Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 15, No. 2, 2024

Exploration on the Development of Photovoltaic Power Generation Path in China Based on the Goal of "Double Carbon"

Yan Zhang, Liyuan Wang, Siying Jing, Baike Wang

School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, 233030, China

Abstract: Electric energy is an economic, practical, clean, and easy to control and transform the form of energy, power industry in the national economy occupies a very important position. This paper focuses on the prediction of the development trend of power supply, analyzes the current situation of domestic power supply by selecting relevant indicators, predicts the development trend of power supply in 2024-2060, and discusses the main influencing factors of its development. At the same time, a dynamic programming model is established, particle swarm optimization is used to initially explore the maximum value of China's total photovoltaic power generation and the impact of related economic and policy factors on the total power generation, and the relevant conclusions are given.

Keywords: Power supply forecast, Photovoltaic power generation, Dynamic programming model, Particle swarm optimization.

1. Introduction

Electric energy is an economical, practical, clean and easy to control and transform energy form, and electric power industry occupies a very important position in the national economy. With the continuous improvement of people's living standards, the demand for electric energy in various countries is also increasing. In the past, China mainly used traditional energy sources such as coal, oil and natural gas to generate electricity. Now, based on the continuous improvement of China's scientific and technological level, renewable energy generation has gradually become the main power supply mode. Photovoltaic power generation as an emerging renewable energy generation technology, can convert solar energy into electricity, is a clean energy, will not have a negative impact on the environment, but also in line with the current energy transformation needs. China put forward the call to strengthen the construction of ecological civilization, in the background of coal, oil and other traditional energy pollution, limited resource potential, photovoltaic power generation is undoubtedly one of the best alternative energy, vigorously develop solar photovoltaic represented by renewable energy, promote renewable energy as the leading new round of energy revolution, can effectively reduce carbon emissions. Addressing the root causes of growing environmental and climate problems.

2. Literature Review

Wu Ping and other scholars on the energy allocation mode, focusing on the analysis of China's power and energy situation in recent years, from the current energy problems in China, power problems and other aspects of the discussion, from the long-distance transmission, power imports, electric vehicle development and grid intelligence of these aspects of the analysis, and for the current situation of China's power energy, put forward solutions [1]. Yang Minying analyzes the main factors involved and the problems caused by some measures, through the change of the power supply and demand situation in recent years from the supply exceeding demand in the late Ninth Five-Year Plan period to the short supply in the early Tenth Five-Year Plan period. Explore the objective law of the occurrence of supply and demand contradiction to promote the development of electric power industry, and explore how to further reduce the contradiction of supply and demand in the future development process [2]. Based on the research on the relationship between power supply, power demand and economic development in China, Zhou Dan conducted in-depth research and drew relevant conclusions from three perspectives: theoretical methods, empirical analysis and prediction [3]. Aiming at the problem of power generation prediction performance caused by non-stationary photovoltaic power generation data, scholars such as Qiu Shuqi and others proposed a photovoltaic power generation prediction method based on improved variational mode decomposition and ensemble learning, and showed through experiments that the proposed method had higher prediction accuracy and smaller error compared with other methods [4]. Wang Denghai et al. proposed a hybrid model of convolutional neural network (CNN) and long short-term memory (LSTM) to predict photovoltaic power generation, aiming at the problems such as inaccurate prediction of photovoltaic power generation caused by weather timing changes in different regions [5].

3. Forecast the Development Trend of Power Supply in China

3.1. Screening Index

The power energy industry is closely related to economic conditions, household consumption level, urbanization rate, marketization, and other factors. Therefore, this paper selects relevant factors from four aspects: economy and industry, energy structure and consumption, population and society, and environment. In this case, power production is taken as an indicator of power supply. Based on relevant literature and network data, this paper uses crawler technology to screen 17 indicators that are closely related to power supply [6], as shown in Figure 1.

