**A Hybrid Forecasting Approach for Short-Term Photovoltaic Power Generation Based on WOACO and LSSVM**

.

ABSTRACT

|  |
| --- |
| Due to significant random volatility of photovoltaic power generation, a short-term prediction method for photovoltaic power generation based on Chaos Whale Algorithm(WOACO) optimized Least Squares Support Vector Machine (LSSVM) was designed to further improve prediction accuracy of photovoltaic power generation. Firstly, meteorological factors that play a crucial role in photovoltaic power generation are extracted through Pearson correlation coefficient; Secondly, using WOACO to optimize parameters of LSSVM prediction model; Finally, simulate and analyze actual photovoltaic power generation in a certain region in northwest China. Whether in summer or winter, the prediction accuracy of the prediction model has significantly improved under three different evaluation indicators. Results demonstrate that WOACO-LSSVM combination model has higher accuracy in predicting photovoltaic power generation. |

*Keywords: photovoltaic power forecasting; least squares support vector machine; whale optimization algorithm; Pearson correlation coefficient; Combination Model*

1. INTRODUCTION

In response to call to accelerate construction of a new energy system, domestic photovoltaic power generation systems have shown a rapid development trend and become one of most promising new energy sources today[1,2]. However, instability and random fluctuations of photovoltaic power generation systems can also lead to a series of problems such as reduced system moment of inertia and decreased frequency and voltage tolerance after grid connection, which can have a very adverse impact on power system. Therefore, accurately predicting output of photovoltaic power generation systems is of great practical significance in enhancing stability of photovoltaic power stations, improving risk resistance of power grid, and expanding its application scope[3].

At present, many scholars at home and abroad are focusing on improving prediction accuracy of photovoltaic power generation, resulting in a series of related research methods, mainly including time series prediction method, autoregressive moving average model method, wavelet analysis method, artificial neural network, combination prediction method, and support vector machines (SVM), etc[4]. Among them, Least Square Support Vector Machines (LSSVM), as an improved algorithm of SVM, is favored by many scholars due to its superiority in photovoltaic power generation prediction and fast problem-solving. However, determination of its kernel function parameter *σ* and penalty parameter *γ* usually depends on human selection[5], which greatly reduces accuracy of model. Therefore, Whale Optimization Algorithm based on Chaos Optimization (WOACO) is used to optimize above two parameters to improve overall prediction ability of combined model. Reference [6] decomposes historical wind power models using wavelet analysis to obtain approximate and detailed sequences. However, wavelet analysis often generates redundant subsequences, and selection of wavelet bases depends on human selection; Reference [7] optimized kernel function parameters and penalty parameters of least squares support vector machine the butterfly optimization algorithm, but there is room for improvement in convergence speed of this algorithm; Reference [8] uses Spearman correlation coefficient to select meteorological characteristic quantities in photovoltaic power generation. However, if two variables are not monotonic, Spearman correlation coefficient may also be too low to determine the relationship between power generation and meteorological characteristic quantities; Reference [9] designed a combination model combining Complete Ensemble Empirical Mode Decomposition(CEEMD) and sparrow search algorithm to optimize SVM, which greatly improved its short-term photovoltaic power prediction accuracy in different months, but selection of input variables was too arbitrary.

Based on above literature, it is found that combination prediction model has characteristics of wider applicability and higher accuracy in predicting output of new energy. Therefore, based on selection of meteorological feature quantities, this article establishes a WOACO-LSSVM photovoltaic power generation combination prediction model. Due to the random fluctuations and intermittency of the original photovoltaic power generation, meteorological factors that play a key role in photovoltaic power generation before July 21 in summer and December 21 in winter were first extracted through Pearson correlation coefficient, and meteorological factors with high correlation were used as input data. Secondly, chaotic whale algorithm WOACO was used to optimize the parameters of LSSVM model. Finally, actual photovoltaic power generation in a certain region in northwest China was used for simulation analysis to demonstrate that the combination prediction model designed in study is more universal and accurate.

