Prediction model of photovoltaic power generation based on improved granular computing and neural networks

Zhibo Yang, Suhan Wang

College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Abstract. The application potential of solar energy generation is immense, yet its volatility poses significant challenges to power supply. With the aim of enhancing the accuracy of photovoltaic power generation prediction to better address its variability and stochastic nature, this study proposes a method based on enhanced granular computing and neural networks. The model utilizes wavelet transform to analyze the original time series in the frequency domain and employs fast Fourier transform to extract major periodic components. These components are then used to construct granularity matrices for training a backpropagation neural network (BPNN), aiming to achieve precise prediction of photovoltaic power generation. The research findings demonstrate that the model achieves an average error rate of around 17% on the test dataset, exhibiting outstanding performance compared to other classical time series prediction models. This study provides an effective method and reference for photovoltaic power generation prediction, contributing significantly to the field.

Keywords: photovoltaic power prediction; granular computing; BP neural network; frequency domain analysis; periodic extraction.

1. Introduction

Sustainable energy plays a crucial role in national development. Despite China's vast land and abundant resources, the richness of resources does not imply unrestricted exploitation. In the present era, human energy demands are increasing with technological progress, making energy scarcity a significant issue for China's development. Without sufficient energy support, national progress is unattainable. Particularly in China, where energy consumption is relatively high, there is a necessity to actively pursue the development of new renewable energy sources [1]. Solar energy is considered an ideal renewable energy source due to its abundant reserves and environmentally friendly characteristics, making it highly competitive for the future. China's photovoltaic (PV) installed capacity has nearly tripled in the past five years and is expected to continue growing in the coming years. As of October 2023, China's cumulative PV installed capacity has reached 53.6 billion kilowatts, accounting for 38.2% of the country's total sustainable energy generation capacity, with over half of the nation's new installed capacity coming from PV. Furthermore, analysis suggests that by 2030, renewable energy capacity will increase by 2.4 times, with 95% of this growth coming from solar PV and wind sources.

PV generation, as an emerging renewable energy technology, is heavily influenced by solar radiation intensity and other meteorological conditions, resulting in high volatility and randomness that pose significant challenges to power supply. To mitigate the impact of PV generation uncertainty on the power system and enhance overall system reliability, accurate prediction of PV power output is essential. Many experts have researched PV output forecasting, which can be broadly categorized into two main types: physics-based direct measurement methods, utilizing mathematical models based on the operational principles of PV systems and empirical weather data to predict power output, and statistical machine learning and deep learning models. With the advancement of artificial intelligence technology, the latter category is increasingly applied by scholars, including typical methods such as Long Short-Term Memory (LSTM) neural network models for prediction [2], convolutional neural networks (CNNs), recurrent neural networks (RNNs) [4], and support vector machines [5]. Some researchers also consider the periodic characteristics of PV power and combine...
frequency domain decomposition with neural networks to achieve short-term PV power prediction [6].

Granular computing is a novel method in the field of artificial intelligence research for simulating human thinking and solving complex problems [7]. In time series forecasting, some scholars have achieved precise fitting of time trends by combining granular computing with neural network models. For instance, M. Jafarian [8] and others have developed models combining fuzzy modeling techniques with artificial neural networks to estimate annual wind turbine power generation in different regions, demonstrating improved performance over traditional methods. Shen [9] proposed a BP neural network based on granularity computing (BP GCC), which effectively predicts financial trends by considering the periodic granularity characteristics of financial data.

This paper builds upon existing granular computing and neural network models proposed by scholars and introduces data preprocessing to enhance prediction model accuracy. The data preprocessing part mainly involves identifying the periodic characteristics of the original time series. Since PV generation time series exhibit periodicity and volatility, past periods can influence future periods. Thus, this study utilizes frequency domain transformation to fully explore the sequence’s periodic features, identifying several significant time windows to guide the selection of granularity matrices in the prediction model. Using data from Zhoushan City, Zhejiang Province in 2020 as a case study, the research results demonstrate that the improved prediction model achieves an average error rate of around 17% on the test set, outperforming the original granularity-based BP neural network and other classical time series prediction models. This study provides a new reference for PV power prediction in time series scenarios.