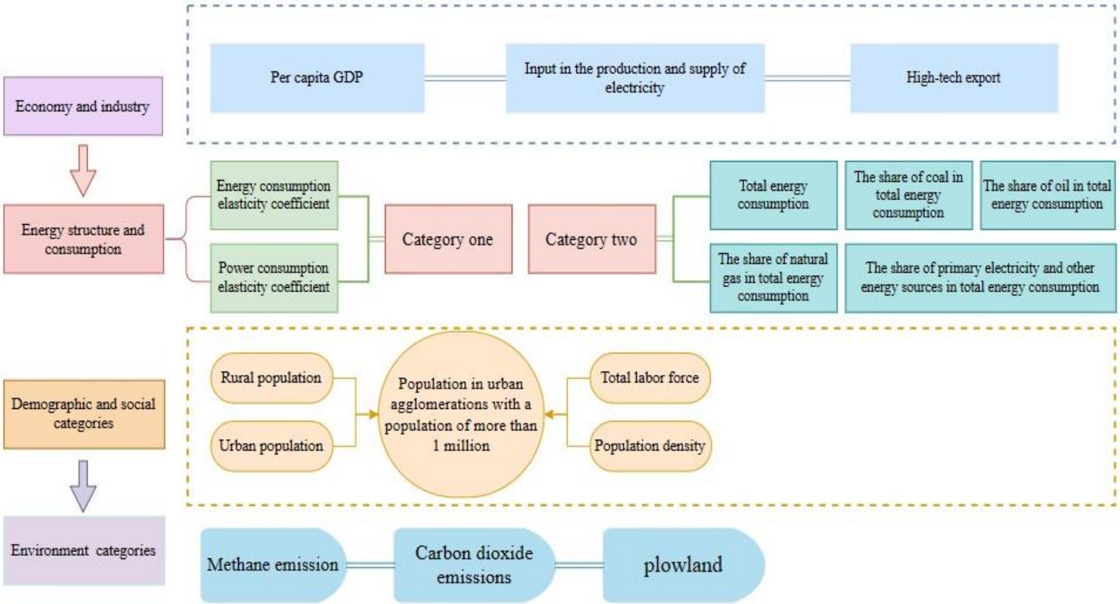

Figure 1. Relationship diagram of influencing factors

3.2. Data Preprocessing

3.2.1. Outlier Processing Based On Boxplot

First of all, the impact of different attribute data dimensions should be eliminated before data modeling and other work. Due to the different measurement scales and methods of the relevant data of each indicator, the different attribute dimensions of the data will have an impact on the data analysis. In this paper, the Min-Max normalization method is adopted to transform the original data and map the data to [0, 1], to eliminate the influence of dimension. The change formula of the Min-Max normalization method is as follows:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

The Min-Max normalization process was carried out according to the relevant data of the selected influencing factors, and the obtained results were partially displayed as shown in Table 1.

Table 1. Normalized processing of indicator data.

Index\Year	2000	2001	2002	2003	2004
Electricity production	0	0.0159	0.0380	0.0707	0.1080
Population and Social Category	0	0.0835	0.1715	0.2555	0.3341
Methane emissions	0.0280	0.01419	0	0.0497	0.1424
Carbon dioxide emissions	0	0.0187	0.0462	0.1107	0.1814
Farmland	0.8133	0.8606	0.8919	0.9074	0.9228
Per capita GDP	0	0.0173	0.0380	0.0629	0.0905
Total investment in electricity production and supply industry	0.0004	0.0002	0	0.0383	0.0698
High tech exports	0	0.0408	0.0816	0.1225	0.1633
Energy Structure and Consumption 1	0.8798	0.9689	0.9078	1	0.9288
Energy Structure and Consumption 1	0	0.0650	0.2293	0.3966	0.5232

*Note: The normalized values are reserved for three decimal places for display.

When there are obvious outliers in the data, it will bring unnecessary errors to data analysis and model building, so it is necessary to detect and eliminate the outliers. Because the value distribution of the problem is not uniform, it does not conform to the characteristics of normal distribution, so the boxplot which does not require the data distribution is used for outlier detection.

The principle of using boxplot to detect outliers of data is as follows: by calculating the quartile plus or minus 1.5 times the quartile distance, that is, calculating the values of Q1-1.5IQR and Q3+1.5IQR, the data outside this interval is specified as the outlier. In the box plot, you can see the median, upper quartile, lower quartile, upper and lower edges, and potential outliers of the variable data. In this paper, the upper quartile is used to replace the data with values greater than Q3+1.5IQR, and the lower quartile is used to replace the data with values smaller than Q1-1.5IQR, and the box plot of outliers is drawn, as shown in Figure 2. As can be seen from the figure, the middle line represents the median, the upper and lower edges of the box represent the upper and lower quartile respectively, the horizontal lines above and below the figure represent the upper and lower edges, and the top and bottom points are potential outliers. There are relatively few outliers in each indicator data, indicating that the data are relatively few outliers and the data quality is relatively good.