2. METEOROLOGICAL FACTORS AFFECTING PHOTOVOLTAIC POWER GENERATION

Factors that cause instability and intermittency in actual power generation of photovoltaics mainly include component temperature, ambient temperature, air pressure, humidity, total radiation, direct radiation, and scattered radiation. However, literature [10,11] has a small screening range for input variables, which may lead to significant errors in the prediction of the combined model. Excessive input variables can increase the complexity of the model and prolong the prediction time. Therefore, the Pearson correlation coefficient method is used in this paper to select important factors that affect photovoltaic power generation, and its expression is[12]:



In the formula, *λi* is Pearson correlation coefficient between the *i*-th meteorological characteristic and output of new energy;  is size of the *i*-th meteorological feature quantity at time *t;*  is average value of the *i*-th meteorological characteristic quantity;  is output of new energy at time *t*;  is average output of new energy.

Pearson correlation coefficient method usually divides the correlation between variables into the following five regions: if 0.8≤|*λ*|<1.0, it is a very strong correlation; If 0.6≤|λ|<0.8, it is a strong correlation degree; If 0.4≤|λ|<0.6, it is a moderate degree of correlation; If 0.2≤|λ|<0.4, it is a weak correlation degree; If 0≤|λ|<0.2, it is an extremely weak correlation.

The data selected in the article are various meteorological data and actual power generation of a photovoltaic power station in a northwest region of China during the first 21 days of July in summer 2019 and the first 21 days of December in winter 2019. Pearson correlation coefficients between various meteorological influencing factors and the actual power generation of the photovoltaic power station are shown in Table 1. It is not difficult to find that:

Regardless of summer or winter, total radiation, direct radiation, and scattered radiation of various meteorological influencing factors are very strongly correlated with actual output of photovoltaics, with only scattered radiation showing strong correlation characteristics in summer. Although correlation between component temperature, ambient temperature, and photovoltaic output cannot reach a strong correlation region, it reaches a moderate correlation in summer and even shows strong correlation characteristics in winter. However, during most periods, air pressure and humidity show a weak or even extremely weak correlation with actual power of photovoltaics.

**Table 1. Pearson coefficients of various meteorological influencing factors and photovoltaic power generation**

|  |  |  |
| --- | --- | --- |
| **Meteorological influencing factors** | **Pearson correlation coefficient** | |
| **Before**  **July**  **21st** | **Before December 21st** |
| component temperature | 0.5208 | 0.7564 |
| ambient temperature | 0.5144 | 0.6047 |
| air pressure | -0.0173 | 0.0625 |
| humidity | -0.2912 | -0.4173 |
| total radiation | 0.8985 | 0.9504 |
| direct radiation | 0.8977 | 0.9394 |
| scattered radiation | 0.7091 | 0.9478 |

Based on above analysis, in process of predicting combination of photovoltaic power generation in winter and summer in this region, input data in selected sample set are total radiation, direct radiation, scattered radiation, ambient temperature, and component temperature, and output data is actual output of photovoltaic power.

3 Chaos Whale Algorithm

WOA is an optimization algorithm inspired by whale hunting process and officially proposed by Mirjalili from Griffith University[13]. WOA drew inspiration from hunting methods of whales in sea, namely random search hunting, encirclement hunting, and foam net hunting[14]. Population positions in WOA are randomly generated, and blind spots of most peak functions are difficult to search for. In response to proposed problem, WOA combines ergodicity of logistic algorithm[15] and designs a chaotic whale optimization algorithm based on chaos optimization(WOACO). In the WOA, position of each humpback whale represents a feasible solution. Detailed process of the three hunting methods is as follows:

1) Randomly search for hunting. Humpback whales freely explore sea by randomly selecting a whale as their target to find optimal hidden solution, as shown in equations (2) and (3):





In the formula: *J* is distance between humpback whale and random target whale; *t* is number of iterations;  is position vector of random target whale;  is position vector of humpback whale; *A* and *C* are coefficient vectors, respectively.





In formula: *a* is parameter vector that linearly decreases from 2 to 0; r1 and r2 are both random vectors between [0,1].

