt-hours.

For the power industry to achieve the carbon neutral goal, 75% of the electricity needs to be provided by new energy generation, based on this goal, this paper proposes relevant countermeasures from three aspects: resource development, system security, and economic constraints.

Funding

Technology Project of China Southern Power Grid Company Limited (ZBKJXM20220025).

References

[1] Bouziane, S. E., Khadir, M. T., & Dugdale, J. (2021). A collaborative predictive multi-agent system for forecasting carbon emissions related to energy consumption. Multiagent and Grid Systems, 17(1), 39-58.

[2] Liu, B., & Huo, X. (2024). Prediction of Photovoltaic power generation and analyzing of carbon emission reduction capacity in China. Renewable Energy, 222, 119967.

[3] Leerbeck, K., Bacher, P., Junker, R. G., Goranović, G., Corradi, O., Ebrahimy, R., ... & Madsen, H. (2020). Short-term forecasting of CO2 emission intensity in power grids by machine learning. Applied Energy, 277, 115527.

[4] LUO, B., ZHANG, Y., ZHANG, L., ZHANG, W., & CHENG, Y. (2023). Carbon system structure optimization and carbon emission prediction method and case verification in energy field. Journal of China Coal Society, 48(7), 2657-2667.

[5] Zhang, X., Zhu, H., Cheng, Z., Shao, J., Yu, X., & Jiang, J. (2025). A review of carbon emissions accounting and prediction on the power grid. Electrical Engineering, 1-14.

[6] Li, M., Wang, W., De, G., Ji, X., & Tan, Z. (2018). Forecasting carbon emissions related to energy consumption in Beijing-Tianjin-Hebei region based on grey prediction theory and extreme learning machine optimized by support vector machine algorithm. Energies, 11(9), 2475.

[7] Shi, W., Yang, J., Qiao, F., Wang, C., Dong, B., Zhang, X., ... & Wang, W. (2024). CO2 emission prediction based on carbon verification data of 17 thermal power enterprises in Gansu Province. Environmental Science and Pollution Research, 31(2), 2944-2959.

[8] Li, G., Chen, X., & You, X. Y. (2023). System dynamics prediction and development path optimization of regional carbon emissions: A case study of Tianjin. Renewable and Sustainable Energy Reviews, 184, 113579.

[9] Zhang, Y., Li, X., & Zhang, Y. (2023). A novel integrated optimization model for carbon emission prediction: A case study on the group of 20. Journal of environmental management, 344, 118422.

[10] Zhao, Y., Liu, R., Liu, Z., Liu, L., Wang, J., & Liu, W. (2023). A review of macroscopic carbon emission prediction model based on machine learning. Sustainability, 15(8), 6876.

[11] Wang, H., Wei, Z., Fang, T., Xie, Q., Li, R., & Fang, D. (2024). Carbon emissions prediction based on the GIOWA combination forecasting model: A case study of China. Journal of Cleaner Production, 445, 141340.

[12] Yu, H., Yang, Y., Li, B., Liu, B., Guo, Y., Wang, Y., ... & Meng, R. (2023). Research on the community electric carbon emission prediction considering the dynamic emission coefficient of power system. Scientific Reports, 13(1), 5568.

[13] Wu, Z., Wang, Z., Yang, Q., & Li, C. (2024). Prediction model of electric power carbon emissions based on extended system dynamics. Energies, 17(2), 472.

[14] Luo, J., Zhuo, W., Liu, S., & Xu, B. (2024). The optimization of carbon emission prediction in low carbon energy economy under big data. IEEE Access, 12, 14690-14702.

[15] Bao, X., Xie, T., & Huang, H. (2021). Prediction and control model of carbon emissions from thermal power based on system dynamics. Polish journal of environmental studies, 30(6), 5465-5477.

[16] Wang, X., Yan, C., Liu, W., & Liu, X. (2022). Research on carbon emissions prediction model of thermal power plant based on SSA-LSTM algorithm with boiler feed water influencing factors. Sustainability, 14(23), 15988.

[17] Chen, Z., Zhao, W., Lin, X., Han, Y., Hu, X., Yuan, K., & Geng, Z. (2024). Load prediction of integrated energy systems for energy saving and carbon emission based on novel multi-scale fusion convolutional neural network. Energy, 290, 130181.

[18] Wang, H., Li, B., & Khan, M. Q. (2022). Prediction of Shanghai electric power carbon emissions based on improved STIRPAT model. Sustainability, 14(20), 13068.

[19] Haoyu Chen,Xi Chen,Guanwen Zhou,Linghong Zheng,Ming Xu,Li Yu & Hong Zhang. (2025). Carbon emission accounting method for coal-fired power units of different coal types under peak shaving conditions. Energy,320,135314-135314.

[20] Xing Wang,Haitao Wang,Zhenliang Dong,Yingfei Xiong,Huili Shi & Ping Wang. (2025). An NOMA-VLC power allocation scheme for multi-user based on sparrow search algorithm. Optoelectronics Letters,21(5),278-283.

[21] Niaz Muhammad Shahani & Xigui Zheng. (2025). Predicting backbreak due to blasting using LSSVM optimized by metaheuristic algorithms. Environmental Earth Sciences,84(6),156-156.

[22] Wei Xiang Huang,Chuan Zuo Gao,Bo Wu,Tan Chen,Ting Yang & Bing Zhang. (2025). [Development Path of Guangxi to Reach the Carbon Emission Peak Based on STIRPAT Model]. Huan jing ke xue= Huanjing kexue,46(2),682-695.