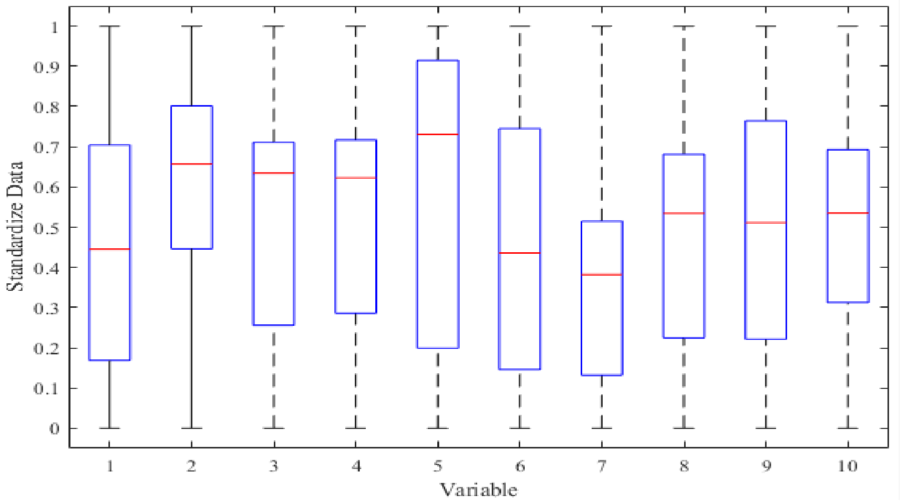

Figure 2. Box Diagram

3.2.2. Missing Value Processing

The missing data is mainly caused by two parts, one is the missing value of the data itself, and the other is the operation of eliminating the outlier of the data, which leads to the missing data. Based on the missing data caused by these two reasons, if it is directly eliminated, it will have a certain impact on the analysis and prediction results. Therefore, the problem is processed by interpolation and filling, and the missing data is filled by linear interpolation, as shown in Figure 3.

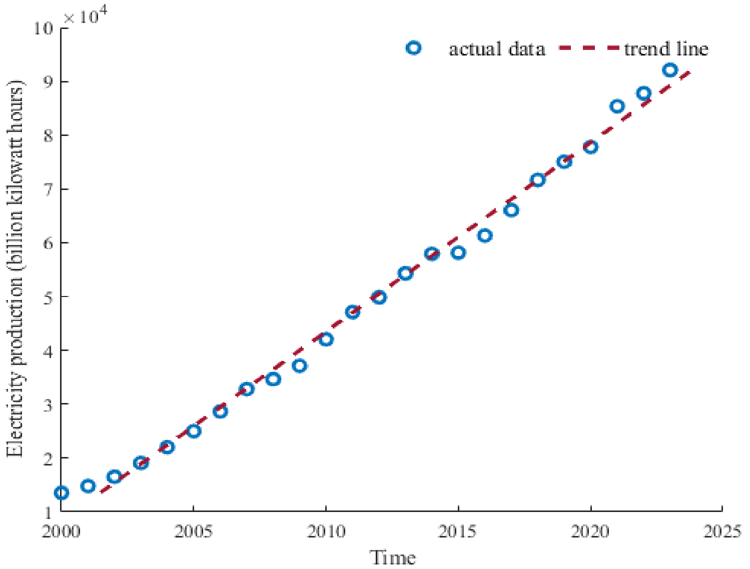

Figure 3. Alignment of linear interpolation values

3.2.3. Descriptive Statistical Analysis

To better analyze the interaction between power supply and various factors, this paper draws the development trend chart of each indicator, as shown in figure 4, which can more intuitively reflect the distribution of each variable in different values.