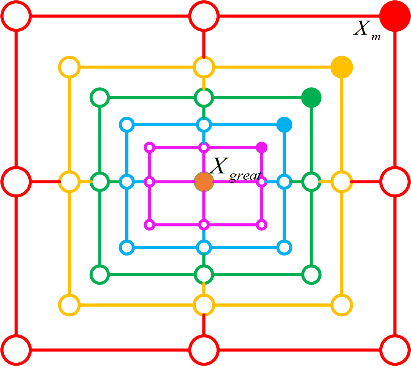
2) Surrounding hunting. Hunting route for humpback whales is shown in equations (6) and (7):





In formula:  is current optimal whale position vector.

Hunting route of the humpback whale is shown in Figure 1, and other solutions will continuously update their position around the current optimal position, while narrowing down the hunting range for more efficient global exploration.



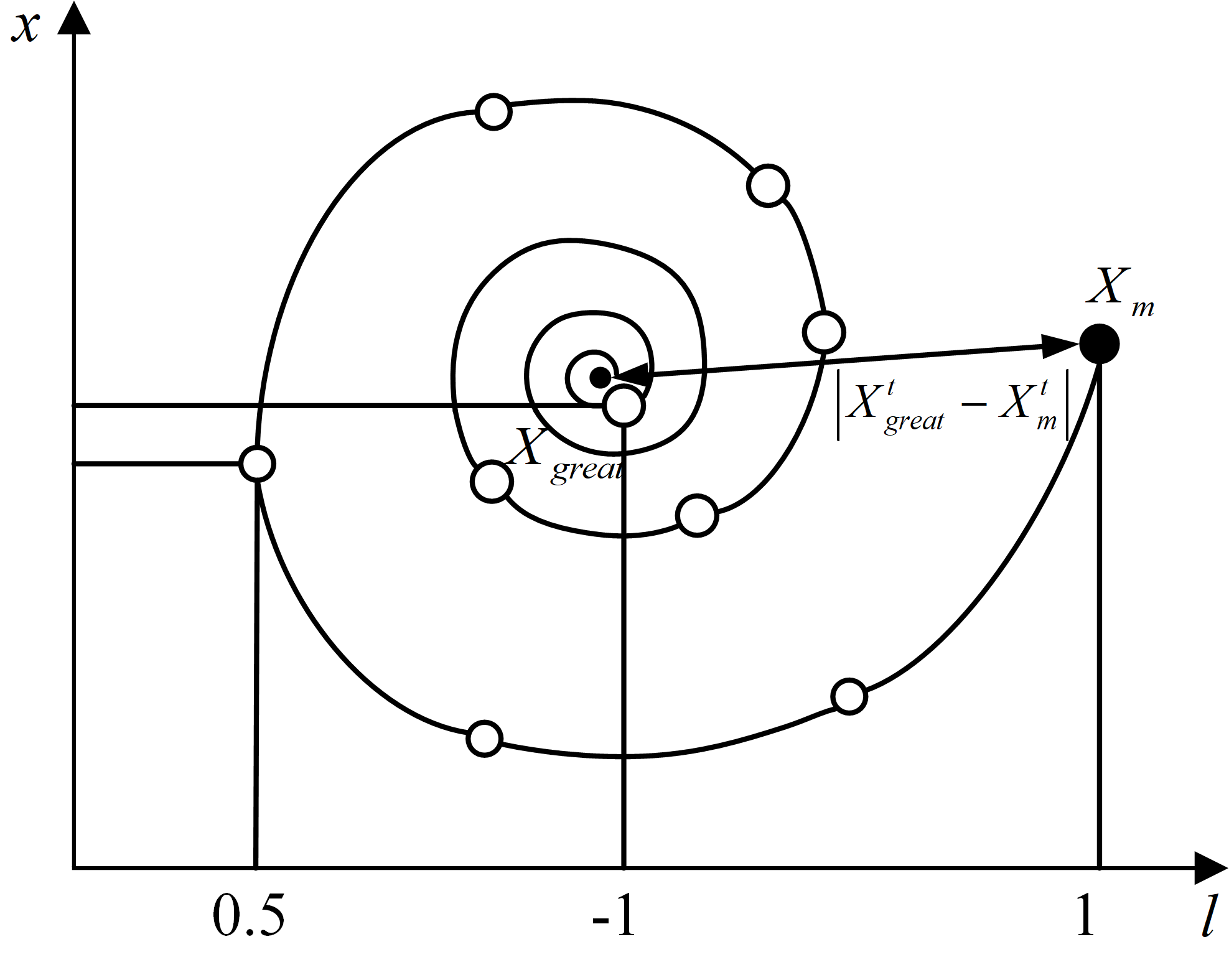
**Fig.1 Hunting route surrounded by humpback whales**