2. Basic Theory

2.1. Wavelet transform and Fourier transform

Wavelet transform is a time-frequency analysis method that exhibits distinct advantages in analyzing non-stationary time series, thus becoming an effective tool for current time series analysis. The definition of wavelet transform is as follows [10]:

Let \( f(t) \) be a square integrable function, denoted as \( f(t) \in L^2(R) \), and \( \varphi(t) \) be the mother wavelet. If \( \varphi(t) \) satisfies the condition:

\[
C_\varphi = \int_{-\infty}^{\infty} \frac{|\hat{\varphi}(\omega)|^2}{\omega} d\omega < \infty
\]  

(1)

Then, the wavelet transform of \( f(t) \) is defined as:

\[
Wf(a, b) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{a}} \varphi^* \left( \frac{t-b}{a} \right) dt
\]  

(2)

In equation (2), \( a \) represents the scale factor, and \( b \) represents the translation factor. When performing wavelet transform, it is necessary to choose a base function. Common choices include Haar wavelets, dbN wavelets, symN wavelets, etc. Considering the purpose of this paper is to explore the periodic characteristics of time series, dbN wavelets are selected as the base function. However, wavelet transform only allows for approximate observation of partial periodic components of a time series, and it can be challenging to determine longer or shorter periodic components. Therefore, this paper also comprehensively considers Fourier transform to explore periodic features.

Fourier transform can convert signals between the time domain and the frequency domain, essentially decomposing the original signal \( f(t) \) in the time domain into a sum of different periodic cosine and sine waves. The definition is as follows [11]:

\[
F(j\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt
\]  

(3)
Here, \( F(j\omega) \) is the result of the fast Fourier transform. Through this transformation, a signal in the time domain can be transformed into a signal in the frequency domain.

### 2.2. Granular Computing

A grain refers to a category of elements, and granulation of information involves the process of partitioning objects of the same category into different grains, where each grain is a combination of objects based on similarity, indistinguishability relationships, and functional aggregation. The average measure of the size of grains is called granularity. In information representation, granularity can be used to analyze data information and knowledge abstraction from different perspectives [12].

For representing the size of a grain: Let the given domain be \( U \), and the granularity partition be \( U = \bigcup_{i \in \tau} G_i \). Then, the size of the grain \( G \) is defined as \( d[G] = \int_G dx \). If the domain is discrete, this formula represents the total number of individuals in the information grain \( G \); if the domain is continuous, this expression measures the length of the information grain. Essentially, finer granularity leads to more complex search processes, while coarser granularity simplifies the search process. The process of ascending from fine granularity to coarse granularity is an abstraction of information process that significantly reduces the complexity of data processing [13].

In intuitive terms, the basic components of granular computing mainly include particles, granule layers, and granule structures. A particle is a vague term and is the most basic element that constructs a granular computing model. For example, in the context of this study on photovoltaic power generation, the numerical values of daily power generation are small particles, whereas the values of weekly power generation over seven days form a larger grain.

### 2.3. BP neural network

Artificial Neural Networks (ANNs) [14] are computational models inspired by the biological neural system, designed to mimic the information processing and transmission mechanisms between neurons in the human brain. The Backpropagation (BP) algorithm is a typical learning algorithm used in artificial neural networks. Its learning process involves two key stages: forward propagation and backward propagation.

During the forward propagation stage, input information propagates from the input layer through hidden layers to the output layer, where the output values are computed based on the activation functions of neurons. The computed output values are then compared with the expected values. If there is an error present, the process moves to the backward propagation stage.

In the backward propagation stage, the error information is propagated back along the original connection pathways. This involves adjusting the weights of neurons layer by layer to minimize the error. This iterative process continues until the network’s output results meet the desired accuracy requirements.
In this study, the first column of the granularity matrix is used as input, and it is compared with the first value of the second column. The error is then propagated backward, and the parameters of the layers in the backpropagation (BP) network are adjusted based on this error. Subsequently, forward propagation is performed again. This training process continues iteratively until the error reaches a predefined range, at which point the model training concludes. During this training process, the BP network adjusts its parameters to minimize the error between the predicted output and the desired output, gradually improving the network's performance until the specified error threshold is met.

