Published by Avanti Publishers

Global Journal of Energy Technology Research Updates

ISSN (online): 2409-5818

Photovoltaic Power Generation Forecasting Based on the ARIMA-BPNN-SVR Model

Guo-Feng Fan[1], Hui-Zhen Wei[1], Meng-Yao Chen[1] and Wei-Chiang Hong[2,3,*]

[1]School of Mathematics & Statistics, Ping Ding Shan University, Ping Ding Shan 467000, Henan, China

[2]Department of Information Management, Asia Eastern University of Science and Technology, New Taipei 22064, Taiwan

[3]Department of Information Management, Yuan Ze University, Zhongli 32003, Taiwan

ARTICLE INFO

Article Type: Research Article

Keywords:
Combination forecasting
Photovoltaic power generation
Support vector regression (SVR)
Back propagation neural network (BPNN)
Auto regression integrate moving average (ARIMA)

Timeline:
Received: May 09, 2022
Accepted: July 04, 2022
Published: August 05, 2022

Citation: Fan G-F, Wei H-Z, Chen M-Y, Hong W-C. Photovoltaic Power Generation Forecasting Based on the ARIMA-BPNN-SVR Model. Glob J Energ Technol Res Updat. 2022; 9: 18-38.

DOI: https://doi.org/10.15377/2409-5818.2022.09.2

ABSTRACT

With the continuous expansion of the capacity of photovoltaic power generation systems, accurate power generation load forecasting can make grid dispatching more reasonable and optimize load distribution. This paper proposes a combined forecasting model based on Auto Regression Integrate Moving Average (ARIMA), back propagation neural network (BPNN), and support vector regression (SVR), namely ARIMA-BPNN-SVR model, aiming at the problem of low accuracy of a single model and traditional forecasting model. Through the complementary advantages of ARIMA, BPNN, and SVR models, the model has good anti-noise ability, nonlinear mapping, and adaptive ability when processing photovoltaic power generation data. Data experiments are carried out on solar photovoltaic power generation in the United States, and the accuracy of model forecasting is evaluated according to MAE, MSE, RMSE, and MAPE. The experimental results show that the proposed ARIMA-BPNN-SVR outperforms the forecasting performance of the single models ARIMA, BPNN, and SVR. Its MAE, MSE, RMSE and MAPE are 0.53, 0.41, 0.64 and 0.84 respectively. In the Wilcoxon sign-rank test, the p-value of the proposed model reached 0.98, indicating the effectiveness of the ARIMA-BPNN-SVR model.

*Corresponding Author
Email: samuelsonhong@gmail.com
Tel: +(886) 936072140

©2022 Fan *et al*. Published by Avanti Publishers. This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited. (http://creativecommons.org/licenses/by-nc/4.0/)

1. Introduction

Improving the forecast level of short-term photovoltaic power generation is a key issue for integrating solar photovoltaic power stations into the existing grid system and developing and utilizing solar photovoltaic power.

1.1. Photovoltaic Power Generation Forecasting Method

1.1.1. Traditional Photovoltaic Power Generation Forecasting Methods

The traditional photovoltaic power generation forecasting methods include time series [1,2], grey theory [3-5], regression analysis [6-8], and neural network methods [9-11].

For example, Zhang et al. [2] proposed an adaptive hybrid forecasting model which combines improved variational mode decomposition (IVMD), an autoregressive integrated moving average (ARIMA) model, and an improved deep confidence network (IDBN) model. Compared with other models, the forecasting performance of this model is verified. Ding et al. [3] proposed a new discrete grey model with time-varying parameters to deal with various PPG time series with nonlinear, periodicity, and fluctuation. The experiment shows that the forecasting accuracy is improved remarkably. AlShafeey et al. [8] study multiple regression (MR) or artificial neural network (ANN) methods. The results of the performance comparison show that the forecasting accuracy of the neural network model is better than that of the MR model. Natarajan et al. [11] proposed a feature selection and forecasting model that uses the radial confidence neural network (RBNN) model to automate the energy optimal forecasting process. The RBNN model has better forecasting power and less error than other models.

1.1.2. Modern Photovoltaic Power Generation Forecasting Methods

The modern photovoltaic power generation forecasting methods start to use artificial intelligence algorithms such as support vector machines [12-14], random forests [15,16] and other machine learning algorithms [17], convolutional neural networks [18,19], long short-term memory artificial neural networks [20-22] and other deep learning algorithms [23].

Pan et al. [14] proposed to construct a support vector machine based on data processing and use an ant colony algorithm to optimize the parameters of the support vector machine. The results show that the R2 of the mixed model reaches 0.997. Niu et al. [15] constructed a hybrid forecasting model called RF-CEEMD-DIFPSO-BPNN. The radio frequency method was used to calculate the importance and eliminate the unimportant factors. Then the importance calculated by RF is transferred to the IGIVA model as a weight value to screen the similar days of different weather types and improve the data quality of the training set. Wang et al. [17] proposed to design a comprehensive system based on the automatic optimization of the variable mode decomposition mechanism and determine the weight of the system by the multi-objective intelligent optimization algorithm. In particular, it is proved theoretically that the developed forecasting system can achieve Pareto's optimal solution. Zang et al. [18] proposed a new data preprocessing method to construct input feature maps for two new CNNS of historical PV power series, meteorological elements, and numerical weather forecasts. A meta-learning strategy based on multiple loss function networks is proposed to train the two-deep networks to ensure the high robustness of the extracted convolutional features. Ahmed et al. [21] used the integration-based long short-term memory (LSTM) algorithm, which consists of 10 LSTM models. The method compares the effects of seasonal and periodic variations on time series data and PV output forecasting.

Although modern photovoltaic forecasting technology has made significant progress, there is still much space for exploration. People have studied the combined and mixed methods of photovoltaic power generation forecasting. Das et al. [31] made a comprehensive and systematic review of the direct forecasting of PV power generation. The importance of the correlation of the input-output data and the preprocessing of model input data are discussed.

1.2. Research Motivation

Due to the instability and randomness of solar energy, photovoltaic power generation is affected by various factors such as weather, temperature, and radiation. ARIMA model can effectively reflect the trend of photovoltaic power generation and whether it is stable and depicts the linearity within the feature data. The BPNN model can describe nonlinear functions, so it can also be used to describe the nonlinear characteristics of photovoltaic power generation. Support vector regression (SVR) training algorithms excel on nonlinear problems, distinguishing training examples into two classes and forecasting the class of subsequent examples. The parameters and kernel parameters of the support vector regression are optimized by using the PSO algorithm to improve the forecasting accuracy of the support vector regression. It can be seen that different methods can reflect the characteristics of the data from different angles. Therefore, we combine the three to form the proposed ARIMA-BPNN-SVR model to complement each other's advantages.

The first part of this paper introduces the ARIMA, BPNN, and SVR models in detail; the second part, based on the photovoltaic power generation data in the United States, uses the ARIMA, BPNN, SVR models and the proposed ARIMA-BPNN-SVR model to forecast and analyze the photovoltaic power generation data; The third part, on this basis, uses the sign consistency test to test the validity of the model, and briefly summarizes the research results of this paper.

2. Method

2.1. Autoregressive Integrated Moving Average (ARIMA) Model

The ARIMA model is essentially the result of the combination of the ARMA model and the different operations. The ARIMA model can be divided into AR, MA, ARMA, and ARIMA models [24]. The ARIMA model can differentiate the unstable time series and convert it into a stable series.

1). AR(*p*) model

The expression of the AR(p) model can be formalized as Eq. (1),

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t \\ \phi_p \neq 0 \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, \forall s \neq t \\ E(x_s x_t) = 0, s < t \end{cases} \tag{1}$$

Let this model be AR(*p*). When all the characteristic roots of AR(*p*) are in the element circle, it is a stationary sequence.

2). MA(*q*) model

The expression of the MA(*q*) model can be formalized as Eq. (2),

$$\begin{cases} x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \\ \theta_q \neq 0 \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \end{cases} \tag{2}$$

Denote the model as MA(*q*). The value of the MA(*q*) model at any time is only related to the weighted average term, so the MA(*q*) model is always stationary.

3). ARMA(*p*,*q*) model

The expression of the MA(*q*) model can be formalized as Eq. (3),

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \theta_q \neq 0 \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \tag{3}$$

Let this model be ARMA(*p*,*q*).