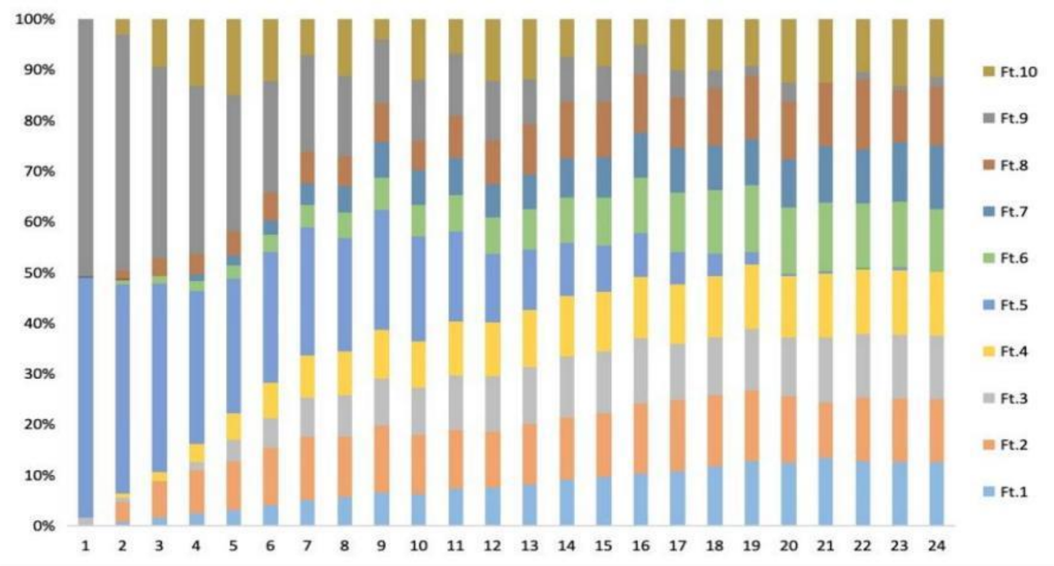

Figure 4. Histogram distribution

As can be seen from the figure, the 10 factors related to power supply also show an upward trend with the increase of years, which is closely related to social progress, scientific and technological development, economic growth, and the improvement of people's living standards.

3.3. Correlocation Analysis Based On the Spearman Coefficient

After the data were preprocessed, the Spearman correlation coefficient was used to characterize the correlation between different indicators. Spearman correlation coefficient is used to measure the nonlinear relationship between two variables, which is usually used to process data with non-normal distribution or ordered data. Its calculation formula is as follows:

$$\rho_s = \frac{\sum_{i=1}^{N}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{N}(R_i - \bar{R})^2 (S_i - \bar{S})^2}} \quad (1)$$

In the formula, Ri and Si are the grades of the values of observation i respectively, $\bar{R}$ and $\bar{S}$ are the average grades of variables x and y, for N is the total number of observations.

Spearman's correlation coefficient ranges from -1 to 1, indicating that there is a complete positive correlation between the ranks of two variables, that is, when one variable increases, the other variable also increases.

$\rho = -1$ is a completely negative correlation, indicating that there is a complete negative correlation between the ranks of two variables, that is, when one variable increases, the other decreases. $\rho = 0$ means when there is no correlation, it means that there is no linear or monotonic relationship between the ranks of two variables.

The thermal map of correlation coefficient of each index is shown in the figure 5.It can be seen from the thermal map that power production is highly correlated with per capita GDP, energy structure and consumption, population density, methane emissions, and other aspects, with the correlation reaching up to 98%, and the correlation between most factors and the power production can reach more than 90%. It indicates that the selected indicators can be used to make a more accurate prediction of the development trend of power supply in the next 40 years.