2.4. Improved prediction model

In the context of granular computing, the different periodic trends in photovoltaic (PV) power generation represent different granularity levels within the dataset. In the case study presented in this paper, the daily PV power generation trend is considered a smaller grain, while the weekly power generation trend represents a larger grain. In simple terms, granular computing can be viewed as a method of categorizing data.

In real-life applications such as power generation forecasting, the effects of various granularity levels are taken into consideration. Therefore, this paper chooses to integrate granular computing with neural networks to simulate how the human brain processes information. The model for this integration is illustrated in Figure 2.

![Prediction flowchart](image.png)
In practical power generation forecasting, different time window periods are often considered. For example, predicting the next day’s data based on the previous day’s data, predicting the next day’s data based on the past 7 days, or using a 30-day window, etc. The choice of time window often impacts the accuracy of the prediction model. Time series data can be decomposed into trend components, periodic components, and random components. If a large number of time windows are used to construct a granularity matrix as selected in this paper, the randomness or noise in the data can significantly influence the prediction, leading to neural networks learning substantial biases and deviating from the original trend or periodic features.

Therefore, this paper first utilizes wavelet transform and Fourier transform to explore the periodic characteristics of this time series. Based on the discovered time windows from this exploration, a granularity matrix is constructed. The main steps involved in constructing this matrix are as follows [9]:

1. Construct a one-dimensional matrix \( M \) using all the data points \( y_1, y_2, \ldots, y_n \) with \( n \) data points. Here are the steps involved:
   1.1. Construct a one-dimensional matrix \( M \) using all the data points \( y_1 \) to \( y_n \), denoted as \( M \) with a length \( L_M \).
   1.2. Create a total granularity matrix \( G \) consisting of \( K \) rows, where each row \( n \) of the transition matrix \( G \) (from \( L_M-K \) columns) is defined as: \( G(n,n: L_M) = M(1,1:L_M - n - 1) \).

   This step involves using the data from \( M \) to populate each row of \( G \) based on the specified transition pattern.

2. Select a portion of the transition matrix \( G \) constructed by \( K \) grains to form the final matrix \( G_k \), where: \( G_k(1:1: L_M - k) = G(1:1: k - 1: L_M) \). Here, \( G_k \) is a submatrix of \( G \), and the size and content of \( G_k \) are determined by the choice of \( K \) grains.

3. The resulting matrix, as depicted in Figure 3, represents the data points arranged according to the specified granularity. Each entry in the matrix corresponds to a specific data value \( y_k \) from the original dataset. The rows and columns of the matrix are both related to the chosen granularity, making the size and structure of the constructed matrix dependent on the selected grains. Different choices of grains will result in different matrices with varying dimensions and contents.

\[
\begin{array}{cccc}
\text{Line 1} & k & \ldots & n \\
\text{Line 2} & k-1 & k & \ldots & n-1 \\
\text{Line 3} & k-2 & k-1 & k & \ldots & n-2 \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
\text{Line K} & 1 & 2 & \ldots & n-k+1 \\
\end{array}
\]

Fig 3. Constructed matrix

After completing the matrix construction, considering the objectives of prediction accuracy and efficiency, this study proceeds with further normalization of the constructed matrix to scale the values within the range \((0, 1)\). The normalization formula and its inverse, represented by equations (4) and (5), are as follows:

\[
y_i' = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \quad (4)
\]

\[
y_i = y_i' \cdot (y_{\max} - y_{\min}) + y_{\min} \quad (5)
\]

During the training of the neural network using the matrix as input, a specific column of the matrix serves as input, and the output is the first data point of the subsequent column. This process calculates the error between the predicted value and the original data for each day. Therefore, different granularity levels selected for the matrix will yield varying prediction values from the neural network.
In this study, a weighted approach is adopted based on the errors associated with each granularity level. By calculating the weighted average of predictions from different granularity levels, the final output result is obtained. The formula for computing the relative error rate, as shown in equation (6) where $y_i$ represents the true value for the $i$-th day, is as follows:

$$Er_i = \frac{|\hat{y}_i - y_i|}{y_i}$$  \hspace{1cm} (6)