4). ARIMA(*p*,*d*,*q*) model

The expression of the ARIMA(*p*,*d*,*q*) model can be formalized as Eq. (4),

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \tag{4}$$

In Eq. (4), $\nabla^d = (1-B)^d$ Can be abbreviated as Eq. (5),

$$\nabla^d x_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \tag{5}$$

In Eq. (5), $\{\varepsilon_t\}$ is a zero-mean white noise sequence.

2.2. Back-Propagation Neural Networks (BPNN) Model

The BPNN model is a multi-layer feed-forward neural network, which can form a three-layer network by optimizing the function weight to generate corresponding errors, which is used in the fields of function approximation [25-27]. The main feature of this network is forward transmission and reverse transmission. In the forward transmission process, the input signal is processed layer by layer from the input layer through the hidden layer until the output layer. The neuron state of each level only affects the neurons of the next level. When the output layer does not get the expected output, reverse transmission is introduced to adjust the weight and threshold of the network through forecasting errors so that the forecasting results of the BPNN model are consistent with the expected output.

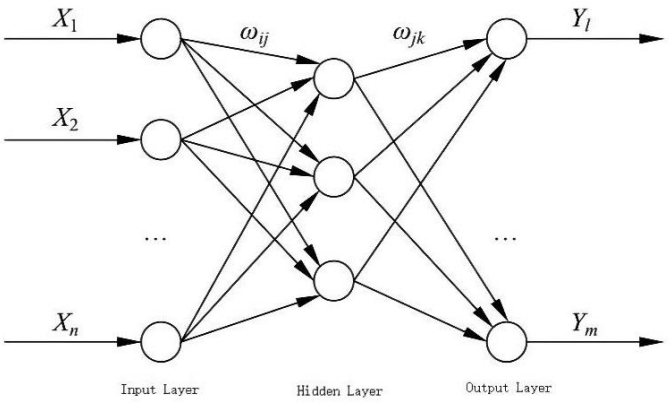


Figure 1: Topological structure of the BPNN model.

Figure **1** demonstrates the BPNN topology diagram, X_1, X_2,…, X_n. The input value of the BPNN model Y_1, Y_2,…, Y_m is the forecasting value of the BPNN whereas ω_{ij} and ω_{jk} are the weights of the BPNN model. As seen in Figure **1**, the BPNN can be regarded as a nonlinear function, and the network's input value and forecasting value are the function's independent and dependent variables, respectively. When the number of input nodes is n and the number of output nodes is m, the BP neural network expresses the function mapping relationship from n independent variables to m dependent variables.

2.3. Support Vector Regression (SVR) Model

Vapnik first proposed the support vector regression (SVR) model. It is based on the principle of structural risk minimization and the VC dimension theory. Compared with other machine learning algorithms, it can effectively improve the learning performance of the model. Compared with artificial neural networks and fuzzy logic theory, the SVR model is considered a good alternative and has a stronger generalization ability. The SVR model can be used for nonlinear regression, and the main research object is small sample data. When forecasting samples, the support vector regression algorithm has a fast operation speed, high forecasting accuracy, and fewer parameters to be adjusted. So, its application field and development prospects are extensive [28-30].

The SVR model adopts the support vector regression method to introduce the structural risk function into the SVR model. The SVM model maximizes the “distance” at the samples closest to the hyperplane, while the SVR model minimizes the “distance” at the farthest samples, as shown in Figure **2**.

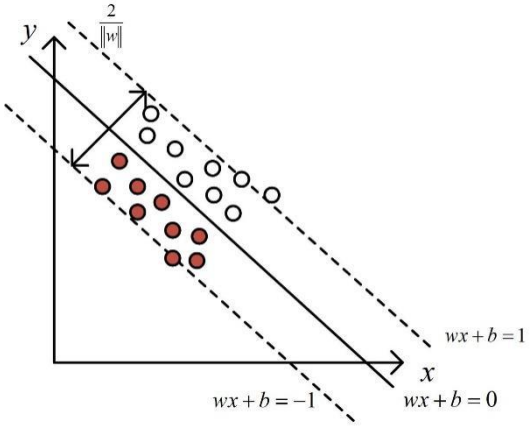


Figure 2(a): SVM schematic.

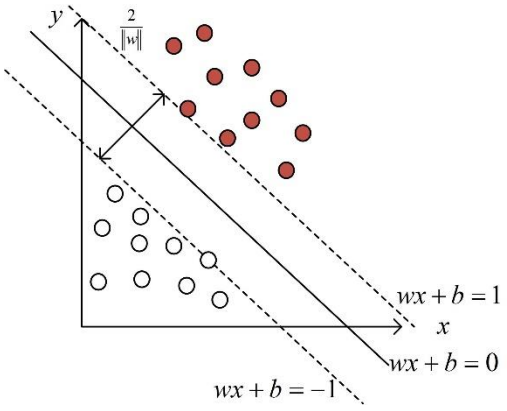


Figure 2(b): SVR schematic.

The SVM optimization objective as Eq. (6),

$$L(w,b,x,y)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{N}\xi_i \tag{6}$$

Its constraints are the Eqs. (7) and (8),

$$y_i(w\cdot x_i+b)\geq 1-\xi_i(i=1,\cdots,N) \tag{7}$$

$$\xi_i\geq 0(i=1,\cdots,N) \tag{8}$$

By introducing relaxation coefficients ξ_i,ξ_i^*, the regression problem becomes the optimal solution. Then the standard form of the SVM regression can be written as Eq. (9),

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2} w^T w + C\sum_{i=1}^{l} \xi_i + C\sum_{i=1}^{l} \xi_i^* \tag{9}$$

Its constraint is Eq. (10),

$$\begin{aligned} & w^T \varphi(x_i) + b - y_i \le \varepsilon + \xi_i \\ & y_i - w^T \varphi(x_i) - b \le \varepsilon + \xi_i^* \\ & \xi_i \ , \xi_i^* \ge 0, i = 1, 2, ..., l \end{aligned} \tag{10}$$

Its dual representation is Eq. (11),

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2} \left(\alpha - \alpha^*\right)^T Q \left(\alpha - \alpha^*\right) + \varepsilon \sum_{i=1}^{l} \left(\alpha - \alpha^*\right) + \sum_{i=1}^{l} y_i \left(\alpha - \alpha^*\right) \tag{11}$$

Its constraint is Eq. (12),

$$\sum_{i=1}^{l} \left(\alpha_i - \alpha_i^*\right) = 0, 0 \le \alpha_i, \alpha_i^* \le C \tag{12}$$

Here $Q_{ij} = K(x_i, x_j) \equiv \varphi(x_i)^T \varphi(x_j)$ is the kernel function.

2.4. ARIMA-BPNN-SVR

The principle of ARIMA-BPNN-SVR is to first use the ARIMA, BPNN, and SVR models to forecast the photovoltaic power generation data in the next 14 days. Then, the three models are regressed to obtain the final forecast result. Photovoltaic power generation data is a non-stationary time series, so traditional ARIMA is used for forecasting. The ARIMA model is a time series model established after differential processing of non-stationary signals based on the ARMA model. The ARIMA model is very simple, requiring only endogenous variables to make forecasting without resorting to exogenous variables. The photovoltaic power generation is subjected to differential operation to eliminate the factors that cause sequence fluctuations, and then the ARMA model is fitted to extract the inherent information of the sequence fully. However, the ARIMA model can only capture the linear features of the sequence and cannot reflect the nonlinear features.

Moreover, the difference operation will cause a certain loss of information to make the sequence stationary. Based on this, selecting BPNN and SVR models to forecast the sequence can fully extract the nonlinear features of the sequence. The BPNN model has a strong nonlinear mapping ability and can approximate any nonlinear continuous function with arbitrary precision. Meanwhile, the BPNN model has self-learning and self-adaptive ability, strong generalization ability, and certain fault tolerance ability, and can better deal with sequences containing noise. However, the BPNN model has a slow convergence speed and is easy to fall into a minimum point. The SVR model uses preset nonlinearity to map project input data to high-dimensional space and to minimize the risk based on the structural risk minimization criterion. Therefore, the forecasted object has good generalization performance. At the same time, when solving the regression problem. The final problem is equivalent to a convex optimization problem, and the results obtained by the SVR model have high fitting accuracy, but the selection of the model kernel function is entirely based on experience. Determining kernel parameters and penalty parameters is optimized by particle swarm optimization and genetic algorithm, but there are still problems such as complex optimization process.

Therefore, this paper performs generalized linear regression on ARIMA, BPNN, and SVR models and also seeks the optimal linear combination model, ARIMA-BPNN-SVR.

The ARIMA-BPNN-SVR model forecasting process is shown in Figure **3**. The specific modeling process is as follows:

Step 1: By establishing an ARIMA model for forecast, a forecasting result y_1 is obtained.