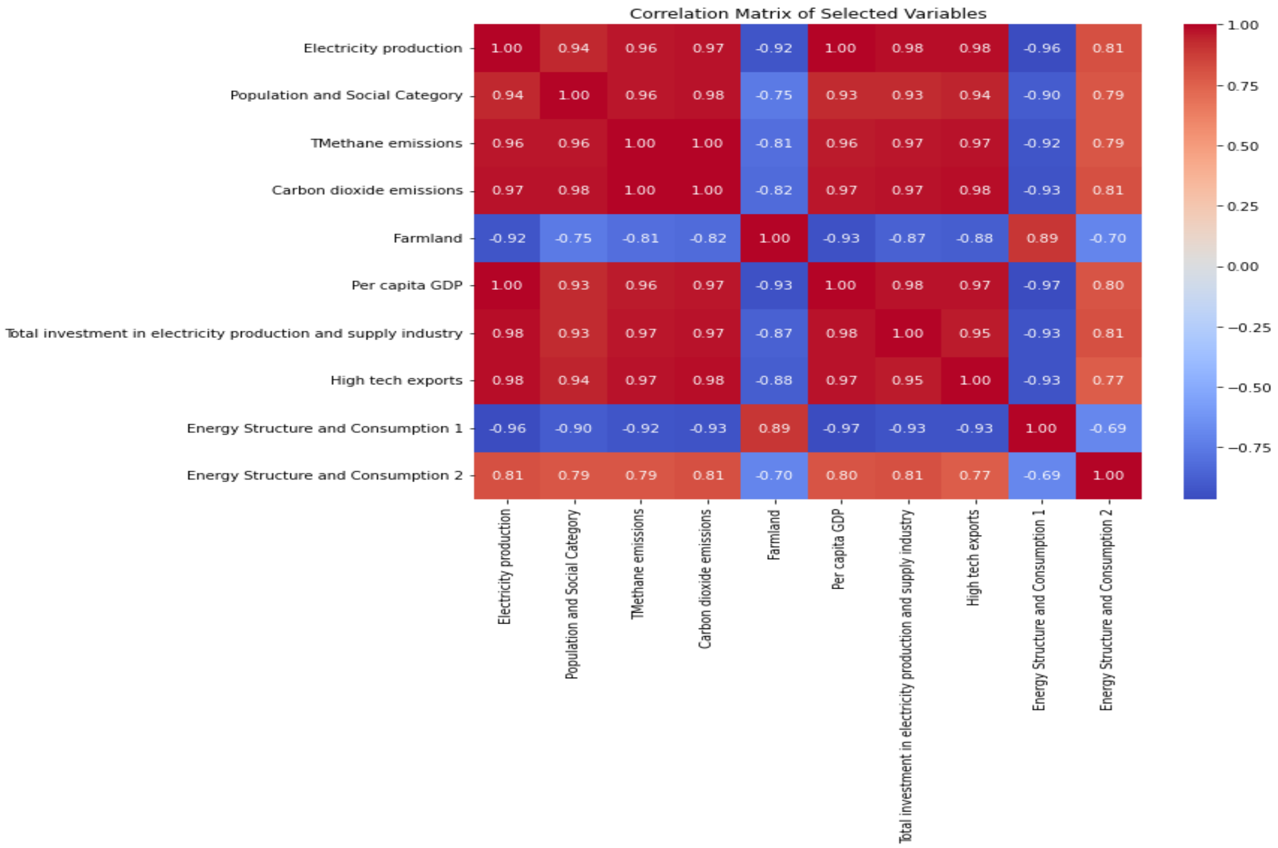

Figure 5. Heat map of the correlation coefficient between indexes.

3.4. Power Supply Trend Prediction Based On the LSTM-SHAP Model

3.4.1. LSTM-SHAP Model

LSTM neural network is a variant of recurrent neural network (RNN). LSTM structure adopts a control gate mechanism and is composed of four parts: memory cell, input gate, output gate, and forgetting gate. Its core concept lies in the state of the cell and the structure of the "gate". The cell state can transmit the relevant information in the sequence processing process all the time, which overcomes the influence of short-term memory. The structure of the "gate" determines which information should be saved or forgotten [7], as shown in Figure 6.

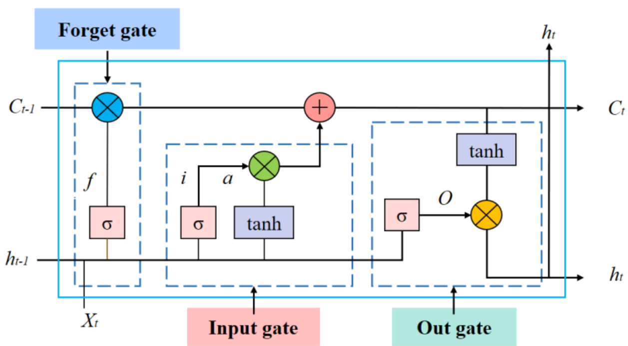

Figure 6. LSTM flow chart.

The input gate is used to determine how much of the input at the current moment can be saved to the cell state. The input gate receives the input at the current moment and the output at the previous moment as input. The sigmoid activation function is used to calculate the input, and the tanh activation function is used to obtain the new candidate memory cell calculation.

$$x_t = \delta(W_{ix}x_t + W_{ih}h_{t1} + b_i) \tag{2}$$

$$\tilde{c}_t = \tanh(W_{\tilde{c}_t x}x_t + W_{\tilde{c}_t h}h_{t1} + b_{\tilde{c}_t}) \tag{3}$$

The forget gate ft is used to evaluate the importance of the memorized information in the current moment in the cell state of the previous moment, and a forget gate value of 0 indicates that no information is transmitted to the calculation Ct, and a value of 1 indicates that all information is transmitted to the calculation Ct.

$$f_t = \delta(W_{fx}x_t + f_{ih}h_{t-1} + b_f) \tag{4}$$

After the processing of the forgetting gate and the input gate, the cell state is updated to form a long-term memory. The updating formula is as follows:

$$C_t = f_t C_{t-1} + i_t \tilde{c}_t \tag{5}$$

The output gate is used to select the output of the cell state at the current moment, and the final output state is ht:

$$o_t = \delta(W_{\delta x}x_t + f_{\delta h}h_{t-1} + b_o) \tag{6}$$

$$h_t = o_t \tanh(C_t) \tag{7}$$

The SHAP model is a method used to explain how any model comes up with a prediction, which can explain the importance of features in a machine learning model, and its core is the Shapley value. The Shapley value of a feature represents the contribution of the feature to the model's prediction results when the value of the feature is compared to the expected value of the feature, and the feature importance of the entire model is calculated by calculating the Shapley value of each feature value in all possible cases.