Furthermore, considering that the models established under different granularities predict different total numbers of days, this study employs the modified Mean Absolute Percentage Error (MAPE) to represent the prediction error rate for a specific granularity. The calculation formula for MAPE, denoted as $\delta$ in equation (7), incorporates $n$ as the number of days predicted under that granularity:

$$\delta = \frac{\sum Er_i}{n}$$ \hspace{1cm} (7)

Subsequently, the calculation formula for the weight $Q_k$ corresponding to each granularity matrix is derived in equation (8):

$$Q_k = \frac{1 - \delta_k}{\sum (1 - \delta_k)_i}$$ \hspace{1cm} (8)

Finally, based on the weights $Q_k$ assigned to each granularity, the final prediction result $\bar{y}$ is computed using a weighted average approach, as shown in equation (9):

$$\bar{y} = \sum_i^k Q_k y_k$$ \hspace{1cm} (9)

By utilizing the calculated weights $Q_k$ to combine the predictions from different granularity levels, the final prediction $\bar{y}$ is obtained, offering an integrated forecast result that accounts for the varying prediction accuracies associated with different granularities. This approach ensures an optimized and robust prediction outcome that reflects the overall trend of the predicted data while considering the performance of individual granularity models.

3. Example analysis

3.1. Data description

The study utilizes hourly distribution load data and meteorological data from Zhoushan City, Zhejiang Province, for the year 2020, comprising 8760 data points including wind speed, ground radiation intensity, temperature, and load information. The raw data may contain duplicates, inconsistencies, anomalies, and other issues requiring preprocessing to transform it into usable data.

Photovoltaic (PV) power generation is influenced by various factors in real-life scenarios, including the properties of PV components such as solar panels, as well as weather factors like solar radiation and temperature [15]. Given available solar radiation and temperature data, this study calculates the PV power generation per unit area of a PV panel using equation (10). Assuming a photovoltaic conversion efficiency $\eta$, panel area $S$, solar radiation $I$, and temperature $t_0$, the PV power generation is represented as:

$$P_s = \eta S I [1 - 0.005(t_0 + 25)]$$ \hspace{1cm} (10)

In this study, the photovoltaic conversion efficiency $\eta$ is set to 0.2. Using the above formula, the hourly PV power generation is calculated. By summing the PV power generation for each hour of the day, the daily PV power generation data for the entire year is obtained.
3.2. Result analysis

Some scholars have pointed out that wavelet transforms can effectively analyze time-frequency characteristics in many time series data with periodic patterns and have significant applications in areas such as predicting power generation time series [16]. Wavelet analysis decomposes the analyzed time series into low-frequency and high-frequency components using appropriate wavelet basis functions and decomposition levels, projecting them onto corresponding time scales. This decomposition reveals frequency characteristics in the time domain, accurately identifying major periodic components, trends, and abrupt changes, while also providing insights into the lengths and distributions of various time scales (periods).

In this study, the original time series is first decomposed using the db4 wavelet into 5 levels of detail coefficients and 1 level of approximation coefficients. The detail coefficients represent the high-frequency components of the signal, which typically contain information about local variations, transitions, or noise within the signal. On the other hand, the approximation coefficients correspond to the low-frequency part of the signal, capturing long-term trends and periodicities [17]. The decomposition results are illustrated in Figure 4.

As shown in Figure 4, different periodic components can be observed from the high-frequency and low-frequency parts of the wavelet decomposition. From Figure (c) and Figure (d), a periodicity of approximately 20 days is evident, while Figure (f) suggests a periodicity of around 60 days. Therefore, two periodic ranges are initially identified based on these observations.

However, wavelet transforms may not exhibit a clear advantage in identifying shorter periodic components. To extract more periodic components, especially shorter ones, beyond what is evident in the wavelet analysis, a fast Fourier transform (FFT) is applied to the high-frequency components. The partial results obtained after applying FFT to the high-frequency part are illustrated in Figure 5.