Step 2: Using the BPNN model to forecast the data and get the forecasting result y_2.

Step 3: Establish the SVR model again for forecast and use the PSO particle swarm algorithm to optimize the SVR model to obtain the forecasting result y_3.

Step 4: Use the three forecasting results to establish a regression equation $Y = a_1 y_1 + a_2 y_2 + a_3 y_3 + c$.

Step 5: Get the forecasting result Y.

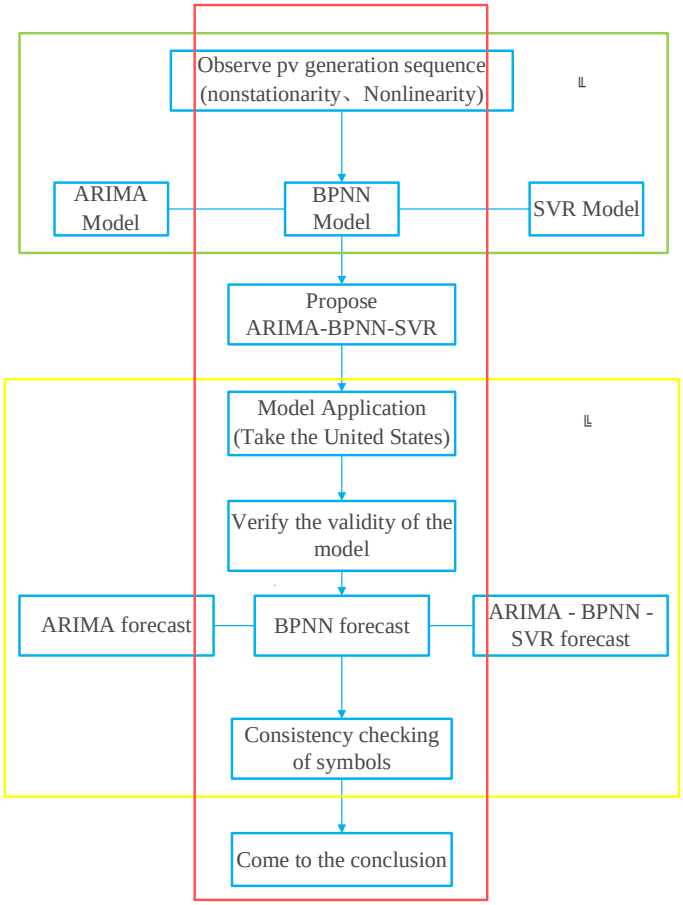


Figure 3: Forecasting flow chart of ARIMA-BPNN-SVR model.

3. Empirical Research

3.1. Data Description and Data Processing

This paper uses the photovoltaic power generation data of 11 regions in the United States from 0:00 on January 1st, 2019, to 23:00 on February 7th, 2019, for a total of 38 days to preprocess the data. By analyzing the characteristics of the data, combined with the corresponding theories and methods, the proposed ARIMA-BPNN-SVR model is proposed to forecast the data. The optimal combination model is obtained using the power generation data of 31 days to forecast the photovoltaic power generation of the next 7 days, and finally,

comparing the accuracy of each model.

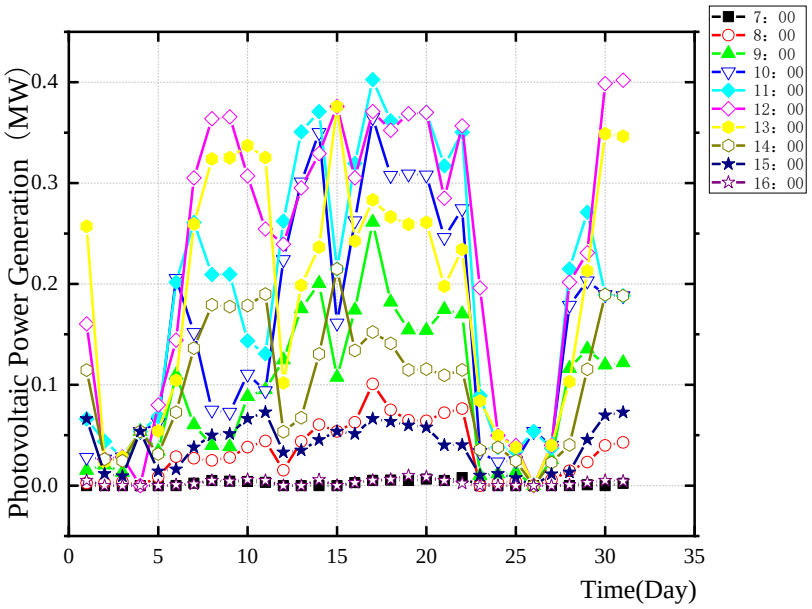


Figure 4: Photovoltaic power generation timing chart.

It can be seen from Figure 4 that the photovoltaic power generation is the lowest at 7:00 and 8:00 in the morning and at 15:00 and 16:00, which is lower than the power generation at other time points, while the photovoltaic power generation at 11:00, 12:00 and 13:00 is relatively high. The power generation is relatively high. The United States is located in the Western Hemisphere, where 7:00 and 8:00 am is the sunrise. Since the sun has just risen, the light is not very sufficient, so the photovoltaic power generation is lower than at other times. Around noon is the longest time of illumination. At noon, the sun directly hits the ground, and the radiation is the largest, so the photovoltaic power generation also reaches the maximum. At 15:00 and 16:00, due to the decrease in sunlight, the photovoltaic power generation also decreased. Every January, the United States is in the winter season, the temperature is low, and extremely cold weather often occurs. It also affects the amount of photovoltaic power generation, resulting in unstable photovoltaic power generation.

In addition, it can be seen from Figure **4** that the two lowest peaks in photovoltaic power generation were in January 2019, i.e., for those days on January 2nd and January 3rd, January 24th, January 25th, and January 26th. According to the US Meteorological Department, since late January 2019, the United States has experienced rare cold weather. Temperatures fell to the lowest levels in decades in much of the region, with the extreme cold causing frostbite to many people and turning more than half of the area into an “ice bank”. A cold snap brought significant cooling and snow to the US Midwest and Northeastern states. Due to the influence of temperature, solar radiation intensity, pressure, humidity, and other factors, photovoltaic power generation was affected. Therefore, this “cold wave” also brought great losses to the power generation of photovoltaic systems in the United States. Extremely cold weather across much of the United States has disrupted the nation's power grid, with rare power outages across much of the country. The sudden extreme cold weather has greatly impacted the energy industry in the United States. Whether it is wind power generation, photovoltaic power generation, other power systems, or oil and gas, they have all suffered great damage. The first to be affected is the US power system, with power companies forecasting a “cold wave” will destroy more than 30 gigawatts of electricity.

Since the United States is at night from 0:00 to 6:00 and 17:00 to 23:00, there is no light, and the photovoltaic power generation is 0MW. Therefore, the data is preprocessed, and the photovoltaic power generation of 24 hours a day is added together, and a total of 38 days of data are obtained. This paper uses the photovoltaic power generation data of the first 31 days as a training set and the photovoltaic power generation data of the next seven days as a validation set to forecast the photovoltaic power generation of the next seven days.

3.2. Experimental Evaluation Indicators

This paper is based on the forecasting of Australian weekly load and demand. MAE, MSE, RMSE, and MAPE are selected as evaluation criteria, and the expressions are Eqs. (13) to (16).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{14}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{15}$$

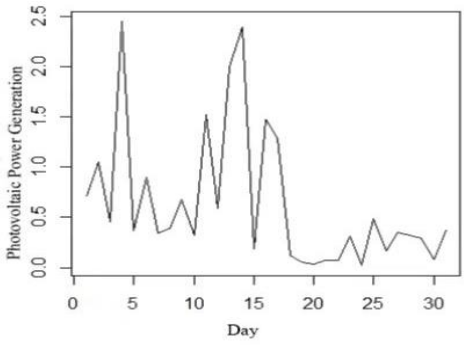
$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{16}$$

where, y_i is the true value, $\hat{y}_i$ is the forecast value, and n is the number of selected forecast points.