3)Humpback whale foam net hunting simulates the process of whales swimming in a circular motion and spraying bubbles to drive away prey. This process aims to improve accuracy of optimal solution, as shown in equation (8):



In the formula: *b* is a constant used to define shape of spiral, set to 1 by default;*l* is a random number between [-1,1].

Hunting process of humpback whale foam net is shown in Figure 2. Purpose of through spiral foam nets is to attack prey and enhance local development.



**Fig.2 Hunting process of foam net**

During hunting process of a humpback whale, probability of surrounding and attacking prey is equal, as shown in equation (9):



WOA algorithm has a simple principle and two main parameters. Vector parameter |*A*| determines whether search method is random search or surround search, and vector parameter |*C*| determines direction of movement and step size during update process. *ρ* > 0.5 ，Algorithms tend to favor local development. When *ρ* ≤ 0.5, it leans towards global exploration to avoid getting stuck in local optima.

4. LEAST SQUARES SUPPORT VECTOR MACHINE PREDICTION MODEL BASED ON WOACO

**4.1 least squares support vector machine**

On the one hand, LSSVM transforms inequality constraints in traditional support vector machines into equality constraints, and on the other hand, selects sum of squared errors loss function in experience loss of the training set, thus achieving a transition from solving quadratic programming problems to solving linear equation systems, greatly improving convergence accuracy and solving speed[16,17]. Setting dataset as , *xi* ∈*Rk* as input variable, *yi* ∈*R* as corresponding output variable, and *n* as number of samples, regression model g(*x*) is obtained as follows:



In formula: *α* is weight in feature space; *λ*(*xi*) is a nonlinear mapping; *b* is bias amount.

In process of solving regression function, LSSVM needs to adhere to principle of minimizing structural risk, and objective function in optimization model is:



In formula: *γ* is punishment parameter; *ωi* is error vector.

Constraints in optimization model are:



Lagrange function is:



In formula,***τ***is Lagrange multiplier.

By using KKT condition, Lagrange function is defined and its variables are differentiated to obtain multiple equation conditions, resulting in following linear equation system[18]:



In formula: *L* is unit column vector; *E* is an n-dimensional identity matrix; *Ω* is kernel matrix.

Final LSSVM regression function obtained is:





In the formula: *K(x, xi)* is radial basis kernel function; *σ* is width parameter of radial basis kernel function.

When using RBF function as kernel function of LSSVM, penalty parameter *γ* and kernel function width parameter *σ* mainly affect its prediction accuracy and convergence speed. Therefore, WOACO is selected to optimize above two parameters to minimize uncertainty caused by manual parameter setting[19,20].

**4.2 Optimization of LSSVM parameters using WOACO**

Specific optimization process for penalty parameter *γ* and kernel function width parameter *σ* of LSSVM using the WOACO algorithm is as follows:

1) Set population size *N* of whale population, maximum iteration times *T*max, dimensionality *dim*, upper limits *Ub1* and *Ub2* of the search space, and lower limits *lb1* and *lb2*.

2) Using logistic chaotic mapping to perform chaos initialization on initial position of humpback whale population.

3) Calculate fitness values of individuals in whale population and record optimal solutions of individuals in population.

4) Generate a random number *ρ* within interval range [0,1], update parameter vector *a* with each iteration, and synchronously update coefficient vectors *A* and *C* according to equations (4) and (5).

5) If random number ρ<0.5 and coefficient vector |A|≥1, then a random search strategy is used to update position of next generation of individuals; If random number ρ<0.5 and coefficient vector |A|<1, then next generation of individual positions will be updated using a surround search strategy centered on optimal solution position of individual mentioned above; Otherwise, bubble net strategy will be used to update position of next generation of individuals, with aim of further optimizing accuracy of solution for local development.