Fig 4. Low frequency and high frequency parts

As shown in Figure 4, different periodic components can be observed from the high-frequency and low-frequency parts of the wavelet decomposition. From Figure (c) and Figure (d), a periodicity of approximately 20 days is evident, while Figure (f) suggests a periodicity of around 60 days. Therefore, two periodic ranges are initially identified based on these observations.

However, wavelet transforms may not exhibit a clear advantage in identifying shorter periodic components. To extract more periodic components, especially shorter ones, beyond what is evident in the wavelet analysis, a fast Fourier transform (FFT) is applied to the high-frequency components. The partial results obtained after applying FFT to the high-frequency part are illustrated in Figure 5.
Through Fourier transform analysis, the main periodicities corresponding to the peaks are identified as 33 days for Figure 1 and 50 days for Figure 2. By performing fast Fourier transforms (FFT) on various high-frequency components and combining the results with wavelet analysis, this study preliminarily identifies five major periodicities: 62 days, 37 days, 58 days, 12 days, and 46 days.

After determining these periodicities, the study constructs granularity matrices for the dataset at each granularity level, normalizes them, and proceeds with neural network training to establish a prediction model. Subsequently, weighted values corresponding to different granularities are calculated using the formula mentioned earlier, and the final prediction results are obtained by averaging the weighted predictions.

In selecting the prediction time window, the study first considers the limitations of the prediction model. Despite having a large amount of sample data, precise long-term predictions are challenging. Therefore, like many scholars focusing on medium to short-term forecasting, this study avoids overly long time windows. The established prediction model includes granularity matrices with periodicities of 12 days, 37 days, 46 days, 58 days, and 62 days, enabling the capture of medium-term trends and cyclical fluctuations in photovoltaic power generation data.

Considering the completeness of the time window’s periodicity, this study chooses a 7-day prediction window. The last 7 days of the year’s data are used as the test set, while the remaining data is utilized as the training set for model development and evaluation.

Taking the 12-day granularity period model as an example, this paper determined the parameters of the neural network using grid search and cross-validation methods. The settings for each parameter are shown in Table 1.

---

Fig 5. Partial results after Fourier transform
Table 1. Neural network parameter value setting

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of hidden layers</td>
<td>1</td>
</tr>
<tr>
<td>Number of hidden layer neurons</td>
<td>42</td>
</tr>
<tr>
<td>Learning speed</td>
<td>0.01</td>
</tr>
<tr>
<td>Activation function</td>
<td>Relu</td>
</tr>
<tr>
<td>Optimization algorithm</td>
<td>Adam</td>
</tr>
<tr>
<td>Maximum number of iterations</td>
<td>1000</td>
</tr>
</tbody>
</table>

Following the algorithmic procedure, this paper sequentially trained models for 5 granularity matrices and then predicted the values for the last 7 days of the dataset. The weights for each granularity were determined based on the error values obtained during prediction.

Fig 6. Prediction results

The prediction error rates for the last 7 days under various granularities are shown in Figure 6. It can be observed that for the 7-day prediction period, the error rates mostly range between 5% to 20%, showing overall stable fluctuations in error rates. For smaller granularities, such as the 12-day period model, the overall error rate may be slightly higher due to the smaller amount of input data per prediction, which may limit the neural network model's ability to fully capture underlying patterns.

Each model exhibits fluctuations in prediction accuracy when forecasting different numbers of days. Therefore, by calculating weights using Equation (8) and (9) for each model and then averaging them, the final result is obtained. The calculated weights reflect the contribution of each model to the prediction results, effectively reducing the impact of individual granularity model errors and improving overall prediction accuracy, leading to more reliable and robust forecasts. The weights computed based on errors for each granularity are presented in Table 2:

Table 2. Weights for each granularity

<table>
<thead>
<tr>
<th>Granularity</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Granularity1 12-day period</td>
<td>0.1805</td>
</tr>
<tr>
<td>Granularity1 37-day period</td>
<td>0.1967</td>
</tr>
<tr>
<td>Granularity1 46-day period</td>
<td>0.1990</td>
</tr>
<tr>
<td>Granularity1 58-day period</td>
<td>0.2084</td>
</tr>
<tr>
<td>Granularity1 62-day period</td>
<td>0.2151</td>
</tr>
</tbody>
</table>
When calculating the weights for each granularity, it is important not only to consider the prediction error rate but also to understand the specific meanings associated with selecting each granularity. For example, in this study, various periodic terms were discovered through wavelet and Fourier transforms, with some derived from decomposed low-frequency components and others from high-frequency components. In frequency domain analysis, low-frequency components reflect the overall characteristics of the time series, whereas high-frequency components better capture detailed information about the time series, with the low-frequency components having a greater impact on periodic trends [18]. Therefore, the weights corresponding to the periodic terms of the low-frequency components should be larger, while those corresponding to the high-frequency components that have less impact on periodicity should be smaller.

From Table 2, it can be observed that the weight corresponding to the periodic term of 62 days, associated with the low-frequency component, is the largest, indicating superior performance in prediction accuracy. Hence, no further weight adjustments were made in this study.

Additionally, Table 2 demonstrates minimal differences in weights across various granularities, indicating that the granularity selections were appropriate and each granularity contributes uniformly to the final prediction result. If significant differences in prediction accuracy were observed across different granularities, it would suggest that the selected periodic terms have minimal impact on the time series and are not suitable as prediction time windows. The relatively small differences in accuracy among the periodic terms corresponding to each granularity also demonstrate the suitability and adaptability of the methods used for preprocessing and exploring periodic terms in this study.

After obtaining the weights for each granularity, the final prediction results on the test set were derived by averaging the weighted predictions from each granularity. To assess whether the method of exploring periodic terms through frequency domain analysis improves the prediction model accuracy, this study also calculated the prediction model accuracy based solely on empirically selected periodic terms. The empirically chosen five periodic terms were 1 day, 7 days, 14 days, 30 days, and 60 days. The comparison of prediction results is shown in Figure 7.

This study also compared the improved prediction model with other time series forecasting models, such as the ARIMA model and LSTM (Long Short-Term Memory) model. The comparison of Mean Absolute Percentage Error (MAPE) among different models is shown in Table 3, where "A" represents the improved neural network prediction model under the enhanced granularity computing approach, and "B" represents the previous model.
Table 3. Model error comparison

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>LSTM</th>
<th>ARIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE</td>
<td>0.1732</td>
<td>0.1920</td>
<td>0.2248</td>
<td>0.2432</td>
</tr>
</tbody>
</table>

The results indicate that all four time series forecasting models performed relatively well, with average error rates not exceeding 24%, suggesting the dataset exhibits certain regular patterns. The improved neural network prediction model under the enhanced granularity computing approach developed in this study showed the best performance, with an average error rate of around 17%. This model demonstrated higher accuracy, indicating that combining preprocessing to extract cycles with a neural network model under the granularity computing approach can enhance prediction accuracy and improve model utility.

4. Conclusion

To reduce the impact of photovoltaic (PV) generation uncertainty on power systems and enhance overall system reliability, this study proposes an improved granularity computing approach-based PV power prediction model. Firstly, the model employs wavelet transform to analyze the original time series in the frequency domain, decomposing it into 5 layers of high-frequency components and 1 layer of low-frequency component, followed by fast Fourier transform (FFT) to identify main periodic terms. Secondly, based on the identified five main periodic terms, granularity matrices are constructed for each frequency component, and each granularity matrix is inputted into a BP neural network for training until the error reaches the specified range, indicating the end of model training.

Subsequently, this study calculates the average error rates on the final 7 days of the test set for the five granularities and computes the corresponding weights based on these error rates. Observing the weights of each granularity, it is found that the low-frequency components carrying cycle information correspond to higher weights, thus obviating the need for further weight adjustment indices. Finally, the study obtains the model’s prediction results on the test set by averaging the predicted values of each granularity weighted, and compares them with the pre-improved model and other classical time series forecasting models like ARIMA and LSTM.

The results show that the proposed improved granularity computing approach-based neural network prediction model performs optimally, with an average error rate of around 17%. This indicates that combining frequency domain analysis to identify data periodic terms with granularity computing neural network prediction models has certain advantages, providing useful insights for PV generation forecasting.

References