3.3. Modeling Analysis

3.3.1. ARIMA Model

According to the ARIMA modeling process, the stationarity test of photovoltaic power generation is first carried out. Figure **5** is a time sequence diagram, an autocorrelation function, and a partial autocorrelation function diagram of photovoltaic power generation. The autocorrelation function is larger than 0 in the first three orders, and it starts to approach 0 in the fourth order. Therefore, it does not show a rapid trend to 0, indicating that the sequence is not stationary. Moreover, the non-stationary sequence is verified by the unit root test.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.190	0.190	1.2317	0.267
		2	0.335	0.311	5.2023	0.074
		3	0.301	0.227	8.5040	0.037
		4	-0.023	-0.216	8.5236	0.074
		5	0.014	-0.159	8.5318	0.129
		6	-0.161	-0.188	9.5863	0.143
		7	0.002	0.182	9.5866	0.213
		8	-0.121	0.030	10.241	0.249
		9	0.071	0.188	10.477	0.313
		10	0.126	0.088	11.256	0.338
		11	-0.096	-0.251	11.723	0.385
		12	0.100	-0.151	12.260	0.425
		13	-0.050	-0.037	12.400	0.495
		14	-0.129	-0.010	13.398	0.495
		15	-0.127	-0.001	14.425	0.494
		16	-0.166	-0.015	16.315	0.431

Figure 5(a): Photovoltaic power generation timing chart. **Figure 5(b):** Autocorrelation and partial autocorrelation function plot.

Therefore, a first-order difference is performed on the original sequence. Figure **6** is the time sequence diagram after the first-order difference, the autocorrelation function diagram, and the partial autocorrelation function diagram. Looking at the timing diagram after the first difference, the sequence fluctuates around the constant 0. Its autocorrelation coefficient has the characteristics of rapidly tending to 0, and it is approximately 0 after the second order. The partial autocorrelation coefficient shows the characteristics of tailing. Therefore, the series is stationary. In addition, the unit root test is performed on the difference series. Moreover, the test results

show that the values of the τ statistic of various models are all less than the significance level $(\alpha = 0.05)$, so it can be considered that the series is significantly stationary.

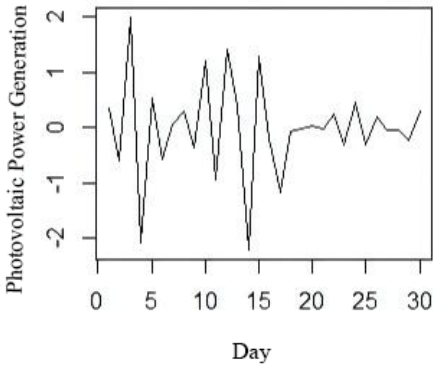


Figure 6(a): The plot of the original data.

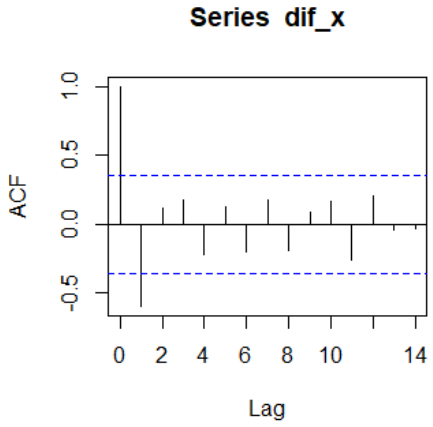


Figure 6(b): Timing diagram after first-order difference.

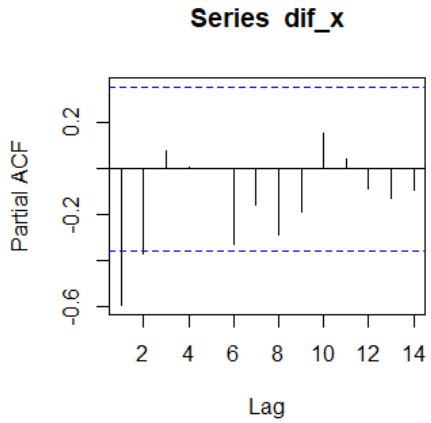


Figure 6(c): Autocorrelation and partial autocorrelation function diagram after first-order difference.

The sequence after the first difference is stationary, so the ARMA model is fitted. The specific identified ARIMA models are shown in Table **1**. It can be found that each evaluation index of ARIMA (2,1,0) is lower than that of ARIMA (0,1,2), which means the ARIMA (2,1,0) model has a better effect. The forecasting results of the ARIMA (2,1,0) model are illustrated in Figure **7**.

Table 1: ARIMA model

	MAE	MSE	RMSE	MAPE
ARIMA (0,1,2)	1.09	2.04	1.43	0.92
ARIMA (2,1,0)	0.91	1.64	1.28	0.61

3.3.2. BPNN Model

1) Number of hidden layer nodes

The number of hidden layer nodes in the BPNN model greatly influences the forecasting accuracy of the BPNN model. Too few nodes in the network will lead to a decrease in the learning efficiency of the network and an increase in the number of training times of the network, resulting in over-training of the network.

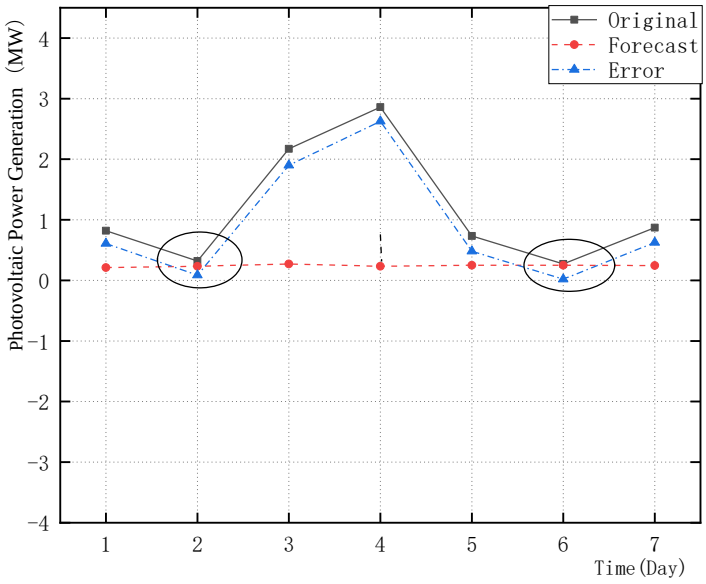


Figure 7: ARIMA (2,1,0) model forecasting effect diagram.

$$l < n-1 \tag{17a}$$

$$l < \sqrt{(m+n)} + a \tag{17b}$$

$$l = \log_2 n \tag{17c}$$

Equation (17) represents the number of nodes in the input layer, the number of nodes in the hidden layer, and the number of nodes in the output layer; it is a constant from 0 to 10. When using the model for forecasting, firstly, the coverage area of the hidden layer nodes is obtained according to the formula, and then the optimal number of nodes is determined by the experimental method.

2). Training and forecasting

Firstly, the 38-day historical solar energy in 11 regions is used, and the 31-day historical data in January is used as a learning sample to forecast the solar energy in the first week of February for seven days.

The number of nodes in the input layer n=11, the number of nodes in the output layer m=1, and the number of nodes in the hidden layer can be determined by the above three methods. Finally, when the number of nodes in the hidden layer is determined to be 3, the training result is shown in Figure **8**. In Figure **8**, you can see that the training set and the validation set almost overlap. Figure **9** is the BPNN model forecasting effect diagram.

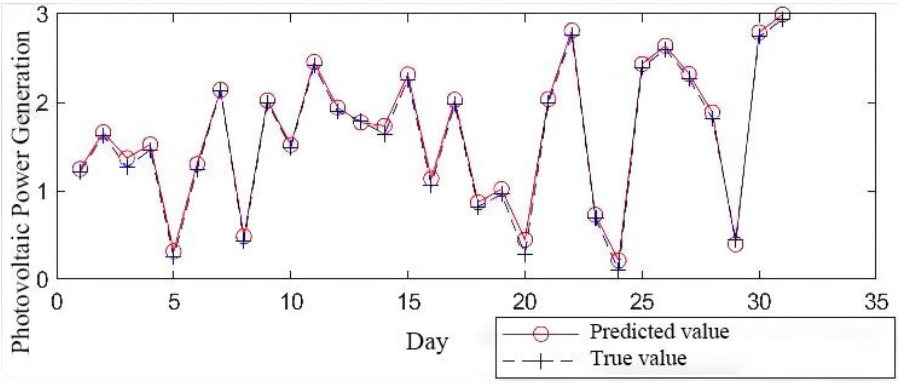


Figure 8: Fitting effect of the BPNN model's training set.