6) Using only mean square error as fitness function fitness[21], if fitness(*x*t+1)<fitness(*x*t), optimal position of whale individual is updated, otherwise it remains unchanged.

7) Determine whether maximum number of iterations has been reached. If maximum iteration requirement is met, output penalty parameter *γ* and kernel function width parameter *σ* of LSSVM; otherwise, return to step 4 to continue iterative optimization.

8) Finally, based on obtained LSSVM parameters, a short-term photovoltaic power generation combination prediction model based on WOACO-LSSVM was established.

Specific flowchart of optimizing LSSVM parameters using WOACO is shown in Figure 3.



**Fig.3 WOACO Optimization LSSVM Parameter Flow Chart**

5. Case Studies Analysis

This article uses actual photovoltaic power generation of a certain region in northwest China for simulation analysis, with a time span of 21 days and a sampling interval of 15 minutes. The sampling time is from July 1st to July 21st in summer and from December 1st to December 21st in winter, with 2016 data points collected. Actual photovoltaic power generation data curves collected are shown in Figure 4 and Figure 5, and the test sets selected are July 22nd and December 22nd.



**Fig.4 PV power generation for first 21 days of July**



**Fig.5 PV power generation for first 21 days of December**

In process of using a multiple linear regression model, different ranges of eigenvalues result in some eigenvalues having a much greater impact on results than others. Therefore, normalization is used to keep each eigenvalue within same measurement scale. The expression is:



In formula, *W*unit represents normalized data; *W*max and *W*min are maximum and minimum values of input and output data, respectively.

**5.1 Short term photovoltaic power prediction results and analysis**

Considering fluctuation and intermittency of photovoltaic output, chaotic whale algorithm WOACO is used to optimize penalty parameter *γ* and kernel function width parameter *σ* of LSSVM, thus constructing a short-term photovoltaic power generation combination prediction model that can complete 24-hour prediction of photovoltaic output. To demonstrate superior performance of WOACO-LSSVM combined prediction model, LSSVM algorithm, PSO-LSSVM, and WOA-LSSVM were directly used for prediction. Comparison chart of photovoltaic output prediction on July 22 in summer and the comparison chart of absolute error in photovoltaic prediction are shown in Figures 6 and 7, respectively. The comparison chart of photovoltaic output prediction on December 22 in winter and comparison chart of absolute error in photovoltaic prediction are shown in Figures 8 and 9, respectively.



**Fig.6 Comparison of PV Output Forecast on July 22nd**



**Fig.7 Comparison of Absolute Error in PV Forecasting on July 22nd**



**Fig.8 Comparison of PV Output Forecast on** **December 22nd**



**Fig.9 Comparison of Absolute Error in PV Forecasting on December 22nd**

According to output prediction and absolute error comparison in Figures 6, 7, 8, and 9, it can be found that WOACO-LSSVM combined prediction model has higher accuracy and superior performance in predicting short-term actual photovoltaic power generation compared to LSSVM, PSO-LSSVM, and WOA-LSSVM.

**5.2 Evaluation indicators for prediction results**

In order to more efficiently and accurately evaluate prediction accuracy of prediction model, it has been decided to use evaluation indicators to measure final results. The evaluation indicators used are root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE).







In formula, *N* is number of data sets in test set,  represents true value of photovoltaic output in *i*-th time period, and represents predicted value of photovoltaic output in *i*-th time period.

Comparison of evaluation indicators for predicting short-term photovoltaic output in July of summer under different methods using four prediction models: LSSVM, PSO-LSSVM, WOA-LSSVM, and WOACO-LSSVM is shown in Table 2.

**Table 2 Evaluation Index Values under Different Prediction Models (July)**

|  |  |  |  |
| --- | --- | --- | --- |
| **prediction model** | **RMSE/MW** | **MSE/MW** | **MAE/MW** |
| LSSVM | 5.5752 | 31.0824 | 3.7775 |
| PSO-LSSVM | 4.5934 | 21.0998 | 2.8021 |
| WOA-LSSVM | 4.5644 | 20.8335 | 2.4474 |
| WOACO-LSSVM | 4.4363 | 19.6804 | 2.3340 |

Combining Figure 6, Figure 7, and Table 2, it can be concluded that prediction accuracy of combined prediction model has been greatly improved compared to a single prediction model, with WOACO-LSSVM prediction model showing best improvement effect; Error in RMSE has been reduced by 20.43%; In terms of MSE, error decreased by 36.68%, and in terms of MAE, error decreased by 38.21%.