3.4.2. Prediction Results of Power Supply Development Trend Based On the LSTM-SHAP Model

The LSTM model is used to forecast the development trend of power supply in 2024-2060. According to trend chart, as shown in Figure 7, in the next 40 years, China's power supply will reach a peak value of 106794.04 in 2029 after a short rising period. After that, there will be a period of decline. When the current level of power supply is reduced to the current level, it will be maintained at a level value after that, which is related to carbon peak, carbon neutrality, etc., and interact to maintain the power supply within a range.

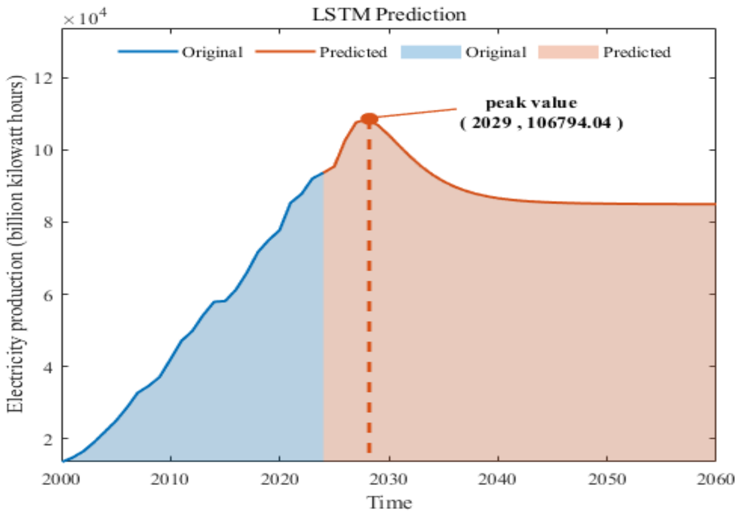

Figure 7. LSTM model prediction results.

Specific forecast results are shown in Table 2.

Table 2. Power supply forecast results 2024-2060.

Year	Electricity production	Year	Electricity production	Year	Electricity production
2024	93621.961	2037	88671.797	2050	85090.711
2025	95398.563	2038	87790.578	2051	85060.922
2026	102675.65	2039	87117.063	2052	85037.047
2027	107574.21	2040	86605.75	2053	85017.859
2028	108397.14	2041	86219.055	2054	85002.43
2029	106794.04	2042	85926.922	2055	84990.016
2030	104046.73	2043	85705.898	2056	84980.047
2031	100960.77	2044	85538.047	2057	84972.063
2032	97985.703	2045	85409.875	2058	84965.688
2033	95339.578	2046	85311.281	2059	84960.617
2034	93101.258	2047	85234.852	2060	84670.288
2035	91271.844	2048	85175.109	—	—
2036	89813.336	2049	85128.047	—	—