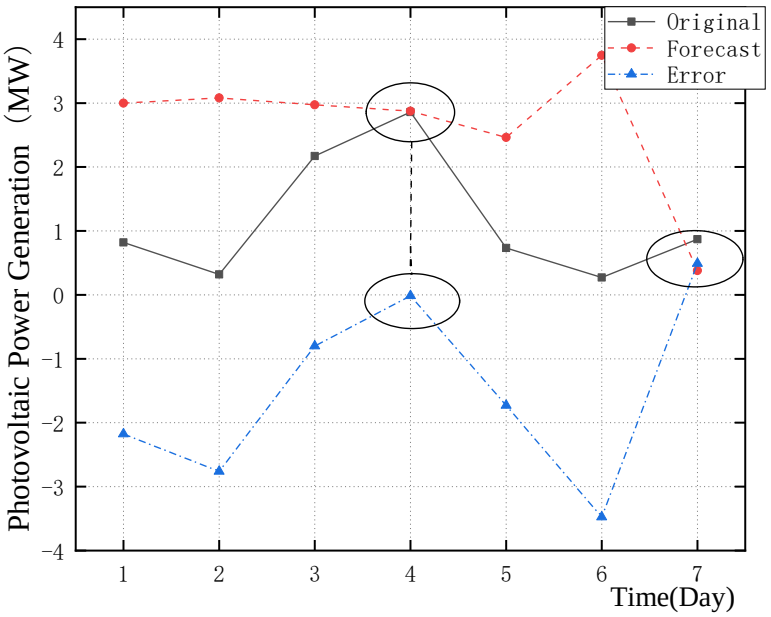


Figure 9: BPNN model forecasting effect diagram.

3.3.3. SVR Model

1). Select training set and test set

Among the 418 samples, rows 1~31 are used as the training set, and rows 32~38 are used as the test set. Among them, rows 1-31, columns 1-10 are used as the input layer of the training set, and column 11 is the output layer of the training set; rows 32-38, columns 1-10 are the input layer of the test set, and column 11 is the input layer of the test set. The output layer for the test set.

The SVR is trained using the training set, and the obtained model is used to make forecasting, which is compared with the forecasting results of the training set and the test set. From Figures **10** and **11**, we can see $R^2 = 0.96$ and $MSE = 0.011$ in the comparison of the training set forecasting results. $R^2 = 0.78$ and $MSE = 0.062$ in the comparison of the test set forecasting results. In principle, the higher the value of E (closer to 1), the better the fit. Generally, if the goodness of fit is greater than 0.7, the model can be considered to meet the requirements. The smaller the error result, the better. The smaller the error result, the higher the fitting degree of the model, and the expected recommendation effect can be achieved. The model achieves the expected forecasting effect from the numerical values of the goodness of fit and the mean square error.

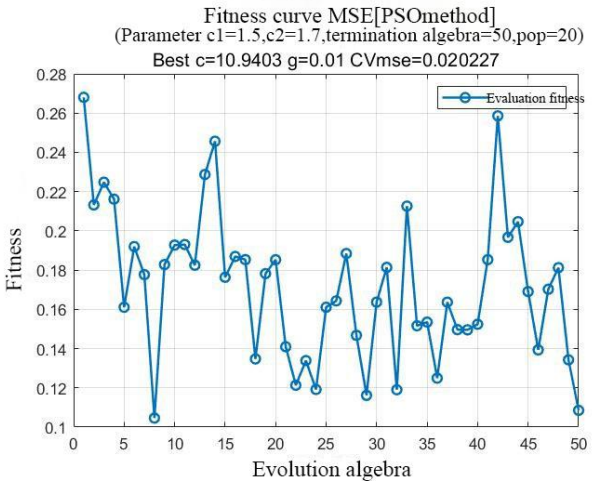


Figure 10(a): Fitness curve.

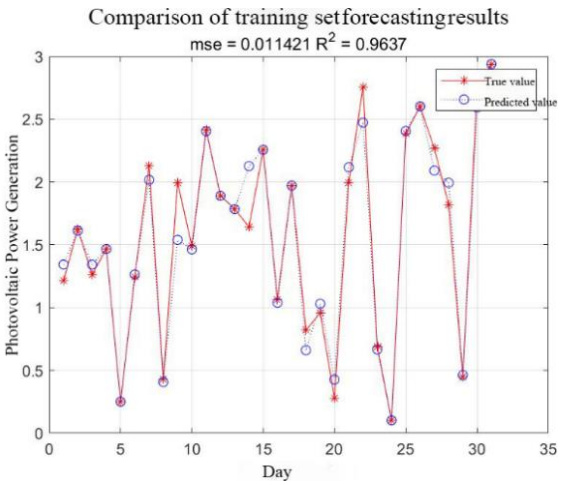


Figure 10(b): Comparison chart of the training set forecasting results.

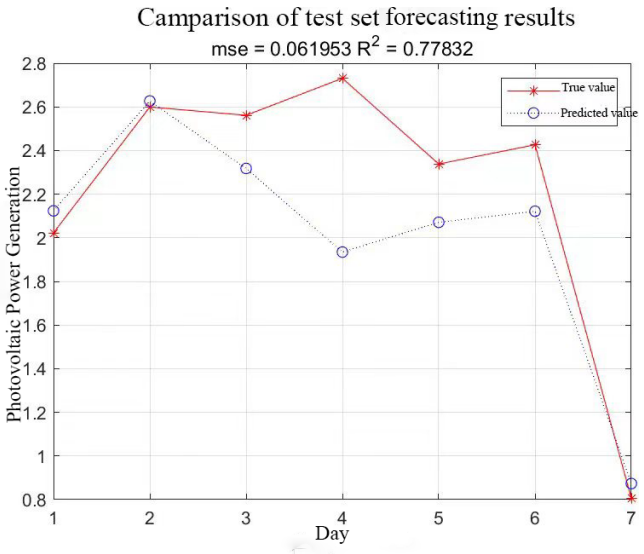


Figure 11: Comparison of SVR forecasting results.

As shown in Figures **10** and **11**, the forecasting comparison of the training set is the same as the actual value, with an error of 1.1%, and the error of the forecasting result of the test set is 6.2%, but the comparison of the forecasting results of the test set has some inaccurate forecasting problems, but within the acceptable error range. The reason for this may be that the overall temperature in some parts of the United States rose sharply from February 3rd to February 5th, the illumination increased, and the photovoltaic power generation also increased, which is consistent with the previous forecasting. It may also be that the model itself still has certain defects and needs further optimization. The model can achieve better forecasting results by appropriately increasing the data size. By comparison, it can be concluded that the error of the SVR model is smaller than that of the other two models, and the forecasting effect of the SVR model is better than that of the ARIMA and BPNN models.

3.3.4. ARIMA-BPNN-SVR

From the modeling of ARIMA, BPNN, and SVR, we found that the forecasting and fitting effect of ARIMA on the sequence is not ideal, and the forecasting error of BPNN and SVR on the sequence is not significant. However, the forecasting results of the sequence are not ideal. Therefore, we established the ARIMA-BPNN-SVR combination model to forecast the sequence and observe the forecasting results.

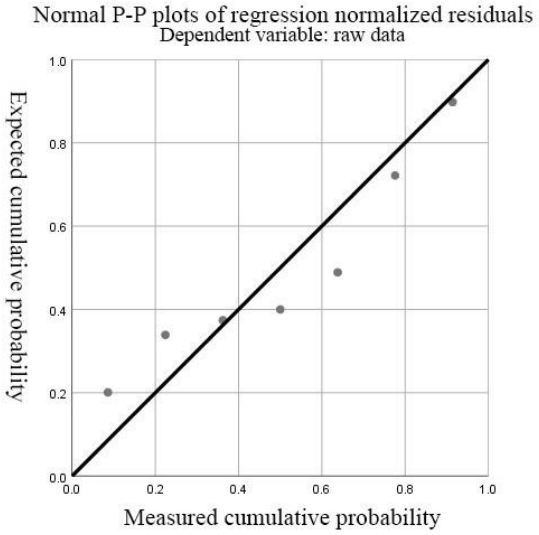


Figure 12: Normal P-P plot of regression normalized residuals.