Comparison of evaluation indicators for predicting short-term photovoltaic output in December winter under different methods using four prediction models, namely LSSVM, PSO-LSSVM, WOA-LSSVM, and WOACO-LSSVM, is shown in Table 3.

**Table 3 Evaluation Index Values under Different Prediction Models (**[**December**](file:///C:\Users\99630\AppData\Local\Programs\baidu-translate-client\resources\app.asar\app.html)**)**

|  |  |  |  |
| --- | --- | --- | --- |
| **prediction model** | **RMSE/MW** | **MSE/MW** | **MAE/MW** |
| LSSVM | 3.8425 | 14.7651 | 1.5029 |
| PSO-LSSVM | 1.5800 | 2.4964 | 1.0267 |
| WOA-LSSVM | 1.5071 | 2.2715 | 0.9674 |
| WOACO-LSSVM | 1.4209 | 2.0190 | 0.8490 |

Combining Figure 8, Figure 9, and Table 3, it can also be observed that prediction accuracy of combined prediction model has still been greatly improved. And use of WOACO to optimize LSSVM parameters also achieved better results, with a 10.07% reduction in RMSE compared to PSO-LSSVM and a 5.72% reduction in error compared to WOA-LSSVM; Compared to PSO-LSSVM, MSE decreased by 19.12%, and compared to WOA-LSSVM, the MSE decreased by 11.12%; In terms of MAE, compared to PSO-LSSVM, error has decreased by 17.31%, and compared to WOA-LSSVM, error has decreased by 12.24%.

Based on above analysis, it can be found that the WOACO-LSSVM combined prediction model not only avoids high errors caused by manual parameter selection in LSSVM, but also has higher accuracy compared to other algorithms that optimize LSSVM parameters.

6. Conclusion

Considering volatility and intermittency of photovoltaic power generation, this paper proposes a short-term photovoltaic power generation prediction method based on WOACO-LSSVM combination prediction model, which can effectively improve prediction accuracy of photovoltaic power generation. After simulation analysis, designed combination prediction model can draw the following conclusions compared to other models:

(1) Pearson correlation coefficient method was used to obtain five important meteorological factors that affect photovoltaic output, namely component temperature, ambient temperature, total radiation, direct radiation, and scattered radiation, greatly improving prediction accuracy and reducing the complexity of the prediction process.

(2) WOACO algorithm is used to optimize the penalty parameter *γ* and kernel function width parameter *σ* in LSSVM, which can effectively handle a series of problems such as complex weather conditions and low prediction accuracy that are difficult to cope with.

Therefore, WOACO-LSSVM combination prediction model designed in article can efficiently and accurately complete short-term forecasting of photovoltaic output, and has good application prospects and value.

Competing interests

Authors have declared that no competing interests exist.

Disclaimer (Artificial intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Reference