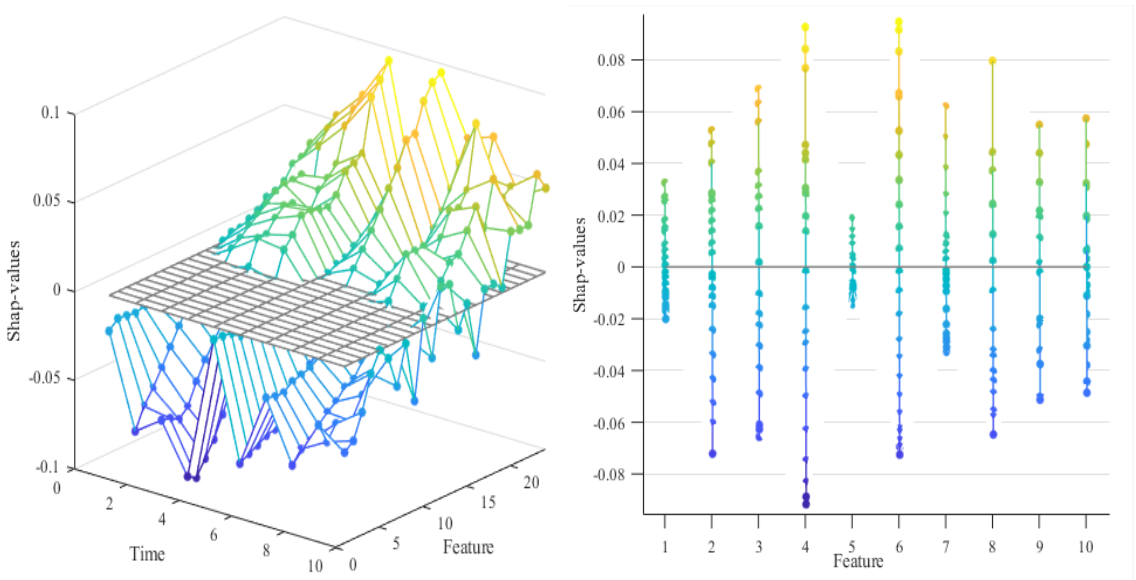

Figure 8. Feature importance

From the SHAP value of each of the above indicators, as shown in Figure 8, the fourth factor, carbon dioxide emissions, has the greatest impact on prediction results and the greatest contribution to the prediction results, while cultivated land has the least impact on the prediction results, but still has a certain degree of impact. Among them, cultivated land belongs to the environmental factors, and the third and fourth indicators of the same type of factors are listed. That is, both methane emissions and carbon dioxide emissions have an important impact on the predicted results.

4. Explore the Maximum Potential of Photovoltaic Power Generation in China

4.1. Establish a Dynamic Programming Model

According to the results of questions 1 and 2, the variables involved in this question are hypothesized, and based on the assumed conditions and set values, the maximum photovoltaic power generation is taken as the objective function, and the maximum photovoltaic power generation is calculated under constrained conditions.

Assumed conditions:

(1) The starting year is 2024 and the ending year is 2060.

(2) The initial installed capacity will be 100 megawatts (MW), with a random annual installed capacity growth rate of between 1% and 5%.

(3) Power generation efficiency varies between 15% and 25% per year.

(4) The amount of solar radiation varies between 1500 and 2000KWH/m2 per year.

(5) The number of effective generating hours per year is assumed to be 1825 hours (5 hours per day on average). It is divided into 37 indicators according to the year, and the objective function is:

$$E_T = \sum_{i=2024}^{2060} (C_i \cdot E_i \cdot R_i \cdot H) \tag{8}$$

The decision variables are:

Ci: Installed capacity in Year i (MW); Ei: Generation efficiency for Year i (%); Si: Solar radiation in year i(KWH/m2).

Constraints mainly include economic constraints, technical constraints, environmental constraints, and policy constraints in four aspects:

$$\begin{cases} C_{Ti} = C_I + C_{OM}, \\ C_{Ii} = C_i * (C_{panel} + C_{install} + C_{grid}), \\ YC_{OMi} = C_i * C_{OMi}, \\ C_I \le budget_i, \\ 0 \le E_i \le 25\%, \\ S_i \le S_{\max}, \\ C_i \le C_r, \\ G_T \le G_C. \end{cases} \tag{9}$$

Where Symbols and meanings is shown as Table 3.

Table 3. Symbols and meanings.

Symbol	Implication	Symbol	Implication
E_T	Total power generation (KWh)	H	Number of effective generating hours per year
C_{Ti}	Total cost for year i (Yuan)	YC_{OMi}	Operation and maintenance cost for Year i (Yuan)
C_{Ii}	The investment cost in year i (Yuan)	C_{panel}	Photovoltaic panel cost (Yuan)
E_i	Year i Photovoltaic power efficiency Photovoltaic power efficiency	$C_{install}$	Installation cost (Yuan)
S_i	Solar radiation in the year i (KWh/m^2)	C_{grid}	Grid access cost (Yuan)
C_i	Installed capacity in year i (MW)	T_{grid}	Grid capacity

4.2. Model Solving Based On Particle Swarm Optimization Algorithm

To improve the accuracy of the algorithm, we adopt particle swarm optimization [8]. Particle swarm optimization (PSO) is simple and easy, with convergence speed and few parameters. Based on the traditional particle swarm optimization algorithm, this paper improves the problem. Through the above programming model, each particle determines the optimal solution of each particle individual and finds a global optimal value from these individual optimal solutions. The main steps are shown in Figure 9.