The y_1 forecasting result was obtained using the ARIMA model, the y_2 forecasting result was obtained using the BPNN model, and the y_3 forecasting result was obtained by using the SVR model. The regression equation $Y = 10.173 * y_1 + 0.205 * y_2 - 0.446 y_3 - 0.955$ is established, the forecasting value Y is obtained, and the forecasting effect graph is drawn. Figure **12** (the Normal P-P plot) assumes that the error terms are normally distributed, and from Figure **13**, we can see that the forecasting effect of the ARIMA-BPNN-SVR model fits the original data much better than the other three, and the sequence trends are consistent, especially on the fifth and seventh days. The error is almost zero, which further illustrates the effectiveness of the model proposed in this paper. From February 3rd to 5th, large errors were also analyzed in the previous article. Affected by the "cold wave" weather, the temperature in the United States plummeted, and blizzards raged, resulting in reduced sunshine and reduced photovoltaic power generation. From the 3rd to the 5th, the overall temperature in different parts of the United States rose sharply, and the weather was cloudy and sunny. Therefore, the photovoltaic power generation in these three days increased significantly. Although the temperature continued to rise from the 6th to the 8th, the power generation showed a downward trend. This is because these days are cloudy and rainy, and the conversion of the photovoltaic power generation system is unstable. Therefore, the forecasted value shows an opposite trend to the actual power generation. The temperature was stable from the 10th to the 14th, and the weather did not change much, so the forecasting effect was very good. Furthermore, after the extreme cold in February and a rare blackout in the US, many are turning to battery storage as a major component of their solar system. Therefore, more and more people are installing solar power systems.

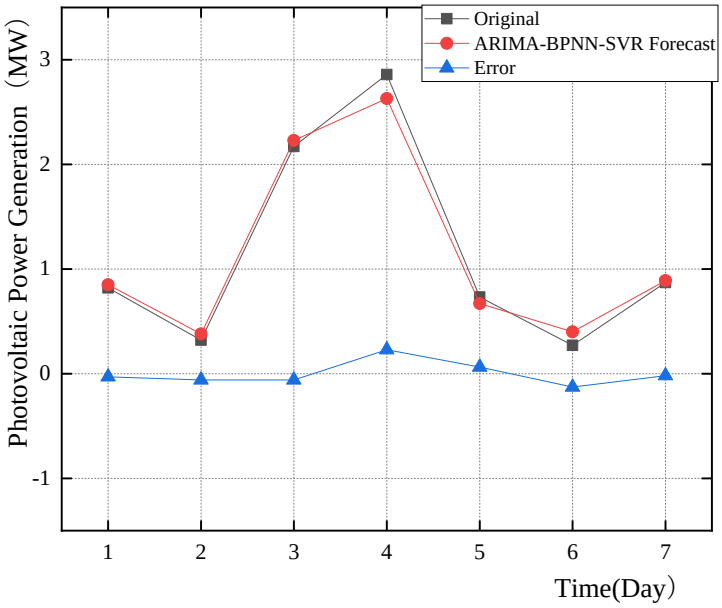


Figure 13: Forecasting effect of the proposed ARIMA-BPNN-SVR model.

4. Analysis and Discussion of Results

4.1. Error Analysis of Mixed Model

To choose which of the methods proposed in this paper is optimal, three methods, ARIMA, BPNN, and SVR model, are compared, as shown in Table **2**.

Table **2** shows the evaluation indicators of each forecasting model. It can be seen that the ARIMA-BPNN-SVR model proposed in this paper has significantly improved MAE, MSE, and RMSE compared with the ARIMA, BPNN, and SVR models. However, we see that the forecasting effects of ARIMA and BPNN, both single models, are quite different. There are many variables in the hidden layer of the neural network. When the training time is too long, the network will be over-trained. The network has high accuracy in the training set, but when it is taken out of it

and used on other data, its accuracy is significantly reduced. BPNN model also has some disadvantages, such as easy to fall into local optimum, and its training effect depends on the initial random weight. In forecasting the ARIMA model, because it is a non-stationary sequence, first-order and second-order differences are performed on it. ARIMA handles the randomness of the sequence, so compared with BPNN, it has a good forecasting effect. Although the forecasting effect is not as good as that of SVR and ARIMA-BPNN-SVR, the overall forecasting accuracy is within an acceptable range. SVR has unique advantages in solving small-sample, nonlinear, and high-dimensional pattern recognition. By finding the smallest structural risk, the generalization ability of the learning machine can be improved to reduce the empirical risk and the maximum confidence interval so that a good statistical law can be obtained to ensure a small number of samples, thereby improving the forecasting effect. However, the forecasting effect is still not as good as that of ARIMA-BPNN-SVR. It can be seen that using different methods for different data helps extract sequence information and improve the forecasting effect.

Table 2: Error comparison results

	ARIMA	BPNN	SVR	ARIMA-BPNN-SVR
MAE	0.91	1.64	1.13	0.72
MSE	1.65	4.05	1.87	0.79
RMSE	1.28	2.01	1.37	0.89
MAPE	0.61	3.93	2.56	1.13

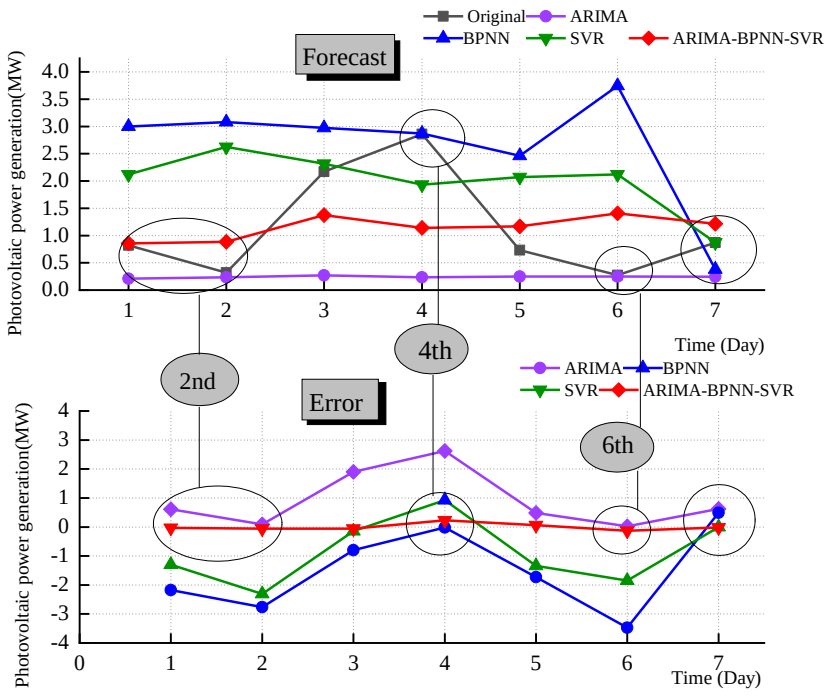


Figure 14: Forecasting values and errors of each model.

Combined with the error comparison results in Table **2** and the forecasting values and error maps of each model in Figure **14**, the analysis can be performed more intuitively and vividly. We can see that ARIMA's forecast effect is not ideal, the error is large, and the overall trend and fluctuation range of photovoltaic power generation are not reflected. For the BPNN model, it is found that the forecasting error of these points on the 4th and 7th days is very low, especially for the 4th day when the other three methods do not achieve the forecasting accuracy. Although the forecasting effect of PSO-SVR is more accurate than that of the BPNN model, there are deviations in the forecasting of individual points, the error fluctuates greatly, and the forecasting effect is unstable. The

forecasting value of the day is different from the actual value. The forecasting effect of the ARIMA-BPNN-SVM model proposed in this paper is relatively stable overall. Furthermore, the forecasting error is smaller than that of the other three methods, and only a few points have large errors.

By analyzing the different photovoltaic power generation at different times, in the time range of 0:00~6:00, the United States is at night, and there is no sunlight, so the photovoltaic power generation is 0MW. From 7:00 to 16:00, the United States is in the daytime, there is sunlight, and photovoltaic power generation increases, and as previously analyzed, 11:00 am, noon, and 13:00, 14:00 are all the most abundant moments, so the photovoltaic power generation reaches its maximum value. After 14:00, the sun began to decline slowly, the amount of solar radiation decreased, and the photovoltaic power generation decreased until the sun finally went down, and the photovoltaic power generation decreased to 0MW. In January 2019, the United States was in winter and experienced a rare "cold wave" in 30 years. The snowstorm raged, and the temperature dropped to extremely low. Therefore, the amount of light in January was not high. From February 2nd to February 4th, the overall temperature in the United States has risen sharply, the illumination has increased, and the photovoltaic power generation has also increased. However, from February 1st to February 14th, the weather was mostly cloudy and rainy, with strong winds. As the temperature and weather are constantly changing, the power generation effect of the photovoltaic power generation system becomes poor.