1. Lin, Y., Ming, P., Zhuo, L., Xuri, S., Yadi, L., & Yong, T. (2023). Framework for joint wind and photovoltaic power forecasting and pre-dispatch. High Voltage Engineering, 50(9), 3823-3836.
2. He, W., Ziyue, L., Jing, B., & Yuhang, P. (2024). Research on two-stage risk scheduling for transmission and distribution power grids with high proportion of distributed energy access. Power System and Clean Energy, 40(1), 119-129.
3. Yichen, H., Wei, H., Changli, S., Xiaoqiang, G., & Tao, H. (2024). Photovoltaic power prediction algorithm based on parameter optimazation of multi-kernel svm. Acta Energiae Solaris Sinica, 45(9), 394-404.
4. Youjun, Y., Jinlin, L., Hui, Z., & Hongjun, W. (2023). Short-term pv power prediction considering feature extraction and optimised lssvm. Journal of Shaanxi University of Science & Technology, 41(3), 200-207.
5. Yufang, C., Ziming, Z., Yang, T., & Wencong, H. (2023). Wind power prediction model based on ceemd and improved SSA-LSSVM. Transducer and Microsystem Technologies, 42(10), 130-134.
6. Jiamin, L., Congrui, L., Zhihao, Z., & Sheng, L. (2019). Combination ultra-short-term prediction of wind power based on WD-CS-SVM. Electric Power Engineering Technology, 38(5), 24-29.
7. Pengzhen, S., Xia, W., Chunmei, Z., Lirong, X., Jiahao, Y., & Jialiang, Y. (2024). Short-term wind power prediction based on VMD-BOA-LSSVM-AdaBoost. Acta Energiae Solaris Sinica, 45(1), 226-233.
8. Jiangang, T., Pei, Z., Chunhua, P., Min, S., Tieqiang, W., & Rui, Y. et al. (2020). Ultra short-term forecast of photovoltaic generation based on time-division long short-term memory neural networks. Modern Electric Power, 37(6), 629-638.
9. Pengfei, W., Xincong, S., Yongming, Z., Long, H., Yang, L., & Xiaomin, J. et al. (2024). Short-term photovoltaic power prediction based on CEEMD-SSA-SVM. Water Power, 50(4), 87-94.
10. Rui, W., Yingzhou, W., & Jing, L. (2023). Prediction of photovoltaic power generation based on ICEENDAN-DTW and ISMA-WLSSVM. Journal of Engineering for Thermal Energy and Power, 38(9), 131-140.
11. Bin, L., Yifan, Z., Shiyе, Y., & Jilei, Y. (2022). Research on photovoltaic power generation prediction method based on improved extreme learning machine (elm). Journal of Engineering for Thermal Energy and Power, 37(10), 207-214.
12. Quoc-Thang, P., Yuan-Kang, W., & Quoc-Dung, P. (2024). Enhancing one-day-ahead probabilistic solar power forecast with a hybrid transformer-lube model and missing data imputation. [IEEE Transactions on Industry Applications](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=28), 60(1), 1396-1408.
13. Mirjalili S, Lewis A. (2016). The whale optimization algorithm. Advances in engineering software, 95, 51-67.
14. Lirong, X., Bin, W., Hongyin, B., Wuxing, L., & Matireyimu, M. (2021). Super-short-term wind power forecasting based on EEMD-WOA-LSSVM. Acta Energiae Solaris Sinica, 42(7), 290-296.
15. Hanbin, W., Min, S., Huankun, Z., Jixin, Z., & Huaming, Z. (2023). Short-term interval prediction of photovoltaic power based on eemd-aloco-svm model. Acta Energiae Solaris Sinica, 44(11), 64-71.
16. Zuoxun, W., Yangyang, K., & Jian, L. (2024).The power load forecasting model of combined sade-elm and fa-cawoa-svm based on cssa. [IEEE Access](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6287639), 12, 41870-41882.
17. Quanbo, G., Chen, G., Haoyu, J., Zhenyu, L., Gang, Y., & Jianmin, Z. (2022). Industrial power load forecasting method based on reinforcement learning and PSO-LSSVM. IEEE Transactions on Cybernetics, 52(2), 1112-1124
18. Guomin, X., & Tianxiang, C. (2024). Short-term prediction of photovoltaic power generation by improved chimpanzee optimization algorithm. Proceedings of the CSU-EPSA, 36(2), 135-143.
19. Yongli, W., Minhan, Z., Suhang, Y., Mengju, W., Yang, Y., & Zhao, L. (2022). Multi-load cooperative prediction model of integrated energy system based on multi-energy coupling mechanism. Journal of North China Electric Power University(Natural Science Edition), 49(2), 118-126.
20. Zhonghua, H., Tao, Z., Yuo, H., Fuqiang, L., Liang, H., & Bingkang, L.(2019). Short-term electric load forecasting model considering the influence of air pollution prevention and control policy. Smart Power, 47(5), 1-9.
21. Xiaohua, C., Zhiping, W., Jiekang, J., Haiwen, X., Shengyu, C., & Xunxiang, Z. (2023). Short-term photovoltaic power forecasting based on coot-svm. Sichuan Electric Power Technology, 46(6), 28-33,40.