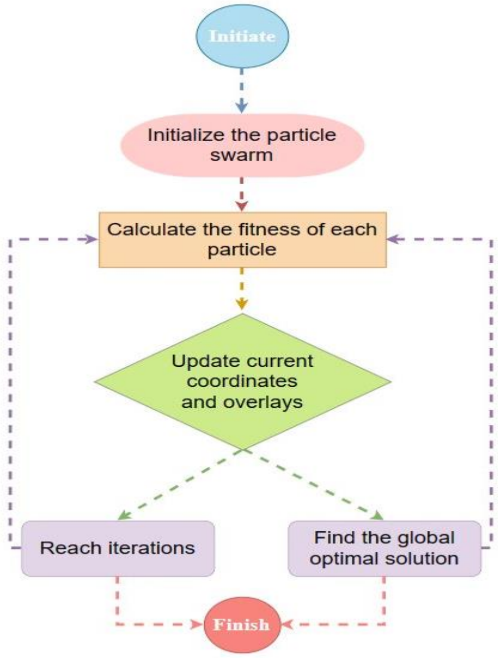

Figure 9. Schematic diagram of particle swarm optimization

MATLAB is used to solve the problem, and the optimal result is 4817085532.91 kWh under various assumptions and constraints, that is, the maximum total power generation is 4817085532.91 kWh. During the simulation period 2024-2060, the result is affected by some parameter Settings in the model assumptions, and the result is more accurate when the parameter Settings are close to the true values. When using the model to solve, many factors such as economy, technology, environment, and policy are considered. In the future, with the progress of technology, the total power generation will also increase, but at the same time, it is also restricted by economic factors such as cost and market price, as well as policy changes and the degree of impact on the environment. This result is an estimate of the maximum power generation; in practice, more external factors should be taken into account.

5. Conclusion

In the next 40 years, China's power supply will first rise and then decline, and will peak in 2029, with the predicted result of 106,794.04 kWh. Finally, when the power supply is equivalent to the current power supply, there will be a gradual trend to flatten out, which is related to the carbon peak and carbon neutrality, which interact with each other and maintain the power supply within a range.

Among the factors affecting the development trend of power supply, carbon dioxide emission has the greatest impact on the forecast results, while arable land has the least impact on the forecast results, but it still has a certain degree of impact. Among the factors, cultivated land belongs to the environmental category, and methane emission and carbon dioxide emission have important impacts on the forecast results.

The maximum total power generation of photovoltaic power generation is 4817085532.91 kWh. During the simulation period 2024-2060, many factors such as economy, technology, environment and policy are considered. In the future, with the progress of technology, the total power generation will also increase, but at the same time, it will be restricted by economic factors such as cost and market price, as well as policy changes and the degree of impact on the environment. This result is an estimate of the maximum power generation; in practice, more external factors should be taken into account.

Acknowledgements

This work is supported by the Undergraduate Research and Innovation Fund Program of Anhui University of Finance and Economics (Project No.:XSKY24223).

References

[1] Wu Ping, Guo Wei, Jin Chunjie, et al. Discussion on Current Situation and Solutions of Electric power and energy in China [J]. Electrical Technology, 2018,19(05):1-4+14.

[2] Yang Minying. Thinking on the change of power supply and demand situation [J]. Electric Power Technical Economics, 2004(02):8-10.

[3] Zhou D. An empirical study on power supply and demand in China [D]. Sichuan University, 2008.

[4] Qiu Shuqi, JIAN Zhaomin, Fang Lixiong, et al. Pv power generation prediction based on variational mode decomposition and ensemble learning [J]. Smart Power, 2019, 52(03):32-38. (in Chinese)

[5] Wang Denghai, An Yuexin, Liao Chenbo, et al. Research on photovoltaic power generation prediction method based on CNN-LSTM hybrid neural network [J]. Journal of Xi 'an Shiyou University (Natural Science Edition), 2019, 39(01):129-134.

[6] Editorial Department of this Journal. Prediction and Analysis of the Difficulties of Electric Power Development in 2018 -- An interview with the Industry Development and Environmental Resources Department of China Electricity Council [J]. China Electric Power Enterprise Management, 2018(07):52-55.

[7] Hao H Q. Study on the hydrologic process of Urumqi No. 1 Glacier in Tianshan Mountains based on interpretable deep learning [D]. Tianjin normal university, 2022.

[8] Chen Jian, Huang Zhi, Xu Tingliang et al. Based on improved binary particle swarm algorithm to optimize bearing fault diagnosis of DBN [J]. Journal of combination machine tools and automatic processing technology, 2024 (01): 168-173.