Therefore, the ARIMA-BPNN-SVR model proposed in this paper only uses historical data to predict the future photovoltaic power generation, and the prediction effect is easily affected by external environments such as bad weather. Moreover, ARIMA requires the sequence to be stationary, and for non-stationary data, it needs to be stationary through different operations. However, the randomness of photovoltaic data is large, and multiple differences will lead to excessive information loss. BPNN and SVR still do not have good theoretical guidance in parameter selection to achieve optimal results. Therefore, ARIMA-BPNN-SVR still has limitations in dealing with more random data and seeking the optimal prediction effect. In the future, further in-depth research on data preprocessing and optimization algorithms can be used to deal with severe weather changes.

4.2. Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test of two paired samples analyzes two paired samples to infer whether there is a difference between the population distributions. The null hypothesis is that the two population distributions from the two paired samples are not significantly different. The error comparison chart shows that the MSE, RMSE, and MAPE of the SVR model and the ARIMA-BPNN-SVR model are the closest to the actual value. Therefore, the SVR model, the ARIMA-BPNN-SVR model, and the real value are selected as paired samples for testing, and the other two A signed-rank tests were performed. The null hypothesis proposed in this paper is that there is no significant difference between the actual value and the forecasting value. The significance level was selected $\alpha = 0.05$, and the test results were obtained, as shown in Table **3**.

As can be seen from Table **3**, the *p* values of the signed-rank test between the SVR model and the ARIMA-BPNN-SVR model are 0.075 and 0.735, respectively, which are greater than 0.05, accepting the null hypothesis, indicating that there is no significant difference between the two. When $p < 0.05$, there is a significant difference between the two groups, and when $p < 0.01$, there is a very significant difference between the two groups, and it is reasonable to think that the larger the *p*-value, the stronger the reason for accepting the null hypothesis. The *p*-value of the ARIMA-BPNN-SVR model is the largest, and it can be considered that the ARIMA-BPNN-SVR model is closer to the original value, which means that the forecasting result of the ARIMA-BPNN-SVR model is better, and the forecasting effect of the method selected in this paper is better.

Table 3: Signed rank test results

	BPNN	**SVR**	**ARIMA**	**ARIMA-BPNN-SVR**
Signed rank test (*p*-values)	0.04	0.08	0.02	0.74

4.3. Compare with Recent Models

Table **4** lists the details of comparing results between the recent and proposed models. In Table 4, the authors listed the comparisons among the recently proposed models in the literature and the proposed model in this paper. The first benchmark model for comparison was proposed by VanDeventer et al. [32] by using a genetic algorithm-based support vector machine (GASVM) model for short-term power forecasting of residential-scale photovoltaic systems. The proposed GASVM model initially classifies the historical weather data using an SVM classifier and is later optimized by the genetic algorithm using an ensemble technique. Experimental results demonstrated that the proposed GASVM model outperforms the conventional SVM model by a difference of about 669.624 W in the RMSE value and 98.7648% of the MAPE error.

The second benchmark model for comparison was proposed by Seyedmahmoudian et al. [33], conducting a comparative study of standard particle swarm optimization (PSO) and differential evolution (DE). Results of the 1-h time horizon demonstrate that the RMSE, MRE, MAE, MBE, WME, and VAR of the DEPSO-based forecasting is 4.4%, 3.1%, 0.03, −1.63, 0.16, and 0.01, respectively. Results also indicate that the proposed DEPSO approach is more efficient and accurate compared with the PSO and DE.

Table 4: Comparisons among the recent models and the proposed model

Ref.	Year	Method used	Location	Horizon	Forecast error
VanDeventer *et al*. [32]	2019	GASVM	Geelong	1 day	RMSE 11.226W MAPE 1.705%
Seyedmahmoudian *et al*. [33]	2018	DEPSO	Victoria	1 day	RMSE 4.4% MRE 3.1% MAE 0.03 MBE -1.63 WME 0.16 VAR 0.01
The proposed model	2022	ARIMA-BPNN-SVR	USA	1 day	MAE 0.53 MSE 0.41 RMSE 0.64 MAPE 0.84

This paper proposes the ARIMA-BPNN-SVR model for photovoltaic power generation forecasting. The experimental results show that the proposed ARIMA-BPNN-SVR outperforms the single models ARIMA, BPNN, and SVR forecasting performance. Its MAE, MSE, RMSE and MAPE are 0.53, 0.41, 0.64, and 0.84, respectively.

The proposed ARIMA-BPNN-SVR model combines the advantages of ARIMA, ARIMA, and SVR models. In addition, using data preprocessing and optimization algorithms, the above models also give good forecasting results. Therefore, in the future, further research can be done here.

5. Conclusion

The study of photovoltaic power generation forecast is one of the important contents of photovoltaic power station research. Due to the instability, intermittency, and randomness of solar energy, the current solar power generation forecasting models have problems such as significant forecasting errors and weak generalization ability.

The hybrid forecasting model based on ARIMA, BPNN, and SVR proposed in this paper is used for the short-term power generation forecast of the photovoltaic power generation system. The proposed ARIMA-BPNN-SVR model makes ARIMA, BPNN, and SVR form complementary advantages. When dealing with non-stationary power

generation data, ARIMA fully extracts the linear features of the series. BPNN and SVR fully extract nonlinear sequence features. Both BPNN and SVR have strong nonlinear mapping and generalization capabilities. BPNN has self-learning and adaptive ability, strong generalization ability, and certain fault tolerance ability. The final solution of SVR is a convex quadratic programming problem, which, in theory, will be the optimal global solution. It solves the local extremum problem that cannot be avoided in BPNN. Therefore, the proposed ARIMA-BPNN-SVR model can fully extract sequence information, and its generalization ability is strong, which can reduce the interference of outliers and avoid falling into the problem of local optimum, so the prediction effect is greatly improved. In the forecast of photovoltaic power generation in the United States, the four evaluation indicators of the ARIMA-BPNN-SVR model proposed in this paper, i.e., MAE, MSE, RMSE, and MAPE are 0.53, 0.41, 0.64, and 0.84, respectively. Using ARIMA, BPNN and SVR are reduced by about 50%, and MSE is reduced by about 70%, which means the prediction effect of the mixed model is significantly improved. In the Wilcoxon sign-rank test, the *p*-value of the proposed ARIMA-BPNN-SVR model reaches 0.98, indicating that the forecasted value is infinitely close to the original power generation data, which proves the validity and general applicability of the proposed ARIMA-BPNN-SVR model.

Acknowledgments

Guo-Feng Fan is thankful for the support from the project grants: Teaching Reform Project of Pingdingshan University (2021-JY29). Wei-Chiang Hong thanks the support from the Ministry of Science and Technology, Taiwan (MOST 110-2410-H-161-001 & MOST 111-2410-H-161-001).

Nomenclature

Symbol	Symbol Description
$ARIMA$	The differential autoregressive moving average process
SVR	Support vector regression
$BPNN$	BP neural network
p	The order of the autoregressive model
x_t	The time series $\{x_t\}$
ϕ_0	The coefficient of the AR model
ω	The weight of the BP neural network
L	Lagrange function
y_1, y_2, y_3	Forecasting results
ε_t	Random disturbance term
E	Mathematical expectation
Var	Variance
m	Number of nodes in the output layer
σ_ε^2	The variance of the random disturbance term
q	The order of the moving average model
α	Lagrange multiplier
μ	The constant mean of the moving average model

θ_1	Moving average coefficient
d	The differential order
y_i	True value
MAE	Mean absolute error
MSE	Mean square error
MRE	Mean relative error
WME	Weekly mean error
c	Constant term
$\Phi(B)$	Polynomial autoregressive coefficients of stationary reversible ARMA models
∇^d	$(1-B)^d$
$\Theta(B)$	Moving average coefficient polynomial of stationary reversible ARMA model
X_n	The input value of the BPNN
m	The number of the output node
Y_m	The forecasting value of the BPNN
n	The number of the input node
w	The normal vector to the hyperplane
$\|w\|$	The length of the parameter w to be optimized
C	Regularization parameter
a_1, a_2, a_3	Coefficient of the regression equation
b	The intercept of the hyperplane
$Q_{ij} = K(x_i, x_j)$	Kernel function
l	Number of nodes in the hidden layer
n	Enter the number of layer nodes
ξ_i	Stretching factor
α^*	Lagrange multiplier
a	Constant term
Y	Regression equation
$\hat{y}_i$	Forecast value
$RMSE$	Root mean square error
$MAPE$	Mean absolute percentage error
MBE	Mean bias error
VAR	The variance of the prediction errors

References

[1] Li YT, Su Y, Shu LJ. An ARMAX model for forecasting the power output of a grid connected photovoltaic system. *Renewable Energy*, 2014; 66: 78-89. https://doi.org/10.1016/j.renene.2013.11.067.

[2] Zhang JL, Tan ZF, Wei YM. An adaptive hybrid model for day-ahead photovoltaic output power prediction. Journal of Cleaner Production, 2020; 244: 118858. https://doi.org/10.1016/j.jclepro.2019.118858.

[3] Ding S, Li RJ, Tao Z. A novel adaptive discrete grey model with time-varying parameters for long-term photovoltaic power generation forecasting. Energy Conversion and Management, 2021; 227: 113644. https://doi.org/10.1016/j.enconman.2020.113644

[4] Zhong ZF, Yang CX, Cao WY, Yan CY. Short-term photovoltaic power generation forecasting based on a multivariable grey theory model with parameter optimization. Mathematical Problems in Engineering, 2017; 2017: 5812394. https://doi.org/10.1155/2017/5812394

[5] Yu L, Ma X, Wu WQ, Xiang XW, Wang Y, Zeng B. Application of a novel time-delayed power-driven grey model to forecast photovoltaic power generation in the Asia-Pacific region. Sustainable Energy Technologies and Assessments, 2021; 44:100968. https://doi.org/10.1016/j.seta.2020.100968.

[6] Yadav AK, Chandel SS. Identification of relevant input variables for prediction of 1-minute time-step photovoltaic module power using artificial neural network and multiple linear regression models. Renewable and Sustainable Energy Reviews, 2017; 77: 955-969. https://doi.org/10.1016/j.rser.2016.12.029.

[7] Zhong JQ, Liu LY, Sun Q, Wang XY. Prediction of photovoltaic power generation based on general regression and back propagation neural network. Energy Procedia, 2018; 152: 1224-1229. https://doi.org/10.1016/j.egypro.2018.09.173.

[8] AlShafeey M, Csáki C. Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods. Energy Reports, 2021; 7: 7601-7614. https://doi.org/10.1016/j.egyr.2021.10.125.

[9] Wu YX, Guo TT, Wang H. Photovoltaic power prediction method based on similar day and BP neutral network. IOP Conference Series: Earth and Environmental Science, 2018; 170: 042016. https://doi.org/10.1088/1755-1315/170/4/042016.

[10] Zhang Y, Kong LQ. Photovoltaic power prediction based on hybrid modeling of neural network and stochastic differential equation. ISA Transactions, 2021. https://doi.org/10.1016/j.isatra.2021.11.008.

[11] Natarajan Y, Kannan S, Selvaraj C, Mohanty SN. Forecasting energy generation in large photovoltaic plants using radial belief neural network. Sustainable Computing: Informatics and Systems, 2021; 31: 100578. https://doi.org/10.1016/j.suscom.2021.100578.

[12] Eseye AT, Zhang JH, Zheng DH. Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information. Renewable Energy, 2018; 118: 357-367. https://doi.org/10.1016/j.renene.2017.11.011.

[13] Xun T, Lei SH, Ding XC, Chen K, Huang K, Nie YX. Photovoltaic power forecasting method based on adaptive classification strategy and HO-SVR algorithm. Energy Reports, 2020; 6: 921-928. https://doi.org/10.1016/j.egyr.2020.11.108.

[14] Pan MZ, Li C, Gao R, Huang YT, You H, Gu TS, Qin F. Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization. Journal of Cleaner Production, 2020; 277: 123948. https://doi.org/10.1016/j.jclepro.2020.123948.

[15] Niu DX, Wang KK, Sun LJ, Wu J, Xu XM. Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: A case study. Applied Soft Computing 2020; 93: 106389. https://doi.org/10.1016/j.asoc.2020.106389.

[16] Meng M, Song CG. Daily Photovoltaic Power Generation Forecasting Model Based on Random Forest Algorithm for North China in Winter. Sustainability 2020, 12: 2247. https://doi.org/10.3390/su12062247.

[17] Wang JZ, Zhou YL, Li ZW. Hour-ahead photovoltaic generation forecasting method based on machine learning and multi objective optimization algorithm. Applied Energy, 2022; 312: 118725. https://doi.org/10.1016/j.apenergy.2022.118725.

[18] Zang HX, Cheng LL, Ding T, Cheung KW, Wei ZN, Sun GQ. Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning. International Journal of Electrical Power & Energy Systems, 2020; 118: 105790. https://doi.org/10.1016/j.ijepes.2019.105790.

[19] Wang F, Zhang ZY, Liu C, Yu YL, Pang SL, Duić N, Shafie-khah M, Catalão JPS. Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting. Energy Conversion and Management, 2019; 181: 443-462. https://doi.org/10.1016/j.enconman.2018.11.074.

[20] Sharma N, Mangla M, Yadav S, Goyal N, Singh A, Verma S, Saber T. A sequential ensemble model for photovoltaic power forecasting. Computers & Electrical Engineering, 2021; 96: 107484. https://doi.org/10.1016/j.compeleceng.2021.107484.

[21] Ahmed R, Sreeram V, Togneri R, Datta A, Arif MD. Computationally expedient Photovoltaic power Forecasting: A LSTM ensemble method augmented with adaptive weighting and data segmentation technique. Energy Conversion and Management, 2022; 258: 115563. https://doi.org/10.1016/j.enconman.2022.115563.

[22] Huang XQ, Li Q, Tai YH, Chen ZQ, Liu J, Shi JS, Liu WM. Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM. Energy, 2022; 246: 123403. https://doi.org/10.1016/j.energy.2022.123403.

[23] Mellit A, Massi Pavan A, Lughi V. Deep learning neural networks for short-term photovoltaic power forecasting. Renewable Energy, 2021; 172: 276-288. https://doi.org/10.1016/j.renene.2021.02.166.

[24] Wang ZP, Qu JF, Fang XY, Li H, Zhong T, Ren H. Prediction of early stabilization time of electrolytic capacitor based on ARIMA-Bi_LSTM hybrid model. Neurocomputing, 2020; 403: 63-79. https://doi.org/10.1016/j.neucom.2020.03.054.

[25] Liu LY, Liu DR, Sun Q, Li HL, Wennersten R. Forecasting power output of photovoltaic system using a BP network method. Energy

Procedia, 2017; 142: 780-786. https://doi.org/10.1016/j.egypro.2017.12.126.

[26] Li YQ, Zhou L, Gao PQ, Yang B, Han YM, Lian C. Short-Term power generation forecasting of a photovoltaic plant based on PSO-BP and GA-BP neural networks, Frontiers in Energy Research, 2022. https://doi.org/10.3389/FENRG.2021.824691.

[27] Hu KY, Wang LD, Li WJ, Cao SH, Shen YY. Forecasting of solar radiation in photovoltaic power station based on ground-based cloud images and BP neural network. IET Generation, Transmission & Distribution, 2021; 16: 333-350. https://doi.org/10.1049/GTD2.12309.

[28] Lin GQ, Li LL, Tseng ML, Liu HM, Yuan DD, Tan RR. An improved moth-flame optimization algorithm for support vector machine prediction of photovoltaic power generation. Journal of Cleaner Production, 2020; 253: 119966. https://doi.org/10.1016/j.jclepro.2020.119966.

[29] Boegli M, Stauffer Y. SVR based PV models for MPC based energy flow management. Energy Procedia, 2017; 122: 133-138. https://doi.org/10.1016/j.egypro.2017.07.317.

[30] Lin JM, Li HM. A Hybrid K-Means-GRA-SVR model based on feature selection for day-ahead prediction of photovoltaic power generation. Journal of Computer and Communications, 2021; 9: 91-111. https://doi.org/10.4236/JCC.2021.911007.

[31] Das UK, Tey KS, Seyedmahmoudian M, Mekhilef S, Idris MYI, Van Deventer W, Horan B, Stojcevski A. Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, 2018; 8: 912-928. https://doi.org/10.1016/j.rser.2017.08.017.

[32] VanDeventer W, Jamei E, Thirunavukkarasu, GS, Seyedmahmoudian M, Soon TK, Horan B, Mekhilef S, Stojcevski A. Short-term PV power forecasting using hybrid GASVM technique. *Renewable Energy*, 2019; 140:367-379. https://doi.org/10.1016/j.renene.2019.02.087.

[33] Seyedmahmoudian M, Jamei E, Thirunavukkarasu GS, Soon TK, Mortimer M, Horan B, Stojcevski A, Mekhilef S. Short-Term Forecasting of the output power of a building-integrated photovoltaic system using a metaheuristic approach. *Energies*, 2018, 11:1260. https://doi.org/10.3390/en11051260.